Residual Stopping and Pointwise Inference for Nonuniform Entropic Self-Transport

Scott Schubr
Luminary Oddities

October 18, 2026

Abstract

We study observable residual stopping for the half-self-Sinkhorn iteration with iid samples from a fixed positive smooth density on the one- or two-dimensional torus. For the normalized periodic Gaussian, assume that $\varepsilon \to 0$, $n\varepsilon^{3d/2} \to \infty$, and $n\varepsilon^{d/2+3} \to 0$. Starting from the constant scaling, we stop at the first projective square-root-mass residual below $t_n = n^{-1/2}\varepsilon^{-\gamma}$, where $0 < \gamma < d/4$ is fixed. The stopped score has the same pointwise Gaussian limit, centered at the true density score, as the exactly scaled empirical estimator. Its attained residual lies between ct_n and t_n with probability tending to one for every fixed $0 < c < 1/2$, although $\sqrt{n}\, t_n \to \infty$. The stopping index is asymptotic to $(d/4 - \gamma)\log(1/\varepsilon)/\log 2$ in probability. For the coefficient-one threshold, uniform sampling already gives a different score limit when $\gamma \geq d/4$. The proof retains the iteration-dependent empirical-norm error and uses a normalized-row cancellation to control nonuniform initialization in dimension two.

Keywords: entropic self-transport; Sinkhorn iteration; statistical stopping; score estimation; pointwise inference.

1 Introduction

Symmetric kernel scaling gives an empirical entropic self-transport score through its smooth off-sample extension. A numerical solver is normally terminated using a normalization residual, whereas statistical inference depends on the error in a spatial derivative. This paper identifies a family of residual tolerances that preserves the pointwise limiting law along the actual sampled iteration.

The tolerance can exceed $n^{-1/2}$ by a diverging factor. We prove this for a first-hit rule checked from the initial scaling, rather than for a prescribed iteration number. We also determine the leading stopping count, control the residual attained at that hit, and give a counterexample at the upper tolerance exponent. The same observations determine the scale, iterates, residual and queried score.

1.1 Related work

Kokot's dissertation [2], Chapter 3, gives Gaussian-RKHS scaling differentiability, the self-diagonal plus-resolvent derivative, and a fixed-regularization Hilbert central limit theorem. Its Section 4.4.1 also gives first-order population matching after two symmetric updates. We use these established mechanisms and provide the joint bandwidth–sample–iteration estimates needed for residual stopping. Mordant [3] studies decreasing-regularization self-transport score asymptotics and the half-self iteration in a Euclidean quadratic-cost model. Our statistical limit is proved for the exact periodic Gaussian.

Wormell and Reich [4] analyze empirical bistochastic normalization, resolvents and spectral convergence, including an alternative symmetric scaling iteration. Cheng and Landa [5] show that approximate nonuniform normalization at statistical accuracy suffices for graph-Laplacian consistency. Maeda, Yao and Nitanda [7] combine sampling and iteration errors for two-sample Schrödinger bridges. Spectral discrepancy stopping is developed in, among other works, Celisse and Wahl [6]. These results establish the general statistical motivation for early termination.

The present conclusions concern the specific nonlinear first residual hit: its admissible exponent range, attained residual and leading count, together with preservation of the exact estimator's true-score limit. The earlier residual-certificate paper [8] gives bounds for arbitrary positive candidates and empirical sharpness witnesses. The bandwidth windows below are sufficient.

2 Model and main theorem

Let $d \in \{1, 2\}$ and let $f \in C^6(\mathbb{T}^d)$ be a fixed probability density with $\min f > 0$. Set $\mu = f\,dx$ and let $\mu_n = n^{-1}\sum_{i=1}^n \delta_{X_i}$ for iid $X_i \sim \mu$. Lebesgue measure on the unit torus has total mass one. The kernel is

$$k_\varepsilon(z) = \sum_{\ell\in\mathbb{Z}^d} (2\pi\varepsilon)^{-d/2} e^{-|z+\ell|^2/(2\varepsilon)}, \qquad \lambda_k = e^{-2\pi^2\varepsilon|k|^2}. \tag{1}$$

We also write $k_\varepsilon(x, y) = k_\varepsilon(x - y)$. Here λ_k are the Fourier multipliers and $H_\varepsilon = \exp(\varepsilon\Delta/2)$ is convolution against Lebesgue measure. Write $K_\nu b(x) = \int k_\varepsilon(x - y)b(y)\,d\nu(y)$ and $K_n = K_{\mu_n}$. Throughout,

$$\varepsilon = \varepsilon_n \to 0, \qquad n\varepsilon^{3d/2} \to \infty, \qquad n\varepsilon^{d/2+3} \to 0. \tag{2}$$

Constants may depend on this fixed f and on fixed exponents.

Let $\widehat{a}$ be the positive empirical self-scale, $\widehat{a}K_n\widehat{a} = 1$. At the sample sites it exists uniquely: the strictly convex coercive function

$$\Phi(u) = \frac{1}{2n^2}\sum_{i,l} k_\varepsilon(X_i - X_l)e^{u_i+u_l} - \frac{1}{n}\sum_i u_i$$

has a unique minimizer and its gradient equations are the scaling equations. Coercivity and strict convexity follow already from the positive diagonal terms. Extend the scale by $\widehat{a}(x) = 1/K_n\widehat{a}(x)$ away from the sites. Starting with $b_0 = 1$, use the averaged symmetric iteration [1, 3]

$$b_{j+1} = \left\{b_j/(K_n b_j)\right\}^{1/2}, \qquad s_j = 2\nabla \log K_n b_j, \qquad \widehat{s} = 2\nabla \log K_n\widehat{a}. \tag{3}$$

Only site values are needed for each convolution. Define

$$F_j = (b_j K_n b_j)^{1/2}, \quad S_j = \mu_n F_j, \quad r_j = \left\|F_j/S_j - 1\right\|_n, \qquad \|u\|_n^2 = \mu_n\,|u|^2\,. \tag{4}$$

Both r_j and s_j are invariant under positive scalar rescaling of b_j. For fixed $\gamma > 0$, put

$$\begin{aligned} t_n &= n^{-1/2}\varepsilon^{-\gamma}, \qquad L_n = \max\{3, \lceil 4\log n\rceil\}, \\ \tau_\gamma &= \min\big(\{j \in \{0, \dots, L_n\} : r_j \le t_n\} \cup \{L_n\}\big). \end{aligned} \tag{5}$$

The budget makes the rule finite on every sample. Set

$$A_n = \sqrt{n}\,\varepsilon^{d/4+1/2}, \qquad J_d = \int_{\mathbb{R}^d} \frac{(2\pi\xi_1)^2}{(1 + e^{2\pi^2|\xi|^2})^2}\,d\xi > 0. \tag{6}$$

Theorem 2.1. *Assume* (2) *and fix* $0 < \gamma < d/4$. *Then*

$$\frac{\tau_\gamma}{\log(1/\varepsilon)} \longrightarrow \frac{d/4 - \gamma}{\log 2} \quad \textit{in probability}. \tag{7}$$

For every fixed $0 < c < 1/2$,

$$\mathbb{P}\{\tau_\gamma < L_n,\ ct_n \le r_{\tau_\gamma} \le t_n\} \to 1. \tag{8}$$

For each fixed $x \in \mathbb{T}^d$,

$$A_n\left\|s_{\tau_\gamma}(x) - \widehat{s}(x)\right\| \to 0 \qquad \textit{in probability}, \tag{9}$$

$$A_n\{s_{\tau_\gamma}(x) - \nabla \log f(x)\} \Rightarrow N_d\left(0, \frac{4J_d}{f(x)}I_d\right). \tag{10}$$

The exact empirical score $\widehat{s}$ *satisfies the same centered limit.*

In particular $\sqrt{n}\, r_{\tau_\gamma} \to \infty$ in probability. For $\varepsilon = n^{-\kappa}$ the sufficient power windows are

$$d = 1: \quad 2/7 < \kappa < 2/3, \qquad d = 2: \quad 1/4 < \kappa < 1/3.$$

For fixed $0 < \gamma_1 < \gamma_2 < d/4$, subtraction on the same sample gives

$$\frac{\tau_{\gamma_1} - \tau_{\gamma_2}}{\log(1/\varepsilon)} \longrightarrow \frac{\gamma_2 - \gamma_1}{\log 2} \quad \text{in probability.} \tag{11}$$

This compares update counts. The upper exponent obstruction is proved in Section 7.

3 Population geometry and the empirical scaling branch

Throughout this section, $d \in \{1, 2\}$, f is the fixed density in the main theorem, and $\mu = f\, dx$. Constants can depend on f and d. Write H_ε for convolution against the kernel in (1) with respect to Lebesgue measure and write

$$K_\mu u = H_\varepsilon(fu), \qquad K_n u = \frac{1}{n}\sum_{i=1}^{n} k_\varepsilon(\,\cdot\,, X_i)u(X_i).$$

Let a denote the positive population scale, $aK_\mu a = 1$, and put

$$v = a^{-1} = K_\mu a, \qquad w = H_\varepsilon f, \qquad q = a\sqrt{f}, \qquad d_1 = \log v - \tfrac{1}{2}\log w.$$

All bounds below hold as $\varepsilon \downarrow 0$.

Lemma 3.1 (Population approximation)**.** *The positive population scale is unique and satisfies*

$$\|q - 1\|_{C^2} \le C\varepsilon, \qquad \|d_1\|_{C^2} \le C\varepsilon, \qquad \|-2\nabla \log a - \nabla \log f\|_\infty \le C\varepsilon. \tag{12}$$

In particular, a, v, w are bounded above and below, and their C^2 norms are bounded uniformly for small ε.

Proof. Set $g_f = \sqrt{f}$. The scaling equation becomes

$$\mathcal{F}_\varepsilon(q) := qH_\varepsilon(qg_f) - g_f = 0.$$

We solve it in the Banach algebra $C^2(\mathbb{T}^d)$. First, $\|H_\varepsilon g_f - g_f\|_{C^2} \le C\varepsilon \|g_f\|_{C^4}$ by the heat semigroup identity. Second, $(I + H_\varepsilon)^{-1}$ is uniformly bounded on C^2. Indeed, its Fourier multiplier is $1 - F(\sqrt{\varepsilon}k)$, where

$$F(\xi) = \frac{1}{1 + \exp(2\pi^2|\xi|^2)}.$$

The function F is Schwartz. Periodization of its rescaled inverse Fourier transform gives a convolution kernel whose L^1 norm is bounded independently of ε. The same bound holds after commuting derivatives through convolution.

At $q = 1$,

$$D\mathcal{F}_\varepsilon(1)u = (H_\varepsilon g_f)u + H_\varepsilon(g_f u) = g_f(I + H_\varepsilon)u + E_\varepsilon u,$$

where $\|E_\varepsilon\|_{C^2 \to C^2} = O(\sqrt{\varepsilon})$. To check the last bound, differentiate the commutator $[H_\varepsilon, M_{g_f}]u$ through order two. Each resulting term is a commutator with a derivative of g_f; its kernel contains the difference of that derivative at two points. The Gaussian first absolute moment is $O(\sqrt{\varepsilon})$, so each term is bounded by $C\sqrt{\varepsilon}\|u\|_{C^2}$. The multiplication error $H_\varepsilon g_f - g_f$ has the stronger $O(\varepsilon)$ bound. Consequently $D\mathcal{F}_\varepsilon(1)$ has a uniformly bounded inverse for small ε. The remaining term in $\mathcal{F}_\varepsilon(1 + u)$ is $uH_\varepsilon(g_f u)$, bounded by $C\|u\|_{C^2}^2$ with the corresponding local Lipschitz estimate. A contraction on a ball of radius $C\varepsilon$ therefore yields a positive solution with $q = 1 + O_{C^2}(\varepsilon)$.

For uniqueness, two positive scales a, b obey

$$0 \le \langle a-b, K_\mu(a-b)\rangle_{L^2(\mu)} = -\int \frac{(a-b)^2}{ab}\, d\mu \le 0.$$

Here positivity follows from the nonnegative Fourier multipliers of the kernel. Hence $a = b$. Finally $a = q/\sqrt{f}$, $w = f + O_{C^2}(\varepsilon)$, and the elementary product, reciprocal, and logarithm estimates in C^2 give every conclusion in (12). □

Let $\mathcal{H} = \mathcal{H}_\varepsilon$ be the real reproducing-kernel Hilbert space of k_ε, with feature $k_y = k_\varepsilon(\,\cdot\,, y)$. Define

$$S : L^2(\mu) \longrightarrow \mathcal{H}, \qquad Su = \int a(y) k_y u(y)\, d\mu(y), \qquad S^* h(y) = a(y) h(y),$$
$$C = SS^*, \qquad P = S^*S, \qquad B = \tfrac{1}{2}(I - C), \qquad R = (I + C)^{-1}.$$

For a fixed query x and component $\nu \in \{1, \dots, d\}$, set

$$\theta_{x,\nu} = \partial_{x_\nu}\{a(x) k_x\}, \qquad \alpha_d = \frac{d}{4} + \frac{1}{2}.$$

For a vector section θ_x, inner products with a Hilbert-space element are taken componentwise.

Lemma 3.2 (Feature geometry)**.** *The operator P is a self-adjoint positive Markov operator. Moreover,*

$$\|S\| = \|C\| = 1, \quad \|B\| \le \tfrac{1}{2}, \quad \|R\| \le 1, \qquad \|v\|_{\mathcal{H}} = 1, \quad Cv = v, \quad S^*v = 1, \tag{13}$$
$$\sup_y \|k_y\|_{\mathcal{H}} \le C\varepsilon^{-d/4}, \qquad \sup_x \|\theta_{x,\nu}\|_{\mathcal{H}} \le C\varepsilon^{-\alpha_d}, \qquad \operatorname{tr} C \le C\varepsilon^{-d/2}. \tag{14}$$

For every fixed x and $\nu, \ell \in \{1, \dots, d\}$,

$$\varepsilon^{d/2+1}\langle \theta_{x,\nu}, RCR\theta_{x,\ell}\rangle_{\mathcal{H}} \longrightarrow \frac{J_d}{f(x)}\, \mathbf{1}_{\{\nu=\ell\}}, \quad J_d = \int_{\mathbb{R}^d} \frac{(2\pi\xi_1)^2}{[1 + \exp(2\pi^2|\xi|^2)]^2}\, d\xi. \tag{15}$$

Every fixed interval of positive length strictly inside $(0, 1)$ contains at least $c\varepsilon^{-d/2}$ eigenvalues of C, counted with multiplicity, for all sufficiently small ε.

Proof. The kernel of P relative to μ is $a(x) k_\varepsilon(x, y) a(y)$. It is symmetric, positive semidefinite, and has row sum one by the scaling equation. Jensen's inequality gives $\|P\| \le 1$, and $P1 = 1$ gives equality. The singular-value identities for S yield the asserted norms. Also

$$\|v\|_{\mathcal{H}}^2 = \int a(y) v(y)\, d\mu(y) = 1, \quad Cv = \int a(y)^2 k_y v(y)\, d\mu(y) = v,$$

and $S^*v = av = 1$. Gaussian Fourier sums give

$$k_\varepsilon(y, y) = \sum_{k \in \mathbb{Z}^d} e^{-2\pi^2 \varepsilon |k|^2} = O(\varepsilon^{-d/2}), \qquad \|\partial_{x_\nu} k_x\|_{\mathcal{H}}^2 = \sum_k (2\pi k_\nu)^2 e^{-2\pi^2 \varepsilon |k|^2} = O(\varepsilon^{-d/2-1}).$$

Together with Lemma 3.1, these prove the section bounds. The trace formula is

$$\operatorname{tr} C = k_\varepsilon(0,0) \int a^2\, d\mu = k_\varepsilon(0,0) \int q^2\, dx = O(\varepsilon^{-d/2}).$$

There is a convenient comparison within the same Hilbert space. Put

$$C_H = \int k_y \otimes k_y\, dy, \qquad R_H = (I + C_H)^{-1}.$$

Since $a^2 f = q^2$, comparison of quadratic forms gives

$$\|C - C_H\|_{\mathrm{op}} \le \|q^2 - 1\|_\infty \|C_H\|_{\mathrm{op}} = O(\varepsilon). \tag{16}$$

Here C_H has eigenvalues $\lambda_k = \exp(-2\pi^2\varepsilon|k|^2)$: in the complexification of $\mathcal{H}$, its orthonormal Fourier basis consists of $\sqrt{\lambda_k}\exp(2\pi i k \cdot x)$. In particular $\|C_H\| = 1$. Resolvent identities give

$$\|RCR - R_H C_H R_H\|_{\mathrm{op}} = O(\varepsilon).$$

Also $\theta_{x,\nu} = a(x)\partial_{x_\nu} k_x + (\partial_{x_\nu} a(x))k_x$, where the second term has norm $O(\sqrt{\varepsilon}\,\varepsilon^{-\alpha_d})$. After multiplication by $\varepsilon^{2\alpha_d}$, the operator replacement therefore has error $O(\varepsilon)$ and these section replacements have error $O(\sqrt{\varepsilon})$. Since $a(x)^2 \to 1/f(x)$, the Gram limit reduces to

$$\varepsilon^{d/2+1} \sum_{k\in\mathbb{Z}^d} (2\pi k_\nu)(2\pi k_\ell) \left(\frac{e^{-2\pi^2\varepsilon|k|^2}}{1 + e^{-2\pi^2\varepsilon|k|^2}} \right)^2 .$$

This is a Gaussian-dominated Riemann sum. Its limit is J_d for equal components and zero for unequal components, proving (15).

Finally, by (16) and the compact self-adjoint min–max principle, the ordered eigenvalues of C and C_H differ by at most $C\varepsilon$. Inside a fixed interval $I \subset (0,1)$, choose a smaller interval at positive distance from its endpoints. The heat eigenvalues in that smaller interval correspond to lattice points in a fixed annulus after multiplication by $\sqrt{\varepsilon}$; their number is bounded below by $c\varepsilon^{-d/2}$. All their perturbed eigenvalues belong to I for small ε, proving the count. □

For the empirical statements, all random quantities use the same sample. Define

$$g = (\mu_n - \mu)(ak), \qquad g_0 = (\mu_n - \mu)(k),$$
$$\zeta = n^{-1/2}\varepsilon^{-d/4}, \qquad \eta = n^{-1/2}\varepsilon^{-d/2}, \qquad C_{0,n} = \int a(y)^2 k_y \otimes k_y \, d\mu_n(y).$$

The notation $O_{\mathbb{P},\mathcal{H}}$ below means a bound in Hilbert-space norm.

Lemma 3.3 (Quantitative empirical branch)**.** *Under* (2)*, the exact empirical scale $\widehat{a}$, extended by $\widehat{a}K_n\widehat{a} = 1$, obeys*

$$\widehat{v} - v = Rg + O_{\mathbb{P},\mathcal{H}}(\eta\zeta), \qquad \|\widehat{v} - v\|_{\mathcal{H}} = O_{\mathbb{P}}(\zeta), \qquad \widehat{v} = \widehat{a}^{-1}, \tag{17}$$
$$\|\widehat{a} - a\|_\infty + \sqrt{\varepsilon}\|\nabla(\widehat{a} - a)\|_\infty = O_{\mathbb{P}}(\eta). \tag{18}$$

In addition,

$$\|g\|_{\mathcal{H}} + \|g_0\|_{\mathcal{H}} = O_{\mathbb{P}}(\zeta), \qquad \mathrm{Cov}(g) = \frac{C - v\otimes v}{n}, \qquad \langle v, g\rangle_{\mathcal{H}} = 0, \tag{19}$$
$$\|C_{0,n} - C\|_{\mathrm{op}} \le \|C_{0,n} - C\|_{\mathrm{HS}} = O_{\mathbb{P}}(\eta). \tag{20}$$

Define $S_n, S_n^*, \widehat{C}, P_n$ by replacing a, μ in S, S^*, C, P with $\widehat{a}, \mu_n$. Then

$$\|S_n\| = \|\widehat{C}\| = 1, \qquad \|\widehat{C} - C\|_{\mathrm{op}} = O_{\mathbb{P}}(\eta), \tag{21}$$
$$\sup_x \|\widehat{\theta}_{x,\nu} - \theta_{x,\nu}\|_{\mathcal{H}} = O_{\mathbb{P}}(\eta\varepsilon^{-\alpha_d}), \qquad \widehat{\theta}_{x,\nu} = \partial_{x_\nu}\{\widehat{a}(x)k_x\}. \tag{22}$$

On a common event of probability tending to one, evaluation satisfies $\|h\|_{L^2(\mu_n)} \le C\|h\|_{\mathcal{H}}$ for every $h \in \mathcal{H}$.

Proof. The expectation of $a(X)k_X$ is v, and its second moment operator is C. Hence the covariance identity follows. Moreover $\langle v, a(y)k_y\rangle = a(y)v(y) = 1$, so the claimed orthogonality is exact. The feature norm bound gives $\mathbb{E}\|g\|_{\mathcal{H}}^2 + \mathbb{E}\|g_0\|_{\mathcal{H}}^2 \le C\zeta^2$. Similarly,

$$\mathbb{E}\|C_{0,n} - C\|_{\mathrm{HS}}^2 \le \frac{1}{n}\mathbb{E}\|a(X)^2 k_X \otimes k_X\|_{\mathrm{HS}}^2 \le Cn^{-1}\varepsilon^{-d} = C\eta^2.$$

These are Hilbert-space second-moment bounds; no finite-dimensional operator inequality is needed.

Consider $\mathcal{G}_n(z) = z - K_n(1/z)$ in a neighborhood of v in $\mathcal{H}$. The function v is bounded below, and

$$\mathcal{G}_n(v) = -g, \qquad D\mathcal{G}_n(v) = I + C_{0,n}, \qquad \|(I + C_{0,n})^{-1}\| \le 1.$$

For h small enough to keep $v + h$ positive, the reciprocal remainder is

$$K_n\left(\frac{h^2}{v^2(v+h)}\right).$$

Its $\mathcal{H}$ norm is bounded by $C\varepsilon^{-d/4}\|h\|_{\mathcal{H}}^2$ on an event where $\|C_{0,n}\| \le 2$. Indeed, write the remainder as the feature operator with weights a applied to $(ah)h/(v+h)$; its operator norm is bounded by $\|C_{0,n}\|^{1/2}$, while $\|ah\|_{L^2(\mu_n)} \le \|C_{0,n}\|^{1/2}\|h\|_{\mathcal{H}}$ and $\|h\|_\infty \le C\varepsilon^{-d/4}\|h\|_{\mathcal{H}}$. The same argument gives the local Lipschitz remainder bound. Since $\varepsilon^{-d/4}\zeta = \eta \to 0$, a contraction on a ball of radius $O_{\mathbb{P}}(\zeta)$ yields a positive solution and

$$\widehat{v} - v = (I + C_{0,n})^{-1} g + O_{\mathbb{P},\mathcal{H}}(\eta\zeta).$$

For precision, one can first restrict to $\|g\| \le M\zeta$ for fixed M, perform the contraction for all sufficiently large n, and then let M increase. The resolvent identity and (20) replace $(I+C_{0,n})^{-1}$ by R at cost $O_{\mathbb{P},\mathcal{H}}(\eta\zeta)$.

The positive solution is the exact empirical branch: at sample sites its reciprocal satisfies the finite scaling equations, whose positive solution is unique by the same positive-semidefinite energy identity as in Lemma 3.1. Off sample, $\widehat{v} = K_n\widehat{a}$ fixes the stated extension. There is no extension to arbitrary signed measures in this argument.

The feature and derivative-feature bounds applied to $h = \widehat{v} - v$ show

$$\|h\|_\infty + \sqrt{\varepsilon}\|\nabla h\|_\infty = O_{\mathbb{P}}(\eta).$$

Taking reciprocals, using the deterministic lower bound on v, proves (18). The empirical operator P_n is positive, self-adjoint, and Markov on $L^2(\mu_n)$; hence the norm identities in (21) follow as in the population case. Also $\widehat{a}/a = 1 + O_{\mathbb{P}}(\eta)$ uniformly. Comparing quadratic forms gives

$$\|\widehat{C} - C_{0,n}\|_{\mathrm{op}} \le O_{\mathbb{P}}(\eta)\|C_{0,n}\|_{\mathrm{op}},$$

which proves the remaining operator bound. Differentiating $(\widehat{a} - a)(x)k_x$ and using (18) proves (22). Finally, since a is bounded below,

$$\|h\|_{L^2(\mu_n)}^2 \le C\int a^2h^2\,d\mu_n = C\langle h, C_{0,n}h\rangle_{\mathcal{H}} \le C\|h\|_{\mathcal{H}}^2$$

on the same event. This proves the simultaneous evaluation assertion. □

4 Initialization and the sampled trajectory

All operators below use the same observations. Write $e_j = \log(b_j/\widehat{a})$ and $B_n = (I - P_n)/2$, with empirical operators as in Lemma 3.3. Off sample, $P_nu(x) = \widehat{a}(x)K_n(\widehat{a}u)(x)$ is a Markov row. The exact log update is

$$\Psi_n(e) = \tfrac{1}{2}\{e - \log(P_n(\exp e))\}, \qquad e_{j+1} = \Psi_n(e_j). \tag{23}$$

Its uniform norm is nonincreasing. For small $\|e\|_\infty$, Taylor expansion, Jensen's inequality and the PSD Markov property give

$$\|\Psi_n(e) - B_ne\|_n \le C\,\|e\|_\infty\,\|e\|_n\,, \qquad \|B_n\|_{\mathrm{op}} \le \tfrac{1}{2}. \tag{24}$$

Lemma 4.1 (Two-update initialization)**.** *Let $v = 1/a$, $w = H_\varepsilon f$, and $d_1 = \log v - \frac{1}{2}\log w$. Put $\zeta = n^{-1/2}\varepsilon^{-d/4}$, $\eta = n^{-1/2}\varepsilon^{-d/2}$ and $u = e_1 - d_1$. Then*

$$u = aRg - \tfrac{1}{2}g_0/w + O_{L^2(\mu_n),\mathbb{P}}(\eta\zeta), \tag{25}$$

$$\|u\|_n = O_{\mathbb{P}}(\zeta), \qquad \|u\|_\infty = O_{\mathbb{P}}(\eta),$$

$$\|B_n d_1\|_n = O_{\mathbb{P}}(\varepsilon^2 + \varepsilon^{3/2}\eta), \tag{26}$$

$$e_2 = B_n u + E_2, \qquad \|E_2\|_n = O_{\mathbb{P}}(\varepsilon^2 + \varepsilon^{3/2}\eta + \varepsilon\zeta + \eta\zeta), \tag{27}$$

$$\|e_2\|_n = O_{\mathbb{P}}(\varepsilon^2 + \zeta), \qquad \|e_2\|_\infty = O_{\mathbb{P}}(\varepsilon^2 + \eta). \tag{28}$$

Proof. The first iterate is $b_1 = (K_n 1)^{-1/2}$. Expand $e_1 = \log\widehat{v} - \frac{1}{2}\log K_n 1$, using Lemma 3.3, $K_n 1 = w + g_0$, and the uniform value errors $O_{\mathbb{P}}(\eta)$. A quadratic logarithmic remainder has sample L^2 norm at most its uniform error times its L^2 error. This proves (25).

For (26), Lemma 3.1 gives $\|d_1\|_{C^2} \le C\varepsilon$. Thus

$$|d_1(y) - d_1(x)| \le C\varepsilon \operatorname{dist}(x, y), \qquad \|(I - P)d_1\|_\infty \le C\varepsilon^2.$$

The second estimate follows from the heat-semigroup identity applied to $af\,d_1 - d_1(x)af$, using bounded C^2 norms of af. Replace the normalized row weights $\widehat{a}(y)$ by $a(y)$. Their relative change is $O_{\mathbb{P}}(\eta)$, so its cost is this factor times the empirical row mean of $|d_1(y) - d_1(x)|$. The Gaussian-image bound

$$\operatorname{dist}(x, y)k_\varepsilon(x - y) \le C\sqrt{\varepsilon}\, k_{2\varepsilon}(x - y)$$

and the RKHS value estimate at bandwidth 2ε make this mean $O_{\mathbb{P}}(\varepsilon^{3/2})$, uniformly in x. Denominators are bounded below since $K_n a = K_\mu a + O_{\mathbb{P}}(\eta)$ uniformly. The weight cost is therefore $O_{\mathbb{P}}(\varepsilon^{3/2}\eta)$.

For the remaining fixed-a empirical row, its centered numerator is

$$N_n(x) = (\mu_n - \mu)_y\{k_\varepsilon(x - y)a(y)[d_1(y) - d_1(x)]\}.$$

At $x = X_i$ its own diagonal vanishes. Conditional on $X_i = x$, its mean is $O(\varepsilon^2/n)$ and its variance is at most

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

Averaging conditional second moments gives $\|N_n\|_n = O_{\mathbb{P}}(\varepsilon^{3/2}\zeta + \varepsilon^2/n)$. The uniform denominator error is $O_{\mathbb{P}}(\eta)$ and multiplies a population numerator $O(\varepsilon^2)$. This proves (26) without an independent-query assumption.

Taylor-expand (23) at deterministic d_1. The tilted row $Q_{n,d_1}h = P_n(e^{d_1}h)/(P_n e^{d_1})$ differs from P_n by $O(\varepsilon)$ in sample L^2 operator norm, by bounded multiplication and the contraction of P_n. Its quadratic remainder in u has norm at most $C\,\|u\|_\infty\,\|u\|_n$. Also $\|\Psi_n(d_1) - B_n d_1\|_n \le C\varepsilon^2$. These give (27) and its L^2 consequence. In the uniform norm Ψ_n is Lipschitz with constant one. The coarser row estimate $\|\Psi_n(d_1)\|_\infty = O_{\mathbb{P}}(\varepsilon^2 + \varepsilon\eta)$ uses $\|d_1\|_\infty\, O_{\mathbb{P}}(\eta)$ for weight and empirical-row errors. With $\|u\|_\infty = O_{\mathbb{P}}(\eta)$ this proves (28). □

Lemma 4.2 (Common trajectory and score proxies)**.** *There are nonnegative $E_n = o_{\mathbb{P}}(1)$ and $M_n = O_{\mathbb{P}}(1)$, common to all integers $j \ge 2$, such that*

$$e_j = S_n^* B^j Rg + E_{j,n}, \qquad \sqrt{n}\left\|E_{j,n}\right\|_n \le E_n + M_n\varepsilon^{1/2-d/4}2^{-j}, \tag{29}$$

$$A_n\left\|s_j(x) - \widehat{s}(x) - V_j(x)\right\| \le E_n + M_n\varepsilon^{1/2-d/4}2^{-j}, \qquad V_j(x) = 2\left\langle\theta_x, CB^j Rg\right\rangle. \tag{30}$$

The score remainder is uniform in x; vector-section inner products are componentwise. At each fixed x, for every deterministic $J_n \to \infty$,

$$\sup_{j \ge J_n} A_n\left\|s_j(x) - \widehat{s}(x)\right\| \to 0 \quad \textit{in probability}. \tag{31}$$

Proof. Let $\delta_n = \|e_2\|_\infty$. On one event of probability tending to one, (24) and uniform-norm nonexpansion imply $\|e_j\|_n \le (3/4)^{j-2}\|e_2\|_n$ for all $j \ge 2$. Iterating the remainder yields

$$\sup_{j\ge 2}\left\|e_j - B_n^{j-2}e_2\right\|_n = O_{\mathbb{P}}\{(\varepsilon^2+\eta)(\varepsilon^2+\zeta)\}. \tag{32}$$

Every error in (27) and (32) vanishes after multiplication by $\sqrt{n}$ under (2):

$$\sqrt{n}\varepsilon^2 \to 0, \quad \sqrt{n}\varepsilon^{3/2}\eta = \varepsilon^{3/2-d/2} \to 0, \quad \sqrt{n}\eta\zeta = (n\varepsilon^{3d/2})^{-1/2} \to 0.$$

The remaining terms are bounded by $\varepsilon^{1-d/4}$, $\varepsilon^{2-d/4}$, $\varepsilon^{2-d/2}$ and $\sqrt{n}\varepsilon^4$, all tending to zero.

The commutator

$$q_n(x) = g_0(x) - v(x)g(x) = (\mu_n - \mu)_y k_\varepsilon(x-y)[1 - v(x)a(y)]$$

has zero diagonal at $x = X_i$. Its coefficient is bounded by $C\,\mathrm{dist}(x,y)$, its conditional bias is $O(\varepsilon/n)$, and its conditional variance is $O(\varepsilon^{1-d/2}/n)$. Consequently $\|q_n\|_n = O_{\mathbb{P}}(\sqrt{\varepsilon}\,\zeta)$. Using $v/w - a = O(\varepsilon)$ and $R - \frac{1}{2}I = BR$ in (25) gives

$$u = S_n^* BRg + O_{L^2(\mu_n),\mathbb{P}}(\varepsilon\zeta + \eta\zeta + \sqrt{\varepsilon}\,\zeta).$$

The intertwining $B_n S_n^* = S_n^* \widehat{B}$, with $\widehat{B} = (I - \widehat{C})/2$, is exact. Both half-filters have norm at most $1/2$. Telescoping their powers contributes at most $Cj2^{-j}\eta\,\|g\|_{\mathcal{H}} = O_{\mathbb{P}}(\eta\zeta)$ uniformly in j. The commutator retains its factor 2^{-j}. This proves (29) with common random bounds. In $d = 2$ its fixed-j root-n commutator term is not negligible.

The exact score difference is $2\nabla\log(P_n(\exp e_j))$. The derivative row is $S_n^*\widehat{\theta}_x$, of norm at most $O_{\mathbb{P}}(\varepsilon^{-d/4-1/2})$. Expanding the exponential and the logarithm bounds its nonlinear score remainder by $C\varepsilon^{-d/4-1/2}\delta_n\|e_j\|_n$. After multiplication by A_n, this is bounded by the root-n version of (32). The linear proxy is $2\left\langle \widehat{\theta}_x, \widehat{C}B^j Rg\right\rangle$ up to the trajectory error. Replacing sections and operators costs $O_{\mathbb{P}}(\varepsilon^{-d/4-1/2}\eta\zeta)$, uniformly in j, x. This proves (30). At fixed x, the covariance identity for g gives $\mathbb{E}[A_n^2\|V_j(x)\|^2] \le C4^{-j}$. Summation over $j \ge J_n$ controls the supremum of the proxies. The common error tends to zero on this tail, since $\varepsilon^{1/2-d/4} \le 1$ for $d \le 2$. This proves (31), which is pointwise in x, not a spatial supremum law. □

Lemma 4.3 (Residual comparison)**.** *Put $Q_j = \frac{1}{4}\langle g, CB^{2j}g\rangle$, $q_j = \sqrt{Q_j}$ and $D_n = (n\varepsilon^{3d/2})^{-1/4}$. With common $E_n = o_{\mathbb{P}}(1)$ and $M_n = O_{\mathbb{P}}(1)$, simultaneously for $j \ge 2$,*

$$\sqrt{n}\,|r_j - q_j| \le E_n + M_n\{\varepsilon^{1/2-d/4}2^{-j} + D_n\varepsilon^{-d/8}2^{-j}\}. \tag{33}$$

Proof. The exact mass field is $F_j = \exp\{(e_j + \log P_n(\exp e_j))/2\}$. Its linear part is $(I + P_n)e_j/2$, with remainder bounded in sample L^2 by $C\delta_n\|e_j\|_n$. Dividing by S_j centers this field, with the same remainder order. Intertwining and $(I + C)R = I$ reduce the linear term to $S_n^* B^j g/2$, with the common trajectory error and $O_{\mathbb{P}}(\eta\zeta)$. Its mean obeys the exact identity

$$\langle 1, S_n^* B^j g\rangle_n = \langle \widehat{v} - v, B^j g\rangle = O_{\mathbb{P}}(\zeta^2 2^{-j}),$$

because $S_n 1 = \widehat{v}$ and $Bv = 0$. It is root-n negligible since $\sqrt{n}\zeta^2 = \eta \to 0$. Finally

$$\left|\|S_n^* h\|_n - \|S^* h\|_{L^2(\mu)}\right| \le \left\|\widehat{C} - C\right\|_{\mathrm{op}}^{1/2}\|h\|_{\mathcal{H}}.$$

Apply this to $B^j g$ and use $\sqrt{n}\sqrt{\eta}\,\zeta = D_n\varepsilon^{-d/8}$. This proves (33). □

5 Spectral concentration and the first hit

The proxy Q_j is defined using the deterministic population operator, not a random spectral decomposition of the sample matrix.

Lemma 5.1 (Hilbert quadratic concentration)**.** *Let λ_l be the eigenvalues of C and $\beta_l = (1-\lambda_l)/2$. For $j \geq 1$, put $T_j = \operatorname{tr}(C^2 B^{2j})$. Then*

$$\mathbb{E}Q_j = \frac{T_j}{4n}, \qquad \frac{\operatorname{Var} Q_j}{(\mathbb{E}Q_j)^2} \leq C\frac{\max_l \lambda_l^2 \beta_l^{2j}}{T_j} + \frac{C\varepsilon^{-d/2}4^{-j}}{nT_j}. \tag{34}$$

The same calculation applies to deterministic spectral projections. On a band with $\beta \in [\beta_H, \beta_{H,2}] \subset (0, 1/2)$ of fixed positive width, its relative variance is bounded by

$$C(\varepsilon^{d/2} + n^{-1})(\beta_{H,2}/\beta_H)^{2j}. \tag{35}$$

Proof. For centered iid Hilbert vectors Z_i of covariance Σ, expansion of their squared sample-mean norm gives

$$\operatorname{Var}\left\|n^{-1}\textstyle\sum_i Z_i\right\|^2 = \frac{\mathbb{E}\,\|Z\|^4 - (\operatorname{tr}\Sigma)^2}{n^3} + \frac{2(n-1)\operatorname{tr}(\Sigma^2)}{n^3}. \tag{36}$$

Only equal or reversed off-diagonal index pairs contribute. Terms with a singly occurring independent centered variable vanish by conditioning; the diagonal term in (36) is retained. Take $Z_{j,y} = \frac{1}{2}C^{1/2}B^j[a(y)k_y - v]$. The stationary rank-one term vanishes for $j \geq 1$, so its covariance is $C^2B^{2j}/4$. The feature bound gives $\sup_y \left\|Z_{j,y}\right\|^2 \leq C\varepsilon^{-d/2}4^{-j}$. Bound its fourth moment by this envelope times its second moment, and use $\operatorname{tr}(\Sigma^2) \leq \|\Sigma\|_{\mathrm{op}} \operatorname{tr}\Sigma$. This proves (34).

For a projected band, Lemma 3.2 gives $T_{j,H} \geq c\varepsilon^{-d/2}\beta_H^{2j}$, with its λ values bounded away from zero. The projected individual vector has the sharper envelope $\left\|Z_{j,y,H}\right\|^2 \leq C\varepsilon^{-d/2}\beta_{H,2}^{2j}$. In (36), this bounds the diagonal relative term by $C(\beta_{H,2}/\beta_H)^{2j}/n$, and the covariance term by $C\varepsilon^{d/2}(\beta_{H,2}/\beta_H)^{2j}$. This proves (35). □

Proposition 5.2 (Stopping count)**.** *For fixed $0 < \gamma < d/4$, the rule in (5) satisfies (7) and hits its threshold strictly before L_n with probability tending to one. In particular $\tau_\gamma \to \infty$.*

Proof. Put $u_* = d/4 - \gamma > 0$. Fix a small slack $0 < \delta < \min(u_*, d/8)/2$, write $u_- = u_* - \delta$, and let

$$j_- = \left\lfloor \frac{u_- \log(1/\varepsilon)}{\log 2} \right\rfloor, \qquad j_+ = \left\lceil \frac{(u_* + \delta)\log(1/\varepsilon)}{\log 2} \right\rceil.$$

Choose fixed $\beta_* < 1/2$ close enough to $1/2$ that, with $\sigma = |\log\beta_*|\,/\log 2 > 1$,

$$u_-\sigma < u_*, \qquad u_-(\sigma - 1) < d/8. \tag{37}$$

A fixed band with $\beta \geq \beta_*$ and λ bounded away from zero has $c\varepsilon^{-d/2}$ eigenvalues, hence $T_j \geq c\varepsilon^{-d/2}\beta_*^{2j}$. For every $2 \leq j \leq j_-$, (34) is bounded by

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

The sum of these geometric bounds is $o(1)$: its largest power is at most $C\varepsilon^{-2u_-(\sigma-1)}$, and (37) and $n\varepsilon^{3d/2} \to \infty$ control both terms. A Chebyshev union bound therefore gives, with probability tending to one, simultaneously

$$\sqrt{n}\, q_j \geq c\varepsilon^{-d/4}\beta_*^j, \qquad 2 \leq j \leq j_-. \tag{38}$$

The smallest lower bound, at j_-, dominates $\varepsilon^{-\gamma}$, since $u_-\sigma < u_*$. The empirical-norm term in (33), divided by (38), is bounded in probability by

$$CD_n\varepsilon^{d/8}(2\beta_*)^{-j_-} \leq CD_n\varepsilon^{d/8 - u_-(\sigma-1)} \to 0.$$

The other root-n errors are $o_{\mathbb{P}}(1)$ or $O_{\mathbb{P}}(1)$ for $d \le 2$, whereas the smallest proxy lower bound diverges. Thus no actual threshold hit occurs for $2 \le j \le j_-$, and $r_2/t_n \to \infty$.

For the upper bracket use $T_j \le 4^{-j}\operatorname{tr}(C^2) \le C\varepsilon^{-d/2}4^{-j}$. At j_+, $\mathbb{E}[nQ_{j_+}]/\varepsilon^{-2\gamma} \le C\varepsilon^{2\delta} \to 0$, so Markov's inequality gives $q_{j_+}/t_n \to 0$. The empirical-norm error divided by t_n is $O_{\mathbb{P}}(D_n\varepsilon^{u_*+\delta+\gamma-d/8}) = O_{\mathbb{P}}(D_n\varepsilon^{d/8+\delta}) = o_{\mathbb{P}}(1)$. The other terms also vanish relative to t_n, proving a hit by j_+. Moreover $j_+ < L_n$ eventually: the sampling condition in (2) bounds $\log(1/\varepsilon)$ above by $2\log n/(3d)$ for large n.

It remains to exclude $j = 0, 1$ without imposing a delay on the rule. For small log oscillation, replace e by $e - \mu_n e$, which leaves the residual unchanged. Its linear centered mass field is $(I + P_n)e/2$. On the mean-zero subspace its norm lies between $\|e\|_n/2$ and $\|e\|_n$. The nonlinear error is at most $C\|e\|_\infty\|e\|_n$. Consequently there are fixed local constants such that $r(e)$ is comparable to $\|e\|_n$ after centering, and

$$r(\Psi_n(e)) \le Cr(e). \tag{39}$$

Indeed Ψ_n kills constants, centering is an L^2 contraction, and (24) controls its centered output. No nonlinear mean-preservation or global residual monotonicity is assumed. Since $\|e_1\|_\infty = o_{\mathbb{P}}(1)$ and $r_2/t_n \to \infty$, (39) implies $r_1/t_n \to \infty$ and excludes $j = 1$.

If the fixed f is nonconstant, uniform consistency $K_n 1 \to f$ and the law of large numbers give

$$r_0 \longrightarrow \frac{\{\mathrm{Var}_\mu(\sqrt{f})\}^{1/2}}{\mu\sqrt{f}} > 0.$$

As $t_n \to 0$, this excludes $j = 0$. If $f = 1$, Lemma 3.3 gives $\|e_0\|_\infty = o_{\mathbb{P}}(1)$, so (39) gives $r_0/t_n \to \infty$ and excludes $j = 0$ too. Thus the unrestricted first hit lies between j_- and j_+ with probability tending to one. Letting the arbitrary slack δ tend to zero proves (7). □

Proposition 5.3 (Attained residual)**.** *For each fixed* $0 < c < 1/2$*,* (8) *holds.*

Proof. For each realized g, Q_j is a positive spectral sum. Its log-convexity implies that q_{j+1}/q_j is nondecreasing and at most $1/2$ when defined. Let $J = j_-$ for the slack above. Fix $0 < \beta_L < 1/2$, and choose $\beta_L < \beta_H < \beta_{H,2} < 1/2$ with sufficiently small positive width that

$$2u_- \log(\beta_{H,2}/\beta_H)/\log 2 < d/2.$$

The low-β trace at J is at most $C\varepsilon^{-d/2}\beta_L^{2J}$, whereas the high band has trace at least $c\varepsilon^{-d/2}\beta_H^{2J}$. The latter concentrates by (35), including its diagonal term. The ratio of the two expectations tends to zero. Markov's inequality for the low part and concentration for the high part show that modes $\beta \le \beta_L$ contribute $o_{\mathbb{P}}(Q_J)$. Thus $q_{J+1}/q_J \ge \beta_L - o_{\mathbb{P}}(1)$. Letting β_L approach $1/2$ and using monotonicity gives

$$\sup_{j \ge J} \left| q_{j+1}/q_j - 1/2 \right| \to 0 \quad \text{in probability.} \tag{40}$$

These ratios are defined on an event of probability tending to one, since the retained high band has positive energy there.

Uniformly for $j \ge J$, (33) divided by t_n is $o_{\mathbb{P}}(1)$. Its empirical-norm term is bounded by $O_{\mathbb{P}}(D_n\varepsilon^{u_-+\gamma-d/8}) = O_{\mathbb{P}}(D_n\varepsilon^{d/8-\delta})$, and the commutator term is $O_{\mathbb{P}}(\varepsilon^{1/2-\delta})$. At the first hit, $r_{\tau_\gamma - 1} > t_n$, while (40) gives

$$r_{\tau_\gamma} \ge (1/2 - o_{\mathbb{P}}(1))t_n - o_{\mathbb{P}}(t_n).$$

The definition supplies $r_{\tau_\gamma} \le t_n$, proving the claim. Discrete crossing may retain a rounding phase in r_{τ_γ}/t_n. □

6 Score limits and pointwise inference

Proposition 6.1 (Exact and stopped scores)**.** *Under* (2)*, at each fixed* x *the exact score satisfies*

$$A_n\{\widehat{s}(x) - \nabla \log f(x)\} \Rightarrow N_d(0, [4J_d/f(x)]I_d).$$

The same limit holds at the stopped iterate in Theorem 2.1.

Proof. The reciprocal-branch expansion of Lemma 3.3 gives

$$\widehat{s}(x) - s_{\mathrm{pop}}(x) = 2\,\langle \theta_x, Rg\rangle + o_{\mathbb{P}}(A_n^{-1}), \qquad s_{\mathrm{pop}} = -2\nabla \log a. \tag{41}$$

Indeed $\widehat{s} = 2\nabla \log \widehat{v}$ and the linear derivative is $2\nabla[a(\widehat{v} - v)]$. The logarithmic nonlinear remainder is bounded by $O_{\mathbb{P}}(\eta\varepsilon^{-d/4-1/2}\zeta)$, and replacing $\widehat{v} - v$ by Rg has the same derivative bound. After multiplication by A_n both are $O_{\mathbb{P}}(\sqrt{n}\eta\zeta) = o_{\mathbb{P}}(1)$.

The leading term in (41) is a centered iid triangular array. Its normalized covariance is

$$4\varepsilon^{d/2+1}\left\langle \theta_{x,\nu}, R(C - v\otimes v)R\theta_{x,\ell}\right\rangle.$$

The rank-one subtraction is zero, because $Rv = v/2$ and $\langle \theta_x, v\rangle = \nabla[a(x)v(x)] = 0$. Lemma 3.2 gives the limiting covariance $(4J_d/f(x))I_d$. A normalized summand is bounded in every fixed direction by

$$\frac{C\varepsilon^{d/4+1/2}}{\sqrt{n}}\,\|\theta_x\|_{\mathcal{H}}\sup_y \left\|a(y)k_y - v\right\|_{\mathcal{H}} \le \frac{C}{\sqrt{n}\varepsilon^{d/4}} \to 0.$$

The multivariate Lindeberg theorem therefore applies. Finally the population bias is $O(\varepsilon)$ and $A_n\varepsilon = \sqrt{n\varepsilon^{d/2+3}} \to 0$, proving the exact-score limit.

Proposition 5.2 places the stopping index beyond a deterministic sequence tending to infinity with probability tending to one. Lemma 4.2 then gives (9) at that same-sample random index. Slutsky's theorem gives the stopped limit. □

Proof of Theorem 2.1. Propositions 5.2, 5.3 and 6.1 prove the count, attained-residual and score assertions. The joint event in (8) follows by intersection of their events of probability tending to one. Its lower bound also implies $\sqrt{n}\,r_{\tau_\gamma} \to \infty$. □

Corollary 6.2 (Feasible pointwise inference)**.** *Let $\widetilde{f}(x) = K_n 1(x)$ and let x be fixed. Then*

$$\frac{A_n\sqrt{\widetilde{f}(x)}}{2\sqrt{J_d}}\{s_{\tau_\gamma}(x) - \nabla \log f(x)\} \Rightarrow N_d(0, I_d), \tag{42}$$

$$\frac{A_n^2\widetilde{f}(x)}{4J_d}\left\|s_{\tau_\gamma}(x) - \nabla \log f(x)\right\|^2 \Rightarrow \chi^2_d. \tag{43}$$

Consequently the ball

$$\left\{u \in \mathbb{R}^d : \left\|s_{\tau_\gamma}(x) - u\right\| \le \frac{2\sqrt{J_d\chi^2_{d,1-\alpha}}}{A_n\sqrt{\widetilde{f}(x)}}\right\}$$

has limiting coverage $1-\alpha$ for the fixed true score, for $0 < \alpha < 1$. Predetermined directional intervals follow from (42)*.*

Proof. The periodic KDE is positive, its bias tends to zero, and its variance at fixed x is $O((n\varepsilon^{d/2})^{-1})$. Hence $\widetilde{f}(x) \to f(x) > 0$ in probability. Slutsky's theorem uses no independence between the plug-in and the stopped score. The chi-square law and continuity at its quantile prove coverage. □

For $d = 2$ the constant has the closed form $J_2 = (\pi^2/12 - \log 2)/(2\pi)$: polar coordinates and $u = 2\pi^2 r^2$ reduce it to $(2\pi)^{-1}\int_0^\infty u(1+e^u)^{-2}\,du$. Integration by parts, followed by the geometric-series integral for $\int_0^\infty u/(1+e^u)\,du$, gives $\pi^2/12 - \log 2$.

7 Upper tolerance exponent

The coefficient of the threshold in (5) is one. The following statement concerns that convention and a counterexample in the stated density class.

Theorem 7.1. *Assume* (2) *and* $f = 1$. *If* $\gamma \geq d/4$, *then* $\mathbb{P}(\tau_\gamma = 0) \to 1$. *At each fixed query,*

$$A_n s_{\tau_\gamma}(x) \Rightarrow N_d(0, W_d), \qquad W_d = 2(4\pi)^{-d/2} I_d,$$

whereas $A_n \widehat{s}(x) \Rightarrow N_d(0, 4J_d I_d)$ *and* $W_d - 4J_d I_d$ *is positive definite.*

Proof. For uniform sampling, $a = v = 1$ and C is the heat covariance operator. Write $h = K_n 1 - 1 = g_0$. The actual initial mass field is $F_0 = \sqrt{1+h}$. Its linear part is $h/2$. We have $\|h\|_\infty = O_{\mathbb{P}}(\eta)$ and $\|h\|_{\mathcal{H}} = O_{\mathbb{P}}(\zeta)$, so its nonlinear L^2 error is $O_{\mathbb{P}}(\eta\zeta)$. The empirical-to-population norm change is $O_{\mathbb{P}}(\sqrt{\eta}\,\zeta)$, by the sample heat covariance comparison. The empirical mean is exactly $\mu_n h = \|h\|^2_{\mathcal{H}}$, because $\langle h, 1\rangle = 0$. It is $O_{\mathbb{P}}(\zeta^2) = o_{\mathbb{P}}(\zeta)$. Thus the actual centered and normalized residual satisfies

$$r_0 = \tfrac{1}{2}\,\|h\|_{L^2(dx)} + o_{\mathbb{P}}(\zeta).$$

The squared leading norm has mean $(4n)^{-1}\sum_{k\neq 0}\lambda_k^2$. Its relative variance is $O(\varepsilon^{d/2})$: in (36) the individual squared feature norm is constant by translation invariance, so its diagonal variance vanishes, and $\sum \lambda_k^4/(\sum_{k\neq 0}\lambda_k^2)^2 = O(\varepsilon^{d/2})$. Gaussian lattice sums consequently yield

$$\frac{r_0}{n^{-1/2}\varepsilon^{-d/4}} \longrightarrow c_d = \tfrac{1}{2}(4\pi)^{-d/4} < 1 \quad \text{in probability.} \tag{44}$$

For $\gamma \geq d/4$ the threshold is at least $n^{-1/2}\varepsilon^{-d/4}$, so the actual rule stops at zero with probability tending to one.

Here $s_0 = 2\nabla\log(1+h)$. At a fixed query, $A_n\nabla h$ is tight by its kernel second moment, and $h(x) \to 0$. Its kernel triangular-array CLT has covariance

$$W_d = 4\int_{\mathbb{R}^d} (2\pi\xi)(2\pi\xi)^T e^{-4\pi^2|\xi|^2}\,d\xi = 2(4\pi)^{-d/2} I_d.$$

The Lindeberg bound is again $C/(\sqrt{n}\varepsilon^{d/4}) \to 0$; the logarithmic denominator is negligible by Slutsky. The exact-score limit follows from Proposition 6.1. Writing $\Lambda(\xi) = e^{-2\pi^2|\xi|^2}$, its covariance is

$$4J_d I_d = 4\int (2\pi\xi)(2\pi\xi)^T [\Lambda(\xi)/(1+\Lambda(\xi))]^2\,d\xi.$$

For every nonzero direction, the difference from W_d is the integral of a squared nonzero linear projection against a strictly positive radial weight. It is positive, proving the claim. □

It establishes the upper exponent obstruction for a uniform guarantee over the model class with this rule.

8 Scope of the inference statements

The confidence regions above are pointwise-in-f asymptotic regions at predetermined spatial queries. They do not provide uniform finite-sample coverage over the unrestricted positive C^6 class. A standard discontinuous-functional obstruction [9] makes that distinction necessary. For example, at a query x_0 and each fixed real q,

$$f_{m,q}(x) = 1 + \frac{q}{2\pi m}\sin(2\pi m(x_1 - x_{0,1}))$$

has true score qe_1 at x_0, is smooth and lies between $1/2$ and $3/2$ for large integer m, while its n-sample total variation distance from uniform data is at most $n|q|/(2\pi^2 m)$. For each fixed n this tends to zero.

Any uniformly honest confidence interval would therefore have to cover every fixed q under uniform data with its nominal probability; an almost surely bounded interval cannot do so. Quantitative smoothness information or a specified smoothed target is needed for such finite-sample claims. This does not contradict Theorem 2.1, which fixes the density before taking its limit.

The pointwise result permits confidence regions for a density's logarithmic gradient and tests of a zero gradient at a prespecified point. It does not locate an unknown mode or justify data-selected queries. Uniform spatial limits, boundaries, vanishing densities and diffusion-path error control require additional arguments. At $d = 3$ the two conditions on $n\varepsilon^{9/2}$ in (2) are incompatible; higher dimensions are not covered. The leading update law is an asymptotic iteration count, and finite-sample runtime improvements are not inferred from it.

AI assistance

AI tools assisted preparation and checking. The author is responsible for the content.

References

[1] 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`.

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

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

[4] 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.

[5] 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.

[6] 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`.

[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] S. Schubr. Residual certificates and empirical sharpness for entropic self-transport scores. Luminary Research Archive, edition 1, 18 October 2026. Permanent archive.

[9] J.-M. Dufour. Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica* **65**(6) (1997), 1365–1388. doi:10.2307/2171740.