Normalizing diffusion kernels with optimal transport

Kessler Nathan, Magnet Robin, Feydy Jean

October 28, 2025

A Small Note Before We Begin

I have an **auditory processing disorder**, which means I sometimes need a bit more time to understand spoken questions.

When asking questions, please try to:

- Speak **clearly** and **articulate**;
- Avoid speaking too fast;
- Repeat or rephrase if I ask for clarification.

Thank you for your understanding — it helps ensure a smooth discussion for everyone!

Outline

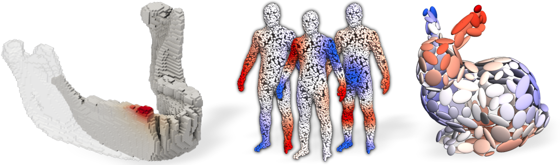
What is a diffusion?

How to diffuse?

Proper framework

Results

Purpose of the paper



Get an operator which smoothes functions on a manifold

Specifications

Should be fast to compute
Should work on a wide variety of modalities (meshes, voxel masks, point clouds, Gaussian mixtures...)

What is a smooth function?

Answer: it's a function with a small derivative!

If we have access to the **differential** d of a function f

Dirichlet energy: $E(f) = \underbrace{\langle df, df \rangle}_{\int |df|^2 \, \mathrm{d}\mu} = \langle f, d^T df \rangle = \langle f, \Delta f \rangle$

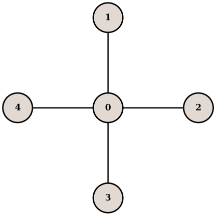
Smooth function: $E(f) \ll \|f\|^2$

Example over a graph

$f = (f_0, f_1, f_2, f_3, f_4)^T$

$df = (f_1 - f_0, f_2 - f_0, f_3 - f_0, f_4 - f_0)^T$

$\langle \Delta f, f \rangle = \langle df, df \rangle = \sum_{i=1}^4 \|f_i - f_0\|^2$



What is the diffusion of f?

Smooth approximation of f

Smooth approximation of f: $\min_g \frac{1}{2}\|f-g\|^2 + \frac{1}{2t}\langle g, \Delta g\rangle$
Deformation from correspondance: $\min_v \frac{1}{2}\|v-v_0\|^2 + \frac{1}{2t}\langle v, \Delta v\rangle$

$g = (\mathrm{Id} - t\Delta)^{-1} f \approx e^{-t\Delta} f$

Continuous point of view: smoothing gradient flow

$\nabla E(f) = \Delta f$
Gradient flow over E: $\partial_t f_t = -\Delta f_t$
Heat diffusion $f_t = e^{-t\Delta} f_0$

What to expect from a diffusion?

Symmetry

$$\Delta^T = (d^T d)^T = d^T d = \Delta \implies \left(e^{-t\Delta}\right)^T = e^{-t\Delta}$$

Diffusion $x \to y$ = diffusion $y \to x$

Constant preservation

$$\Delta 1 = d^T(d1) = 0$$

$$e^{-t\Delta}1 = 1$$

A constant function is already as smooth as possible!

Bonus: mass preservation

$$\underbrace{\left\langle e^{-t\Delta} f, 1 \right\rangle}_{\int_M e^{-t\Delta} f \,\mathrm{d}\mu} = \left\langle f, e^{-t\Delta} 1 \right\rangle = \underbrace{\langle f, 1 \rangle}_{\int_M f \,\mathrm{d}\mu}$$

Problems

Computing a diffusion operator

$(\mathrm{Id} - t\Delta)^{-1} f$: solve a linear system

$e^{-t\Delta} f$: diagonalize a matrix

Isn't there a faster way?

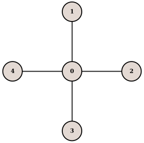
Data modality

d, Δ: objects from differential / riemannian geometry

I don't have any differential structure over my point cloud!

What if I had a proxy

"Adjacency" matrix of the graph: $\begin{bmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$



✓symmetry, ✗ constant preservation, ✗ mass preservation

If I normalize by rows

$$\operatorname{diag}\left(\begin{bmatrix}1/8\\1/2\\1/2\\1/2\\1/2\end{bmatrix}\right)\begin{bmatrix}4&1&1&1&1\\1&1&0&0&0\\1&0&1&0&0\\1&0&0&1&0\\1&0&0&0&1\end{bmatrix}=\begin{bmatrix}1/2&1/8&1/8&1/8&1/8\\1/2&1/2&0&0&0\\1/2&0&1/2&0&0\\1/2&0&0&1/2&0\\1/2&0&0&0&1/2\end{bmatrix}$$

✗ symmetry, ✓constant preservation, ✗ mass preservation

If I normalize by rows and columns

$$\mathrm{diag}\left(\begin{bmatrix}1/\sqrt{20}\\ \sqrt{4/5}\\ \sqrt{4/5}\\ \sqrt{4/5}\\ \sqrt{4/5}\end{bmatrix}\right)\begin{bmatrix}4 & 1 & 1 & 1 & 1\\ 1 & 1 & 0 & 0 & 0\\ 1 & 0 & 1 & 0 & 0\\ 1 & 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0 & 1\end{bmatrix}\mathrm{diag}\left(\begin{bmatrix}1/\sqrt{20}\\ \sqrt{4/5}\\ \sqrt{4/5}\\ \sqrt{4/5}\\ \sqrt{4/5}\end{bmatrix}\right)$$

$$= \begin{bmatrix}1/5 & 1/5 & 1/5 & 1/5 & 1/5\\ 1/5 & 4/5 & 0 & 0 & 0\\ 1/5 & 0 & 4/5 & 0 & 0\\ 1/5 & 0 & 0 & 4/5 & 0\\ 1/5 & 0 & 0 & 0 & 4/5\end{bmatrix}$$

✓constant preservation, ✓symmetry, ✓mass preservation

Can we always find such a diagonal matrix? How?

Sinkhorn algorithm

We want $u, v \in \mathbb{R}^n$ such that, with $\Lambda_u = \mathrm{diag}(u)$ and $\Lambda_v = \mathrm{diag}(v)$

$$\Lambda_u S \Lambda_v 1 = a, \qquad (\Lambda_u S \Lambda_v)^T 1 = b. \tag{1}$$

Dividing by Sv and $S^T u$, we obtain the equations:

$$u = \frac{a}{Sv}, \qquad v = \frac{b}{S^T u},$$

and the alternate scheme:

$$u_{k+1} = \frac{a}{Sv_k}, \qquad v_{k+1} = \frac{b}{S^T u_{k+1}}.$$

Theorem

If $S_{ij} > 0$ pour tout i, j, there exists (up to a scalar factor) a unique couple (u, v) verifying (1), and the alternate scheme converges.

Symmetric case

If S symmetric and $a = b$:

$$(\Lambda_u S \Lambda_v)1 = a, \qquad (\Lambda_u S \Lambda_v)^T 1 = a.$$

Hence $u = v$ and the simplified system:

$$\Lambda S \Lambda 1 = a.$$

Iterative scheme :

$$u_{k+1} = \frac{a}{S u_k}.$$

For better stability, use a geometric mean between old and new value:

$$u_{k+1} = \sqrt{\frac{a}{S u_k}} \sqrt{u_k} \quad \text{(symmetric Sinkhorn).}$$

Our method

Symmetric positive operator S with positive coefficients

$$\downarrow \text{Sinkhorn}$$

Diffusion operator $Q = \Lambda S \Lambda$, $\Lambda = \mathrm{diag}(\lambda)$

Symmetry : $x \to y = y \to x$

Constant preservation : $Q1 = 1$

Spectrum $\subset [0, 1]$: *damping* operator

Positive coefficients : $f \geq 0 \Rightarrow Qf \geq 0$

Symmetry + $Q1 = 1$: $\langle Qf, 1\rangle = \langle f, Q1\rangle = \langle f, 1\rangle \implies$ *mass preservation*

Comparison of a few methods

Row-wise normalization : $S \mapsto D^{-1}S, \quad D_{ii} = \sum_j S_{ij}$

$D^{-1}S1 = 1$: constant preservation

Symmetry ✗, mass ✗

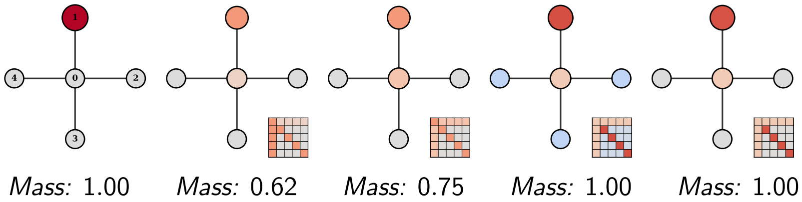
A *left-wise* Sinkhorn iteration

Symmetric normalization : $S \mapsto D^{-1/2}SD^{-1/2}$

Symmetry ✓

Constant ✗, mass ✗

A *symmetric* Sinkhorn iteration



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

$$\text{Initial adjacency matrix: } \begin{bmatrix} 4 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Framework

Let $X \subseteq \mathbb{R}^d$ equipped with a measure μ.
The functional space we consider is

$$L^2_\mu(X) \quad \text{with the scalar product} \quad \langle f, g \rangle_\mu = \int_X f(x) g(x) \, d\mu(x).$$

The **mass** of a function f is

$$\langle f, 1 \rangle_\mu = \int_X f(x) \, d\mu(x).$$

Discrete case

If $\mu = \sum_i m_i \delta_{x_i}$, then $M = \mathrm{diag}(m_i)$ and

$$\langle f, g \rangle_M = f^\top M g, \quad \langle f, 1 \rangle_M = 1^\top M f.$$

The M-**transpose** of an operator A is defined by

$$A^{T_M} = M^{-1} A^\top M, \qquad \langle Af, g \rangle_M = \langle f, A^{T_M} g \rangle_M.$$

An operator S is M-**symmetric** iff

$$S^{T_M} = S \quad \iff \quad S = KM \text{ (ou } S = M^{-1}K\text{) avec } K^\top = K.$$

Recoverable properties of the Laplacian

Définition

The Laplacian is defined by $\Delta = d^\top d$, $C^\infty(\mathcal{M}) \underset{d^\top}{\overset{d}{\rightleftharpoons}} \Gamma(T^*\mathcal{M})$.

Properties

Symmetry: $(d^\top d)^\top = d^\top d$,

Positivity: $\langle f, \Delta f \rangle = \langle df, df \rangle \geq 0$,

Constant cancelation: $d^\top d\, 1 = 0$,

Metzler: $\Delta f(x) = \lim_{r \to 0} \frac{C}{r^{n+1}} \left[\int_{S_r} f(x) - f(y)\, dy \right] \implies \Delta_{ij} \leq 0, \forall i \neq j$.

Laplacian-like operator

Definition

A *Laplacian-like* operator is an endomorphism $\Delta : L^2_\mu(X) \to L^2_\mu(X)$ verifying the following properties:

- **Symmetry:** $\Delta^{\top_\mu} = \Delta$,
- **Constant cancelation:** $\Delta 1 = 0$,
- **Positivity:** $\langle f, \Delta f \rangle \geq 0 \ \forall f \in L^2_\mu(X)$,
- **Metzler:** $\Delta_{ij} \leq 0 \ \forall i \neq j$.

Diffusion operator

We apply the exponential: $\Delta \rightarrow Q = e^{-t\Delta}$

Definition

A *diffusion* operator is an endomorphism $Q : L^2_\mu(X) \rightarrow L^2_\mu(X)$ verifying the following properties:

- **Symmetry:** $Q^{\top_\mu} = Q$,
- **Constant preservation:** $Q1 = 1$,
- **Damping:** Eigenvalues of Q are in $[0, 1]$,
- **Positive coefficients:** $Qf \geq 0 \; \forall f \geq 0$.

Smoothing operator

Definition

A *smoothing* operator is an endomorphism $S : L^2_\mu(X) \to L^2_\mu(X)$ verifying the following properties:

- **Symmetry:** $S^{\top_\mu} = S$,
- **Positivity:** Eigenvalues of S are in $[0, +\infty[$,
- **Positive coefficients:** $Sf \geq 0 \ \forall f \geq 0$.

Symmetric normalization

Require: Smoothing matrix $S \in \mathbb{R}^{N\times N}$

1: Initialize $\Lambda \leftarrow I_N$
2: **while** $\sum_i |\Lambda_{ii} \sum_j S_{ij}\Lambda_{jj} - 1|$ is larger than a tolerance parameter **do**
3: $\quad d_i \leftarrow \sum_j S_{ij}\Lambda_{jj}$
4: $\quad \Lambda_{ii} \leftarrow \sqrt{\Lambda_{ii}/d_i}$
5: **end while**
6: **return** $Q = \Lambda S \Lambda$

Theorem

Let X be a space equipped with a discrete measure $\mu = \sum_i m_i \delta_{x_i}$, and let S be a smoothing operator on $L^2_\mu(X)$. The symmetric Sinkhorn algorithm converges to a unique positive diagonal matrix Λ such that $\Lambda S \Lambda$ is a diffusion operator.

Fixed-scale convergence

Hypotheses

$X \subset \mathbb{R}^d$ bounded, $(\mu^n)_{n\in\mathbb{N}}$ sequence of discrete measures with $\mathrm{supp}(\mu^n) \subseteq X$

$$\mu^n = \sum_{i=1}^{N_n} m_i^n \, \delta_{x_i^n} \rightharpoonup_{\text{weakly}} \mu$$

$k(x, y)$ Gaussian / exponential kernel with σ **fixed**

$$(S^n)_{ij} = k(x_i^n, x_j^n) m_j^n, \; Q^n = \Lambda^n S^n \Lambda^n$$

$$Q^n(f)(x) = \sum_{j=1}^{N_n} \left[\lambda^n(x) k(x, x_j^n) \lambda^n(x_j)\right] m_j^n f(x_j^n)$$

Fixed-scale convergence

Result

Then there exists a continuous function $\lambda : \mathbb{R}^+ \to X$ such that Q^n converges pointwise to

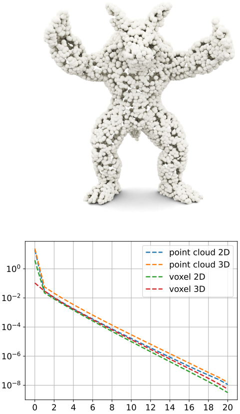
$$Q : f \in L^2_\mu(X) \to (Qf)(x) = \int_X [\lambda(x)k(x,y)\lambda(y)]\, f(y) d\mu(y)$$

in the following sense: for all f continuous,

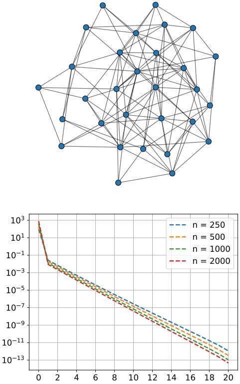
$$Q^n(f) \xrightarrow{\text{unif}} Q(f).$$

Quickness of convergence

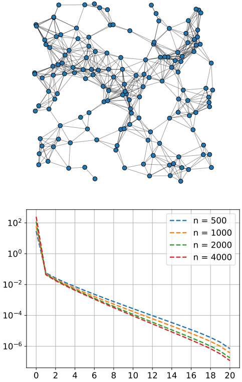
In practice, 5-10 iterations are enough



(a) Armadillo



(b) Erdős–Rényi



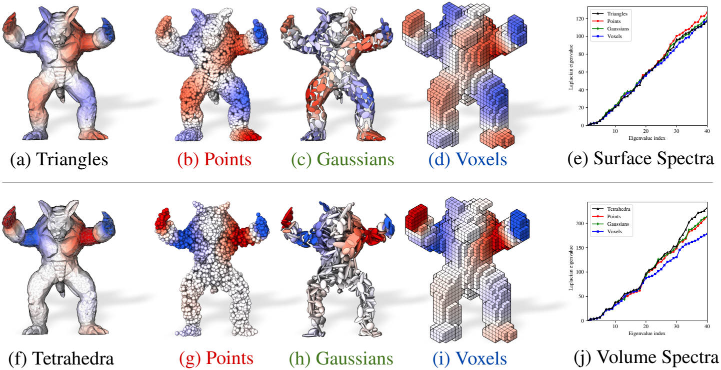
(c) Geometric

Quickness of computation

Left: 5 Sinkhorn iterations with S a point cloud kernel, using PyKeOps
Right: approximation $e^{-t\Delta} = (\mathrm{Id} - t\Delta)^{-1}$ by computing a sparse LU factorization, where Δ is a **sparse** Laplacian (ideal case)

N	**GPU Sinkhorn**	**CPU LU**
10 000	3	65
50 000	21	393
100 000	89	1030
250 000	448	3510
500 000	1817	9100
1 000 000	6789	23 600

Eigenvectors and eigenvalues



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

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

Kernel used in experiments:

$$S_{ij} = m_j \exp\left[-\frac{1}{2}(x_i - x_j)^T(\sigma^2 \,\mathrm{Id} + \Sigma_i + \Sigma_j)^{-1}(x_i - x_j)\right]$$

Sobolev metric "correction"

Framework

We work here with **vector fields** (and not scalar fields) defined over a point cloud.

Considered objects

M : **Wasserstein** metric, $M_{ij} = \delta_{ij}\,\mathrm{Id}$

K : **LDDMM / Sobolev** cometric, $K_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)\mathrm{Id}$

Gradient descent

We want to perform gradient descent over a functional f :

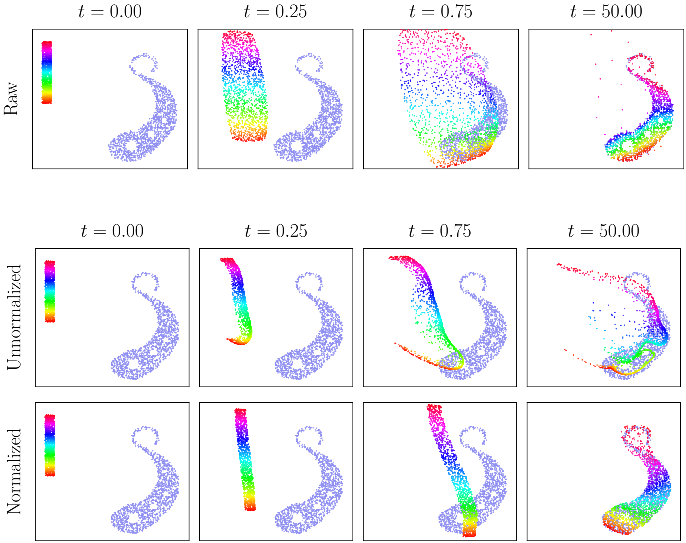
$$\nabla_{K^{-1}}(df) = K(df) = \underbrace{KM}_{\text{smoothing operator}} \Big(\underbrace{M^{-1}df}_{\nabla_M(df)} \Big).$$

$$\underbrace{KM}_{\text{smoothing}} \longrightarrow \underbrace{\Lambda KM\Lambda}_{\text{diffusion}}, \quad \tilde{K} = (\Lambda KM\Lambda)\, M^{-1} = \Lambda K\Lambda.$$

New metric

Diffusion allows to obtain a new metric $\tilde{K}^{-1}$ which has the same notion of mass as M, and still penalize rough vector fields more than M:

$$\langle v, 1\rangle_{\tilde{K}^{-1}} = \langle v, 1\rangle_M\,, \quad \|v\|^2_{\tilde{K}^{-1}} \geq \|v\|^2_M\,.$$



$t = 0.00$
$t = 0.25$
$t = 0.75$
$t = 50.00$
Raw
Unnormalized
Normalized

Smoothing operator

Definition

A *smoothing* operator is an endomorphism $S : L^2_\mu(X) \to L^2_\mu(X)$ verifying the following properties:

- **Symmetry:** $S^{\top_\mu} = S$,
- **Positivity:** Eigenvalues of S are in $[0, +\infty[$,
- **Positive coefficients:** $Sf \geq 0 \; \forall f \geq 0$.

Diffusion operator

Definition

A *diffusion* operator is an endomorphism $Q : L^2_\mu(X) \to L^2_\mu(X)$ verifying the following properties:

- **Symmetry:** $Q^{\top_\mu} = Q$,
- **Constant preservation:** $Q1 = 1$,
- **Damping:** Eigenvalues of Q are in $[0, 1]$,
- **Positive coefficients:** $Qf \geq 0 \; \forall f \geq 0$.

Smoothing cometric

$$\underbrace{S}_{\text{operator}} \underset{(-)M}{\overset{(-)M^{-1}}{\rightleftharpoons}} \underbrace{K}_{\text{cometric}}$$

Definition

A *smoothing* cometric is a bilinear form $K : L^2_\mu(X)^* \otimes L^2_\mu(X)^* \to \mathbb{R}$ verifying the following properties:

Symmetry: $K(f^*, g^*) = K(g^*, f^*)$,

Positivity: $\langle f^*, f^* \rangle_K \geq 0$,

Positive coefficients: $\langle f^*, g^* \rangle_K \geq 0$ for all $f, g \geq 0$.

Diffusion cometric

$$\underbrace{Q}_{\text{operator}} \underset{(-)M}{\overset{(-)M^{-1}}{\rightleftharpoons}} \underbrace{\tilde{K}}_{\text{cometric}}$$

Definition

A *diffusion* cometric is a bilinear form $\tilde{K} : L^2_\mu(X)^* \otimes L^2_\mu(X)^* \to \mathbb{R}$ verifying the following properties:

- **Symmetry:** $\tilde{K}(f^*, g^*) = \tilde{K}(g^*, f^*)$,
- **Constant preservation:** $\langle f, 1 \rangle_{\tilde{K}^{-1}} = \langle f, 1 \rangle_M$,
- **Damping:** $\|f\|^2_{\tilde{K}^{-1}} \geq \|f\|^2_M$,
- **Positive coefficients:** $\langle f^*, g^* \rangle_{\tilde{K}} \geq 0$ for all $f^*, g^* \geq 0$.

Smoothing metric

$$\underbrace{S}_{\text{operator}} \underset{M^{-1}(-)}{\overset{M(-)}{\rightleftharpoons}} \underbrace{G}_{\text{metric}}$$

Definition

A *smoothing* metric is a bilinear form $G : L^2_\mu(X) \otimes L^2_\mu(X) \to \mathbb{R}$ verifying the following properties:

Symmetry: $G(f, g) = G(g, f)$,

Positivity: $\langle f, f \rangle_G \geq 0$,

Positive coefficients: $\langle f, g \rangle_G \geq 0$ for all $f, g \geq 0$.

Diffusion metric

$$\underbrace{Q}_{\text{operator}} \underset{M^{-1}(-)}{\overset{M(-)}{\rightleftharpoons}} \underbrace{\tilde{G}}_{\text{metric}}$$

Definition

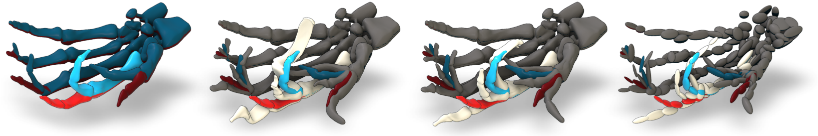
A *diffusion* metric is a bilinear form $\tilde{G} : L^2_\mu(X) \otimes L^2_\mu(X) \to \mathbb{R}$ verifying the following properties:

Symmetry: $\tilde{G}(f, g) = \tilde{G}(g, f)$,

Constant preservation: $\langle f, 1 \rangle_{\tilde{G}} = \langle f, 1 \rangle_M$,

Damping: $\|f\|^2_{\tilde{G}} \leq \|f\|^2_M$,

Positive coefficients: $\langle f, g \rangle_{\tilde{G}} \geq 0$ for all $f, g \geq 0$.



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

Conclusion

Unnormalized operator $\rightarrow$ smoothing operator
Fast / works for many modalities
Behaviour similar to the solution of the heat equation
Appears in rather unexpected contexts (LDDMM, diffusion of (co)-metrics)

Take-away

Maybe you work with an unnormalized operator?
Normalize it! It costs only a few Sinkhorn iterations

Thank you for your attention!