Normalizing diffusion kernels with optimal transport

How to encode smoothness?

Nathan Kessler, Robin Magnet, Jean Feydy
HeKA team, Inria Paris, Inserm, Université Paris-Cité

IPAIM, Bath – Monday, 30 June 2025

Who am I?

Background in **mathematics** and **data sciences**:

2012–2016 ENS Paris, mathematics.

2014–2015 M2 mathematics, vision, learning at ENS Cachan.

2016–2019 PhD thesis in **medical imaging** with Alain Trouvé at ENS Cachan.

2019–2021 **Geometric deep learning** with Michael Bronstein at Imperial College.

2021+ **Medical data analysis** in the HeKA INRIA team (Paris).

HeKA: a translational research team for public health

Hospitals

Inria

Inserm

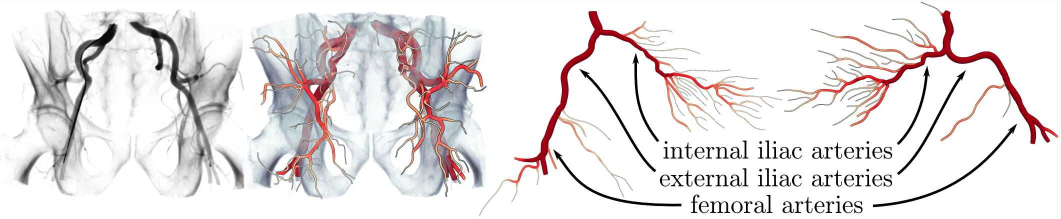
Universities

My main motivation

Develop **robust and efficient** software that **stimulates other researchers**:

1. Speed up **geometric machine learning** on GPUs:
   $\Longrightarrow$ **pyKeOps** library for distance and kernel matrices, 800k+ downloads.
2. Scale up **pharmacovigilance** to the full French population:
   $\Longrightarrow$ **survivalGPU**, a fast re-implementation of the R survival package.
3. Ease access to modern statistical **shape analysis**:
   $\Longrightarrow$ **GeomLoss**, truly scalable optimal transport in Python.
   $\Longrightarrow$ **scikit-shapes**, beta release in September.

A vessel map that preserves vessel lengths and curvatures [HBAF25]



Our new visualization method, tailored to **endovascular** interventions.

A consistent theory of smoothness for diverse data structures

1. The **clean** method: **Laplacians** and heat **diffusions**

2. The **fast** method: **smoothing** with local averages

3. **Sinkhorn normalization** : **fast** smoothing $\mapsto$ **clean** diffusion

4. **Applications**

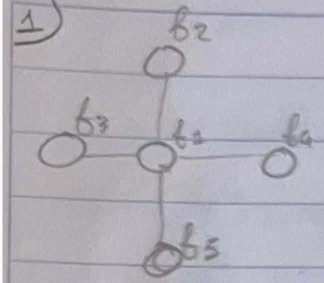
Laplacians and heat diffusions

On the blackboard

- Smoothness on graphs
- Laplacian operators
- Heat diffusion – in theory
- Heat diffusion – in practice

On the blackboard

1)



$$\|f\|^2_{smooth} = (f_1-f_2)^2 + \dots + (f_1-f_5)^2$$

$$\delta = \begin{array}{cc} & \text{nodes} \\ \text{edges} & \begin{bmatrix} 1 & -1 & & & \\ 1 & & -1 & & \\ 1 & & & -1 & \\ 1 & & & & -1 \end{bmatrix} \end{array}, \quad \delta^T\delta = \begin{bmatrix} 4 & -1 & -1 & -1 & -1 \\ -1 & 1 & & & \\ -1 & & 1 & & \\ -1 & & & 1 & \\ -1 & & & & 1 \end{bmatrix}$$

$$\|f\|^2_{smooth} = \|\delta f\|^2 = f^T\delta^T\delta f = f^T\Delta f$$

On the blackboard

By construction of the Laplacian Δ:

i) $\Delta^T = \Delta$

ii) $\Delta \mathbb{1} = 0$

iii) $f^T \Delta f \geq 0$

iv) $i \neq j \Rightarrow \Delta_{ij} \leq 0$

Fourier analysis:

$f = \sum \hat{f}_i e_i$

$\Rightarrow \Delta f = \sum \lambda_i \hat{f}_i e_i$

Can be generalized to weighted, continuous domains, cotan Δ, etc.

On the blackboard

3) Heat diffusion – in theory

$$\dot{f} = -\Delta f \Rightarrow f_t = e^{-t\Delta} f_0$$

$$f = \sum \hat{f}_i e_i$$

$$\Rightarrow e^{-t\Delta} f = \sum e^{-t\lambda_i} \hat{f}_i e_i$$

The diffusion operator $Q = e^{-t\Delta}$ is:

i) $Q^T = Q$

ii) $Q\mathbb{1} = \mathbb{1}$

iii) $\lambda_i^Q = e^{-t\lambda_i^\Delta} \in [0;1]$

iv) $f \geq 0 \Rightarrow Qf \geq 0$

Mass conservation:

$$\langle \mathbb{1}, Qf \rangle = \langle Q\mathbb{1}, f \rangle = \langle \mathbb{1}, f \rangle$$

On the blackboard

Heat diffusion – in practice

1. Diagonalize $\Delta = \sum_i \lambda_i e_i e_i^T$

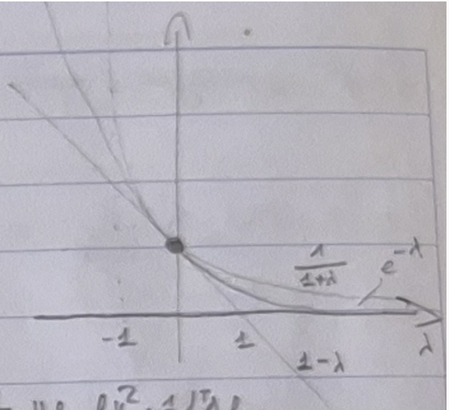
→ expensive, ringing "Gibbs" artefacts

2. Explicit Euler, $f_t = f_0 - t \Delta f_0$

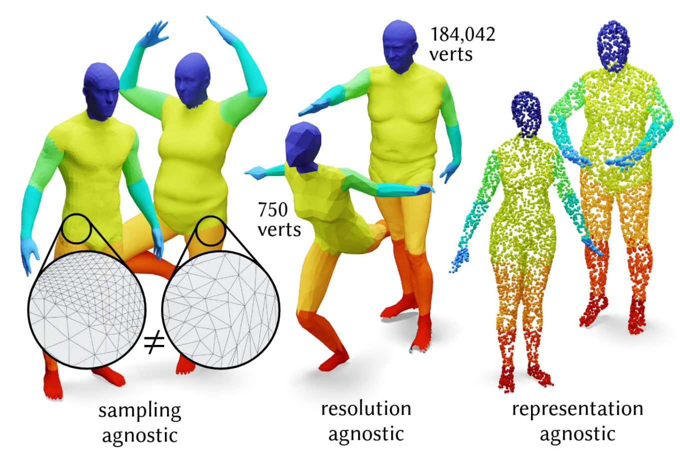
→ unstable, $\hat{f}_i \mapsto (1 - t\lambda_i) \hat{f}_i$

3. Implicit Euler, $f_t = f_0 - t \Delta f_t$

→ Linear solver, $f_t = (I + t\Delta)^{-1} f_0 = \text{argmin} \frac{1}{2t} \|f - f_0\|^2 + \frac{1}{2} f^T \Delta f$



DiffusionNet [SACO22]



Super neat, but requires a **pre-factorization** of Δ.

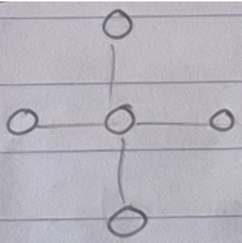
Not GPU or real-time friendly.

Smoothing with local averages

On the blackboard

- Adjacency matrices
- Smoothing operators
- Normalizing smoothing operators

On the blackboard



$$S = \text{Degree} + \text{Adjacency} = \begin{pmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 1 & & & \\ 1 & & 1 & & \\ 1 & & & 1 & \\ 1 & & & & 1 \end{pmatrix}$$

S is a smoothing operator:

i) $S^T = S$ but $S\mathbb{1} \neq \mathbb{1}$

ii) $f^T S f \geq 0$

iii) $f \geq 0 \Rightarrow Sf \geq 0$

On the blackboard

Normalizing smoothing operators

1. use $Q=(S1)^{-1}S \Rightarrow Q1=1$

   problem: $Q^T \neq Q$, $\langle 1, Qf\rangle \neq \langle 1, f\rangle$

2. use $Q=(S1)^{-1/2}\, S\, (S1)^{-1/2}$

   problem: $Q1 \neq 1$

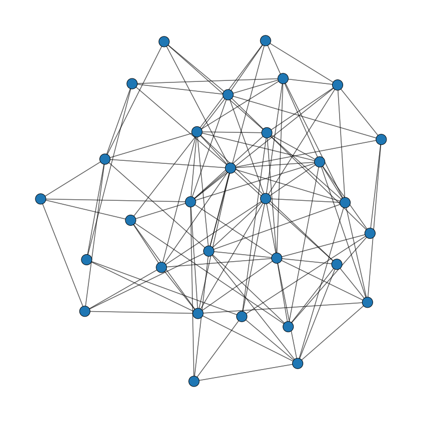
3. iterate step 2!

   $\rightarrow$ find Λ such that $Q1 = \Lambda S \Lambda 1 = 1$

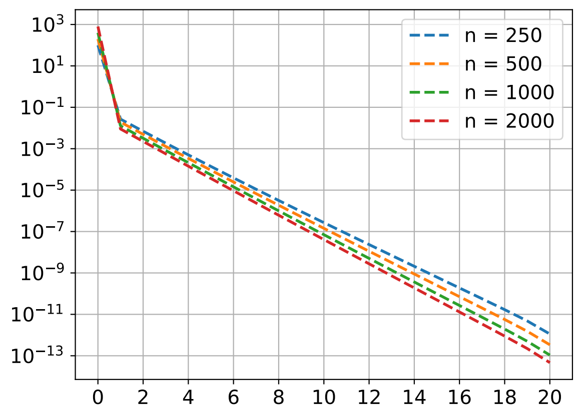
Key result: this turns any smoothing operator $S>0$ into a genuine diffusion operator.

Sinkhorn normalization

Fast convergence: monitoring the average value of $|Q1 - 1|$

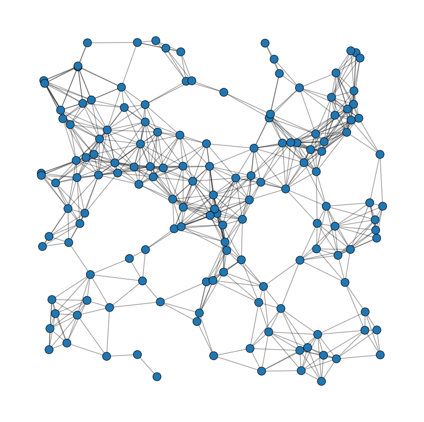


Random graph with n nodes.

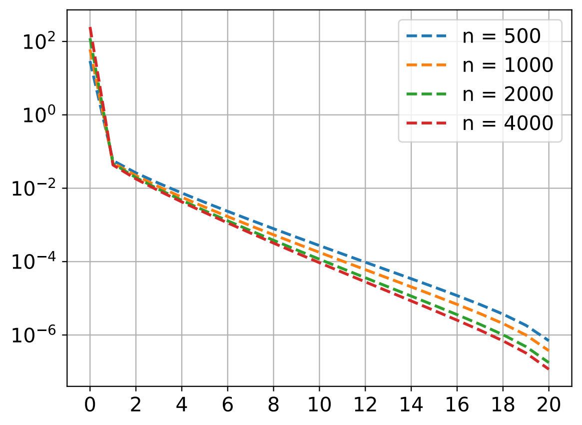


After 5 iterations: 0.01% error.

Fast convergence: monitoring the average value of $|Q1 - 1|$



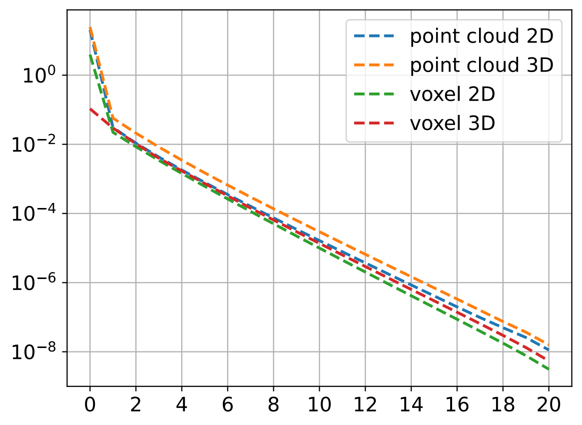
Geometric graph with n nodes.



After 5 iterations: 0.5% error.

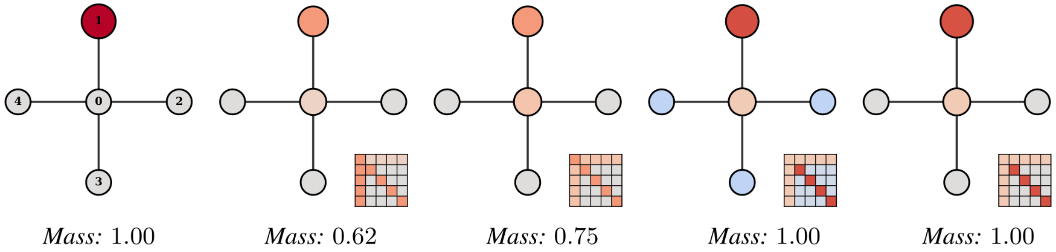
Fast convergence: monitoring the average value of $|Q1 - 1|$

Armadillo surface – 5,000 points.



After 5 iterations: 0.1% error.

Fixing the "central node bias"



(a) Input Dirac (b) Row (c) Symmetric (d) Spectral (e) Sinkhorn

Well-posedness

Focus on **convolution** with a Gaussian or exponential kernel on $\mathbb{R}^d$:

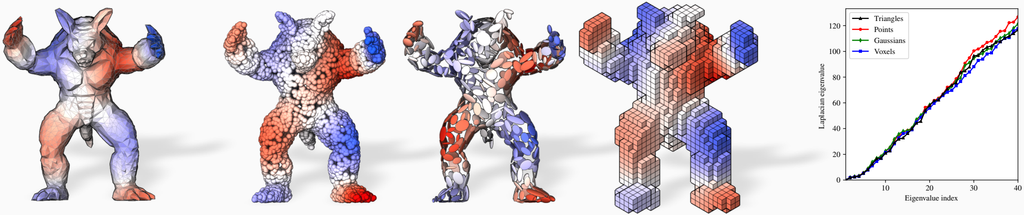
$$S_\mu f \,:\, x_i \mapsto \sum_j k(x_i, x_j) m_j f(x_j)$$

We can interpret the diagonal **scaling** matrix Λ for $Q_\mu = \Lambda S_\mu \Lambda$
as pointwise **multiplication** with a positive, **continuous** function λ.

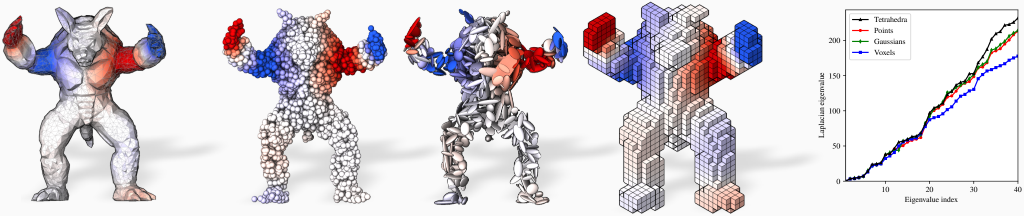
Using standard lemmas from optimal transport theory, we show that:

$$\mu^t \rightharpoonup \mu \implies \lambda^t \xrightarrow{\|\cdot\|_\infty} \lambda \implies Q_\mu f = \lambda^t S_\mu \lambda^t f \xrightarrow{\|\cdot\|_\infty} \lambda S \lambda f = Q f\,.$$

Spectral convergence – 10th eigenvector on the Armadillo

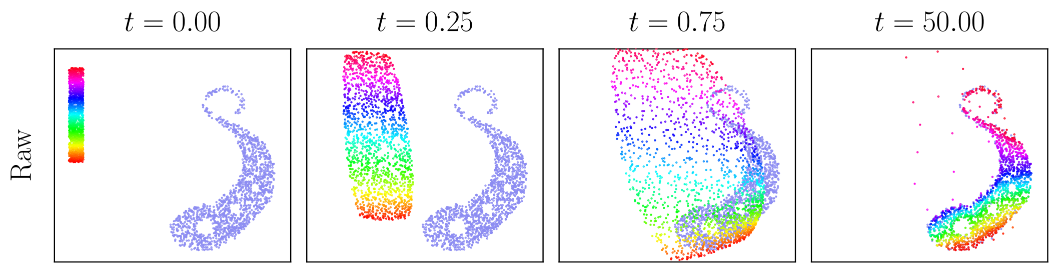


(a) Triangles (b) Points (c) Gaussians (d) voxels (e) Surface Spectra



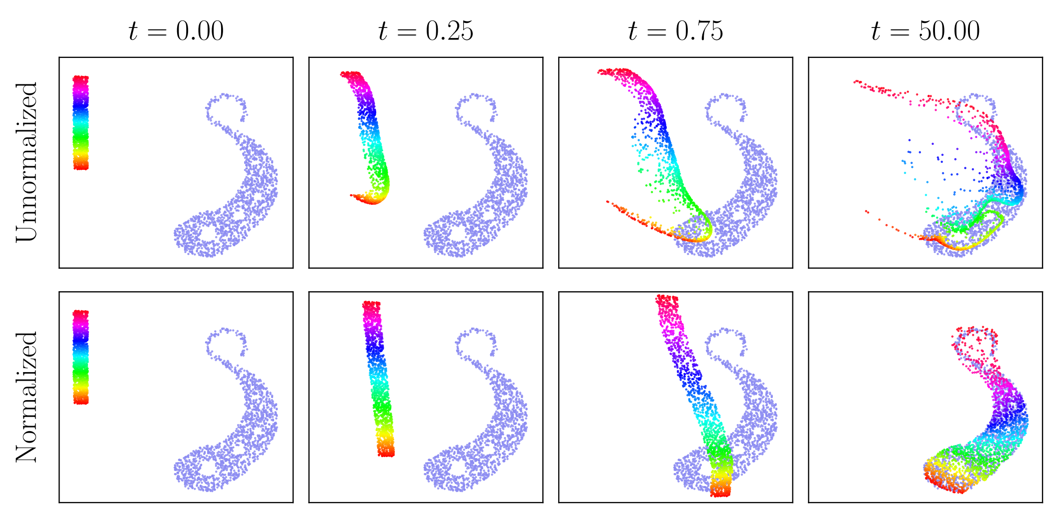
(f) Tetrahedra (g) Points (h) Gaussians (i) Voxels (j) Volume Spectra

Gradient flows – without regularization

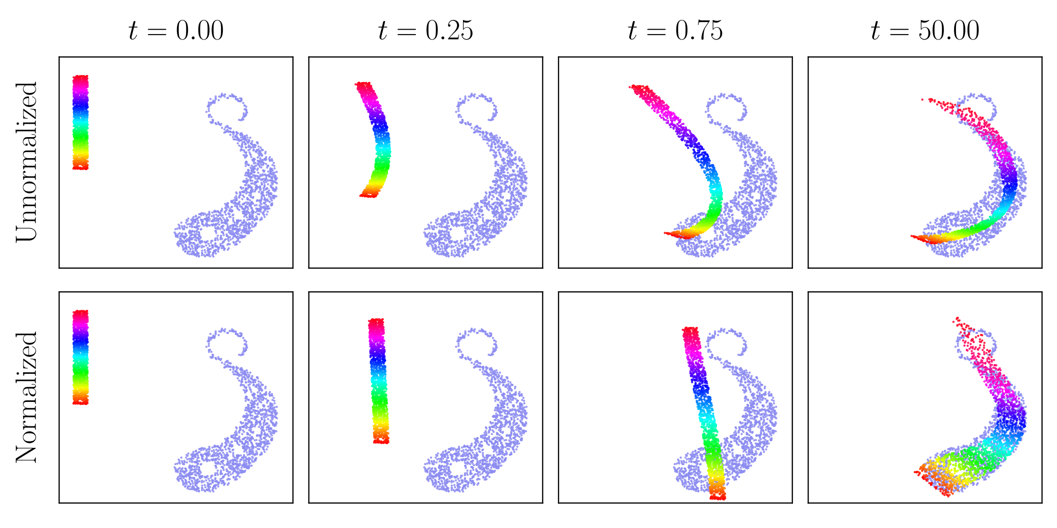


Wasserstein gradient flow of the Energy Distance.

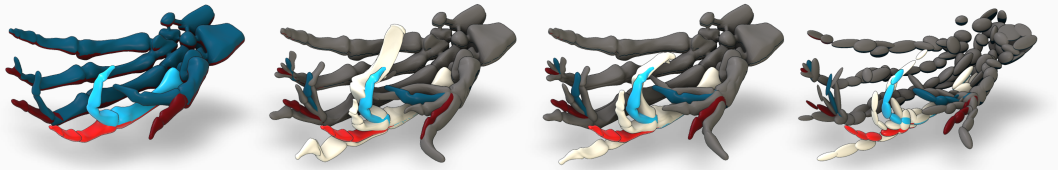
Gradient flows – with a Gaussian regularization at scale $\sigma = 0.07$



Gradient flows – with a Gaussian regularization at scale $\sigma = 0.20$



Shape metrics – geodesic interpolation and extrapolation



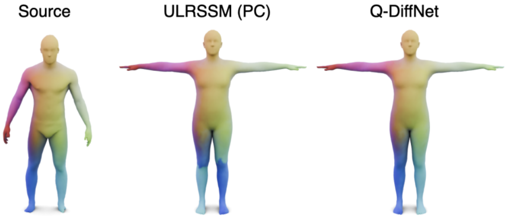
(a) Input Data (b) LDDMM (c) Normalized (d) Gaussian Mixtures

Normalizing LDDMM kernel metrics fixes the “**exploding geodesics**” problem.
We obtain a versatile and topology-preserving metric for shape analysis.

Use a normalized Gaussian convolution instead of a pre-factored Laplacian

Train	FAUST + SCAPE		
Test	FAUST	SCAPE	S19
DiffNet [78]	**1.6**	2.2	4.5
ULRSSM [19]	**1.6**	**2.1**	4.6
DiffNet (PC) [78]	3.0	2.5	7.5
ULRSSM (PC) [19]	2.3	2.4	5.1
Q-DiffNet (QFM)	2.5	3.1	4.1
Q-DiffNet	2.1	2.4	**3.5**

(a) Mean Geodesic Error



(b) Visualization of Correspondences

Figure 6: Evaluation of Q-DiffNet for shape matching. (a) Mean geodesic error comparison. “PC” denotes retraining on point clouds. (b) Predicted correspondences on SHREC19 [59] visualized via color transfer. ULRSSM trained on point clouds confuses symmetric parts (e.g., legs).

Conclusion

We used to face **dilemmas**:

- **Smooth** with **Laplacians** (expensive) or with **local averages** (biased).
- **Normalize** operators with **row-wise** or **symmetric** scaling.

A simple trick – **iterate** the symmetric scaling update:

- Cheap and **versatile**.
- Turn **convolutions** into genuine **diffusion** operators.
- Fix the **“central node bias”**.

Non-intrusive method to enforce theoretical axioms.

Ideally suited to modern parametric models.

References

References i

Guillaume Houry, Tom Boeken, Stéphanie Allassonnière, and Jean Feydy.

Untangling vascular trees for surgery and interventional radiology.

In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, page to appear. Springer, 2025.

Nicholas Sharp, Souhaib Attaiki, Keenan Crane, and Maks Ovsjanikov.

Diffusionnet: Discretization agnostic learning on surfaces.

ACM Transactions on Graphics (TOG), 41(3):1–16, 2022.