Applications of No-Collision Transportation Maps in Manifold Learning

Elisa Negrini
Department of Mathematics, UCLA

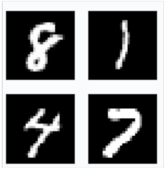
Joint with Levon Nurbekyan (Emory University)

Women in OT meeting, UBC Vancouver
April 18, 2024

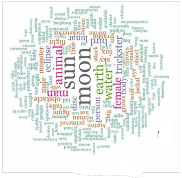
Learning using Distributions

We will work with data that come in the form of distributions.

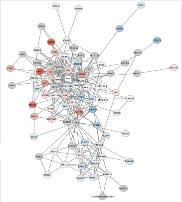
Images as distributions



Text as distribution



Gene Expression data



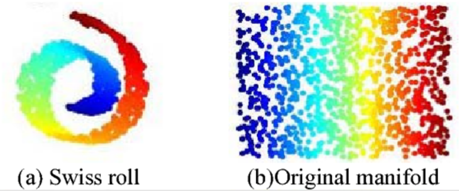
Optimal Transport provides a natural geometry to compare probability measures.

Unsupervised Learning

Given distributions $\mu_i \in \mathcal{P}$, $i = 1, \ldots, N$ discover the underlying structure or patterns in the data.

Manifold Learning

Find a low-dimensional representation of high-dimensional data that preserves the underlying structure or geometry of the data.
May require all the pairwise distances $d(\mu_i, \mu_j)$.



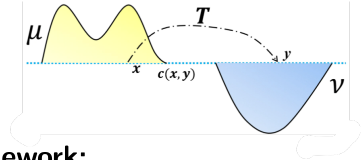
Gu, Rui-jun, and Wenbo Xu. "An Improved Manifold Learning Algorithm for Data Visualization."

We use optimal transport-like maps called *no-collision transportation maps* [8] to solve manifold learning tasks.

[8] Nurbekyan L., Iannantuono A., Oberman A., 2020

Optimal Transport: Monge Formulation

Optimal transport is the general problem finding the most efficient way to move one distribution of mass to another. (Monge 1781)



Mathematical framework:
Find $T : \mathbb{R}^d \to \mathbb{R}^d$ that minimizes the cost $c(x, y)$ to move μ into ν:

$$\inf_T \left\{ \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x) \quad : \quad \nu(B) = \mu(T^{-1}(B)) \ \forall \text{ Borel sets } B \right\}$$

A common choice for the cost is $c(x, T(x)) = \|T(x) - x\|_2^2$. In this case the minimum is known as the squared **2-Wasserstein distance,** W_2.

Pros and Cons of Optimal Transport Distances

Pros:

- W_2 defines a distance and Riemannian structure on the space of probability measures [2].
- The OT distance is sensitive to geometric features of the measures being transported (e.g. the OT map between translated measures is the translation).
- We have a good understanding of theoretical properties [12, 10, 13].

Cons:

- OT maps are expensive to calculate and normally require global optimization.

---

[12] Villani, C. 2009, [2] Ambrosio L., Gigli N., Savaré G. 2017 [8] Peyré, G., Cuturi, M. 2019, [13] Villani, C. 2021

Our Goal

Questions:

1. Can we come up with transport-like maps and distances that are cheaper to compute but retain advantageous properties of optimal ones?
2. Can we use these maps in learning tasks [7]?

Prior Work:

- Linear Optimal Transport (LOT) [14, 5]
- Cumulative Distribution Transform (CDT) [9]
- The Radon cumulative distribution transform (Radon-CDT) [6]
- No-Collision Transportation maps [8]

---

[14] Wang W., Slepčev D., Basu S., Ozolek J.A., Rohde G.K. 2013;
[6] Kolouri S., Park S.R., Rohde G.K. 2015;
[7] Kolouri S., Park S.R., Thorpe M., Slepčev D., Rohde G.K. 2016;
[9] Park S.R., Kolouri S., Kundu S., Rohde G.K. 2018;
[8] Nurbekyan L., Iannantuono A., Oberman A., 2020;
[5] Khurana V., Kannan H., Cloninger A., Moosmüller C. 2023

This Work

- Inspired by Wasserstein Isometric Mapping (Wassmap) [4] and by its linearized version [3, 5], we perform manifold learning using Multidimensional Scaling (MDS) on no-collision distances.

- We prove that no-collision distances accurately capture translations and dilations of a given probability measure.

- In contrast, we prove that OT, LOT and no-collision maps are not able to capture rotations.

---

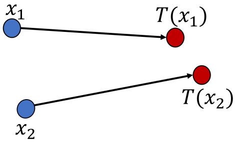
[14] Wang W., Slepčev D., Basu S., Ozolek J.A., Rohde G.K. 2013;
[4] Hamm K., Henscheid N., Shujie K. 2022;
[5] Khurana V., Kannan H., Cloninger A., Moosmüller C., 2023;
[3] Cloninger A., Hamm K., Khurana V., Moosmüller C., 2023

No-Collision Transport Maps

Assume that $X \subseteq \mathbb{R}^d$, and $T : X \to \mathbb{R}^d$.

Definition: We say that T has the *no-collision* property if $\forall\, x_1, x_2 \in X$ such that $x_1 \neq x_2$:

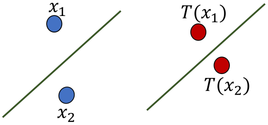
$$(1-s)x_1 + sT(x_1) \neq (1-s)x_2 + sT(x_2)\ \forall s \in (0,1).$$



Definition: We say that T is *half-space preserving* if $\forall\, x_1, x_2 \in X$ such that $x_1 \neq x_2$ there exists $v \in \mathbb{R}^d$ such that

$$(x_2 - x_1)\cdot v \leq 0, \quad (T(x_2) - T(x_1))\cdot v \leq 0,$$

and at least one of the inequalities is strict.



Remark (Ambrosio et al. 2008 [2])

OT maps with $c(x,y) = |x-y|^p,\ p > 1$ *have the no-collision property.*

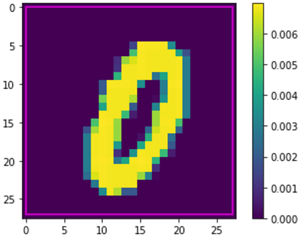
Theorem (Nurbekyan et al. 2020 [8])

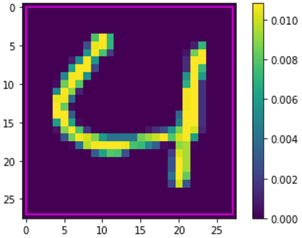
T *has the no-collision property if and only if it is half-space-preserving.*

No-Collision Transport Maps: the Algorithm

Goal: Build no-collision maps between distributions μ and ν based on the half-space preserving property.

1. Let $\mu \in \mathcal{P}(\Omega)$, define $\Omega_0 = \Omega$ and $\mathcal{C}_0 = \{\Omega_0\}$
   Let $\nu \in \mathcal{P}(\Omega)$, define $\Omega_0' = \Omega$ and $\mathcal{C}_0' = \{\Omega_0'\}$

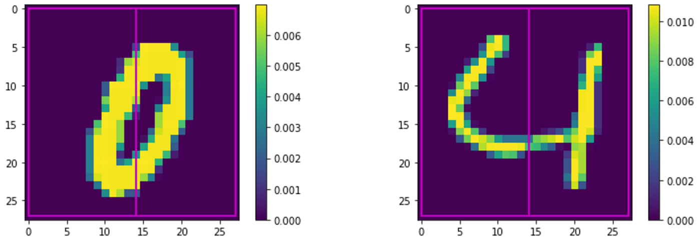




No-Collision Transport Maps: the Algorithm

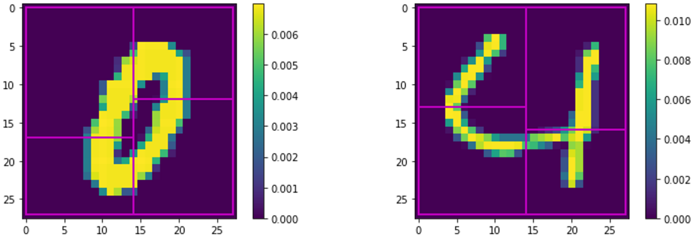
2. Choose a slicing direction $s_1 \in \mathbb{S}^{d-1}$ and find an hyperplane that divides Ω_0 into two parts Ω_{00} and Ω_{01} such that $\mu(\Omega_{00}) = \mu(\Omega_{01}) = \frac{1}{2}$. Define $\mathcal{C}_1 = \{\Omega_{00}, \Omega_{01}\}$.

   Using the same slicing direction, do the same for ν to obtain $\Omega'_{00},\ \Omega'_{01},\ \mathcal{C}'_1 = \{\Omega'_{00}, \Omega'_{01}\}$



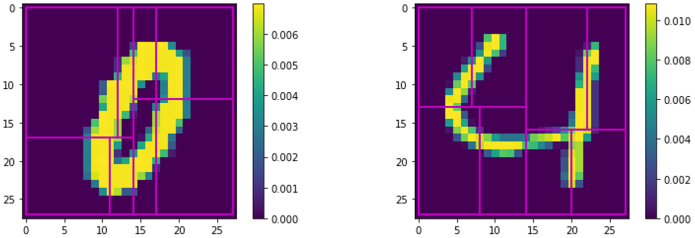
No-Collision Transport Maps: the Algorithm

3. Continue this slicing procedure by slicing each set in $\mathcal{C}_i$ and $\mathcal{C}'_i$ into two parts with equal masses.
   At each step use the same slicing direction for μ and ν.



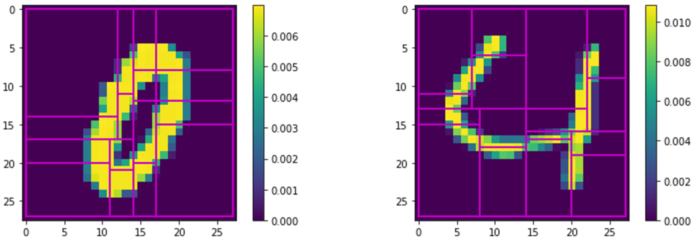
No-Collision Transport Maps: the Algorithm

3. Continue this slicing procedure by slicing each set in $\mathcal{C}_i$ and $\mathcal{C}'_i$ into two parts with equal masses.
   At each step use the same slicing direction for μ and ν.



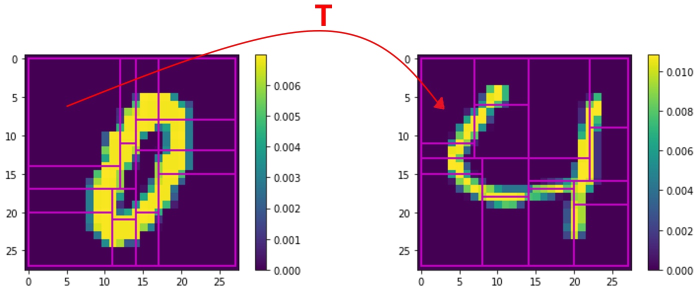
No-Collision Transport Maps: the Algorithm

4. At step N we obtain $N+1$ subsets $\mathcal{C}_i = \{\Omega_b\}$ and $\mathcal{C}'_i = \{\Omega'_b\}$ which form a partition of Ω and for which $\mu(\Omega_b) = \nu(\Omega'_b) = \frac{1}{2^N}$



No-Collision Transport Maps: the Algorithm

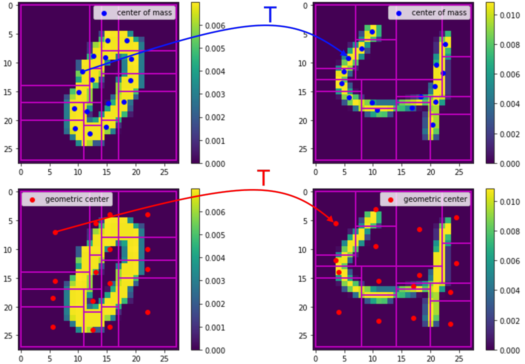
5. In the limit as $N \to +\infty$ we define a no-collision map $T : \Omega \to \Omega$ so that it respects the resulting partitions by matching corresponding leaves in *supp*(μ) and *supp*(ν) that is $T(\Omega_b) \subset \tilde{\Omega}_b$ for all b.



No-Collision Transport Maps: the Algorithm

6. In the discrete setting, for each Ω_b and Ω'_b we denote by c_b and c'_b their "center". In this way we obtain collections $C = \{c_b\}$ and $C' = \{c'_b\}$ that represent respectively the features of μ and ν.

   $T : C \to C'$ such that $T(c_b) = c'_b, \forall b$ is an approximation of the no-collision map.



Pros and Cons

Pros:

- The construction does not involve optimization: only a median search. [1, 11]
- No-collision maps provide comparable results as other optimal transport based methods using less computational time [4, 5].

Cons:

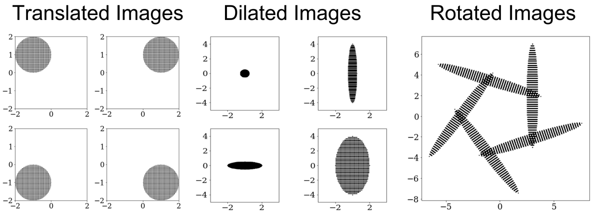
- Since no optimization is involved, these maps are not optimal in general. In some cases, however, the sub-optimality is not severe. Some examples later and in [8, Section 4].
- It is unclear how to pick the number and direction of the cuts.

---

[1] M. Ajtai, J. Komlós, G. Tusnády, 1984; [11] N. G. Trillos, D. Slepčev, 2015; [8] Nurbekyan L., Iannantuono A., Oberman A., 2020; [4] Hamm, K., Henscheid, N., Kang, S. 2022; [5] Khurana, V., Kannan, H., Cloninger, A., Moosmüller, C. 2023

Manifold Learning

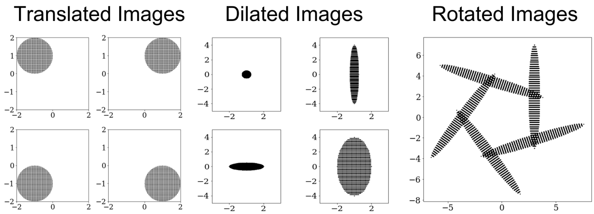
- Take μ_0 the uniform measure on a unit disc and consider its translations and dilations, get $\{\mu\}_{i=1}^{N}$.



[4] Hamm K., Henscheid N., Shujie K. 2022;

Manifold Learning

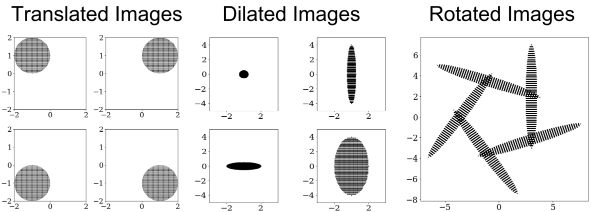
- Take μ_0 the uniform measure on a unit disc and consider its translations and dilations, get $\{\mu\}_{i=1}^N$.
- Take μ_0 the uniform measure on an ellipse and consider its rotations, get $\{\mu\}_{i=1}^N$.



[4] Hamm K., Henscheid N., Shujie K. 2022;

Manifold Learning

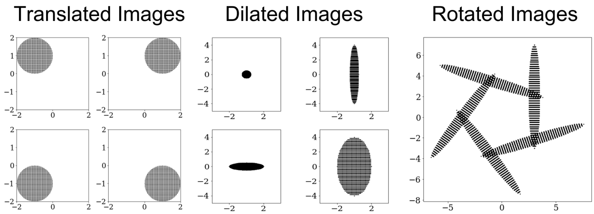
- Take μ_0 the uniform measure on a unit disc and consider its translations and dilations, get $\{\mu\}_{i=1}^{N}$.
- Take μ_0 the uniform measure on an ellipse and consider its rotations, get $\{\mu\}_{i=1}^{N}$.
- Given $\{\mu\}_{i=1}^{N}$ choose a distance and build a distance matrix $D = (d(\mu_i, \mu_j))_{i,j=1}^{N}$



[4] Hamm K., Henscheid N., Shujie K. 2022;

Manifold Learning

- Take μ_0 the uniform measure on a unit disc and consider its translations and dilations, get $\{\mu\}_{i=1}^N$.
- Take μ_0 the uniform measure on an ellipse and consider its rotations, get $\{\mu\}_{i=1}^N$.
- Given $\{\mu\}_{i=1}^N$ choose a distance and build a distance matrix $D = (d(\mu_i, \mu_j))_{i,j=1}^N$
- Run a manifold learning algorithm such as MDS on D.



[4] Hamm K., Henscheid N., Shujie K. 2022;

Theoretical Results: Translations

Let $\mathcal{P}_{ac}(\mathbb{R}^d)$ the set of Borel probability measures over $\mathbb{R}^d$ that are absolutely continuous with respect to the Lebesgue measure.

Theorem (Translation Manifold (Negrini-Nurbekyan'23))

Assume that $\mu_0 \in \mathcal{P}_{ac}(\mathbb{R}^d)$. *Let* $\mu_\theta = (x+\theta)\sharp\mu_0$ *for* $\theta \in \mathbb{R}^d$. *Then for every slicing schedule* $\mathcal{S}$ *we have that*

$$W_{\mathcal{S},p}(\mu_\theta, \mu_{\theta'}) = |\theta - \theta'|, \quad \forall \theta, \theta' \in \mathbb{R}^d,\ p \geq 1.$$

In particular, $(\{\mu_\theta\}, W_{\mathcal{S},p})$ *is isometric to* $(\Theta, |\cdot|)$.

A similar result can be proven for Dilations.

Theoretical Results: Rotations

Denote by R_t the counter-clockwise rotation by angle t around the origin; that is, $R_t x = (x_1 \cos t - x_2 \sin t, x_1 \sin t + x_2 \cos t)$.

Theorem (Rotation Manifold (Nurbekyan, Negrini '23))

Assume that μ_0 is a uniform measure over an elliptical domain

$$\mathcal{E} = \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{(x_1 - u_1)^2}{a^2} + \frac{(x_2 - u_2)^2}{b^2} \leq 1\right\},$$

where $u = (u_1, u_2) \neq 0$, and $a, b > 0$. Furthermore, assume that $\mu_t = (R_t x)\sharp\mu_0$, and $\mathcal{S}$ is a slicing schedule. Then $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_{\mathcal{S},2}\right)$ is isometric to a circle if and only if $a = b$.

Similar results hold if one uses OT or LOT distances.

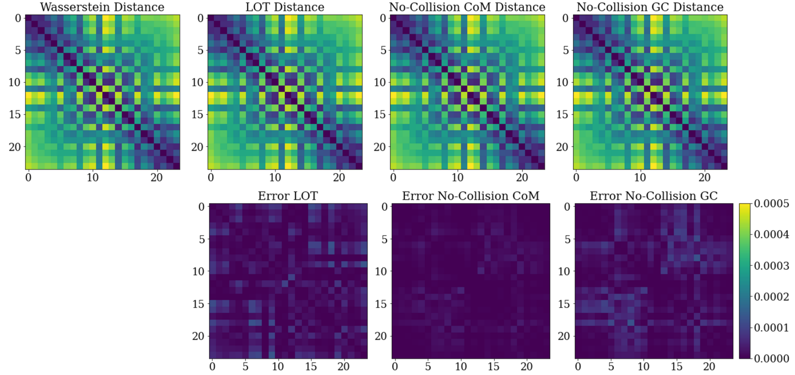
Manifold Learning: Translation

Goal: Reconstruct the underlying grid governing a translation manifold using MDS on OT, LOT, no-collision (with 2 cuts) and Euclidean distances.

Manifold Learning: Translation

Goal: Reconstruct the underlying grid governing a translation manifold using MDS on OT, LOT, no-collision (with 2 cuts) and Euclidean distances.

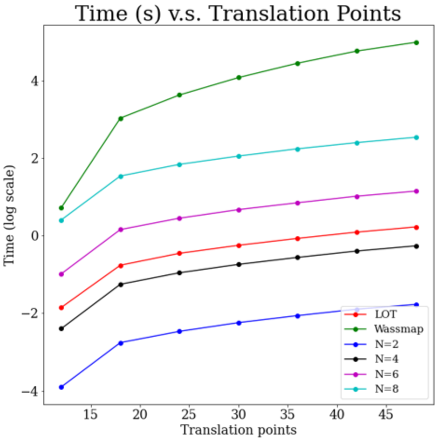
- How well do LOT and no-collision approximate OT distance?



Relative Error: **0.879%** **0.493%** **0.861%**

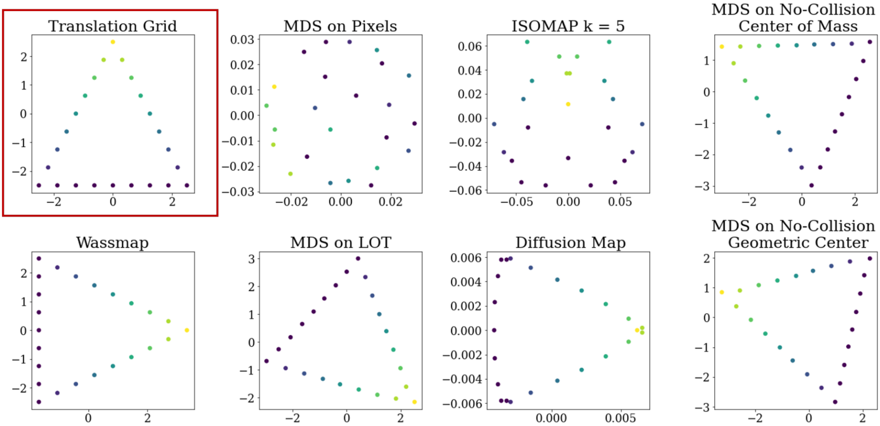
Manifold Learning: Translation

- How fast are the different distance matrix computations?



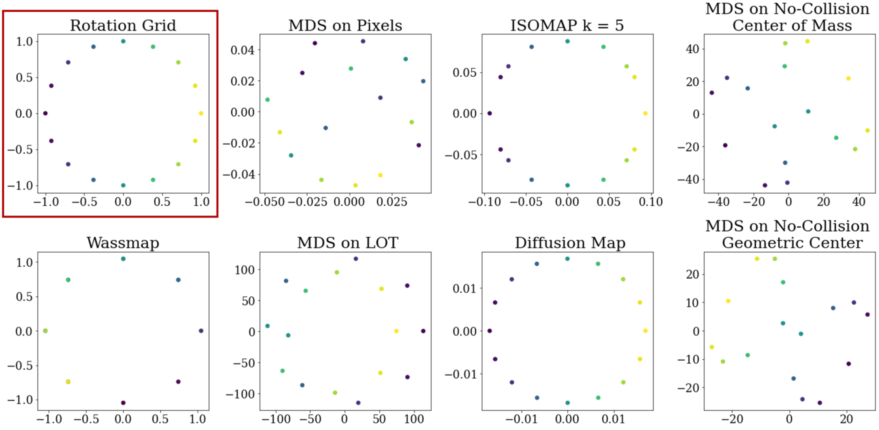
Manifold Learning: Translation

- How good is the manifold reconstruction?



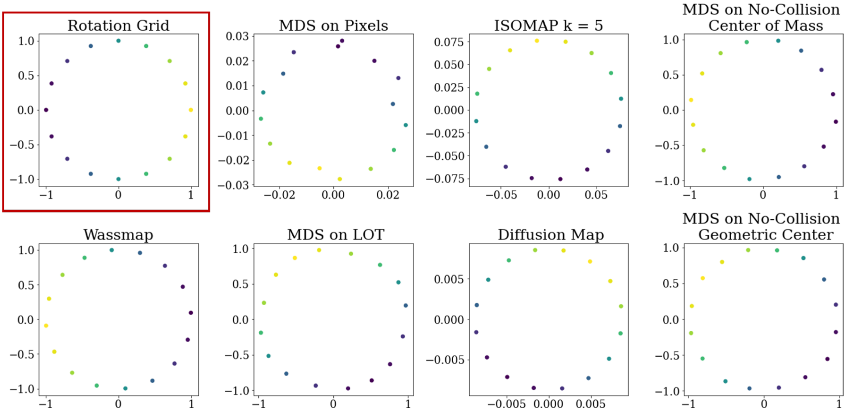
Manifold Learning: Rotation

In general we have no isometry in the case of rotations...



Manifold Learning: Rotation

... Unless we are rotating a circular domain



Conclusion:

- No-collision maps are fast to compute and in certain cases attain nearly optimal costs.
- They attain similar results on manifold learning tasks as other optimal transport based methods, but require less computational time and in some cases attain better OT distance approximations

Future Work:

- Use different cut directions (choose angles randomly at each step).
- Optimize cut directions to minimize transportation cost.
- Explore other learning problems such as classification and clustering using no-collision distances

References I

[1] Miklós Ajtai, János Komlós, and Gábor Tusnády.
On optimal matchings.
Combinatorica, 4:259–264, 1984.

[2] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré.
Gradient flows in metric spaces and in the space of probability measures.
Lectures in Mathematics ETH Zürich. Birkhäuser Verlag, Basel, second edition, 2008.

[3] Alexander Cloninger, Keaton Hamm, Varun Khurana, and Caroline Moosmüller.
Linearized wasserstein dimensionality reduction with approximation guarantees.
arXiv preprint arXiv:2302.07373, 2023.

[4] Keaton Hamm, Nick Henscheid, and Shujie Kang.
Wassmap: Wasserstein isometric mapping for image manifold learning, 2022.

[5] Varun Khurana, Harish Kannan, Alexander Cloninger, and Caroline Moosmüller.
Supervised learning of sheared distributions using linearized optimal transport.
Sampling Theory, Signal Processing, and Data Analysis, 21(1):1–51, 2023.

[6] Soheil Kolouri, Se Rim Park, and Gustavo K Rohde.
The radon cumulative distribution transform and its application to image classification.
IEEE transactions on image processing, 25(2):920–934, 2015.

[7] Soheil Kolouri, Serim Park, Matthew Thorpe, Dejan Slepčev, and Gustavo K Rohde.
Transport-based analysis, modeling, and learning from signal and data distributions.
arXiv preprint arXiv:1609.04767, 2016.

[8] Levon Nurbekyan, Alexander Iannantuono, and Adam M. Oberman.
No-collision transportation maps.
Journal of Scientific Computing, 82(2):45, 2020.

References II

[9] Se Rim Park, Soheil Kolouri, Shinjini Kundu, and Gustavo K Rohde.
The cumulative distribution transform and linear pattern classification.
Applied and computational harmonic analysis, 45(3):616–641, 2018.

[10] Gabriel Peyré, Marco Cuturi, et al.
Computational optimal transport: With applications to data science.
Foundations and Trends® in Machine Learning, 11(5-6):355–607, 2019.

[11] Nicolás Garcia Trillos and Dejan Slepčev.
On the rate of convergence of empirical measures in *infty*-transportation distance.
Canadian Journal of Mathematics, 67(6):1358–1383, 2015.

[12] Cédric Villani.
Optimal transport: old and new, volume 338.
Springer, 2009.

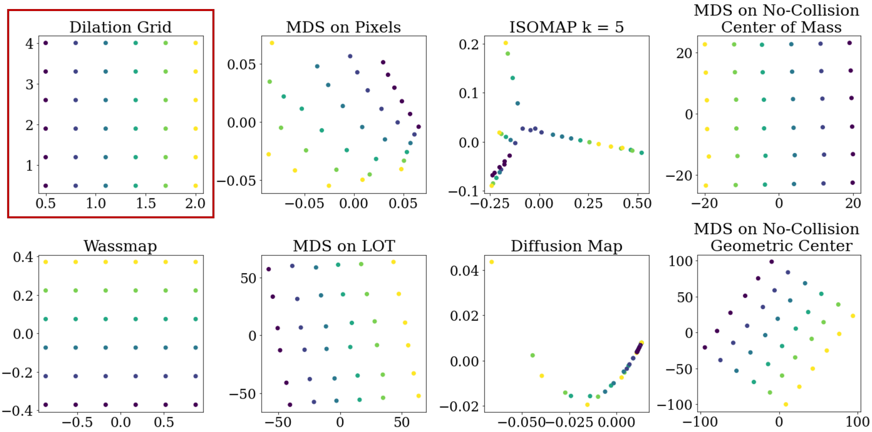
[13] Cédric Villani.
Topics in optimal transportation, volume 58.
American Mathematical Soc., 2021.

[14] Wei Wang, Dejan Slepčev, Saurav Basu, John A. Ozolek, and Gustavo K. Rohde.
A linear optimal transportation framework for quantifying and visualizing variations in sets of images.
International Journal of Computer Vision, 101(2):254–269, 2013.

Manifold Learning: Dilation

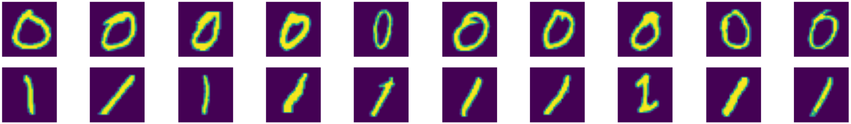
Goal: Reconstruct the underlying grid governing a dilation manifold.

We compare the embeddings given by Wassmap, Multidimensional Scaling (MDS) on pixels, on LOT features and on the no-collision features for 3 cuts.



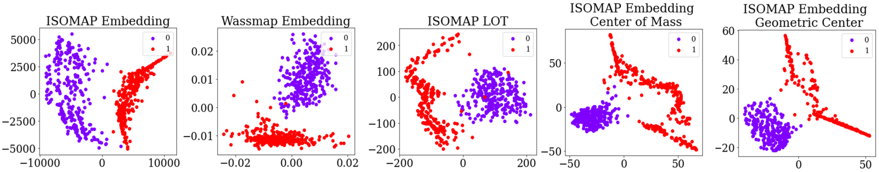
Clustering: MNIST digits

We randomly sample 300 handwritten 0s and 1s from MNIST and compare 2D embeddings for Wassmap, ISOMAP on pixels and LOT and no-collision features for 5 no-collision cuts.



Clustering: MNIST digits

The points are colored according to their class label.

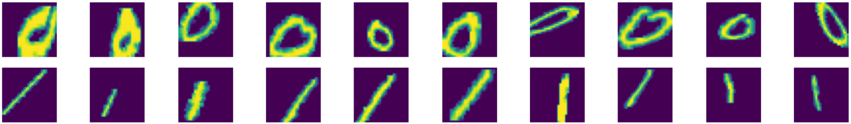


We also compare the computational time for the different methods:

Method	*Wassmap*	*LOT* *1 Gaussian Reference*	*No-collision* $N = 5$
Time (s)	443.3	8.4	9.5

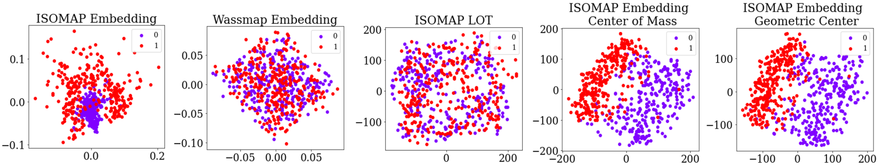
Clustering: Sheared MNIST digits

We randomly sample 300 sheared 0s and 1s from MNIST and compare 2D embeddings for Wassmap, ISOMAP on pixels and LOT and no-collision features for 8 no-collision cuts.



Clustering: Sheared MNIST digits

The points are colored according to their class label.



We also compare the computational time for the different methods:

Method	*Wassmap*	*LOT* *5 Gaussian References*	*No-collision* $N = 8$
Time (s)	443.3	18.3	71.7