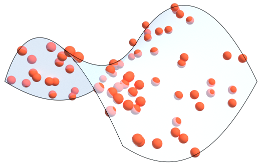
Sinkhorn Normalization of Diffusion Kernels

Nathan Kessler* [1,2]   Robin Magnet* [2]   Jean Feydy [2]

[1] Centre Borelli, ENS Paris-Saclay [2] Inria, Université Paris Cité, Inserm, HeKA, Paris

The problem

Smoothing a signal over local neighborhoods is a fundamental geometric operation.



On manifolds, **heat diffusion** $e^{-t\Delta}$ is the principled tool but requires a Laplacian Δ.

On unstructured data, there is **no natural Laplacian** and simpler convolution kernels are used, despite being **biased by sampling density**

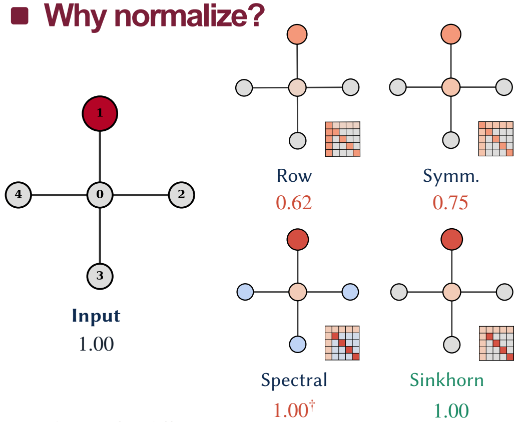
Idea

Correct **any** symmetric smoothing operator into a heat-diffusion-like operator via **symmetric Sinkhorn normalization**. No connectivity, factorization, or Laplacian.

$$\mathbf{S} \longrightarrow \mathbf{Q = \Lambda S \Lambda}$$

smoothing op. *diffusion op.*

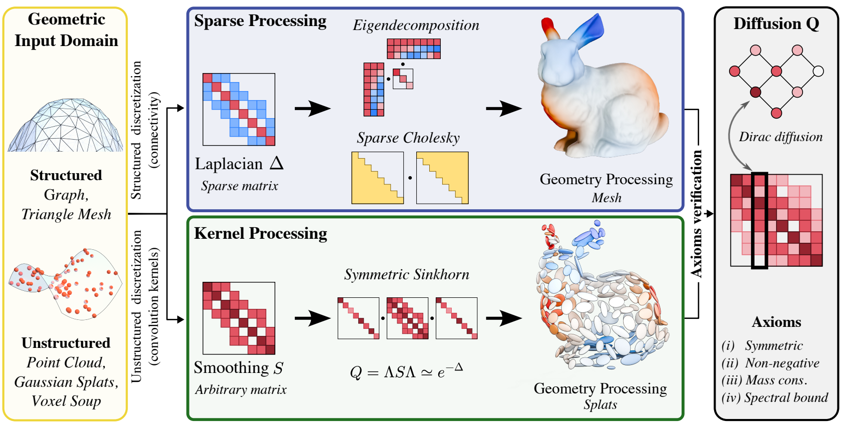
Why normalize?



Signal mass after diffusion.

Row / symm scalings **lose mass** ; spectral truncation adds negative ringing (†). Sinkhorn keeps **both** mass and positivity.

The pipeline



An axiomatic structure

Smoothing, Laplacian and diffusion operators share a small set of axioms w.r.t. the mass-weighted inner product $\langle\cdot,\cdot\rangle_M$

Smoothing S
M-symmetric · PSD · positivity-preserving ($f \geq 0 \Longrightarrow Sf \geq 0$).

Laplace-like Δ
M-symmetric · PSD · $\Delta\mathbf{1}=0$ · Metzler ($\Delta_{ij} \leq 0,\ i \neq j$).

Diffusion Q — *our target*
M-symmetric · $Q\mathbf{1}=\mathbf{1}$ · spectrum $\subseteq [0,1]$ · positivity-preserving.

Guarantees

Existence & uniqueness

A **unique** positive diagonal Λ makes $Q = \Lambda S \Lambda$ a valid diffusion operator.

Stability under refinement

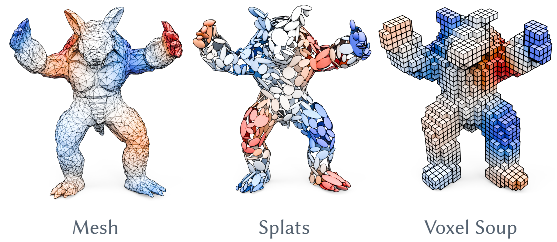
As samples densify ($\mu^s \to \mu$) with **fixed** kernel scale, the normalized operators converge **uniformly** to a bounded continuous diffusion operator, unlike raw kernels whose norm explodes.

Symmetric Sinkhorn normalization

```
input : smoothing S (S_ij ≥ 0)
output : Λ with Q = ΛSΛ
λ ← 1
repeat
  λ ← √(λ ⊘ (Sλ))
until converged // 5-10 iters
return Λ = diag(λ)
```

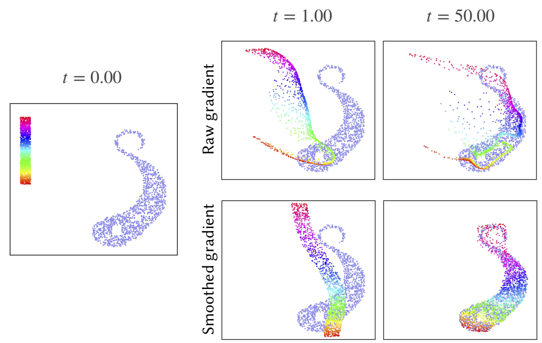
Compute a single positive scaling λ. S is a black-box *matrix–vector product*, GPU-friendly with no factorization. Scales to millions of points via KeOps or attention

Spectral consistency



Leading eigenvectors of Q recover the **low-frequency Laplacian modes** for generic data representations

Beyond spectra



Gradient flow of the energy distance can be stabilized by smoothing the gradient with our Q operator.

Shape correspondence. We introduce Q-DiffNet, an adaptation of DiffusionNet that allows to learn features on arbitraty representation

Runtime. 5 GPU Sinkhorn iterations normalize at *train time* , where sparse Laplacian solvers lack GPU support.

CODE
github.com/RobinMagnet/SinkhornKernels