Residual Certificates and Empirical Sharpness for Entropic Self-Transport Scores

Scott Schubr
Luminary Oddities

October 18, 2026

Abstract

We study how numerical errors in symmetric kernel scaling affect the spatial score obtained from its smooth kernel extension. A full square-root-mass residual gives a sharp kernel-energy bound for every positive candidate scaling, including singular positive semidefinite Gram matrices. For Gaussian kernels this yields computable, scale-invariant pointwise score certificates. For iid samples from a strictly positive C^3 density f on $\mathbb{T}^d$, we prove a simultaneous nonlinear score bound and determine the exact local modulus in a shrinking-bandwidth regime. Its leading constant is $C_d/\sqrt{f(x)}$. When $\sqrt{n}\varepsilon^{d/4+1/2} \to \infty$, residual $o(n^{-1/2})$ suffices for negligible score error at that normalization; positive sample-dependent candidates with residual asymptotic to $c/\sqrt{n}$ attain a nonzero normalized error. The sampling condition $n\varepsilon^d/\log n \to \infty$ is sufficient. A reproducible solver experiment illustrates the computable certificate on actual iterates. The results concern numerical accuracy.

Keywords: entropic self-transport; matrix scaling; residual error bounds; score estimation; empirical sharpness.

1 Introduction

Symmetric entropic self-transport determines a positive scaling of a kernel matrix. Its smooth extension also determines a spatial score. In computation, the scaling equation is solved only approximately, so a small normalization residual must be translated into an error in that score. The translation depends on bandwidth and on local kernel mass, and scalar rescaling of a candidate does not affect its score. These features motivate a residual that respects this invariance.

We first prove a deterministic energy inequality using the full vector of square-root row masses. It gives pointwise Gaussian score certificates that use only computed quantities and no lower bound on the unknown exact scaling. We then consider the periodic Gaussian model with a nonuniform sampling density. The empirical theorem controls all positive candidate scalings on one event of high probability, computes the sharp local amplification, and constructs positive numerical perturbations attaining the limiting residual order. The finite certificate and the asymptotic local modulus serve different purposes: the former is computable before convergence, while the latter identifies the limiting sensitivity.

1.1 Relation to existing work

Kernel-energy geometry for self-transport is established in the theory of Sinkhorn divergences [1]. Bistochastic normalization and its nonuniform-density geometry are studied by Marshall and Coifman [7], and Wormell and Reich [4] develop empirical normalization, resolvent and spectral perturbation estimates. Cheng and Landa [2] relate normalization residuals to scaling errors under factor bounds and choose approximate scaling tolerances from statistical errors. Thus residual-based termination and statistical error budgeting are existing ideas.

Mordant [3] studies decreasing-regularization self-transport scores, including their linearization and statistical fluctuations. The inverse filter and inverse-density scale used below have precedents in

that work. Maeda, Yao and Nitanda [5] combine sampling and finite-iteration errors for two-sample Schrödinger bridge estimation, including normalized exponential drift extensions. Their bounds concern bridge and drift errors along specified iterations. Here the deterministic input is a measured square-root-mass residual, and the output is a pointwise spatial-score certificate for arbitrary positive candidates. The empirical theorem adds a simultaneous nonlinear upper bound and positive sample-adaptive witnesses for its residual order.

Gradient convergence along Sinkhorn iterations is studied by Greco et al. [6]. Belomestny [8] and González-Sanz, Nutz and Stromme [9] give neighboring statistical stability and finite-sample potential bounds. The contribution here is the specific residual-to-score guarantee and its empirical sharpness statement, using these established sensitivity mechanisms. The periodic model is stated explicitly throughout; limit theorems for other costs or domains require their own hypotheses.

2 A finite residual certificate

Let $p_i > 0$, $\sum_{i=1}^N p_i = 1$, and let K be a symmetric positive semidefinite matrix with nonnegative entries and $K_{ii} > 0$. Write

$$(K_p z)_i = \sum_j p_j K_{ij} z_j, \qquad \langle z, w\rangle_p = \sum_i p_i z_i w_i, \qquad \|z\|_K^2 = \left\langle z, K_p z\right\rangle_p .$$

The last expression can be a seminorm on coefficient vectors. For a Gram matrix it equals the squared RKHS norm of $\sum_i p_i z_i k(\cdot, x_i)$. The exact self-scaling a solves

$$a_i (K_p a)_i = 1, \qquad a_i > 0. \tag{1}$$

It exists uniquely: the function

$$F(z) = \tfrac{1}{2}\,\|z\|_K^2 - \sum_i p_i \log z_i$$

has positive-definite Hessian $\operatorname{diag}(p) K \operatorname{diag}(p) + \operatorname{diag}(p_i / z_i^2)$. Its sublevel sets stay away from zero and infinity because the quadratic term bounds $\sum_i p_i^2 K_{ii} z_i^2$ from below and dominates the logarithms. Its unique minimizer satisfies (1).

For a positive computed vector b, define

$$m_i = b_i (K_p b)_i, \quad M = \sum_i p_i m_i, \quad S = \sum_i p_i \sqrt{m_i}, \quad \rho^2 = \sum_i p_i (\sqrt{m_i} - S)^2 = M - S^2, \quad r = \rho / S. \tag{2}$$

Multiplying b by a positive scalar multiplies S and ρ by that scalar and leaves r unchanged.

Theorem 2.1 (Full-residual energy bound)**.** *For every $\lambda > 0$,*

$$\|a - \lambda b\|_K^2 \le \sum_i p_i (\lambda \sqrt{m_i} - 1)^2. \tag{3}$$

The coefficient one and the dependence on the full residual vector are sharp over this matrix class. In particular,

$$\|a - b/S\|_K \le r. \tag{4}$$

Proof. Put $t_i = a_i / b_i$ and $c = \left\langle a, K_p b\right\rangle_p$. Symmetry and (1) give

$$\|a\|_K^2 = 1, \quad \|b\|_K^2 = M, \quad c = \sum_i p_i / t_i = \sum_i p_i t_i m_i.$$

Weighted Cauchy–Schwarz gives $S \le \sqrt{(\sum_i p_i t_i m_i)(\sum_i p_i / t_i)} = c$. Consequently

$$\|a - \lambda b\|_K^2 = 1 + \lambda^2 M - 2\lambda c \le 1 + \lambda^2 M - 2\lambda S,$$

which is (3). Its gap is exactly

$$2\lambda(c - S) = \lambda \sum_i p_i \left(\sqrt{t_i m_i} - \frac{1}{\sqrt{t_i}}\right)^2.$$

Given arbitrary positive masses m_i, choose positive diagonal K and $a_i = (p_i K_{ii})^{-1/2}$, $b_i = \sqrt{m_i}\, a_i$. Then $c = S$ and equality holds for every λ. Strictly positive matrices $K + \eta \mathbf{1}\mathbf{1}^\top$ approach this equality as $\eta \downarrow 0$; continuity follows from the invertible Hessian of F. Finally, substitute $\lambda = 1/S$ in (3) to obtain (4). □

Corollary 2.2. *The best squared bound in* (3) *is* $1 - S^2/M$*, attained at* $\lambda = S/M$*. Relative to the computed function, the bound divided by* λ^2 *is minimized at* $\lambda = 1/S$*. If* $\delta = \max_i |m_i - 1| < 1$*, then*

$$\|a - b\|_K \le 1 - \sqrt{1 - \delta}. \tag{5}$$

Proof. Minimize the two scalar quadratics in λ and $1/\lambda$, respectively. For (5), set $\lambda = 1$ and use $\left|\sqrt{m_i} - 1\right| \le 1 - \sqrt{1 - \delta}$. □

2.1 Computable Gaussian score bounds

For sites $x_i \in \mathbb{R}^d$ take $k_\varepsilon(x, y) = \exp(-\|x - y\|^2/(2\varepsilon))$, and for positive z define

$$v_z(x) = \sum_i p_i z_i k_\varepsilon(x, x_i), \quad g_z(x) = \nabla \log v_z(x), \quad s_z(x) = 2 g_z(x). \tag{6}$$

These are the smooth kernel extensions.

Theorem 2.3 (Pointwise score certificate)**.** *At any point where* $v = v_b(x) > \rho$*, put* $G = \|g_b(x)\|$*. Then*

$$\|s_a(x) - s_b(x)\| \le \frac{2\rho\sqrt{\varepsilon^{-1} + G^2}}{v - \rho}, \tag{7}$$

$$\|s_a(x) - s_b(x)\| \le \frac{2\{\rho^2 G + \rho\sqrt{v^2 G^2 + (v^2 - \rho^2)/\varepsilon}\}}{v^2 - \rho^2}. \tag{8}$$

Both expressions are unchanged by positive rescaling of b*. They vanish when* $\rho = 0$*. If* $v \le \rho$*, these certificates are inconclusive.*

Proof. In the Gaussian RKHS $\mathcal{H}_\varepsilon$, set $h = v_a - v_b/S$. Theorem 2.1 gives $\|h\|_{\mathcal{H}_\varepsilon} \le r$. The value section has squared norm one; its derivative sections are orthogonal to it and have Gram matrix $\varepsilon^{-1} I_d$. Bessel's inequality therefore gives

$$h(x)^2 + \varepsilon \|\nabla h(x)\|^2 \le r^2. \tag{9}$$

Let $V = v/S$ and $g = g_b(x)$. The logarithmic-gradient identity is

$$g_a - g_b = \frac{\nabla h - g h}{V + h}, \qquad V + h \ge V - r > 0.$$

The joint evaluation bound implies $\|\nabla h - g h\| \le r\sqrt{\varepsilon^{-1} + G^2}$, proving (7). To optimize within (9), put $u = |h(x)| \in [0, r]$. The ratio is at most

$$\frac{\sqrt{r^2 - u^2}/\sqrt{\varepsilon} + G u}{V - u}.$$

Its maximum R solves $RV = r\sqrt{(R + G)^2 + \varepsilon^{-1}}$: after rearrangement the maximum of $(R + G)u + \sqrt{r^2 - u^2}/\sqrt{\varepsilon}$ is that right-hand side by Cauchy–Schwarz, attained at a point of $[0, r]$. Solving the quadratic,

$$R = \frac{r^2 G + r\sqrt{V^2 G^2 + (V^2 - r^2)/\varepsilon}}{V^2 - r^2}.$$

Twice this expression gives (8). The case $r = 0$ follows directly. The optimization is sharp for the relaxed RKHS ball. □

Proposition 2.4 (Periodic evaluation constants)**.** *Suppose the kernel is the integral-one periodic Gaussian on $\mathbb{T}^d$ (normalization is with respect to Lebesgue measure, not its diagonal), with Fourier multipliers $\lambda_k = e^{-2\pi^2\varepsilon\|k\|^2}$. Put*

$$K_{0,\varepsilon} = \sum_{k\in\mathbb{Z}^d} \lambda_k, \qquad K_{1,\varepsilon} = \sum_{k\in\mathbb{Z}^d} (2\pi k_1)^2 \lambda_k.$$

If $v_b(x) > \rho\sqrt{K_{0,\varepsilon}}$, then

$$\|s_a(x) - s_b(x)\| \le \frac{2\rho\sqrt{K_{1,\varepsilon} + K_{0,\varepsilon}\,\|g_b(x)\|^2}}{v_b(x) - \rho\sqrt{K_{0,\varepsilon}}}. \tag{10}$$

Proof. The periodic kernel has value-section squared norm $K_{0,\varepsilon}$ and derivative-section Gram matrix $K_{1,\varepsilon} I_d$, orthogonal to the value section. Apply (4) and the joint evaluation argument in the preceding proof with these constants. □

2.2 Why local information enters the certificate

For two sites $-L/2, L/2$ in $\mathbb{R}$, equal weights, and fixed $\varepsilon > 0$, put $k = e^{-L^2/(2\varepsilon)}$, $a_1 = a_2 = \sqrt{2/(1+k)}$, and $b_1 = a_1(1+\eta)$, $b_2 = a_2(1-\eta)$ with $0 < \eta < 1$. The masses are

$$m_{1,2} = 1 \pm \frac{2\eta}{1+k} + \eta^2 \frac{1-k}{1+k},$$

so $\max_i |m_i - 1| \le 2\eta + \eta^2$ independently of L. At the midpoint, $s_a(0) = 0$ and $s_b(0) = -\eta L/\varepsilon$. Thus no finite bound depending only on bandwidth and a positive fixed row-residual tolerance controls all pointwise scores over all sample geometries. The local kernel mass at the midpoint tends to zero. This elementary example explains the local denominator in the certificate.

Corollary 2.5 (Conditional limit-law transfer)**.** *Let A_n be a deterministic nonnegative scalar sequence and let B_n be any valid score certificate above, with value infinity when its denominator condition fails. If an exact estimator for the same model satisfies $A_n(s_a - \theta_n) \Rightarrow Z$ and $A_n B_n \to 0$ in probability, then $A_n(s_b - \theta_n) \Rightarrow Z$. The computed candidate may depend on the data and on a stopping rule.*

Proof. The deterministic bound conditional on the data implies $A_n(s_b - s_a) \to 0$ in probability. Apply Slutsky's theorem. □

3 Sharp residual accuracy for nonuniform empirical self-transport

Fix a positive integer d. We work on the flat torus $\mathbb{T}^d = \mathbb{R}^d/\mathbb{Z}^d$, equipped with normalized Lebesgue measure. Let $f \in C^3(\mathbb{T}^d)$ be a probability density with $\min f > 0$, and put $\mu = f\,dx$. All constants in this section may depend on d, $\min f$, and $\|f\|_{C^3}$. Consider the integral-one periodic Gaussian kernel

$$k_\varepsilon(x,y) = \sum_{\ell\in\mathbb{Z}^d} (2\pi\varepsilon)^{-d/2} \exp\left(-\frac{\|x-y+\ell\|^2}{2\varepsilon}\right). \tag{11}$$

Its integral operator on Lebesgue measure is denoted by H_ε. The Fourier multipliers are $\lambda_k = \exp(-2\pi^2\varepsilon\,\|k\|^2)$, $k \in \mathbb{Z}^d$. Thus the kernel is positive semidefinite, strictly positive, and smooth. The associated entropic cost is $-\varepsilon \log k_\varepsilon$; on the whole torus it is not the squared geodesic cost.

Let $\mathcal{H}_\varepsilon$ be the real reproducing kernel Hilbert space of k_ε. Absolute convergence of its Fourier series and their derivatives gives the diagonal and derivative-section identities

$$K_{0,\varepsilon} = k_\varepsilon(x,x) = \sum_k \lambda_k \sim (2\pi)^{-d/2}\varepsilon^{-d/2}, \tag{12}$$

$$\left\langle \partial_{x_j} k_\varepsilon(\cdot,x), \partial_{x_l} k_\varepsilon(\cdot,x)\right\rangle_{\mathcal{H}_\varepsilon} = K_{1,\varepsilon}\delta_{jl}, \qquad K_{1,\varepsilon} = \sum_k (2\pi k_1)^2 \lambda_k \sim (2\pi)^{-d/2}\varepsilon^{-d/2-1}. \tag{13}$$

Value sections and derivative sections at the same point are orthogonal.

Let $X_1, \dots, X_n$ be independent with law μ, and write $\langle u, v\rangle_n = n^{-1}\sum_i u_i v_i$ and $\|u\|_n^2 = \langle u, u\rangle_n$. The exact empirical scale $\widehat{a}_i > 0$ satisfies

$$\widehat{a}_i \frac{1}{n}\sum_j k_\varepsilon(X_i, X_j)\widehat{a}_j = 1. \tag{14}$$

It exists uniquely for every sample, including samples with repeated sites. Indeed, on the positive orthant the function

$$z \longmapsto \frac{1}{2n^2}\sum_{i,j} z_i k_\varepsilon(X_i, X_j) z_j - \frac{1}{n}\sum_i \log z_i$$

has compact sublevel sets: its nonnegative entries and positive diagonal control escape to infinity, and its logarithmic barrier controls the boundary. Its Hessian is positive definite, since its quadratic term is positive semidefinite and the barrier has positive definite Hessian. Its unique critical point gives (14). Extend the solution by

$$\widehat{v}(x) = \frac{1}{n}\sum_i k_\varepsilon(x, X_i)\widehat{a}_i, \qquad \widehat{a}(x) = \widehat{v}(x)^{-1}. \tag{15}$$

This extension agrees with $\widehat{a}_i$ at X_i. For a positive candidate vector b, define

$$v_b(x) = \frac{1}{n}\sum_i k_\varepsilon(x, X_i) b_i, \qquad s_b(x) = 2\nabla \log v_b(x), \tag{16}$$

$$m_i(b) = b_i \frac{1}{n}\sum_j k_\varepsilon(X_i, X_j) b_j, \qquad S(b) = \frac{1}{n}\sum_i \sqrt{m_i(b)},$$

$$\rho(b) = \left[\frac{1}{n}\sum_i \left(\sqrt{m_i(b)} - S(b)\right)^2\right]^{1/2}, \qquad r(b) = \frac{\rho(b)}{S(b)}. \tag{17}$$

Both $r(b)$ and s_b are invariant under positive rescaling of b. In particular, replacing b by $b/S(b)$ makes $S = 1$ and $\rho = r$. We write $\widehat{s} = s_{\widehat{a}}$.

Set

$$\alpha = \frac{d}{4} + \frac{1}{2}, \qquad J_d = \int_{\mathbb{R}^d} \frac{(2\pi\xi_1)^2}{[1 + \exp(2\pi^2\|\xi\|^2)]^2}\, d\xi, \qquad C_d = 4\sqrt{J_d} > 0. \tag{18}$$

Throughout, $\varepsilon = \varepsilon_n$ is deterministic and

$$\varepsilon_n \longrightarrow 0, \qquad \frac{n\varepsilon_n^d}{\log n} \longrightarrow \infty. \tag{19}$$

For conclusions involving statistical normalization, we also assume

$$A_n = \sqrt{n}\,\varepsilon_n^\alpha \longrightarrow \infty. \tag{20}$$

Condition (19) is a conservative sufficient sampling condition.

Define the empirical Markov operator, its off-sample extension, and its derivative rows by

$$(P_n u)(x) = \frac{1}{n}\sum_i \widehat{a}(x)\widehat{a}_i k_\varepsilon(x, X_i) u_i, \tag{21}$$

$$r_{n,j,x}(i) = \partial_{x_j}[\widehat{a}(x)\widehat{a}_i k_\varepsilon(x, X_i)],$$
$$R_n = (I + P_n)^{-1}, \qquad G_n(x)_{jl} = \left\langle R_n r_{n,j,x}, R_n r_{n,l,x}\right\rangle_n. \tag{22}$$

Here P_n acts on vectors by evaluation at the sample sites.

Theorem 3.1. *Under* (19)*, the following statements hold.*

1. *There are constants $c_0, C_0 > 0$ and events E_n with $\mathbb{P}(E_n) \to 1$ such that, on E_n, every positive vector b with $r(b) \le c_0 \varepsilon_n^{d/4}$ satisfies*

$$\sup_{x \in \mathbb{T}^d} \|s_b(x) - \widehat{s}(x)\| \le C_0 r(b) \varepsilon_n^{-\alpha}. \tag{23}$$

Consequently, under (20)*, every measurable sample-dependent positive b_n with $\sqrt{n}\, r(b_n) \to 0$ in probability satisfies*

$$A_n \sup_x \left\| s_{b_n}(x) - \widehat{s}(x) \right\| \longrightarrow 0 \quad \text{in probability.} \tag{24}$$

2. *For every $n \ge 2$, the local modulus*

$$L_n(x) = \sup_{u \neq 0,\, \langle u, 1 \rangle_n = 0} \lim_{t \to 0} \frac{\left\| s_{\widehat{a}(1+tu)}(x) - \widehat{s}(x) \right\|}{r(\widehat{a}(1+tu))}$$

is exactly

$$L_n(x) = 4\sqrt{\lambda_{\max}(G_n(x))}. \tag{25}$$

Moreover,

$$\sup_x \left\| \varepsilon_n^{2\alpha} G_n(x) - \frac{J_d}{f(x)} I \right\|_{\mathrm{op}} \longrightarrow 0 \quad \text{in probability.} \tag{26}$$

In particular $\varepsilon_n^{\alpha} L_n(x) \to C_d / \sqrt{f(x)}$ uniformly in x in probability.

3. *Fix $x_0 \in \mathbb{T}^d$ and $c > 0$. Under* (20)*, there are measurable positive vectors b_n, depending on the sample and x_0, for which*

$$\sqrt{n}\, r(b_n) \longrightarrow c, \qquad A_n(s_{b_n}(x_0) - \widehat{s}(x_0)) \longrightarrow \frac{c C_d}{\sqrt{f(x_0)}} e_1 \quad \text{in probability.} \tag{27}$$

The local modulus depends on the exact empirical solution. The last assertion concerns the worst case over positive candidate vectors. We prove the theorem through estimates that keep the bandwidth dependence explicit.

3.1 Population scaling and its derivative Gram matrix

Lemma 3.2. *For sufficiently small ε, there is a unique positive continuous function a_ε satisfying*

$$a_\varepsilon(x) \int k_\varepsilon(x, y) a_\varepsilon(y)\, d\mu(y) = 1. \tag{28}$$

It is smooth for fixed ε, and

$$\left\| a_\varepsilon \sqrt{f} - 1 \right\|_{C^1} \le C\varepsilon. \tag{29}$$

Consequently a_ε, a_ε^{-1}, and ∇a_ε are uniformly bounded as $\varepsilon \to 0$.

Proof. Put $g = \sqrt{f}$ and $q = a_\varepsilon g$. The equation becomes $\mathcal{F}_\varepsilon(q) = q H_\varepsilon(qg) - g = 0$ in the Banach algebra $C^1(\mathbb{T}^d)$. We first establish a uniform bound for $(I + H_\varepsilon)^{-1}$ on this space. Its multiplier is $1 - F(\sqrt{\varepsilon} k)$, where

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

This is a Schwartz function. With the Fourier convention $\exp(2\pi i k \cdot x)$, the multiplier $F(\sqrt{\varepsilon} k)$ has convolution kernel

$$\varepsilon^{-d/2} \sum_{\ell \in \mathbb{Z}^d} \check{F}((x + \ell)/\sqrt{\varepsilon}).$$

Its torus L^1 norm is at most $\left\|\check{F}\right\|_{L^1(\mathbb{R}^d)}$. Convolution commutes with first derivatives, so the corresponding operator is uniformly bounded on C^1. The inverse identity holds first on trigonometric polynomials and then by their density in C^1. In particular, no convergence of H_ε to I in C^1 operator norm is needed.

At $q = 1$ the derivative is

$$\begin{aligned}\mathcal{L}_\varepsilon v &= (H_\varepsilon g)v + H_\varepsilon(gv) = g(I + H_\varepsilon)v + \mathcal{E}_\varepsilon v,\\ \mathcal{E}_\varepsilon v &= (H_\varepsilon g - g)v + [H_\varepsilon, g]v.\end{aligned}$$

Gaussian moments give $\|H_\varepsilon g - g\|_{C^1} \le C\varepsilon \|g\|_{C^3}$. For the commutator, integration of $|g(y) - g(x)|$ against the heat kernel gives its C^0 bound. Differentiating gives

$$\partial_j [H_\varepsilon, g]v = [H_\varepsilon, \partial_j g]v + [H_\varepsilon, g]\partial_j v.$$

Applying the same bound to these two terms yields

$$\|[H_\varepsilon, g]v\|_{C^1} \le C\sqrt{\varepsilon}\, \|g\|_{C^2} \|v\|_{C^1}.$$

The periodic Gaussian moments satisfy these bounds using torus distance. Thus $\|\mathcal{E}_\varepsilon\|_{C^1 \to C^1} = O(\sqrt{\varepsilon})$. Multiplication by g is boundedly invertible on C^1; the preceding inverse bound and a Neumann series show that $\mathcal{L}_\varepsilon^{-1}$ is uniformly bounded for small ε.

Writing $q = 1 + v$ gives

$$\mathcal{L}_\varepsilon v = -(H_\varepsilon g - g) - v H_\varepsilon(gv).$$

After applying $\mathcal{L}_\varepsilon^{-1}$, the right-hand side is a contraction on a C^1 ball of radius $C\varepsilon$: its constant term is $O(\varepsilon)$ and the quadratic term has Lipschitz constant $O(\varepsilon)$ there. The fixed point has $q > 0$ and proves (29). The identity $a_\varepsilon = 1/H_\varepsilon(a_\varepsilon f)$ implies smoothness for fixed ε.

To prove uniqueness, let a, b be positive continuous solutions and let K_μ denote integration against k_ε with measure μ. Positive semidefiniteness and symmetry imply

$$0 \le \left\langle a - b, K_\mu(a - b)\right\rangle_{L^2(\mu)} = \int (a - b)(1/a - 1/b)\, d\mu = -\int \frac{(a - b)^2}{ab}\, d\mu \le 0.$$

Thus $a = b$ almost everywhere and hence everywhere, by continuity and the full support of μ. □

Let P_ε be the operator on $L^2(\mu)$ with kernel $a_\varepsilon(x)a_\varepsilon(y)k_\varepsilon(x, y)$. It is self-adjoint, positive semidefinite, and Markov. Jensen's inequality and invariance of μ give $\|P_\varepsilon\| \le 1$, so its spectrum is contained in $[0, 1]$. Write

$$\begin{aligned}r_{j,x}(y) &= \partial_{x_j}[a_\varepsilon(x)a_\varepsilon(y)k_\varepsilon(x, y)],\\ G_\varepsilon(x)_{jl} &= \left\langle (I + P_\varepsilon)^{-1} r_{j,x}, (I + P_\varepsilon)^{-1} r_{l,x}\right\rangle_{L^2(\mu)}.\end{aligned}$$

The same Markov, symmetry, and positivity statements hold for P_n. Both its sample and off-sample rows sum to one. Differentiating the row sum shows that all derivative rows have mean zero.

For later use, consider either operator P with its exact scale a. For a bounded mean-zero direction u and small t, the candidate $b = a(1 + tu)$ satisfies

$$\begin{aligned}m(b) &= (1 + tu)(1 + tPu),\\ \sqrt{m(b)} &= 1 + \tfrac{t}{2}(I + P)u + O(t^2),\\ \frac{r(b)}{|t|} &\longrightarrow \tfrac{1}{2}\|(I + P)u\|_2.\end{aligned} \tag{30}$$

Here $S(b) = 1 + O(t^2)$ because the linear term has mean zero. The exact score identity is

$$s_b(x) - s_a(x) = \frac{2t\nabla(Pu)(x)}{1 + t(Pu)(x)}. \tag{31}$$

In the empirical setting every direction is bounded. Taking $w = (I + P_n)u$, the row representers of the linear map $w \mapsto \nabla(P_n(I + P_n)^{-1}w)(x)$ are $R_n r_{n,j,x}$. Their mean-zero property shows that restricting to mean-zero vectors loses no direction. Its operator norm is $\sqrt{\lambda_{\max}(G_n(x))}$; (30) and (31) prove (25). The identical argument applies in the population setting: smooth directions are dense and the row-resolvent representers are smooth for fixed ε.

Lemma 3.3. *Uniformly in $x \in \mathbb{T}^d$,*

$$\varepsilon^{2\alpha} G_\varepsilon(x) \longrightarrow \frac{J_d}{f(x)} I. \tag{32}$$

Proof. The unitary map $U : L^2(\mu) \to L^2(dx)$ given by $Uv = gv$, $g = \sqrt{f}$, conjugates P_ε to $Q_\varepsilon = qH_\varepsilon q$, where $q = a_\varepsilon g$. Lemma 3.2 gives

$$\|Q_\varepsilon - H_\varepsilon\|_{\mathrm{op}} \le \|q - 1\|_\infty (2 + \|q - 1\|_\infty) = O(\varepsilon).$$

Both operators are positive semidefinite; their I-plus inverses have norm at most one and differ by $O(\varepsilon)$ by the resolvent identity. The transformed derivative row is

$$U r_{j,x}(y) = a_\varepsilon(x) q(y) \partial_{x_j} k_\varepsilon(x, y) + (\partial_j a_\varepsilon(x)) q(y) k_\varepsilon(x, y).$$

Uniformly in x, its difference from $f(x)^{-1/2} \partial_{x_j} k_\varepsilon(x, \cdot)$ has $L^2(dx)$ norm $O(\sqrt{\varepsilon}\, \varepsilon^{-\alpha})$. Indeed the leading coefficient has error $O(\varepsilon)$, whereas $\|k_\varepsilon(x, \cdot)\|_2 = O(\varepsilon^{-d/4})$ and $\left\|\partial_{x_j} k_\varepsilon(x, \cdot)\right\|_2 = O(\varepsilon^{-\alpha})$. The resolvent comparison therefore gives the same relative row error after applying the inverse operators.

Fourier sign symmetry shows that the Gram matrix of $(I + H_\varepsilon)^{-1} \partial_{x_j} k_\varepsilon(x, \cdot)$ is $D_\varepsilon^2 I$, with

$$D_\varepsilon^2 = \sum_{k \in \mathbb{Z}^d} (2\pi k_1)^2 \left(\frac{\lambda_k}{1 + \lambda_k} \right)^2.$$

Rescaling the lattice by $\sqrt{\varepsilon}$ gives $\varepsilon^{2\alpha} D_\varepsilon^2 \to J_d$. The rescaled integrand in (18) is Schwartz, so the lattice Riemann sum converges. Combining this limit with the uniform row comparison proves the claim. □

3.2 Uniform comparison of the empirical and population scales

By Theorem 2.1, if a is an exact empirical scale, $b > 0$ has masses $m_i = m_i(b)$, and $\lambda > 0$, then

$$\|v_a - \lambda v_b\|_{\mathcal{H}_\varepsilon}^2 \le \frac{1}{n} \sum_i (\lambda \sqrt{m_i} - 1)^2. \tag{33}$$

Here $p_i = 1/n$, and the kernel-energy seminorm is the RKHS norm of the extended function. If $\delta = \max_i |m_i - 1| < 1$, it follows that

$$\|v_a - v_b\|_{\mathcal{H}_\varepsilon} \le 1 - \sqrt{1 - \delta} \le \delta. \tag{34}$$

If instead $S(b) = 1$, (33) gives $\|v_a - v_b\|_{\mathcal{H}_\varepsilon} \le r(b)$.

Lemma 3.4. *Put $\eta_n = \sqrt{\log n / (n\varepsilon_n^d)}$. Under* (19)*,*

$$\|\widehat{a} - a_\varepsilon\|_\infty + \sqrt{\varepsilon}\, \|\nabla \widehat{a} - \nabla a_\varepsilon\|_\infty = O_{\mathbb{P}}(\eta_n). \tag{35}$$

Proof. For the deterministic population scale define

$$Z_n(x) = \frac{1}{n} \sum_i k_\varepsilon(x, X_i) a_\varepsilon(X_i) - \int k_\varepsilon(x, y) a_\varepsilon(y)\, d\mu(y).$$

For each x, the summands for $Z_n(x)$ and for $\sqrt{\varepsilon}\, \partial_j Z_n(x)$ have envelope $O(\varepsilon^{-d/2})$ and second moment $O(\varepsilon^{-d/2})$, by Gaussian integration and the bounds on a_ε and f. Their spatial Lipschitz envelopes are

$O(\varepsilon^{-d/2-1/2})$. Put $h_n = n^{-3}\varepsilon^{d/2+1/2}$ and choose a deterministic regular torus grid whose coordinate spacing lies in $[h_n/2, h_n]$ for all sufficiently large n. Condition (19) implies $\log(1/\varepsilon) = O(\log n)$, so the logarithm of the grid cardinality is $O(\log n)$. Scalar Bernstein inequalities at the grid points, followed by a union bound over the grid and coordinates, and the Lipschitz bound between grid points yield

$$\|Z_n\|_\infty + \sqrt{\varepsilon}\,\|\nabla Z_n\|_\infty = O_{\mathbb{P}}\left(\sqrt{\frac{\log n}{n\varepsilon^{d/2}}} + \frac{\log n}{n\varepsilon^{d/2}}\right). \tag{36}$$

The interpolation error is $O(n^{-3})$; the second displayed term is absorbed by the first under (19). This event is uniform in x before evaluation at any sample site. Hence it also bounds $Z_n(X_i)$ without assuming independence between a sample site and the sample used in the sum. The diagonal contribution, of order $1/(n\varepsilon^{d/2})$, is included.

Take $b_i = a_\varepsilon(X_i)$. Its empirical masses satisfy $m_i(b) - 1 = a_\varepsilon(X_i)Z_n(X_i)$, whence $\delta_n \le \|a_\varepsilon\|_\infty \|Z_n\|_\infty$. On the event $\delta_n < 1$, (34) and the reproducing identities (12)–(13) imply

$$\|\widehat{v} - v_b\|_\infty + \sqrt{\varepsilon}\,\|\nabla\widehat{v} - \nabla v_b\|_\infty \le C\delta_n\varepsilon^{-d/4} = O_{\mathbb{P}}(\eta_n). \tag{37}$$

By (36), the same weighted bound holds between $\widehat{v}$ and $H_\varepsilon(a_\varepsilon f) = 1/a_\varepsilon$. The latter has bounded C^1 norm and is uniformly bounded away from zero. Since $\eta_n \to 0$, reciprocation preserves the value estimate. Differentiating the reciprocal and multiplying the gradient estimate by $\sqrt{\varepsilon}$ proves (35). □

3.3 Empirical covariance and uniform derivative Gram comparison

All operators in the following comparison are placed in the same ε-dependent space $\mathcal{H}_\varepsilon$. Define

$$\phi(y) = a_\varepsilon(y)k_\varepsilon(\cdot, y), \qquad \widehat{\phi}(y) = \widehat{a}(y)k_\varepsilon(\cdot, y),$$
$$C = \int \phi(y)\otimes\phi(y)\,d\mu(y), \qquad \widetilde{C} = \frac{1}{n}\sum_i \phi(X_i)\otimes\phi(X_i),$$
$$\widehat{C} = \frac{1}{n}\sum_i \widehat{\phi}(X_i)\otimes\widehat{\phi}(X_i).$$

Here $(v\otimes v)z = v\,\langle v, z\rangle_{\mathcal{H}_\varepsilon}$. The feature map $S : L^2(\mu) \to \mathcal{H}_\varepsilon$, $Su = \int \phi(y)u(y)\,d\mu(y)$, satisfies $P_\varepsilon = S^*S$ and $C = SS^*$. Consequently $\|C\| = 1$. For $S_n u = n^{-1}\sum_i \widehat{\phi}(X_i)u_i$ the analogous identities are $P_n = S_n^*S_n$ and $\widehat{C} = S_nS_n^*$, and $\left\|\widehat{C}\right\| = 1$.

The estimate $\|\phi(y)\|^2 = a_\varepsilon(y)^2K_{0,\varepsilon} \le C\varepsilon^{-d/2}$ and independence give

$$\mathbb{E}\left\|\widetilde{C} - C\right\|_{\mathrm{HS}}^2 \le \frac{1}{n}\mathbb{E}\,\|\phi(X)\otimes\phi(X)\|_{\mathrm{HS}}^2 \le \frac{C}{n\varepsilon^d}. \tag{38}$$

This second-moment calculation applies for each n in its current Hilbert space; a fixed-kernel limit theorem is not required. It implies $\left\|\widetilde{C} - C\right\|_{\mathrm{op}} = O_{\mathbb{P}}((n\varepsilon^d)^{-1/2})$ and $\left\|\widetilde{C}\right\|_{\mathrm{op}} = O_{\mathbb{P}}(1)$. Lemma 3.4 gives $\max_i |\widehat{a}_i/a_\varepsilon(X_i) - 1| = O_{\mathbb{P}}(\eta_n)$. Since the sample features differ only by these scalar factors, comparison of positive weights gives, in quadratic-form order, with the nonnegative random variable

$$\zeta_n = \max_i \left|\left(\frac{\widehat{a}_i}{a_\varepsilon(X_i)}\right)^2 - 1\right| = O_{\mathbb{P}}(\eta_n),$$

$$-\zeta_n\widetilde{C} \;\le\; \widehat{C} - \widetilde{C} \;\le\; \zeta_n\widetilde{C}.$$

It follows that

$$\left\|\widehat{C} - C\right\|_{\mathrm{op}} = O_{\mathbb{P}}(\eta_n). \tag{39}$$

Set $\theta_j(x) = \partial_j[a_\varepsilon(x)k_\varepsilon(\cdot,x)]$ and define $\widehat{\theta}_j$ with $\widehat{a}$. The section bounds and (35) imply, uniformly in x,

$$\left\|\theta_j(x)\right\| = O(\varepsilon^{-\alpha}), \qquad \left\|\widehat{\theta}_j(x) - \theta_j(x)\right\| = O_{\mathbb{P}}(\eta_n\varepsilon^{-\alpha}). \tag{40}$$

Indeed the value error multiplies a derivative section of norm $O(\varepsilon^{-\alpha})$, and the gradient error $O_{\mathbb{P}}(\eta_n\varepsilon^{-1/2})$ multiplies a value section of norm $O(\varepsilon^{-d/4})$. The derivative rows are $S^*\theta_j$ and $S_n^*\widehat{\theta}_j$. Intertwining S and the resolvents therefore gives

$$\begin{aligned} G_\varepsilon(x)_{jl} &= \left\langle \theta_j(x), F(C)\theta_l(x)\right\rangle_{\mathcal{H}_\varepsilon}, \\ G_n(x)_{jl} &= \left\langle \widehat{\theta}_j(x), F(\widehat{C})\widehat{\theta}_l(x)\right\rangle_{\mathcal{H}_\varepsilon}, \qquad F(C) = (I+C)^{-1}C(I+C)^{-1}. \end{aligned}$$

For positive operators, writing $R = (I+C)^{-1}$ gives $F(C) = R - R^2$. The resolvent identity and $\|R\| \le 1$ imply

$$\left\|F(\widehat{C}) - F(C)\right\|_{\mathrm{op}} \le 3\left\|\widehat{C} - C\right\|_{\mathrm{op}}.$$

Using (39) and (40), and noting that d is fixed, we obtain

$$\sup_x \|G_n(x) - G_\varepsilon(x)\|_{\mathrm{op}} = O_{\mathbb{P}}(\eta_n\varepsilon^{-2\alpha}). \tag{41}$$

The uniformity comes from the uniform feature estimate and the single operator bound, rather than a pointwise covariance argument. Lemma 3.3 and $\eta_n \to 0$ now prove (26), completing part (2) of Theorem 3.1.

3.4 Nonlinear upper bound and positive sharpness witnesses

Proof of Theorem 3.1, part (1). By Lemma 3.4, there are fixed positive constants c, C and events of probability tending to one on which

$$c \le \widehat{v} \le C, \qquad \sqrt{\varepsilon}\,\|\nabla\widehat{v}\|_\infty \le C.$$

Normalize any positive candidate to $S(b) = 1$. The energy estimate gives, on the same event and simultaneously for every such candidate,

$$\|v_b - \widehat{v}\|_\infty \le r(b)\sqrt{K_{0,\varepsilon}}, \qquad \|\nabla v_b - \nabla\widehat{v}\|_\infty \le r(b)\sqrt{K_{1,\varepsilon}}.$$

Choose $c_0 > 0$ sufficiently small that $r(b) \le c_0\varepsilon^{d/4}$ implies $v_b \ge c/2$. Then

$$\frac{\nabla v_b}{v_b} - \frac{\nabla\widehat{v}}{\widehat{v}} = \frac{\nabla v_b - \nabla\widehat{v}}{v_b} + \nabla\widehat{v}\frac{\widehat{v} - v_b}{v_b\widehat{v}}.$$

The two terms are bounded by $Cr(b)\varepsilon^{-\alpha}$, using (12)–(13). This proves (23), with an event independent of the candidate. Under (20), $\sqrt{n}\varepsilon^{d/4} = A_n/\sqrt{\varepsilon} \to \infty$. Thus $\sqrt{n}r(b_n) \to 0$ implies the required residual threshold with probability tending to one. Multiplying (23) by A_n proves (24). □

Proof of Theorem 3.1, part (3). Fix x_0 and put $r_1 = r_{n,1,x_0}$ and $d_n = \|R_n r_1\|_n = \sqrt{G_n(x_0)_{11}}$. When $d_n > 0$, set

$$u_n = \frac{2R_n^2 r_1}{d_n}. \tag{42}$$

The mean-zero property of r_1 and invariance of this subspace under R_n give

$$\langle u_n, 1\rangle_n = 0, \qquad \|(I+P_n)u_n\|_n = 2, \qquad \|u_n\|_n \le 2.$$

By (26), d_n is comparable to $\varepsilon^{-\alpha}$ on events of probability tending to one. Intertwining gives the feature representation

$$u_n = S_n^*\left[\frac{2(I+\widehat{C})^{-2}\widehat{\theta}_1(x_0)}{d_n}\right].$$

Uniformly in i, $\left\|\widehat{\phi}(X_i)\right\| = O_{\mathbb{P}}(\varepsilon^{-d/4})$, whereas $\left\|\widehat{\theta}_1(x_0)\right\|/d_n = O_{\mathbb{P}}(1)$. Consequently

$$\|u_n\|_\infty = O_{\mathbb{P}}(\varepsilon^{-d/4}). \tag{43}$$

Every off-sample row of P_n is a probability vector, so the same bound holds for $(P_n u_n)(x)$ uniformly in x.

Let $t_n = c/\sqrt{n}$. Define a vector on every sample as follows: set $u_n = 0$ if $d_n = 0$, and also replace (42) by zero if $t_n \|u_n\|_\infty > 1/2$. Set

$$b_n^{\mathrm{raw}} = \widehat{a}(1 + t_n u_n), \qquad b_n = b_n^{\mathrm{raw}}/S(b_n^{\mathrm{raw}}).$$

These vectors are positive on every sample. The construction is measurable: the strictly convex minimizer defining $\widehat{a}$ depends continuously on the positive kernel matrix, as also follows from its invertible Hessian; all subsequent inverses, norms, and cutoffs are measurable. The replacements have probability tending to zero by (26), (43), and $\sqrt{n}\varepsilon^{d/4} \to \infty$.

On the retained event let $q_n = (I + P_n)u_n$. Then $\langle q_n, 1\rangle_n = 0$, $\|q_n\|_n = 2$, and

$$m(b_n^{\mathrm{raw}}) = 1 + t_n q_n + t_n^2 u_n P_n u_n.$$

The Taylor expansion of the square root gives

$$\left\|\sqrt{m} - 1 - \tfrac{t_n}{2} q_n\right\|_n = O_{\mathbb{P}}(t_n^2 \varepsilon^{-d/4}). \tag{44}$$

For completeness, the product contribution is bounded by $t_n^2 \|u_n\|_\infty \|P_n u_n\|_n$. The quadratic Taylor remainder is bounded by $C \|m - 1\|_\infty \|m - 1\|_n$. The L^2 contraction of P_n, the bound $\|u_n\|_n \le 2$, and (43) control both by the right side of (44). The cutoff ensures that each factor in $m = (1 + t_n u_n)(1 + t_n P_n u_n)$ lies in $[1/2, 3/2]$, so m lies in $[1/4, 9/4]$; hence the Taylor constant is uniform. Also $t_n \varepsilon^{-d/4} \to 0$, so the bound is an asymptotically smaller term than t_n. Projecting away constants in (44) yields

$$\rho(b_n^{\mathrm{raw}}) = t_n[1 + O_{\mathbb{P}}(t_n \varepsilon^{-d/4})], \qquad S(b_n^{\mathrm{raw}}) = 1 + O_{\mathbb{P}}(t_n^2 \varepsilon^{-d/4}).$$

Thus $\sqrt{n} r(b_n) \to c$ in probability.

Finally (31) and (42) imply the exact vector identity, on the retained event,

$$s_{b_n}(x_0) - \widehat{s}(x_0) = \frac{4 t_n G_n(x_0) e_1}{d_n[1 + t_n (P_n u_n)(x_0)]}. \tag{45}$$

Its denominator factor in brackets tends to one. The full matrix limit (26) gives

$$\frac{\varepsilon^\alpha G_n(x_0) e_1}{d_n} \longrightarrow \sqrt{J_d / f(x_0)}\, e_1.$$

Multiplying (45) by A_n proves (27). The vanishing-probability replacements do not affect these probability limits. □

Corollary 3.5. *Suppose, in addition, that an exact empirical limit law $A_n(\widehat{s}(x_0) - \vartheta_n) \Rightarrow Z$ holds in this periodic model. For any sequence satisfying the residual hypothesis in part (1), $A_n(s_{b_n}(x_0) - \vartheta_n) \Rightarrow Z$. For the witnesses in part (3), the limit is instead $Z + (cC_d/\sqrt{f(x_0)})e_1$.*

Proof. Apply Slutsky's theorem to (24) or (27), respectively. □

Remark 3.6. The corollary assumes the stated statistical limit hypothesis. The witnesses show that residual order $n^{-1/2}$ cannot guarantee negligible normalized error uniformly over all positive candidates. The sampling condition $n\varepsilon^d/\log n \to \infty$ is stronger than divergence of the effective neighborhood count $n\varepsilon^{d/2}$ and is used here only as a sufficient condition for the uniform scaling and covariance comparisons.

4 Actual self-Sinkhorn trajectories

We illustrate the certificate on actual iterates, rather than on the adverse witness construction. In dimension one we draw iid observations from $f(x) = 1 + 0.4\cos(2\pi x)$ by independent rejection sampling on the torus. All sample sizes, bandwidths and random seeds are in Table 1. Starting from $a_i^{(0)} = 1$, we use the symmetric damped update

$$\log a_i^{(j+1)} = \log a_i^{(j)} - \tfrac{1}{2}\log m_i(a^{(j)}).$$

The script uses NumPy's default generator with the specified seed, NumPy 2.3.5, double precision, and evaluates scores on the 512 points $x = l/512$, $l = 0, \dots, 511$. No iid limit law is simulated.

For reproducibility, the numerical kernel is a finite Fourier expansion with positive coefficients of the integral-one periodic heat kernel:

$$k_M(x,y) = 1 + 2\sum_{q=1}^{M} e^{-2\pi^2\epsilon q^2}\cos(2\pi q(x-y)), \qquad M = \left\lceil \sqrt{45/(2\pi^2\epsilon)} \right\rceil.$$

It is PSD by its nonnegative Fourier coefficients; positivity is observed at all evaluated entries. For this finite expansion we use the corresponding constants $K_0 = 1+2\sum_{q=1}^{M}\lambda_q$ and $K_1 = 2\sum_{q=1}^{M}(2\pi q)^2\lambda_q$, rather than a diagonal-one normalization. No rigorous enclosure of the omitted Fourier tail is supplied. Each iterate is canonically rescaled to $S = 1$ for evaluation; its score and projective residual r are unchanged. At each query we evaluate

$$B(x) = \frac{2r\sqrt{K_1 + K_0|s_b(x)/2|^2}}{v_b(x) - r\sqrt{K_0}}$$

when the denominator is positive, and report an inconclusive value otherwise. The reported certificate is the maximum of these values on the query grid.

A separate tight solve from the same initialization supplies the numerical reference, stopping at maximum mass residual 10^{-14}. A looser solve at 4×10^{-14} and a dense Newton correction at the tight solution provide numerical sensitivity checks. Across the cases the tight mass residual is at most 7.8×10^{-15}, the maximum score change from the looser reference is below 2.8×10^{-14}, and that from the dense correction is below 9×10^{-15}. Independent central finite differences of the log potential check score signs and factors at three separate points, with maximum discrepancy below 3.8×10^{-10}.

n	ϵ	seed	j_E	j_B	E_{j_B}	B_{j_B}	j_r	E_{j_r}
256	0.04	2026101801	4	4	0.0032	0.00462	3	0.0105
256	0.01	2026101802	4	4	0.00413	0.0073	3	0.0138
512	0.02	2026101803	4	5	0.00201	0.00409	4	0.00527

Table 1: Actual solver stopping comparisons. E_j is the maximum observed score error against the floating reference on the query grid; B_j is the grid certificate. j_E and j_B are the first iterations reaching $E_j \le 10^{-2}$ and $B_j \le 10^{-2}$, respectively. j_r is the first iteration reaching $r \le 10^{-3}$. Iteration zero is the initialization.

The certificate reaches the prescribed score target at iterations 4, 4, and 5; observed errors first reach it at iteration 4 in all three cases. The fixed residual threshold 10^{-3} stops earlier in two cases, and its associated score errors are displayed rather than treated as an automatic score guarantee. The certificate thus supplies a directly interpretable, moderately conservative stopping diagnostic on these runs. This observation concerns these trajectories and bandwidths; it does not show that the theoretical adverse witnesses occur along Sinkhorn iterates.

The reference is approximate, the grid is finite, and all reported inequality comparisons allow floating numerical discrepancy. In particular, the maximum observed grid error does not measure the continuum supremum, and the grid maximum of a pointwise certificate does not certify unseen queries. These experiments supply no interval certificate, statistical CLT, or iteration-complexity lower bound.

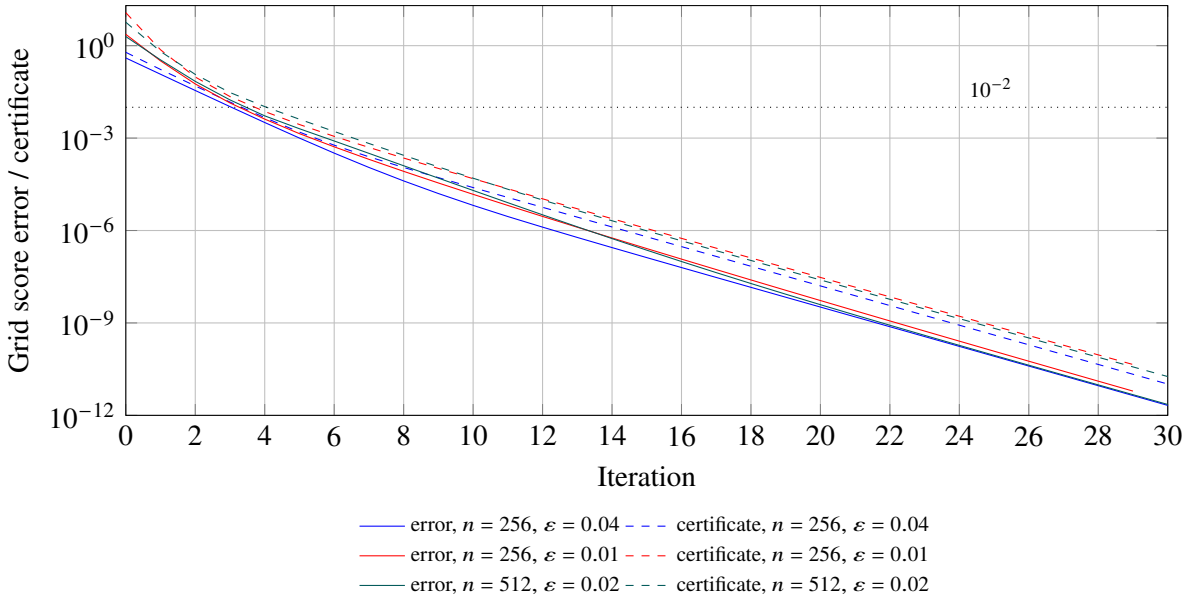


Figure 1: Grid score errors (solid) and a posteriori certificates (dashed) along actual symmetric self-Sinkhorn iterates. The plot is a numerical illustration.

5 Scope and conclusions

The finite certificate converts a measured projective residual into a pointwise error bound for the kernel-extension score. Its proof uses the whole residual and avoids unknown bounds on the exact scaling. The iid periodic theorem identifies the corresponding local bandwidth order and gives positive nonlinear witnesses at residual order $n^{-1/2}$.

The sharpness statement is a worst-case guarantee over positive candidates; the constructed directions need not occur along a particular solver. The exact local constant uses the exact empirical scaling. Any transfer of statistical inference remains conditional on a valid exact-estimator limit theorem for the same model.

Reproducibility. The accompanying source package contains the solver script, its parameters, raw iteration records and figure data. The proofs are analytic and do not depend on floating-point experiments.

AI assistance. AI tools assisted mathematical development, computation and editing. The author is responsible for the final manuscript.

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 Twenty-Second International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research **89**, 2681–2690, 2019. `https://proceedings.mlr.press/v89/feydy19a.html`.

[2] X. Cheng and B. Landa. Bi-stochastically normalized graph Laplacian: convergence to manifold Laplacian and robustness to outlier noise. arXiv:2206.11386v3, 17 July 2024. `https://arxiv.org/abs/2206.11386v3`.

[3] G. Mordant. The entropic optimal (self-)transport problem: Limit distributions for decreasing regularization with application to score function estimation. arXiv:2412.12007v2, 29 April 2026. `https://arxiv.org/abs/2412.12007v2`.

[4] C. L. Wormell and S. Reich. Spectral convergence of diffusion maps: improved error bounds and an alternative normalisation. arXiv:2006.02037v3, 7 April 2021. https://arxiv.org/abs/2006.02037v3. Related journal article: *SIAM Journal on Numerical Analysis*, DOI https://doi.org/10.1137/20M1344093.

[5] I. Maeda, R. Yao, and A. Nitanda. Statistical Analysis of the Sinkhorn Iterations for Two-Sample Schrödinger Bridge Estimation. arXiv:2510.22560v1, 26 October 2025. https://arxiv.org/abs/2510.22560v1.

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

[7] N. F. Marshall and R. R. Coifman. Manifold learning with bi-stochastic kernels. arXiv:1711.06711v2, 27 February 2018. https://arxiv.org/abs/1711.06711v2. Related journal DOI: https://doi.org/10.1093/imamat/hxy065.

[8] D. Belomestny. Uniform Statistical Convergence of Empirical Sinkhorn Potentials with Exponential and Polynomial Dependence on the Regularization Parameter. arXiv:2608.29152v1, 29 August 2026. https://arxiv.org/abs/2608.29152v1.

[9] A. González-Sanz, M. Nutz, and A. J. Stromme. Finite-sample bounds for regularized optimal transport. arXiv:2606.25947v1, 24 June 2026. https://arxiv.org/abs/2606.25947v1.