Applications of No-Collision Transportation Maps in Manifold Learning

Elisa Negrini and Levon Nurbekyan

Abstract

In this work, we investigate applications of no-collision transportation maps introduced in [Nurbekyan et. al., 2020] in manifold learning for image data. Recently, there has been a surge in applying transportation-based distances and features for data representing motion-like or deformation-like phenomena. Indeed, comparing intensities at fixed locations often does not reveal the data structure. No-collision maps and distances developed in [Nurbekyan et. al., 2020] are sensitive to geometric features similar to optimal transportation (OT) maps but much cheaper to compute due to the absence of optimization. In this work, we prove that no-collision distances provide an isometry between translations (respectively dilations) of a single probability measure and the translation (respectively dilation) vectors equipped with a Euclidean distance. Furthermore, we prove that no-collision transportation maps, as well as OT and linearized OT maps, do not in general provide an isometry for rotations. The numerical experiments confirm our theoretical findings and show that no-collision distances achieve similar or better performance on several manifold learning tasks compared to other OT and Euclidean-based methods at a fraction of a computational cost.

1 Introduction

High-dimensional datasets can be very difficult to visualize and may contain redundant information. In particular, while two- or three-dimensional data can be plotted to show its inherent structure, equivalent high-dimensional plots are much less intuitive. One of the fundamental observations in data science is that high-dimensional data often exhibits low-dimensional structure. In fact, a common assumption in machine learning, known as the *manifold hypothesis*, is that observed data lie on a low-dimensional manifold embedded in a high-dimensional space [28, 21, 11, 16]. Manifold learning is a machine learning approach to non-linear dimension reduction based on the manifold hypothesis. Many manifold learning algorithms have been proposed in recent years including Isomap [38], Diffusion Maps [13], Multidimensional Scaling [27], and Laplacian Eigenmaps [6]. These methods have numerous applications in machine learning, neural networks, pattern recognition, image processing, and computer vision [30].

1

While extremely successful for some applications, most nonlinear dimension reduction methods embed data in Euclidean spaces implicitly assuming that Euclidean distances between data points are semantically meaningful. This assumption, however, may not be valid in general. Indeed, in many cases comparing intensities at fixed locations may not reveal the data structure. For instance, in image classification translated copies of a single image can have large Euclidean distance even though they are semantically identical and should be assigned the same label.

Recent works demonstrate that optimal transportation (OT) and related distances are effective in capturing data structure in various applications yielding somewhat more "semantically meaningful" features and distances. In particular, data corresponding to some physical or material displacement, where data points can be modeled as probability distributions, seem to be more amenable to transportation type methods [25, 32, 20, 23, 12].

In this work, we pursue this previous line of research interpreting images as probability distributions on the pixel grid and performing manifold learning on image data using no-collision distances. The latter are transportation-type distances introduced in [33]. We show that no-collision distances achieve similar or better performance compared to other types of OT distances at a fraction of computational cost. More specifically, we study theoretically and numerically the performance of the Multidimensional Scaling (MDS) [27] algorithm on several examples of manifolds in the space of probability measures when the pairwise distance matrix is computed using no-collision distances. Such analysis has been carried out in [20, 12] with Wasserstein (OT) and linearized Wasserstein distances (LOT), respectively.

We demonstrate that no-collision maps faithfully capture spatial displacements similar to OT maps. At the same time, unlike OT distances, no optimization is required for computing no-collision maps. More precisely, we prove that no-collision transportation maps accurately capture translations and dilations of a given probability measure; that is, the no-collision transportation map between a measure and its translation coincides with the translation map, and the no-collision transportation map between a measure and its dilation coincides with the dilation map.

Consequently, no-collision distances provide an isometry between translations of a single probability measure and the translation vectors equipped with a Euclidean distance. Similarly, no-collision distances provide an isometry between dilations of a single probability measure and dilation vectors equipped with an anisotropic Euclidean distance. Since MDS provides an isometric embedding of data in a Euclidean space, these previous results guarantee that MDS with no-collision distances recovers translations and suitably scaled dilations of a single probability measure up to rigid transformations. These results are analogous to the ones proven for OT distances in [20]. In contrast, LOT distances do not generally provide an isometry for dilations unless the reference measure is itself a dilation.

Furthermore, we prove that no-collision transportation maps do not in general provide an isometry for rotations. More precisely, we provide an explicit

example of a curve in the space of probability measures generated by rotating a single measure that does not admit an isometric image lying in a planar circle when equipped with no-collision distances. Interestingly, we prove that OT and LOT do not generally provide isometries for rotations either. The possibility of these distances recovering rotations via MDS was empirically investigated in [20, 32] with some experiments suggesting otherwise.

As mentioned above, the computation of no-collision distances does not require optimization [33], and one expects faster algorithms when utilizing them instead of optimization-based distances. In our numerical experiments, we compare the performance of no-collision maps with those of OT [20] and LOT [32]. We observe that no-collision distances perform similarly or better in terms of accurately recovering OT distances and manifold structures but are faster to compute.

The rest of the paper is organized as follows. Sections 2 and 3 discuss prior work and background on optimal and no-collision transportation maps, respectively. Next, in Section 4, we prove that no-collision maps recover translations and suitably scaled dilations of a single measure up to rigid transformations. Furthermore, in Section 5, we prove that no-collision, OT, and LOT maps do not generally provide isometries for rotations. We present our numerical experiments in Section 6.

2   Prior work

Manifold learning is a non-linear dimension reduction technique based on the assumption that observed data lie on a low-dimensional manifold embedded in a higher-dimensional space. More specifically, assume (Y, d_Y) is a (possibly high dimensional) metric space where we observe data $\{y_1, \cdots, y_m\} \subset Y$. The goal is to find a metric space (Θ, d_Θ) such that, for some $\{\theta_1, \cdots, \theta_m\} \subset \Theta$ we have $d_Y(y_i, y_j) \approx d_\Theta(\theta_i, \theta_j)$ for all i, j. A common choice is to look for $(\Theta, d_\Theta) = (\mathbb{R}^k, \|\cdot\|_2)$ with small k.

Common dimension reduction techniques include Multidimensional Scaling (MDS) [27], Isomap [38], and Diffusion Maps [13]. MDS aims to embed data in a low-dimensional space so that pairwise Euclidean distances between data points are preserved. Isomap builds on MDS but instead of Euclidean distances between data points it preserves pairwise graph-geodesic distances revealing the geometric structure of the data. Finally, Diffusion Maps embed data in a lower-dimensional space through the spectral decomposition of the diffusion operator on the data. With a few exceptions [23], these methods assume data features in a Euclidean space with underlying Euclidean distance generating graph-geodesics and differential operators.

A more recent idea of using optimal transportation distances in machine learning and data science proved successful in many instances such as automatic translation, shape reconstruction, multi-label classification, brain decoding, manifold learning and more (see for instance [15, 26, 25, 3, 20, 32]). Moreover, optimal transportation framework provides a powerful theoretical frame-

3

work due to the substantial work on optimal transportation related to PDEs and other fields [41, 36, 35, 40].

Recall that the square Wasserstein distance between probability measures μ and ν is given by

$$W_2^2(\mu, \nu) := \min_{T:\, \nu = T\sharp\mu} \int_{\mathbb{R}^d} \|x - T(x)\|_2^2 d\mu(x) \tag{1}$$

where $\nu = T\sharp\mu$ means that T *transports* μ into ν; that is, $\nu(B) = \mu(T^{-1}(B))$ for all Borel subsets B. The solution of (1) is called an optimal transportation map.

In [20], the authors propose a variant of the Isomap [38], called Wassmap, which uses Wasserstein distances instead of Euclidean distances when computing parwise distance matrix for data. The authors prove that for manifolds generated by translations or dilations of a fixed probability measure, Wassmap recovers the translation set or a scaled version of the dilation set up to a rigid transformation. While very effective, Wassmap relies on the computation of Wasserstein distance which can be computationally expensive. Indeed, for each distance computation one has to solve an optimization problem (1). Hence, computing pairwise distances between m distributions amounts to solving $\binom{m}{2}$ optimization problems. In [20] the authors propose to trade off accuracy of the embedding with speed of computation by utilizing approximations of the true Wasserstein distance.

Numerous modifications and approximation methods for W_2 distances appeared recently including Linearized Optimal Transportation (LOT) [42, 22, 32], Cumulative Distribution Transform (CDT) [34], Radon-CDT [24], Sinkhorn distances [14, 2, 8, 29], multiscale methods [37, 19], and no-collision transportation maps [33]. The primary goal of these methods is to mitigate the computational cost imposed by computing OT distance while retaining some of the appealing geometric features of OT maps.

LOT embeds the space of probability distributions in $L^2(\mu; \mathbb{R}^d)$, where μ is the reference measure for building LOT maps. More specifically, we associate $\nu \in \mathcal{P}(\mathbb{R}^d)$ with the solution of (1) denoted by T_μ^ν and define

$$W_{2,\mu}^2(\nu_1, \nu_2) = \int_{\mathbb{R}^d} |T_\mu^{\nu_1}(x) - T_\mu^{\nu_2}(x)|^2 d\nu(x).$$

In [12], the authors use $W_{2,\mu}$ instead of W_2 in the MDS to speed up computations which still require solving m Wasserstein distance computations to compare m distributions. Moreover, the result is dependent on the choice of the reference measure, ν.

CDT is the LOT framework for 1D probability measures. In contrast to the higher-dimensional LOT, the Euclidean distance between two transformed (embedded) signals is the exact square Wasserstein distance. Radon-CDT is an extension of the CDT framework to 2D density functions based on sliced Wasserstein distance [9]. The idea of sliced Wasserstein distance is to compute infinitely many linear projections of the high-dimensional distribution to

one-dimensional distributions and then compute the average of the Wasserstein distance between these one-dimensional representations.

In this work, we use no-collision maps introduced in [33]. These maps preserve geometric properties of OT maps, such as no-collision, but are much faster to compute since they do not require optimization. More specifically, the method is based on a partition of probability measures into parts with equal masses via slicing in a sequence of directions chosen beforehand. While not optimal in general, the suboptimality of no-collision maps is often not severe as can be seen in Section 6 and [33, Section 4]. Additionally, we prove in Section 4 that these maps are optimal if the two measures are a translation or dilation of one another.

Partitions with equal masses were also used in [1, 39] for proving optimal rates of convergence of empirical measures in OT distances. One difference is that no-collision maps operate on the continuum level and match points in the supports of two absolutely continuous measures. Additionally, the existence of these maps is guaranteed beyond domains with Lipschitz boundary: see Theorem 2 and [33, Theorem 3].

3 Preliminaries on optimal and no-collision transportation maps

Here we present key definitions and results about optimal transportation and no-collision maps that we use.

3.1 Optimal transportation

We denote by $\mathcal{P}(\mathbb{R}^d)$ the set of Borel probability measures on $\mathbb{R}^d$. Furthermore, we denote by

$$\mathcal{P}_2(\mathbb{R}^d) = \left\{ \mu \in \mathcal{P}(\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} |x|^2 d\mu(x) < \infty \right\}.$$

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

Definition 1. *Let $\mu \in \mathcal{P}(\mathbb{R}^d)$ and $T : \mathbb{R}^d \to \mathbb{R}^d$ is a Borel measurable map. The pushforward of μ via T, denoted by $T\sharp\mu$, is defined as*

$$T\sharp\mu(B) := \mu(T^{-1}(B)) \quad \forall B \subset \mathbb{R}^d \text{ Borel.}$$

For probability measures $\mu, \nu \in \mathcal{P}(\mathbb{R}^d)$ we say that T transports μ into ν if $T\sharp\mu = \nu$

Assume that $c : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ is a Borel measurable nonnegative function so that $c(x, y)$ is the cost of transporting a unit mass from x to y. Then the problem of optimally transporting μ to ν can be written as

$$\min_{T: \nu = T\sharp\mu} \int_{\mathbb{R}^d} c(x, T(x)) d\mu(x). \tag{2}$$

5

Minimizers of this problem are called *optimal transportation maps*. In this paper, we consider the *quadratic* cost $c(x, y) = |x - y|^2$ that yields

$$W_2^2(\mu, \nu) := \min_{T:\, \nu = T \sharp \mu} \int_{\mathbb{R}^d} |x - T(x)|^2 d\mu(x). \tag{3}$$

$W_2(\mu, \nu)$ is called the *square Wasserstein distance* between the measures μ, ν. Since among OT distances we only consider W_2 we refer to it simply as the Wasserstein distance.

In general, (2) and (3) might not have solutions or solutions might not be unique. The following theorem by Brenier provides conditions under which (3) has a unique solution.

Theorem 1 (Brenier [10]). *Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$. If $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ then (3) admits a μ a.e. unique solution. Moreover, there exists a convex function ϕ such that $T(x) = \nabla \phi(x)$ μ a.e..*

Generalizations of this theorem to Riemannian manifolds and with more general cost functions can be found for example in [31, 4, 41].

3.2   No-collision transportation maps

Assume that $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and consider

$$\mathcal{C}_0^\mu = \{\Omega_0^\mu\},$$

where we denote by $\Omega_0^\mu = \mathbb{R}^d$. Next, put $b_1 = 0$, and let $s_1 \in \mathbb{S}^{d-1}$. Since $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ there exists $h_1 \in \mathbb{R}$ such that

$$\mu(x \cdot s_1 \le h_1) = \mu(x \cdot s_1 > h_1) = \frac{1}{2}.$$

Then we set

$$\Omega_{00}^\mu = \{x \cdot s_1 \le h_1\}, \quad \Omega_{01}^\mu = \{x \cdot s_1 > h_1\}, \quad \mathcal{C}_1^\mu = \{\Omega_{00}^\mu, \Omega_{01}^\mu\}.$$

Next, we choose $b_2 \in \{00, 01\}$, and $s_2 \in \mathbb{S}^{d-1}$. There exists $h_2 \in \mathbb{R}$ such that

$$\mu(x \cdot s_2 \le h_2 \mid \Omega_{b_2}^\mu) = \mu(x \cdot s_2 > h_2 \mid \Omega_{b_2}^\mu) = \frac{1}{2}.$$

Then we set

$$\Omega_{b_2 0}^\mu = \{x \cdot s_2 \le h_2\} \cap \Omega_{b_2}^\mu, \quad \Omega_{b_2 1}^\mu = \{x \cdot s_2 > h_2\} \cap \Omega_{b_2}^\mu,$$

and

$$\mathcal{C}_2^\mu = \mathcal{C}_1^\mu \setminus \Omega_{b_2}^\mu \cup \{\Omega_{b_2 0}^\mu, \Omega_{b_2 1}^\mu\}.$$

Going forward, at step $k+1$ we choose b_{k+1} such that $\Omega_{b_{k+1}}^\mu \in \mathcal{C}_k^\mu$, and $s_{k+1} \in \mathbb{S}^{d-1}$. Then there exists $h_{k+1} \in \mathbb{R}$ such that

$$\mu\left(x \cdot s_{k+1} \le h_{k+1} \mid \Omega_{b_{k+1}}^\mu\right) = \mu\left(x \cdot s_{k+1} > h_{k+1} \mid \Omega_{b_{k+1}}^\mu\right) = \frac{1}{2}.$$

Then we set

$$\Omega^{\mu}_{b_{k+1}0} = \{x \cdot s_{k+1} \le h_{k+1}\} \cap \Omega^{\mu}_{b_{k+1}}, \quad \Omega^{\mu}_{b_{k+1}1} = \{x \cdot s_{k+1} > h_{k+1}\} \cap \Omega^{\mu}_{b_{k+1}},$$

and

$$\mathcal{C}^{\mu}_{k+1} = \mathcal{C}^{\mu}_{k} \setminus \Omega^{\mu}_{b_{k+1}} \cup \{\Omega^{\mu}_{b_{k+1}0}, \Omega^{\mu}_{b_{k+1}1}\}.$$

We call the sequence $\mathcal{S} = ((b_k, s_k))_{k=1}^{\infty}$ a *slicing schedule*.

To introduce no-collision distances and maps, we fix a slicing schedule $\mathcal{S}$ and we slice all $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ according $\mathcal{S}$. Loosely speaking, given μ, ν and their partitions $\bigcup_{k=1}^{\infty} \mathcal{C}^{\mu}_k$ and $\bigcup_{k=1}^{\infty} \mathcal{C}^{\mu}_{\nu}$ we then seek to construct a map t^{ν}_{μ} such that $t^{\nu}_{\mu}(\Omega^{\mu}_b) = \Omega^{\nu}_b$ (modulo sets of measure 0) for all $\Omega^{\mu}_b \in \bigcup_{k=1}^{\infty} \mathcal{C}^{\mu}_k$ and $\Omega^{\nu}_b \in \bigcup_{k=1}^{\infty} \mathcal{C}^{\nu}_k$. Once such t^{ν}_{μ} is found, we put

$$W_{\mathcal{S},p}(\mu, \nu) = \left(\int_{\mathbb{R}^d} |t^{\nu}_{\mu}(x) - x|^p d\mu(x) \right)^{\frac{1}{p}}.$$

The rigorous construction of t^{ν}_{μ} is carried out in [33] as follows. Let $\mathcal{S} = ((b_k, s_k))_{k=1}^{\infty}$ be an arbitrary slicing schedule. For $\mu \in \mathcal{P}_{ac}(\mathbb{R}^d)$ denote by $\bigcup_{k=1}^{\infty} \mathcal{C}^{\mu}_k$ the partition of μ according to $\mathcal{S}$. Furthermore, we introduce $\ell^{\mu}_k : \mathbb{R}^d \to \mathbb{R}$ as follows

$$\ell^{\mu}_k(x) = \sum_{\alpha=0}^{m} \frac{i_{\alpha}}{3^{\alpha}}, \quad x \in \Omega^{\mu}_{i_0 i_1 \cdots i_m},$$

for all $\Omega^{\mu}_{i_0 i_1 \cdots i_m} \in \mathcal{C}^{\mu}_k$, where $i_{\alpha} \in \{0, 1\}$. According to [33, Lemma 1], $(\ell^{\mu}_k)_{k=0}^{\infty}$ is a bounded uniformly convergent sequence of measurable functions yielding a Borel measurable limit

$$\ell^{\mu}(x) = \lim_{k \to \infty} \ell^{\mu}_k(x), \quad x \in \mathbb{R}^d.$$

Denote by $\mathcal{K}$ the set of all $d\mu = f(x)dx \in \mathcal{P}_{ac}(\mathbb{R}^d)$ such that

$$\mu\left(\partial(\operatorname{supp}(\mu))\right) = 0,$$

and

$$c \le f(x) \le C, \ \mu \text{ a.e..}$$

Main theoretical results about the existence of no-collision transportation maps in [33] can be summarized in the following theorem.

Theorem 2 ([33]). *Suppose that $\mathcal{S} = ((b_k, s_k))_{k=1}^{\infty}$ is a slicing schedule such that*

$$\{s_k\}_{k=1}^{\infty} \subset \{e_i\}_{i=1}^{d},$$

where $\{e_i\}_{i=1}^{d}$ is an arbitrary basis in $\mathbb{R}^d$. Moreover, suppose that $\mathcal{S}$ is such that each set appearing during the partition gets partitioned in each of $\{e_i\}_{i=1}^{d}$ directions infinitely many times.

Then for every $\mu, \nu \in \mathcal{K}$ we have that

1. *there exists $E_\mu \subset \text{int}(\text{supp}(\mu))$ such that $\mu(E_\mu) = 1$, and for every $x \in E_\mu$ there exits a unique $y \in \text{int}(\text{supp}(\nu))$ such that $\ell^\mu(x) = \ell^\nu(y)$,*

2. $t_\mu^\nu(x) = (\ell^\nu)^{-1}(\ell^\mu(x))$, $x \in E_\mu$ *is measurable, and* $t_\mu^\nu \sharp \mu = \nu$,

3. $(t_\mu^\nu)^{-1}(\Omega_b^\nu) = E_\mu \cap \Omega_b^\mu$ *for all* $\Omega_b^\mu \in \bigcup_{k=1}^\infty \mathcal{C}_k^\mu$ *and* $\Omega_b^\nu \in \bigcup_{k=1}^\infty \mathcal{C}_k^\nu$,

4. *if* $\tilde{t} : \tilde{E} \to \mathbb{R}^d$ *is such that* $\tilde{E} \cap \Omega_b^\mu \subset \tilde{t}^{-1}(\Omega_b^\nu)$ *for all* $\Omega_b^\mu \in \bigcup_{k=1}^\infty \mathcal{C}_k^\mu$ *and* $\Omega_b^\nu \in \bigcup_{k=1}^\infty \mathcal{C}_k^\nu$, *and* $\mu(\tilde{E}) = 1$ *then* $t_\mu^\nu = \tilde{t}$ μ *a.e..*

We use this theorem for computing no-collision transportation maps between translated and dilated measures.

3.3   Discretizing no-collision maps

For discretizing no-collision maps we choose a finite slicing schedule $\mathcal{S}_N = ((b_k, s_k))_{k=1}^N$ and partition given $\mu \in \mathcal{K}$ according to $\mathcal{S}_N$ obtaining sets $\mathcal{C}_N^\mu = \{\Omega_b^\mu\}$. Then we build *no-collision features* by taking either mass or geometric centers of Ω_b^μ; that is,

$$x_b^m(\mu) = \frac{1}{\mu(\Omega_b^\mu)} \int_{\Omega_b^\mu} x \, d\mu(x), \quad x_b^g(\mu) = \frac{1}{|\Omega_b^\mu|} \int_{\Omega_b^\mu} x \, dx, \quad \forall \Omega_b^\mu \in \mathcal{C}_N^\mu.$$

Hence, discrete no-collision maps map no-collision features; that is, for $\mu, \nu \in \mathcal{K}$ we define $\hat{t}^m : \{x_b^m(\mu)\} \to \{x_b^m(\nu)\}$ and $\hat{t}^g : \{x_b^g(\mu)\} \to \{x_b^g(\nu)\}$ as

$$\hat{t}^m(x_b^m(\mu)) = x_b^m(\nu), \quad \hat{t}^g(x_b^g(\mu)) = x_b^g(\nu), \quad \forall b.$$

Furthermore, no-collision distances can be discretized as

$$\hat{W}_{\mathcal{S}_N, p}^m(\mu, \nu) = \left(\sum_b \mu(\Omega_b^\mu) \, |x_b^m(\mu) - x_b^m(\nu)|^p \right)^{\frac{1}{p}},$$

and

$$\hat{W}_{\mathcal{S}_N, p}^g(\mu, \nu) = \left(\sum_b \mu(\Omega_b^\mu) \, |x_b^g(\mu) - x_b^g(\nu)|^p \right)^{\frac{1}{p}}.$$

In practice, choose $\mathcal{S}_N$ so that all Ω_b^μ have the same mass and drop the factors $\mu(\Omega_b^\mu)$ thus computing l_p distances between no-collision features. The algorithm for building no-collision features is as follows.

4   Translation and dilation manifolds

Here, we show that translation and dilation manifolds can be learned by no-collision maps just as with OT maps.

Algorithm 1 Computing no-collision features

Input: $N \geq 1$, $\mathcal{S}_N = ((b_k, s_k))_{k=1}^N$, $\mu \in \mathcal{P}(\Omega)$.

   $k \leftarrow 1$

   $\Omega_0^\mu \leftarrow \Omega$

   $\mathcal{C}^\mu \leftarrow \{\Omega_0^\mu\}$

   while $k < N$ **do**

      Find h_k s.t $\mu\left(x \cdot s_k \leq h_k \mid \Omega_{b_k}^\mu\right) = \mu\left(x \cdot s_k > h_k \mid \Omega_{b_k}^\mu\right)$

      $\Omega_{b_k 0}^\mu = \{x \cdot s_k \leq h_k\} \cap \Omega_{b_k}^\mu$

      $\Omega_{b_k 1}^\mu = \{x \cdot s_k > h_k\} \cap \Omega_{b_k}^\mu$

      $\mathcal{C}^\mu \leftarrow \mathcal{C}^\mu \setminus \Omega_{b_k}^\mu \cup \{\Omega_{b_k 0}^\mu, \Omega_{b_k 1}^\mu\}$

      $k \leftarrow k + 1$

   end while

   for $\Omega_b^\mu \in \mathcal{C}^\mu$ **do**

      $x_b^m(\mu) = \frac{1}{\mu(\Omega_b^\mu)} \int_{\Omega_b^\mu} x \, d\mu(x)$

      $x_b^g(\mu) = \frac{1}{|\Omega_b^\mu|} \int_{\Omega_b^\mu} x \, dx$

   end for

4.1 Translation manifolds

Theorem 3. *Assume that $\mu_0 \in \mathcal{K}$, and let $\mu_\theta = (x + \theta)\sharp\mu_0$ for $\theta \in \mathbb{R}^d$. Then for every slicing schedule $\mathcal{S}$ such as in Theorem 2 we have that*

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

Proof. Firstly, note that it is enough to prove the identity for $\theta' = 0$. Indeed, $\mu_\theta = (x + (\theta - \theta'))\sharp\mu_{\theta'}$.

Next, let $\mathcal{S} = ((b_k, s_k))_{k=1}^\infty$ be a slicing schedule as in Theorem 2. Our goal is to compute ℓ^{μ_θ}, and show that

$$\ell^{\mu_\theta}(x) = \ell^{\mu_0}(x - \theta), \quad \forall x \in \mathbb{R}^d.$$

We have that $\mathcal{C}_0^{\mu_0} = \{\Omega_0\}$, where $\Omega_0^{\mu_0} = \mathbb{R}^d$, and so

$$\ell_0^{\mu_\theta}(x) = 0, \quad \forall x \in \mathbb{R}^d.$$

Next, let h_1 be such that

$$\mu_0\left(\{x \cdot s_1 \leq h_1\}\right) = \mu_0\left(\{x \cdot s_1 > h_1\}\right) = \frac{1}{2}.$$

Then we have that

$$\Omega_{00}^{\mu_0} = \{x \cdot s_1 \leq h_1\}, \quad \Omega_{01}^{\mu_0} = \{x \cdot s_1 > h_1\}.$$

Note that

$$\mu_\theta\left(\{x \cdot s_1 \leq h_1 + \theta \cdot s_1\}\right) = \mu_0\left(\{(x + \theta) \cdot s_1 \leq h_1 + \theta \cdot s_1\}\right)$$

$$= \mu_0\left(\{x \cdot s_1 \leq h_1\}\right) = \frac{1}{2},$$

9

and so

$$\mu_\theta\left(\{x \cdot s_1 \le h_1 + \theta \cdot s_1\}\right) = \mu_\theta\left(\{x \cdot s_1 > h_1 + \theta \cdot s_1\}\right) = \frac{1}{2}.$$

This means that the first two slices for μ_θ are

$$\Omega_{00}^{\mu_\theta} = \{x \cdot s_1 \le h_1 + \theta \cdot s_1\}, \quad \Omega_{01}^{\mu_\theta} = \{x \cdot s_1 > h_1 + \theta \cdot s_1\}.$$

But this means that

$$\Omega_{00}^{\mu_\theta} = \Omega_{00}^{\mu_0} + \theta, \quad \Omega_{01}^{\mu_\theta} = \Omega_{01}^{\mu_0} + \theta.$$

The arguments above apply to all partitions, and we obtain

$$\Omega_b^{\mu_\theta} = \Omega_b^{\mu_0} + \theta$$

for all $\Omega_b^{\mu_\theta} \in \bigcup_{k=1}^\infty \mathcal{C}_k^{\mu_\theta}$ and $\Omega_b^{\mu_0} \in \bigcup_{k=1}^\infty \mathcal{C}_k^{\mu_0}$. Hence $x \in \Omega_b^{\mu_\theta}$ if and only if $x - \theta \in \Omega_b^{\mu_0}$. Since $\ell_k^{\mu_\theta}(x)$ depends only on b for $x \in \Omega_b^{\mu_\theta}$ we obtain that

$$\ell_k^{\mu_\theta}(x) = \ell_k^{\mu_0}(x - \theta), \quad \forall x \in \Omega_b^{\mu_\theta},$$

and so

$$\ell_k^{\mu_\theta}(x) = \ell_k^{\mu_0}(x - \theta), \quad \forall x \in \mathbb{R}^d.$$

Passing to the limit in k we find that

$$\ell^{\mu_\theta}(x) = \ell^{\mu_0}(x - \theta), \quad \forall x \in \mathbb{R}^d.$$

Taking into account Theorem 2 we obtain that

$$t_{\mu_0}^{\mu_\theta}(x) = x + \theta, \quad \mu_0 \text{ a.e.,}$$

and so

$$W_{\mathcal{S},p}(\mu_\theta, \mu_0) = \left(\int_{\mathbb{R}^d} |t_{\mu_0}^{\mu_\theta}(x) - x|^p d\mu_0(x)\right)^{\frac{1}{p}} = |\theta|.$$

$\square$

Corollary 1. *Assume that $\mu_0 \in \mathcal{K}$, $\Theta \subset \mathbb{R}^d$, $p \ge 1$, and $\mathcal{S}$ is a slicing schedule such as in Theorem 2. Then $(\{\mu_\theta\}, W_{\mathcal{S},p})$ is isometric to $(\Theta, |\cdot|)$.*

4.2 Dilation manifolds

Recall that the Hadamard product of $x = (x_1, x_2, \cdots, x_d)$ and $y = (y_1, y_2, \cdots, y_d)$ is defined as

$$x \odot y = (x_1 y_1, x_2 y_2, \cdots, x_d y_d).$$

Theorem 4. *Assume that $\mu_0 \in \mathcal{K}$, and let $\mu_\theta = (\theta \odot x)\,\sharp\mu_0$ for $\theta \in \mathbb{R}_+^d$. Then for every slicing schedule $\mathcal{S}$ such as in Theorem 2 containing only directions parallel to coordinate axes we have that*

$$W_{\mathcal{S},2}(\mu_\theta, \mu_{\theta'}) = |c \odot \theta - c \odot \theta'|, \quad \forall \theta, \theta' \in \mathbb{R}^d,$$

where

$$c_j^2 = \int_{\mathbb{R}^d} x_j^2 d\mu_0(x), \quad 1 \le j \le d.$$

Proof. Let $\mathcal{S} = ((b_k, s_k))_{k=1}^\infty$ be a slicing such as in Theorem 2 with $\{s_k\}$ parallel to coordinate axes, and $\theta, \theta' \in \mathbb{R}_+^d$. Our goal is to show that

$$t_{\mu_\theta}^{\mu_{\theta'}}(x) = \left(\frac{\theta_1'}{\theta_1} x_1, \frac{\theta_2'}{\theta_2} x_2, \cdots, \frac{\theta_d'}{\theta_d} x_d \right), \quad \mu_\theta \text{ a.e..}$$

We have that

$$\ell_0^{\mu_\theta}(x) = \ell_0^{\mu_{\theta'}}(x) = 0, \quad \forall x \in \mathbb{R}^d.$$

Without loss of generality, assume that $s_1 = (1, 0, \cdots, 0)$. Let h_1 be such that

$$\mu_\theta \left(\{x_1 \le h_1\} \right) = \mu_\theta \left(\{x_1 > h_1\} \right) = \frac{1}{2}.$$

But then we have that $\mu_{\theta'} = \left(\frac{\theta_1'}{\theta_1} x_1, \frac{\theta_2'}{\theta_2} x_2, \cdots . \frac{\theta_d'}{\theta_d} x_d \right) \sharp \mu_\theta$, and so

$$\mu_{\theta'} \left(\left\{ x_1 \le \frac{\theta_1'}{\theta_1} h_1 \right\} \right) = \mu_\theta \left(\left\{ \frac{\theta_1'}{\theta_1} x_1 \le \frac{\theta_1'}{\theta_1} h_1 \right\} \right)$$

$$= \mu_\theta \left(\{x_1 \le h_1\} \right) = \frac{1}{2}.$$

Hence,

$$\mu_{\theta'} \left(\left\{ x_1 \le \frac{\theta_1'}{\theta_1} h_1 \right\} \right) = \mu_{\theta'} \left(\left\{ x_1 > \frac{\theta_1'}{\theta_1} h_1 \right\} \right) = \frac{1}{2}.$$

This means that

$$\Omega_{00}^{\mu_\theta} = \{x_1 \le h_1\}, \quad \Omega_{01}^{\mu_\theta} = \{x_1 > h_1\},$$

and

$$\Omega_{00}^{\mu_{\theta'}} = \left\{ x_1 \le \frac{\theta_1'}{\theta_1} h_1 \right\}, \quad \Omega_{01}^{\mu_\theta'} = \left\{ x_1 > \frac{\theta_1'}{\theta_1} h_1 \right\}.$$

So the dilation map

$$(x_1, x_2, \cdots, x_d) \mapsto \left(\frac{\theta_1'}{\theta_1} x_1, \frac{\theta_2'}{\theta_2} x_2, \cdots, \frac{\theta_d'}{\theta_d} x_d \right)$$

maps $\Omega_{00}^{\mu_\theta}$ to $\Omega_{00}^{\mu_{\theta'}}$ and $\Omega_{01}^{\mu_\theta}$ to $\Omega_{01}^{\mu_{\theta'}}$. Similarly, we obtain that this dilation map maps $\Omega_b^{\mu_\theta}$ to $\Omega_b^{\mu_{\theta'}}$ for all $\Omega_b^{\mu_\theta} \in \bigcup_{k=1}^\infty \mathcal{C}_k^{\mu_\theta}$ and $\Omega_b^{\mu_0} \in \bigcup_{k=1}^\infty \mathcal{C}_k^{\mu_0}$. Again, since $\ell_k^{\mu_\theta}(x)$ depends only on b for $x \in \Omega_b^{\mu_\theta}$ we obtain that

$$\ell_k^{\mu_\theta}(x) = \ell_k^{\mu_{\theta'}} \left(\frac{\theta_1'}{\theta_1} x_1, \frac{\theta_2'}{\theta_2} x_2, \cdots, \frac{\theta_d'}{\theta_d} x_d \right), \quad \forall x \in \Omega_b^{\mu_\theta},$$

11

and so

$$\ell_k^{\mu_\theta}(x) = \ell_k^{\mu_{\theta'}}\left(\frac{\theta_1'}{\theta_1}x_1, \frac{\theta_2'}{\theta_2}x_2, \cdots, \frac{\theta_d'}{\theta_d}x_d\right), \quad \forall x \in \mathbb{R}^d.$$

Passing to the limit in k we find that

$$\ell^{\mu_\theta}(x) = \ell^{\mu_{\theta'}}\left(\frac{\theta_1'}{\theta_1}x_1, \frac{\theta_2'}{\theta_2}x_2, \cdots, \frac{\theta_d'}{\theta_d}x_d\right), \quad \forall x \in \mathbb{R}^d.$$

Taking into account Theorem 2 we obtain that

$$t_{\mu_\theta}^{\mu_{\theta'}}(x) = \left(\frac{\theta_1'}{\theta_1}x_1, \frac{\theta_2'}{\theta_2}x_2, \cdots, \frac{\theta_d'}{\theta_d}x_d\right), \quad \mu_\theta \text{ a.e.,}$$

and so

$$W_{\mathcal{S},2}^2(\mu_\theta, \mu_{\theta'}) = \int_{\mathbb{R}^d} \left|t_{\mu_\theta}^{\mu_{\theta'}}(x) - x\right|^2 d\mu_\theta(x) = \int_{\mathbb{R}^d} \sum_{j=1}^d \left(\frac{\theta_j'}{\theta_j}x_j - x_j\right)^2 d\mu_\theta(x)$$

$$= \int_{\mathbb{R}^d} \sum_{j=1}^d \left(\theta_j' x_j - \theta_j x_j\right)^2 d\mu_0(x) = \sum_{j=1}^d c_j^2(\theta_j' - \theta_j)^2.$$

$\square$

Corollary 2. *Assume that $\mu_0 \in \mathcal{K}$, $\Theta \subset \mathbb{R}_+^d$, and $\mathcal{S}$ is an arbitrary schedule with only slicing directions parallel to coordinate axes. Then $(\{\mu_\theta\}, W_{\mathcal{S},2})$ is isometric to $(c \odot \Theta, |\cdot|)$, where*

$$c_j^2 = \int_{\mathbb{R}^d} x_j^2 d\mu_0(x), \quad 1 \le j \le d.$$

5 Rotation manifolds

In [20] the authors investigated whether some structural information can be learned from pairwise distances of the subset

$$\{(Rx)\sharp\mu_0 \ : \ R \in SO(d)\} \subset \mathcal{P}_2(\mathbb{R}^d).$$

For example, can one find an isometry between this subset and some representation of $SO(d)$ in a Euclidean space? One particular question considered in [20] is as follows. Denote by R_t the counter-clockwise rotation by angle t around the origin; that is,

$$R_t x = (x_1 \cos t - x_2 \sin t, x_1 \sin t + x_2 \cos t).$$

Then the question is whether one generically has that $\{(R_t x)\sharp\mu_0 \ : \ t \in [0, 2\pi]\}$ with W_2 distance is isometric to $(\partial B_r, |\cdot|_2)$ for some $r > 0$, where $\partial B_r = \{x \in \mathbb{R}^2 \ : \ |x|_2 = r\}$. In [20], the authors perform a numerical experiment that recovers what looks like a circle when μ_0 is the uniform measure over an ellipse.

Similar question is considered in [12], where the W_2 is replaced by a $W_{2,\nu}$ – the linearized W_2 distance with a reference measure ν.

Here we show that generically neither metric makes the set

$$\{(R_t x)\sharp\mu_0 \ : \ t \in [0, 2\pi]\}$$

isometric to a planar circle.

We achieve this by analytically computing the W_2 and $W_{2,\nu}$ distances between rotations of uniform probability measures over ellipses in plane as considered in a numerical experiments in [20] and between rotations of Gaussians as considered in numerical experiments in [12]. In fact, these examples are somewhat identical in terms of pairwise distances since OT maps and distances between two uniform measures over ellipses match, up to scaling, with OT maps and distances of corresponding Gaussian measures [18] and [35, Remarks 2.31-32].

Finally, we show that no-collision maps do not resolve the isometry problem either.

5.1   Wasserstein distances

We start with preliminary lemmas.

Lemma 1. *Assume that $\mu_0 \in \mathcal{P}_2(\mathbb{R}^2)$, and $\mu_t = (R_t x)\sharp\mu_0$. Then we have that*

$$W_2(\mu_s, \mu_t) = W_2(\mu_0, \mu_{t-s}), \quad 0 \le s < t \le 2\pi.$$

Proof. Recall that for $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$ we denote by $\Pi(\mu, \nu)$ the set of all transportation plans with marginals μ, ν. Additionally, we denote by $\Pi_0(\mu, \nu)$ the set of all optimal transportation plans from μ to ν.

Let $\pi \in \mathcal{P}_2(\mathbb{R}^2 \times \mathbb{R}^2)$. Then we define $\hat{\pi} = (R_s x, R_s y)\sharp\pi$. It is straightforward to check that $\pi = (R_{-s} x, R_{-s} y)\sharp\hat{\pi}$, and $\hat{\pi} \in \Pi(\mu_s, \mu_t)$ if and only if $\pi \in \Pi(\mu_0, \mu_{t-s})$. Furthermore, we have that

$$\int_{\mathbb{R}^2 \times \mathbb{R}^2} |x - y|^2 d\hat{\pi}(x, y) = \int_{\mathbb{R}^2 \times \mathbb{R}^2} |R_s x - R_s y|^2 d\pi(x, y)$$

$$= \int_{\mathbb{R}^2 \times \mathbb{R}^2} |x - y|^2 d\pi(x, y)$$

Choosing $\pi \in \Pi_0(\mu_0, \mu_{t-s})$ we obtain that $W_2(\mu_s, \mu_t) \le W_2(\mu_0, \mu_{t-s})$. Next, choosing $\hat{\pi} \in \Pi_0(\mu_s, \mu_t)$ we obtain that $W_2(\mu_s, \mu_t) \ge W_2(\mu_0, \mu_{t-s})$. Hence $W_2(\mu_s, \mu_t) = W_2(\mu_0, \mu_{t-s})$, and $\pi \in \Pi_0(\mu_0, \mu_{t-s})$ if and only if $\hat{\pi} \in \Pi_0(\mu_s, \mu_t)$. $\square$

Lemma 2. *Assume that $\mu_0 \in \mathcal{P}_2(\mathbb{R}^2)$, and $\mu_t = (R_t x)\sharp\mu_0$. Then $\{\mu_t \ : \ t \in [0, 2\pi]\}$ is an absolutely continuous curve with constant metric derivative.*

Proof. Let $0 \le s < t \le 2\pi$. We have that $(R_{t-s}x)\sharp\mu_0 = \mu_{t-s}$, and so

$$W_2^2(\mu_0, \mu_{t-s}) \le \int_{\mathbb{R}^2} |R_{t-s}x - x|^2 d\mu_0(x) = 4\sin^2\frac{t-s}{2}\int_{\mathbb{R}^2}|x|^2\mu_0(x)$$
$$\le c|t-s|^2.$$

Invoking Lemma 1 we obtain

$$W_2(\mu_s, \mu_t) \le c|t-s|, \quad 0 \le s < t \le 2\pi.$$

So $\{\mu_t \ : \ t \in [0, 2\pi]\}$ is Lipschitz and therefore absolutely continuous. Applying [5, Theorem 1.1.2] we obtain that there exists

$$|\mu_t'| = \lim_{h\to 0}\frac{W_2(\mu_t, \mu_{t+h})}{|h|}$$

for Lebesgue a.e. $t \in [0, 2\pi]$. Invoking Lemma 1 again, we have that

$$W_2(\mu_t, \mu_{t+h}) = W_2(\mu_{t'}, \mu_{t'+h}), \quad \forall t, t' \in [0, 2\pi],$$

and so if the metric derivative exists at one point it exists everywhere, and it is constant; that is,

$$|\mu_t'| = \lim_{h\to 0}\frac{W_2(\mu_0, \mu_h)}{|h|} = c, \quad \forall t \in [0, 2\pi].$$

$\square$

Proposition 1. *Assume that $a, b > 0$, and μ_0 is the uniform probability measure over the elliptical domain*

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

Furthermore, assume that $\mu_t = (R_t x)\sharp\mu_0$. Then we have that

$$W_2^2(\mu_s, \mu_t)$$
$$= 4|u|^2\sin^2\frac{t-s}{2} + \frac{1}{2}\left(a^2 + b^2 - \sqrt{(a^2+b^2)^2\cos^2(t-s) + 4a^2b^2\sin^2(t-s)}\right),$$

$$(4)$$

for all $s, t \in [0, 2\pi]$, where $u = (u_1, u_2)$.

Proof. Taking into account Lemma 1 we only need to prove (4) for $s = 0$. From [18] (see also [35, Remarks 2.31-32]) we have that

$$W_2^2(\mu_0, \mu_t) = |m_0 - m_t|^2 + \text{tr}\left(\Sigma_0 + \Sigma_t - 2\left(\Sigma_0^{1/2}\Sigma_t\Sigma_0^{1/2}\right)^{1/2}\right),$$

14

where m_t and Σ_t are the mean and covariance of μ_t, respectively. A direct computation yields

$$m_0 = u, \quad \Sigma_0 = \begin{pmatrix} \frac{a^2}{4} & 0 \\ 0 & \frac{b^2}{4} \end{pmatrix}.$$

Furthermore, we have that

$$m_t = R_t m_0 = R_t u,$$

and

$$\Sigma_t = R_t \Sigma_0 R_{-t} = \begin{pmatrix} \frac{a^2 \cos^2 t + b^2 \sin^2 t}{4} & \frac{(a^2-b^2)\cos t \sin t}{4} \\ \frac{(a^2-b^2)\cos t \sin t}{4} & \frac{a^2 \sin^2 t + b^2 \cos^2 t}{4} \end{pmatrix}.$$

To conclude the proof, we need to calculate $\mathrm{tr}\left(\Sigma_0^{1/2}\Sigma_t\Sigma_0^{1/2}\right)^{1/2}$. Denoting by $M = \Sigma_0^{1/2}\Sigma_t\Sigma_0^{1/2}$ we have that

$$\mathrm{tr}(M^{1/2}) = \left(\mathrm{tr}(M) + 2\sqrt{\det(M)}\right)^{1/2}$$

since M is a 2×2 positive-definite matrix. Next, we have that

$$\det M = \det(\Sigma_0^{1/2})\det(\Sigma_t)\det(\Sigma_0^{1/2}) = \det \Sigma_0 \cdot \det \Sigma_t = (\det \Sigma_0)^2 = \frac{a^4 b^4}{256},$$

where we used the similarity of Σ_t and Σ_0. Furthermore, using the cyclical invariance of the trace we have that

$$\mathrm{tr}\, M = \mathrm{tr}\left(\Sigma_0^{1/2}\Sigma_t\Sigma_0^{1/2}\right) = \mathrm{tr}\left(\Sigma_t\Sigma_0^{1/2}\Sigma_0^{1/2}\right) = \mathrm{tr}(\Sigma_t\Sigma_0)$$

$$= \mathrm{tr}\left[\begin{pmatrix} \frac{a^2 \cos^2 t + b^2 \sin^2 t}{4} & \frac{(a^2-b^2)\cos t \sin t}{4} \\ \frac{(a^2-b^2)\cos t \sin t}{4} & \frac{a^2 \sin^2 t + b^2 \cos^2 t}{4} \end{pmatrix} \cdot \begin{pmatrix} \frac{a^2}{4} & 0 \\ 0 & \frac{b^2}{4} \end{pmatrix}\right]$$

$$= \frac{(a^4 + b^4)\cos^2 t + 2a^2 b^2 \sin^2 t}{16}.$$

Hence,

$$\mathrm{tr}(M^{1/2}) = \left(\mathrm{tr}(M) + 2\sqrt{\det(M)}\right)^{1/2}$$

$$= \left(\frac{(a^4 + b^4)\cos^2 t + 2a^2 b^2 \sin^2 t}{16} + 2\sqrt{\frac{a^4 b^4}{256}}\right)^{1/2}$$

$$= \left(\frac{(a^4 + b^4)\cos^2 t + 2a^2 b^2 \sin^2 t + 2a^2 b^2}{16}\right)^{1/2}$$

$$= \frac{\sqrt{(a^2 + b^2)^2 \cos^2 t + 4a^2 b^2 \sin^2 t}}{4}.$$

15

Finally, applying the similarity of Σ_t and Σ_0 again we have that

$$\operatorname{tr}\Sigma_t = \operatorname{tr}\Sigma_0 = \frac{a^2 + b^2}{4},$$

and so

$$
\begin{aligned}
W_2^2(\mu_0, \mu_t) =&|m_0 - m_t|^2 + \operatorname{tr}\left(\Sigma_0 + \Sigma_t - 2\left(\Sigma_0^{1/2}\Sigma_t\Sigma_0^{1/2}\right)^{1/2}\right) \\
=&|u - R_t u|^2 + 2\operatorname{tr}\Sigma_0 - 2\operatorname{tr}(M^{1/2}) \\
=&4|u|^2 \sin^2\frac{t}{2} + \frac{a^2 + b^2 - \sqrt{(a^2 + b^2)^2 \cos^2 t + 4a^2b^2 \sin^2 t}}{2}.
\end{aligned}
$$

$\square$

We are now in a position to prove that rotation manifolds are not isometric to circles in general.

Theorem 5. *Assume that $a, b > 0$, and μ_0 is the uniform probability measure over the elliptical domain*

$$\mathcal{E}_0 = \left\{(x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \leq 1\right\}.$$

Furthermore, assume that $\mu_t = (R_t x)\sharp\mu_0$. Then $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_2\right)$ is isometric to a circle if and only if $a = b$.

Proof. When $a = b$ we have that $\mu_t = \mu_0$ for all $t \in [0, 2\pi]$, so $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_2\right)$ is isometric to a (degenerate) circle.

Now assume that $a \neq b$ and suppose by contradiction that there exists $\gamma : [0, 2\pi] \to \partial B_r$ for some $r > 0$ such that

$$W_2^2(\mu_s, \mu_t) = |\gamma(s) - \gamma(t)|^2, \quad \forall s, t \in [0, 2\pi].$$

From Proposition 1 we have that

$$|\gamma(0) - \gamma(\pi/4)| = |\gamma(\pi/4) - \gamma(\pi/2)| = |\gamma(\pi/2) - \gamma(3\pi/4)| = |\gamma(3\pi/4) - \gamma(\pi)|.$$

Additionally, we have that $|\gamma(0) - \gamma(\pi)| = 0$, and so $\gamma(0) = \gamma(\pi)$. This means that the quadrilateral with vertices $\gamma(0), \gamma(\pi/4), \gamma(\pi/2), \gamma(3\pi/2)$ is a rhombus. Since the only rhombus that can be inscribed in a circle is the square, we obtain that $\gamma(0), \gamma(\pi/4), \gamma(\pi/2), \gamma(3\pi/2)$ form a square. The latter means that

$$|\gamma(\pi/2) - \gamma(0)|^2 = 2|\gamma(\pi/4) - \gamma(0)|^2.$$

From Proposition 1 we have that

$$
\begin{aligned}
|\gamma(\pi/2) - \gamma(0)|^2 =&\frac{(a - b)^2}{2}, \\
|\gamma(\pi/4) - \gamma(0)|^2 =&\frac{a^2 + b^2 - \sqrt{\frac{(a^2+b^2)^2}{2} + 2a^2b^2}}{2},
\end{aligned}
$$

16

and so we obtain

$$\frac{(a-b)^2}{2} = a^2 + b^2 - \sqrt{\frac{(a^2+b^2)^2}{2} + 2a^2b^2}.$$

Simplifying the latter yields

$$(a-b)^4 = 0,$$

which is a contradiction. $\qquad\square$

We can also prove a more general version of Theorem 5 with elliptical domains not necessarily centered at the origin.

Theorem 6. *Assume that $a, b > 0$, and μ_0 is the uniform probability measure over the elliptical domain*

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

Furthermore, assume that $\mu_t = (R_t x) \sharp \mu_0$. Then $\left(\{\mu_t\}_{t \in [0, 2\pi]}, W_2 \right)$ is isometric to a circle if and only if $a = b$.

Proof. Denote by $u = (u_1, u_2)$. Assume that $u \neq 0$ for otherwise Theorem 5 applies. If $a = b$ then from (4) we have that

$$W_2^2(\mu_s, \mu_t) = 4|u|^2 \sin^2 \frac{t-s}{2} = |R_s u - R_t u|^2,$$

which means that $\mu_t \mapsto R_t u$ is an isometry; that is, $\left(\{\mu_t\}_{t \in [0, 2\pi]}, W_2 \right)$ is isometric to the circle of radius $|u|$.

Now assume that $a \neq b$, and suppose by contradiction that there exists $\gamma : [0, 2\pi] \to \partial B_r$ for some $r > 0$ such that

$$W_2^2(\mu_s, \mu_t) = |\gamma(s) - \gamma(t)|^2, \quad \forall s, t \in [0, 2\pi].$$

Similar to the proof of Theorem 5 we have that $\gamma(0), \gamma(\pi/2), \gamma(\pi), \gamma(3\pi/2)$ are vertices of a square. Indeed, the only difference is that $u \neq 0$ implies that $\{\mu_t\}$ and $\{\gamma(t)\}$ are 2π-periodic instead of π-periodic. Hence, we must have

$$|\gamma(\pi) - \gamma(0)|^2 = 2|\gamma(\pi/2) - \gamma(0)|^2.$$

Invoking Proposition 1 we find that

$$|\gamma(\pi) - \gamma(0)|^2 = 4|u|^2,$$
$$|\gamma(\pi/2) - \gamma(0)|^2 = 2|u|^2 + \frac{(a-b)^2}{2},$$

and so we obtain

$$4|u|^2 = 4|u|^2 + (a-b)^2,$$

which is a contradiction. $\qquad\square$

Remark 1. *Proposition 1 and Theorems 5, 6 can be of course extended to uniform measures over general elliptical domains*

$$\mathcal{E} = \left\{ x \in \mathbb{R}^2 \ : \ (x-u)^\top \Pi^{-1}(x-u) \le 1 \right\},$$

where $u \in \mathbb{R}^2$, and Π is symmetric and positive definite. Indeed, if μ_0 is the uniform probability measure over E, then $\mu_{t_0} = (R_{t_0}x)\sharp\mu_0$ is the uniform probability measure over

$$E' = \left\{ (x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{(x_1 - (R_{t_0}u)_1)^2}{\lambda_1} + \frac{(x_2 - (R_{t_0}u)_2)^2}{\lambda_2} \le 1 \right\},$$

where $\lambda_1, \lambda_2 > 0$ are the eigenvalues of Π. Hence, we have that

$$
\begin{aligned}
& W_2^2(\mu_s, \mu_t) \\
=& W_2^2(\mu_{t_0}, \mu_{t_0+t-s}) = W_2^2(\mu_{t_0}, (R_{t-s}x)\sharp\mu_{t_0}) \\
=& 4|R_{t_0}u|^2 \sin^2\frac{t-s}{2} + \frac{\lambda_1 + \lambda_2 - \sqrt{(\lambda_1+\lambda_2)^2\cos^2(t-s) + 4\lambda_1\lambda_2\sin^2(t-s)}}{2} \\
=& 4|u|^2 \sin^2\frac{t-s}{2} + \frac{\operatorname{tr}\Pi - \sqrt{(\operatorname{tr}\Pi)^2\cos^2(t-s) + 4\det\Pi\ \sin^2(t-s)}}{2}.
\end{aligned}
$$

5.2 Linearized Wasserstein distances

Let $\nu \in \mathcal{P}_{2,ac}(\mathbb{R}^2)$. Recall that for $\mu_1, \mu_2 \in \mathcal{P}_2(\mathbb{R}^2)$ we define the linearized Wasserstein distance with a reference measure ν as

$$W_{2,\nu}(\mu_1, \mu_2) = \left(\int_{\mathbb{R}^2} |T_2(x) - T_1(x)|^2 d\nu(x) \right)^{\frac{1}{2}},$$

where T_1, T_2 are optimal transportation maps pushing ν forward to μ_1, μ_2, respectively.

Lemma 3. *Assume that $\nu \in \mathcal{P}_{ac,2}(\mathbb{R}^2)$ is rotationally symmetric; that is, $(R_t x)\sharp\nu = \nu$ for all $t \in \mathbb{R}$. Furthermore, assume that $\mu_t = (R_t x)\sharp\mu_0$ for some $\mu_0 \in \mathcal{P}_2(\mathbb{R}^2)$, and denote by T_t the optimal transportation map that pushes ν forward to μ_t. Then we have that*

$$T_t(x) = R_t T_0(R_{-t}x), \ \ a.e. \ in \ \mathbb{R}^2, \ 0 \le t \le 2\pi,$$

and

$$W_{2,\nu}(\mu_s, \mu_t) = W_{2,\nu}(\mu_0, \mu_{t-s}), \quad 0 \le s < t \le 2\pi.$$

Proof. Since T_0 is the optimal transportation map between ν and μ_0 we have that $\pi_0 = (x, T_0(x))\sharp\nu$ is the optimal transportation plan between ν and μ_0. Applying the arguments in Lemma 1 we have that $\pi_t = (R_t x, R_t y)\sharp\pi_0$ is the optimal transportation plan between $(R_t x)\sharp\nu = \nu$ and $(R_t x)\sharp\mu_0 = \mu_t$. Note that

$$\pi_t = (R_t x, R_t y)\sharp\pi_0 = \pi_t = (R_t x, R_t T_0(x))\sharp\nu.$$

Furthermore, taking into account that $\nu = (R_{-t}x)\sharp\nu$ we obtain that

$$\pi_t = (R_t x, R_t T_0(x))\sharp\nu = (x, R_t T_0(R_{-t}x))\sharp\nu,$$

and so π_t is generated by a necessarily optimal transportation map $x \mapsto R_t T_0(R_{-t}x)$. But optimal transportation map from ν to μ_t is unique by Brenier's theorem, and so

$$T_t(x) = R_t T_0(R_{-t}x), \quad \text{a.e. in } \mathbb{R}^2.$$

Next, we have that

$$W_{2,\nu}^2(\mu_s, \mu_t)$$
$$= \int_{\mathbb{R}^2} |T_t(x) - T_s(x)|^2 d\nu(x) = \int_{\mathbb{R}^2} |R_t T_0(R_{-t}x) - R_s T_0(R_{-s}x)|^2 d\nu(x)$$
$$= \int_{\mathbb{R}^2} |R_{t-s} T_0(R_{-t}x) - T_0(R_{-s}x)|^2 d\nu(x)$$
$$= \int_{\mathbb{R}^2} |R_{t-s} T_0(R_{s-t}x) - T_0(x)|^2 d\nu(x) = \int_{\mathbb{R}^2} |T_{t-s}(x) - T_0(x)|^2 d\nu(x)$$
$$= W_{2,\nu}^2(\mu_0, \mu_{t-s}).$$

$\square$

Proposition 2. *Assume that ν is a standard normal distribution, and μ_0 is a Gaussian with mean $u \in \mathbb{R}^2$ and covariance matrix $\Sigma_0 = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}$ for some $a, b > 0$. Furthermore, let $\mu_t = (R_t x)\sharp\mu_0$. Then we have that*

$$W_{2,\nu}^2(\mu_s, \mu_t) = 4|u|^2 \sin^2 \frac{t-s}{2} + 2(a-b)^2 \sin^2(t-s), \quad 0 \le s < t \le 2\pi. \quad (5)$$

Proof. Taking into account Lemma 3 it is enough to prove (5) for $s = 0$. Furthermore, we have that $\mu_t = (R_t x)\sharp\mu_0$ is a Gaussian with mean $R_t u$ and covariance matrix $\Sigma_t = R_t \Sigma_0 R_{-t}$. Hence, from [18] (see also [35, Remarks 2.31-32]) we have that

$$T_t(x) = R_t u + \Sigma_t^{\frac{1}{2}} x, \quad x \in \mathbb{R}^2.$$

Furthermore, recalling that

$$\int_{\mathbb{R}^2} |m + Mx|^2 d\nu(x) = \text{tr}(M^\top M) + |m|^2,$$

we obtain

$$W_{2,\nu}^2(\mu_0, \mu_t)$$
$$= \int_{\mathbb{R}^2} |T_t(x) - T_0(x)|^2 d\nu(x) = \int_{\mathbb{R}^2} \left| R_t(u) - u + (\Sigma_t^{\frac{1}{2}} - \Sigma_0^{\frac{1}{2}} x) \right|^2 d\nu(x)$$
$$= |R_t u - u|^2 + \text{tr}(\Sigma_t^{\frac{1}{2}} - \Sigma_0^{\frac{1}{2}})^2 = 4|u|^2 \sin^2 \frac{t}{2} + \text{tr}\left(\Sigma_t + \Sigma_0 - \Sigma_t^{\frac{1}{2}}\Sigma_0^{\frac{1}{2}} - \Sigma_0^{\frac{1}{2}}\Sigma_t^{\frac{1}{2}} \right)$$
$$= 4|u|^2 \sin^2 \frac{t}{2} + \text{tr}\,\Sigma_t + \text{tr}\,\Sigma_0 - 2\,\text{tr}\left(\Sigma_t^{\frac{1}{2}}\Sigma_0^{\frac{1}{2}} \right)$$
$$= 4|u|^2 \sin^2 \frac{t}{2} + 2\,\text{tr}\,\Sigma_0 - 2\,\text{tr}\left(\Sigma_t^{\frac{1}{2}}\Sigma_0^{\frac{1}{2}} \right),$$

19

where we used the similarity of Σ_t and Σ_0, and the cyclic invariance of the trace. Next, we have that

$$\Sigma_t^{\frac{1}{2}} = (R_t \Sigma_0 R_{-t})^{\frac{1}{2}} = R_t \Sigma_0^{\frac{1}{2}} R_{-t} = \begin{pmatrix} a\cos^2 t + b\sin^2 t & (a-b)\sin t \cos t \\ (a-b)\sin t \cos t & a\sin^2 t + b\cos^2 t \end{pmatrix},$$

and so

$$\mathrm{tr}\left(\Sigma_t^{\frac{1}{2}} \Sigma_0^{\frac{1}{2}}\right) = \mathrm{tr}\left[\begin{pmatrix} a\cos^2 t + b\sin^2 t & (a-b)\sin t \cos t \\ (a-b)\sin t \cos t & a\sin^2 t + b\cos^2 t \end{pmatrix} \cdot \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}\right]$$
$$= (a^2 + b^2)\cos^2 t + 2ab\sin^2 t.$$

Therefore we have that

$$W_{2,\nu}^2(\mu_0, \mu_t) = 4|u|^2 \sin^2 \frac{t}{2} + 2(a^2 + b^2) - 2(a^2 + b^2)\cos^2 t - 4ab\sin^2 t$$
$$= 4|u|^2 \sin^2 \frac{t}{2} + 2(a-b)^2 \sin^2 t.$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\square$

Theorem 7. *Assume that ν is a standard normal distribution, and μ_0 is a Gaussian with mean $u \in \mathbb{R}^2$ and covariance matrix $\Sigma_0 = \begin{pmatrix} a^2 & 0 \\ 0 & b^2 \end{pmatrix}$ for some $a, b > 0$. Furthermore, let $\mu_t = (R_t x)\sharp \mu_0$. Then $\left(\{\mu_t\}_{t \in [0,2\pi]}, W_{2,\nu}\right)$ is isometric to a circle if and only if $u = 0$ or $a = b$.*

Proof. If $u = 0$ then (5) yields that

$$W_{2,\nu}(\mu_s, \mu_t) = 2(a-b)^2 \sin^2(t-s) = \frac{(a-b)^2}{2}|R_{2s}e_1 - R_{2t}e_1|^2,$$

which means that $\mu_t \mapsto \frac{|a-b|}{\sqrt{2}} R_{2t}e_1$ is an isometry, where $e_1 = (1,0)$.

If $a = b$ then (5) yields that

$$W_{2,\nu}(\mu_s, \mu_t) = 4|u|^2 \sin^2 \frac{t-s}{2} = |R_s u - R_t u|^2,$$

which means that $\mu_t \mapsto R_t u$ is an isometry.

Now assume that $u \neq 0$, $a \neq b$, and suppose by contradiction that there exists $\gamma : [0, 2\pi] \to \partial B_r$ for some $r > 0$ such that

$$W_{2,\nu}^2(\mu_s, \mu_t) = |\gamma(s) - \gamma(t)|^2, \quad \forall s, t \in [0, 2\pi].$$

Similar to Theorems 5 and 6 we have that $\gamma(0), \gamma(\pi/2), \gamma(\pi), \gamma(3\pi/2)$ are vertices of a square, and so

$$|\gamma(\pi) - \gamma(0)|^2 = 2|\gamma(\pi/2) - \gamma(0)|^2.$$

From (5) we have that

$$|\gamma(\pi) - \gamma(0)|^2 = 4|u|^2,$$
$$|\gamma(\pi/2) - \gamma(0)|^2 = 2|u|^2 + 2(a-b)^2,$$

and so we obtain

$$4|u|^2 = 4|u|^2 + 4(a-b)^2,$$

which is a contradiction. $\qquad\square$

Remark 2. *As in Remark 1, we note that Proposition 2 and Theorem 7 naturally extend to Gaussians μ_0 with a generic covariance matrix Σ_0. Indeed, for a suitable angle t_0 we have that μ_{t_0} is a Gaussian with mean $m_{t_0} = R_{t_0} u$ and covariance $\Sigma_{t_0} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, where $\lambda_1, \lambda_2 > 0$ are the eigenvalues of Σ_0. But then we have that*

$$
\begin{aligned}
W^2_{2,\nu}(\mu_s, \mu_t) &= W^2_{2,\nu}(\mu_{t_0}, \mu_{t_0+t-s}) = W^2_{2,\nu}(\mu_{t_0}, (R_{t-s}x)\sharp\mu_{t_0}) \\
&= 4|R_{t_0}u|^2 \sin\frac{t-s}{2} + 2(\lambda_1 - \lambda_2)^2 \sin^2(t-s) \\
&= 4|u|^2 \sin\frac{t-s}{2} + 2((\lambda_1 + \lambda_2)^2 - 4\lambda_1\lambda_2)\sin^2(t-s) \\
&= 4|u|^2 \sin\frac{t-s}{2} + 2((\operatorname{tr}\Sigma_0)^2 - 4\det\Sigma_0)\sin^2(t-s).
\end{aligned}
$$

5.3 No-collision distances

Here, we show that no-collision distances in general do not provide an isometry between rotated measures and planar circles either.

Theorem 8. *Assume that μ_0 is a uniform measure over an elliptical domain*

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

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 such as in Theorem 4. Then $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_{\mathcal{S},2}\right)$ is isometric to a circle if and only if $a = b$.

Proof. When $a = b$ we have that μ_t is the uniform measure over the disk of radius a centered at $R_t u$. Therefore, $\mu_t = (R_t u - u + x)\sharp\mu_0$, and by Theorem 3 we have that
$$W_{\mathcal{S},2}(\mu_s, \mu_t) = |R_s u - R_t u|, \quad \forall s, t,$$
which means that $\mu_t \mapsto R_t u$ is an isometry.

Now assume that $a \neq b$, and suppose by contradiction that there exists $\gamma : [0, 2\pi] \to \partial B_r$ for some $r > 0$ such that

$$W^2_{\mathcal{S},2}(\mu_s, \mu_t) = |\gamma(s) - \gamma(t)|^2, \quad \forall s, t \in [0, 2\pi].$$

21

We have that $\mu_{\pi/2}$ is the uniform measure over the elliptical domain

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

Furthermore, denote by ν, ν' uniform measures over elliptical domains

$$\mathcal{E}_0 = \left\{ (x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \le 1 \right\},$$

$$\mathcal{E}_0' = \left\{ (x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{x_1^2}{b^2} + \frac{x_2^2}{a^2} \le 1 \right\},$$

respectively. From Theorems 3 and 4 we have the following no-collision transportation maps

$$t_{\mu_0}^{\nu}(x_1, x_2) = (x_1 - u_1, x_2 - u_2),$$
$$t_{\nu}^{\nu'}(x_1, x_2) = \left(\frac{b}{a} x_1, \frac{a}{b} x_2 \right),$$
$$t_{\nu'}^{\mu_{\pi/2}}(x_1, x_2) = (x_1 - u_2, x_2 + u_1).$$

Hence, by [33, Proposition 2], which follows from Part 1 in Theorem 2, we have that

$$t_{\mu_0}^{\mu_{\pi/2}}(x_1, x_2) = t_{\nu'}^{\mu_{\pi/2}} \circ t_{\nu}^{\nu'} \circ t_{\mu_0}^{\nu}(x_1, x_2) = \left(\frac{b}{a}(x_1 - u_1) - u_2, \frac{a}{b}(x_2 - u_2) + u_1 \right)$$

is the no-collision map pushing forward μ_0 to $\mu_{\pi/2}$. Note that $t_{\mu_0}^{\mu_{\pi/2}} = \nabla\phi$, where
$$\phi(x_1, x_2) = \frac{b(x_1 - u_1)^2}{2a} - u_2 x_1 + \frac{a(x_2 - u_2)^2}{2b} + u_1 x_2.$$

Since ϕ is convex, Brenier's theorem yields that $t_{\mu_0}^{\mu_{\pi/2}}$ is the optimal transportation map pushing μ_0 optimally to $\mu_{\pi/2}$, and so

$$W_{\mathcal{S},2}(\mu_0, \mu_{\pi/2}) = W_2(\mu_0, \mu_{\pi/2}).$$

Similarly, we can prove that

$$W_{\mathcal{S},2}(\mu_{\pi k/2}, \mu_{\pi l/2}) = W_2(\mu_{\pi k/2}, \mu_{\pi l/2}), \quad \forall k, l \in \mathbb{Z}.$$

From Proposition 1 we have that

$$W_2(\mu_0, \mu_{\pi/2}) = W_2(\mu_{\pi/2}, \mu_\pi) = W_2(\mu_\pi, \mu_{3\pi/2}) = W_2(\mu_{3\pi/2}, \mu_0),$$

and so

$$W_{\mathcal{S},2}(\mu_0, \mu_{\pi/2}) = W_{\mathcal{S},2}(\mu_{\pi/2}, \mu_\pi) = W_{\mathcal{S},2}(\mu_\pi, \mu_{3\pi/2}) = W_{\mathcal{S},2}(\mu_{3\pi/2}, \mu_0).$$

The latter yields

$$|\gamma(0) - \gamma(\pi/2)| = |\gamma(\pi/2) - \gamma(\pi)| = |\gamma(\pi) - \gamma(3\pi/2)| = |\gamma(3\pi/2) - \gamma(0)|,$$

22

by the isometry assumption. Thus, $\gamma(0), \gamma(\pi/2), \gamma(\pi), \gamma(3\pi/2)$ are vertices of a rhombus that must necessarily be a square since $\{\gamma(t)\} \subset \partial B_r$. Therefore we have that

$$|\gamma(0) - \gamma(\pi)|^2 = 2|\gamma(0) - \gamma(\pi/2)|^2.$$

But

$$
\begin{aligned}
|\gamma(0) - \gamma(\pi)|^2 &= W_{\mathcal{S},2}^2(\mu_0, \mu_\pi) = W_2^2(\mu_0, \mu_\pi), \\
|\gamma(0) - \gamma(\pi/2)|^2 &= W_{\mathcal{S},2}^2(\mu_0, \mu_{\pi/2}) = W_2^2(\mu_0, \mu_{\pi/2}),
\end{aligned}
$$

and we arrive at a contradiction by repeating the calculation in Theorem 6. $\quad\square$

6 Experimental results

In this section, we provide several experiments [1] that demonstrate our theoretical results. In particular we perform manifold learning on synthetically generated 2-dimensional images using no-collision distances and compare our results with other manifold learning techniques.

For our experiments we generate synthetic data as follows. We fix $\mu_0 \in \mathcal{P}_{ac}(\mathbb{R}^d)$ and consider

$$\mathcal{M}(\mu_0, \Theta) = \{F_\theta \sharp \mu_0 \ : \ \theta \in \Theta\},$$

where F_θ is a parametric mapping. In particular, we have

- $F_\theta(x) = x - \theta$ for translations,

- $F_\theta(x) = \theta \odot x$ for dilations,

- $F_\theta(x) = R_\theta x$ for for rotations.

We then sample $\{\theta_1, \cdots, \theta_m\} \subset \Theta$ and consider

$$\{\mu_1, \mu_2, \cdots, \mu_m\} = \{F_{\theta_1} \sharp \mu_0, F_{\theta_2} \sharp \mu_0, \cdots, F_{\theta_m} \sharp \mu\}.$$

In practice, we work with discrete versions of these distributions: we represent them discretely as an $n \times n$ pixel image. Each point in the support of the image has a pixel value which (after normalization) is the mass associated with the density μ_i.

Once the data is generated, the embedding is performed as follows: given the set of observed transformed distributions, we compute the distance matrix containing all the pairwise distances. We then perform MDS on the distance matrix to find a 2-dimensional embedding. In each example we compute the Euclidean distance matrix, Wasserstein distance matrix, LOT distance matrix (using the Python Optimal Transport (POT) package [17]), and no-collision distance matrices.

---

[1]Code for this work is available at `https://github.com/enegrini/Applications-of-No-Collision-Transportation-Maps-in-Manifold-Learning.git`.

Specifically, we compute the no-collision distance matrix applying Algorithm 1 to each of the observed distributions for generating corresponding no-collision features (geometric centers or centers of mass) and computing pairwise Euclidean distances between these features. For each example, we assume that images have the same size and we choose the number of cuts, N, that gives the best visual results. In particular, we start with $N = 2$ and increase the number of cuts until the result start decreasing in accuracy. For large enough N the problem reduces to performing MDS on the original pixel images, and experiments show that this case provides inaccurate manifold reconstructions.

We compare our results for manifold learning with MDS, Isomap, and Diffusion Maps on the original pixel images using Euclidean distances; with Wassmap; and with MDS on LOT features. For Wassmap and LOT, we generate point-cloud data from the images as done in the original papers. For LOT, we pick the reference distribution to be a Gaussian with given mean and covariance matrix. For Isomap, the k-NN graph is used to estimate geodesics and k is picked to give the best visual results. Finally, for Diffusion maps we use a Gaussian kernel and select the parameter epsilon according to the algorithm proposed in [7].

6.1 Translation Manifold

We fix a base measure μ_0 to be the indicator function of a disc of radius 1 centered at the origin. For a given translation set $\Theta = \{\theta = (\theta_1, \theta_2)\} \subset \mathbb{R}^2$ the corresponding translation manifold is given by:

$$\mathcal{M}^{\mathrm{trans}}(\mu_0, \Theta) := \{(x - \theta)\sharp\mu_0;\ \theta \in \Theta\}$$

In this example, we consider three translation sets: $\Theta_1 = [-1, 1]^2$, sampled on a uniform 4×4 grid, $\Theta_2 = [-5, -1] \times [-2, 2] \cup [1, 5] \times [-2, 2]$ where $[-5, -1] \times [-2, 2]$ is sampled on a 4×4 grid while $[1, 5] \times [-2, 2]$ is sampled on a 6×6 grid, and Θ_3 representing a triangular translation grid.

6.1.1 Translation with parameter set Θ_1

Since both no-collision maps and LOT give an approximation of the true Wasserstein distance, the first question we address is how good these approximations are. In Figure 1 we compare the distance matrices obtained for the two translation examples. In the first row, from left to right we represent the true Wasserstein distance, LOT distance and no-collision distances; in the second row we represent the L^2 error in the approximation of the Wasserstein distance. In this case for no-collision maps we use $N = 2$ cuts since this was the choice that gave the best results both for center of mass and geometric center features. As reference for LOT in this case we picked one Gaussian centered at $(0, 0)$ with covariance matrix $25I$. Picking more than one references gives similar results in this case but requires more computational time. In Table 1 we report the relative error in Frobenius norm for each distance. From these experiments we can see in this case LOT and no-collision provide an accurate approximation of the Wasserstein distance with relative errors that are less than 2%. However,

the no-collision distance approximation using center of mass is the one that provides the best accuracy with relative error less than 1%.
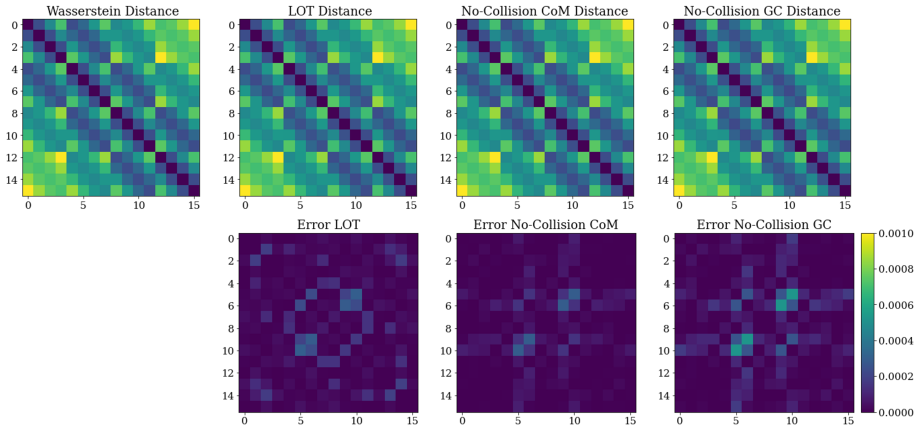

Figure 1: Distance matrix comparison for translation example with parameter set Θ_1. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and Wasserstein distance.

Distance	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	1.17%	0.96%	1.31%

Table 1: Relative error in Frobenius norm of Wasserstein distance approximation using LOT and no-collision distances for translation example with parameter set Θ_1.

Not only no-collision distances give good approximations of the true Wasserstein distance, but they are also much faster to compute than the LOT and true Wasserstein distance since they do not require any optimization. Specifically, in the example with parameter set Θ_1 with 16 translation samples true Wasserstein distance computation took 15.2 seconds, LOT computation took 0.76 seconds while no-collision maps with $N = 2$ cuts took 0.056 seconds which is 13.5 times faster than LOT and 270 times faster than Wassmap. To explore the computational time aspect more in depth we compared in Figure 2 the time (in log scale) needed to obtain the distance matrix for Wassmap, LOT and no-collision for an increasing number of translation points for the two parameter sets. For no-collision $N = 2$ already gave accurate results for all amounts of translation points, but for completeness we also report the time required for larger N that also provided accurate distance approximations. We see that the computational time for Wassserstein distance is always larger than the one needed for LOT and no-collision (for less than 8 cuts) and it also increases faster as the number of

25

translation points increases. Comparing LOT and no-collision we see that LOT computational time is larger than no-collision when $N \leq 4$, however no-collision already provides the best approximation for $N = 2$.
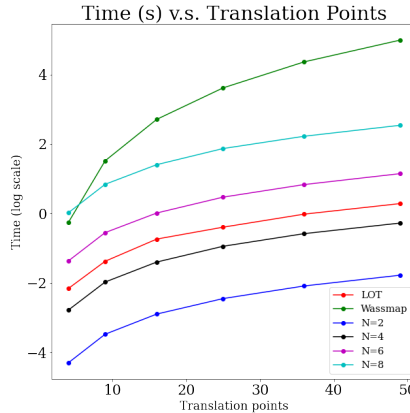

Figure 2: Computational time Wassmap, LOT and no-collision for multiple choices of N and increasing number of translation points for parameter set Θ_1.

Finally, In Figures 3 we show the translation grid obtained by MDS embedding in a 2-dimensional space using the distance matrices provided by the different methods. The results below are in line with our theoretical results in Section 4.1: no-collision maps are able to recover translation manifolds up to rigid transformation and a scaling. The scaling is due to the fact that no-collision maps give only an approximation of the true Wasserstein distance. Wassmap and LOT also produce an accurate embedding up to rigid transformation and a scaling, while Isomap, Diffusion Maps and MDS on the original pixel features using Euclidean distances show significant skewing and overlapping points.

From this example we see that while all methods relying on Wasserstein distance or its approximation provide very accurate embeddings, no-collision map require only a fraction of the computational time but still give very accurate distance approximations and embeddings.
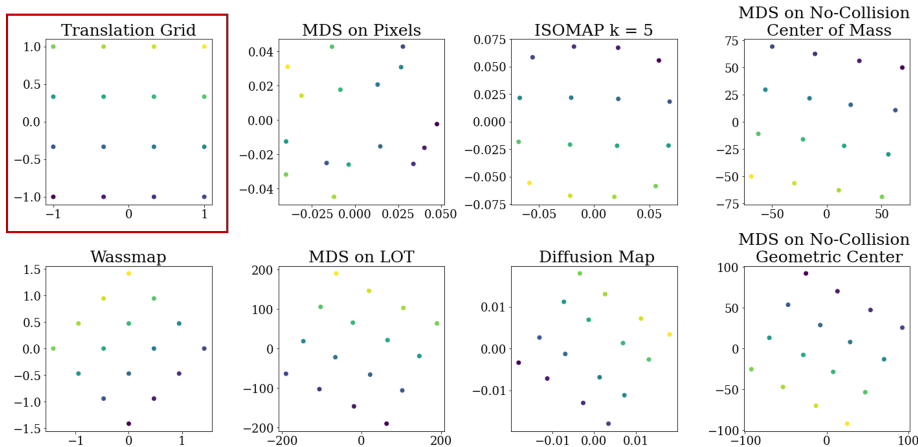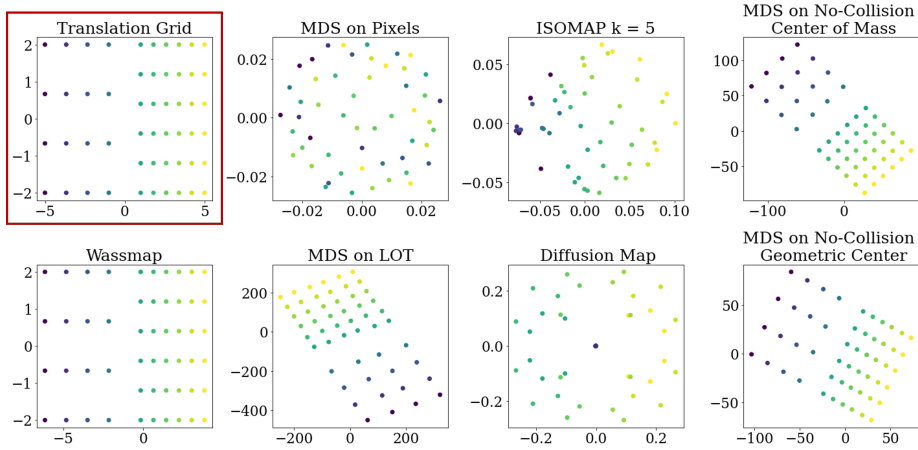
Figure 3: Translation manifold generated by the characteristic function of the unit disk with parameter set Θ_1 sampled on a 4×4 grid. We show the original translation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

6.1.2 Translation with parameter set Θ_2

A similar analysis can be done for parameter set Θ_2. In Figure 4 and Table 2 we compare the distance matrices. As in the previous example we use $N = 2$ cuts for no-collision maps. From this example we see that again both LOT and no-collision provide accurate approximations of the Wasserstein distance matrix, with no-collision with center or mass giving the best approximation with error less than 0.4%. As noted before, not only no-collision gives accurate results but it is also very fast to compute. In this example with 52 translation samples true Wasserstein distance computation took 163.1 seconds, LOT computation took 1.07 seconds while no-collision maps with N = 2 cuts took 0.211 seconds which is 5.1 times faster than LOT and 773 times faster than Wassmap. A similar plot as the once before of computational time v.s. translation points can be obtained in this case. We omit this plot for brevity.
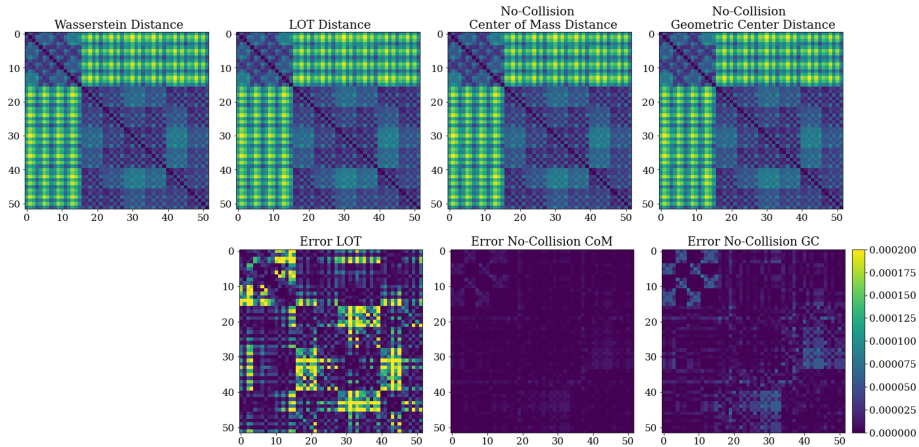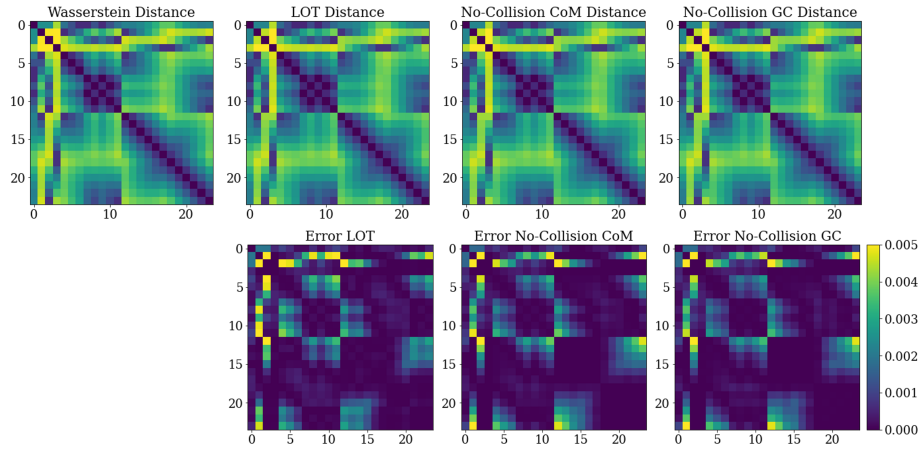
Figure 4: Distance matrix comparison for translation example with parameter set Θ_2. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and Wasserstein distance.

Distance	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	1.58%	0.36%	0.73%

Table 2: Relative error in Frobenius norm of Wasserstein distance approximation using LOT and no-collision distances for translation example with parameter set Θ_2.

Finally, as expected from the theoretical analysis, we can see from Figure 5 that also in this case methods based on Wasserstein distance (or approximations of it) are able to correctly reconstruct the translation grid up to rigid transformation and a scaling, while the other methods cannot.
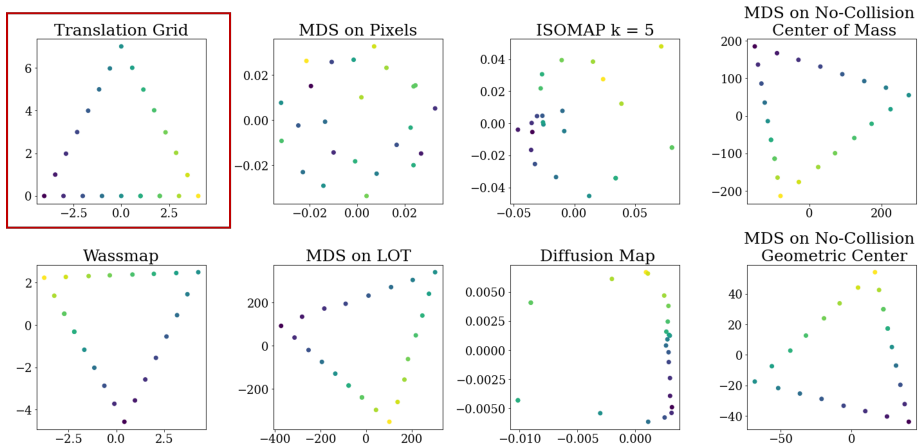
Figure 5: Translation manifold generated by the characteristic function of the unit disk with parameter set Θ_2 sampled on a 4×4 and a 6×6 grid. We show the original translation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

6.1.3 Translation with parameter set Θ_3

A similar analysis can be done for parameter set Θ_3. In Figure 6 and Table 3 we compare the distance matrices. Here we use $N = 4$ and $N = 2$ cuts for no-collision maps respectively with center of mass and geometric center. In this case LOT and no-collision provide comparable approximations of the Wasserstein distance matrix, with no-collision with geometric center giving the best approximation with error less than 4.6%. In terms of computational time in this case with 24 translation samples no-collision maps with $N = 2$ cuts was 6.7 times faster than LOT and 424 times faster than Wassmap.

Figure 6: Distance matrix comparison for translation example with parameter set Θ_3. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and Wasserstein distance.

Distance	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	4.76%	4.62%	4.56%

Table 3: Relative error in Frobenius norm of Wasserstein distance approximation using LOT and no-collision distances for translation example with parameter set Θ_3.

Figure 7 confirms once again out theoretical findings showing that methods based on Wasserstein distance (or approximations of it) are able to correctly reconstruct the triangular translation grid up to rigid transformation and scaling, while the other methods cannot.

Figure 7: Translation manifold generated by the characteristic function of the unit disk with parameter set Θ_3. We show the original translation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

6.2 Dilation Manifold

We fix a base measure μ_0 to be the indicator function of a disc of radius 1 centered at the origin. For a given dilation set $\Theta = \{\theta = (\theta_1, \theta_2)\} \subset \mathbb{R}^2$ the corresponding dilation manifold is given by:

$$\mathcal{M}^{\mathrm{dil}}(\mu_0, \Theta) := \{(\theta \odot x)\sharp\mu_0;\ \theta \in \Theta\}$$

In this example we consider $\Theta = [0.5, 2] \times [0.5, 4]$ and sample it on a uniform 6×6 grid.

As before, in Figure 8 we compare the distance matrices obtained for the dilation case. In the first row, from left to right we represent the true Wasserstein distance, LOT distance and no-collision distances; in the second row we represent the L^2 error in the approximation of the Wasserstein distance. For no-collision maps we use $N = 3$ cuts for center of mass and $N = 6$ cuts for geometric center since these were the choices that gave the best results. As reference for LOT in this case we picked one Gaussian centered at $(0, 0)$ with covariance matrix $25I$. In Table 4 we report the relative error in Frobenius norm for each distance. From these experiments we can see that LOT and no-collision provide an accurate approximation of the Wasserstein distance with relative errors that are less than 7%. However, the no-collision distance approximation using center of mass is the one that provides the best accuracy with relative errors that are less than 1.5%.
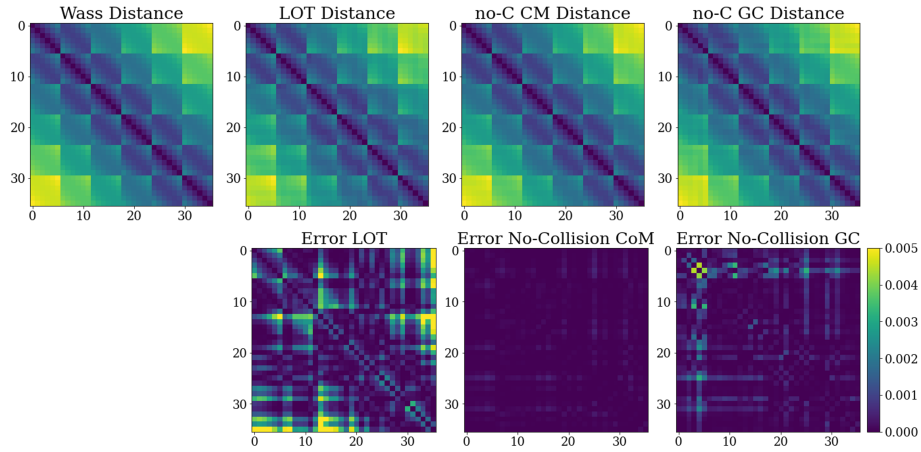
Figure 8: Distance matrix comparison for dilation example. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and Wasserstein distance.

Distance	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	6.30%	1.40%	3.32%

Table 4: Relative error in Frobenius norm of Wasserstein distance approximation using LOT and no-collision distances for dilation example with parameter set Θ.

As in the previous example we can compare the computational time needed for the different methods to compute the Wasserstein distance. In this case, with 36 dilation points Wasserstein distance computation took 72.4 seconds, LOT embedding took 1.51 seconds while no-collision maps took 0.135 seconds which is 11 times faster than LOT and 536 times faster than Wassmap. In Figure 9 we compare the computational time needed to produce the distance matrix for an increasing number of dilation points for Wassmap, LOT with one Gaussian reference and no-collision maps with center of mass for multiple N. As before we observe an faster increase in time for Wassmap as the number of dilation points increase when compared with LOT and no-collision, with no-collision for $N = 3$ being the fastest method for all amounts of dilation points.
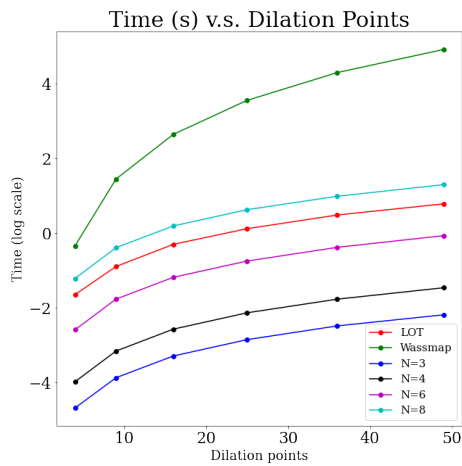
32

Figure 9: Computational time comparison for Wassmap and no-collision for multiple cuts and increasing number of dilation points

In Figure 10 we show the dilation grids obtained by MDS embedding in a 2-dimensional space using the distance matrices provided by the different methods. Again, in line with out theoretical results in Section 4.2, no-collision maps are able to reconstruct the dilation manifold correctly modulo a rigid transformation and a scaling. Wassmap also produces an accurate embedding while LOT is able to capture the correct form of the manifold, but shows some skewing at the boundary of the grid. On the contrary MDS, Isomap and Diffusion Maps on pixel features are not able to reconstruct the underlying dilation manifold. Again this example shows that the advantage of no-collision maps over Wassmap and LOT is that they require only a fraction of the computational time while still producing accurate distance approximation and embeddings.
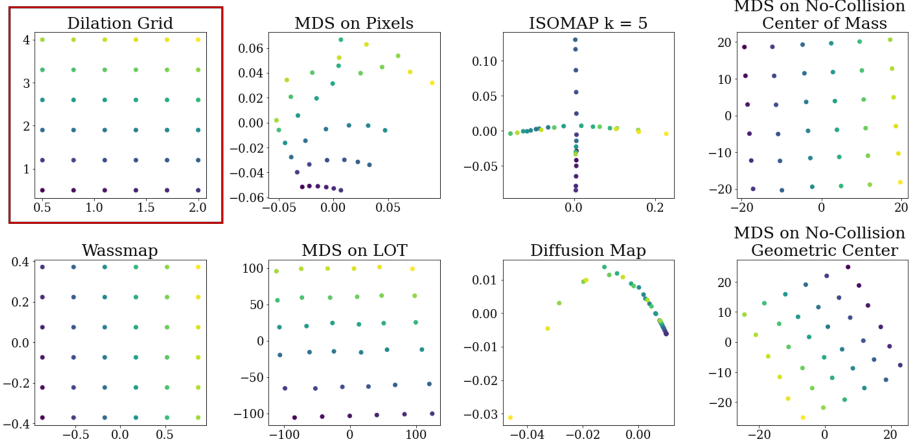
Figure 10: Dilation manifold generated by the characteristic function of the unit disk with parameter set Θ sampled on a 6×6 grid. We show the original dilation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

6.3 Rotation Manifold

As proved in Section 5, in general the rotation manifold

$$\mathcal{M}^{\mathrm{rot}}(\mu_0, \mathbb{S}^1) = \{(R_t x)\sharp\mu_0 \ : \ t \in [0, 2\pi]\}$$

is not isometric to a circle. Specifically, in Section 5 we showed that if $a, b > 0$, $a \neq b$, and μ_0 is the uniform probability measure over the elliptical domain

$$E_0 = \left\{ (x_1, x_2) \in \mathbb{R}^2 \ : \ \frac{x_1^2}{a^2} + \frac{x_2^2}{b^2} \leq 1 \right\}.$$

and $\mu_t = (R_t x)\sharp\mu_0$, then $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_2\right)$ is not isometric to a circle. On the other hand, if $a = b$ then the two spaces are isometric.

In this section we show two rotation examples. In the first one we define μ_0 to be the uniform probability measure over the ellipse of center $(0, 1)$ with $a = b = 2$. In this case the isometry is satisfied and we show that all methods are able to reconstruct the rotation grid correctly in Euclidean space. In the second example we define μ_0 to be the uniform probability measure over the ellipse of center $(0, 1)$ with $a = 5, b = 2$ and in this case, since there is no isometry between $\left(\{\mu_t\}_{t\in[0,2\pi]}, W_2\right)$ and a circle, we will see that the methods fail to reconstruct the ellipse correctly in Euclidean space. Note that this consideration only concerns the accuracy of the embedding in Euclidean space.

34

6.3.1 Rotation of ellipse centered at $(0,1)$ with $a = b = 2$

As in the previous example we start by examining the approximation of the Wasserstein distance matrix given by LOT and no-collision maps. As shown in Section 5, in the case of the rotation of an ellipse we can compute the Wasserstein distance analytically (4). In Figure 11 we compare the distance matrices obtained in this case. In the first row, from left to right we represent the true analytical Wasserstein distance, the computational Wasserstein distance (computed using the POT package in Python), the LOT distance, and no-collision distances; in the second row we represent the L^2 error in the approximation of the Wasserstein distance. For no-collision maps we use $N = 2$ cuts for center of mass and $N = 4$ cuts for geometric center since these were the choices that gave the best results. In Table 5 we report the relative error in Frobenius norm for each distance. From these experiments we can see that, as expected, the most accurate approximation of the true analytical Wasserstein distance is given by the Wasserstein distance computed in Python. However, LOT and no-collision also provide an accurate approximation of the Wasserstein distance with relative errors that are less than 4.5%. In particular, the no-collision distance approximation using center of mass is the one that provides the best accuracy with relative errors that are less than 1.6%.
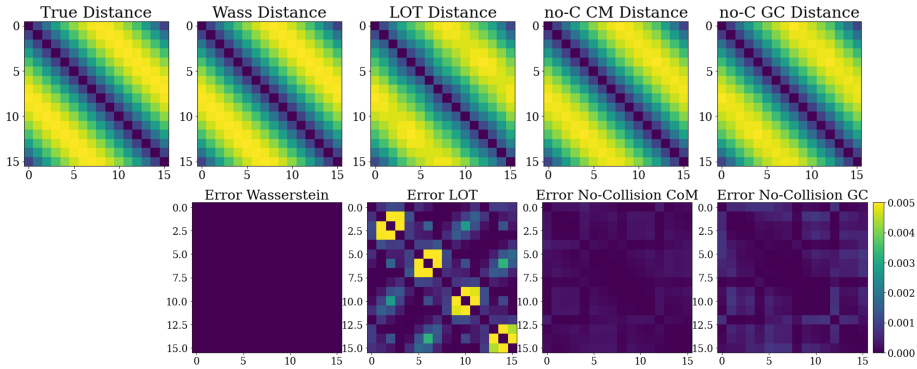

Figure 11: Distance matrix comparison for rotation example with $a = 2, b = 2$. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and True analytical Wasserstein distance.

Distance	POT computed Wasserstein	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	0.143%	4.62%	1.58%	2.27%

Table 5: Relative error in Frobenius norm of Wasserstein distance approximation using Python, LOT and no-collision distances for rotation of ellipse with $a = b = 2$.

Not only no-collision distances provide accurate Wassertein distance approx-

imation in this case, but they are also computationally very fast. In this example, with 16 rotation points Wasserstein distance computation with Python took 14.5 seconds, LOT embedding took 0.39 seconds while no-collision maps took 0.064 seconds which is 6 times faster than LOT and 226 times faster than Wassmap. In Figure 12 we compare the computational time to compute the distance matrix for an increasing number of rotation points for Wassmap, LOT with one Gaussian reference and no-collision maps with center of mass for multiple N. The computational time comparison is similar to the previous cases, with no-collision for $N = 2$ being the fastest method for all amounts of rotation points.
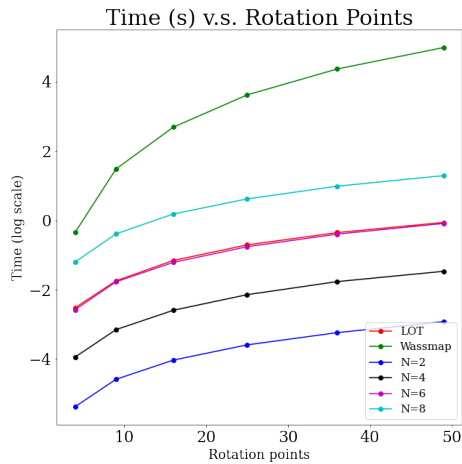

Figure 12: Computational time comparison for Wassmap and no-collision for multiple cuts and increasing number of rotation points $a = 2, b = 2$

In Figure 13 we show the rotation grids obtained by MDS embedding in a 2 dimensional space using the distance matrices provided by the different methods. Since in this example we selected $a = b = 2$ and the isometry is satisfied all methods based on Wasserstein distance (or an approximation of it) are able to reconstruct the rotation grid exactly. However, note that in this case MDS on pixel features did not provide a good approximation.
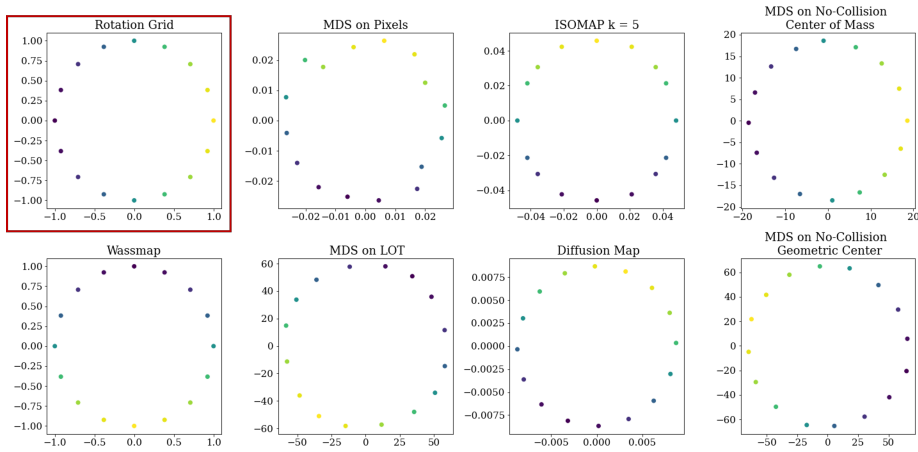
Figure 13: Rotation manifold generated an ellipse centered at $(0, 1)$ with axis $a = b = 2$. We show the original rotation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

6.3.2 Rotation of ellipse centered at $(0, 1)$ with $a = 5, b = 2$

As in the previous example we start by examining the approximation of the Wasserstein distance matrix given by LOT and no-collision maps. In Figure 14 we compare the distance matrices obtained when rotating an ellipse centered at $(0, 1)$ with $a = 5, b = 2$. For no-collision maps we use $N = 3$ cuts for center of mass and $N = 8$ cuts for geometric center since these were the choices that gave the best results. In Table 6 we report the relative error in Frobenius norm for each distance. We can see that like in the previous case the most accurate approximation of the true analytical Wasserstein distance is given by the Wasserstein distance computed in Python. We notice, however, that the approximation of the Wasserstein distance given by LOT and no-collision map is much less accurate than in the previous case with LOT error being less than 10% and no-collision error with center of mass being less than 16%.

We omit here the comparison of the computational time since they are the same as in the previous case.
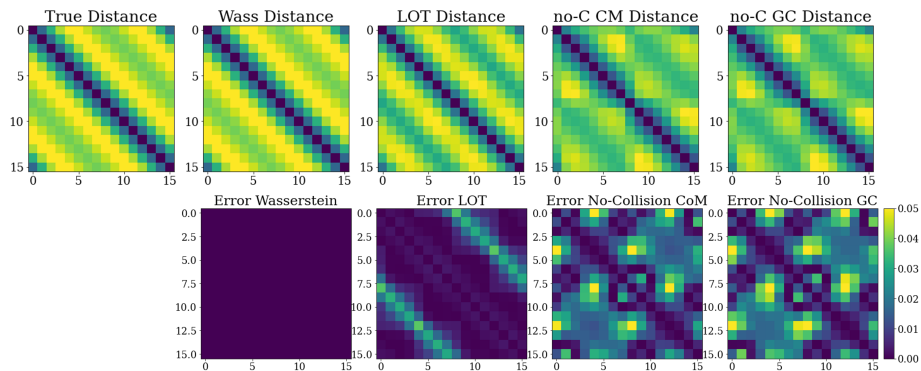
Figure 14: Distance matrix comparison for rotation example with $a = 5, b = 2$. **Top:** Distance matrices. **Bottom:** Squared Error between approximate distance matrices and True Wasserstein distance.

Distance	POT computed Wasserstein	LOT	No-Collision Center or Mass	No-Collision Geometric Center
Relative Error	0.301%	9.77%	15.2%	16.4%

Table 6: Relative error in Frobenius norm of Wasserstein distance approximation using Python, LOT and no-collision distances for rotation of ellipse with $a = 5, b = 2$.

Finally, in Figure 15 we report the 2 dimensional MDS embedding of the rotation. As expected, since in this example we selected $a = 5$, $b = 2$ and the isometry is not satisfied, all methods based on Wasserstein distance (or an approximation of it) are unable to reconstruct the rotation grid exactly. Note however that in this case Diffusion Maps are able to reconstruct the rotation manifold correctly.
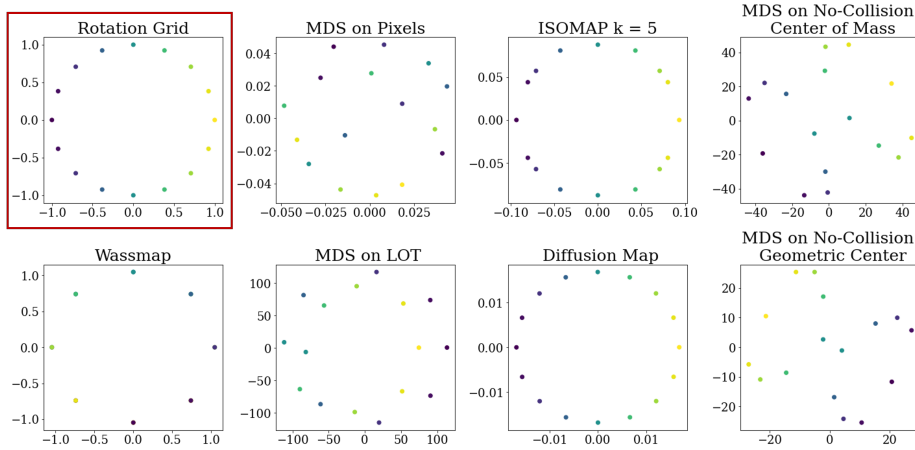
Figure 15: Rotation manifold generated an ellipse centered at $(0,1)$ with axis $a = 5$, $b = 2$. We show the original rotation grid (circled in red), and the embeddings obtained by MDS, Diffusion Maps, Isomap on pixel features, Wassmap, MDS on LOT features and MDS on no-collision features.

7 Acknowledgements

References

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

[2] Jason Altschuler, Francis Bach, Alessandro Rudi, and Jonathan Niles-Weed. Massively scalable sinkhorn distances via the nyström method. *Advances in neural information processing systems*, 32, 2019.

[3] David Alvarez-Melis and Tommi S Jaakkola. Gromov-wasserstein alignment of word embedding spaces. *arXiv preprint arXiv:1809.00013*, 2018.

[4] Luigi Ambrosio, Alberto Bressan, Dirk Helbing, Axel Klar, Enrique Zuazua, Luigi Ambrosio, and Nicola Gigli. A user's guide to optimal transport. *Modelling and Optimisation of Flows on Networks: Cetraro, Italy 2009, Editors: Benedetto Piccoli, Michel Rascle*, pages 1–155, 2013.

[5] Luigi Ambrosio, Nicola Gigli, and Giuseppe Savaré. Gradient flows with metric and differentiable structures, and applications to the Wasserstein

space. *Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl.*, 15(3-4):327–343, 2004.

[6] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[7] Tyrus Berry, Dimitrios Giannakis, and John Harlim. Nonparametric forecasting of low-dimensional dynamical systems. *Physical Review E*, 91(3):032915, 2015.

[8] Mauro Bonafini and Bernhard Schmitzer. Domain decomposition for entropy regularized optimal transport. *Numerische Mathematik*, 149(4):819–870, 2021.

[9] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51:22–45, 2015.

[10] Yann Brenier. Polar factorization and monotone rearrangement of vector-valued functions. *Communications on pure and applied mathematics*, 44(4):375–417, 1991.

[11] Dong Chen and Hans-Georg Müller. Nonlinear manifold representations for functional data. 2012.

[12] Alexander Cloninger, Keaton Hamm, Varun Khurana, and Caroline Moosmüller. Linearized wasserstein dimensionality reduction with approximation guarantees, 2023.

[13] Ronald R Coifman and Stéphane Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.

[14] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. *Advances in neural information processing systems*, 26, 2013.

[15] Julie Digne, David Cohen-Steiner, Pierre Alliez, Fernando De Goes, and Mathieu Desbrun. Feature-preserving surface reconstruction and simplification from defect-laden point sets. *Journal of mathematical imaging and vision*, 48:369–382, 2014.

[16] Charles Fefferman, Sanjoy Mitter, and Hariharan Narayanan. Testing the manifold hypothesis. *Journal of the American Mathematical Society*, 29(4):983–1049, 2016.

[17] Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z. Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, Léo Gautheron, Nathalie T.H. Gayraud, Hicham Janati, Alain Rakotomamonjy, Ievgen Redko, Antoine Rolet, Antony Schutz, Vivien Seguy, Danica J. Sutherland, Romain Tavenard,

Alexander Tong, and Titouan Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

[18] Matthias Gelbrich. On a formula for the l2 wasserstein metric between measures on euclidean and hilbert spaces. *Mathematische Nachrichten*, 147(1):185–203, 1990.

[19] Samuel Gerber and Mauro Maggioni. Multiscale strategies for computing optimal transport. *arXiv preprint arXiv:1708.02469*, 2017.

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

[21] Peter W Jones, Mauro Maggioni, and Raanan Schul. Manifold parametrizations by eigenfunctions of the laplacian and heat kernels. *Proceedings of the National Academy of Sciences*, 105(6):1803–1808, 2008.

[22] 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, 2023.

[23] Joe Kileel, Amit Moscovich, Nathan Zelesko, and Amit Singer. Manifold learning with arbitrary norms. *Journal of Fourier Analysis and Applications*, 27(5):82, 2021.

[24] 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.

[25] Soheil Kolouri, Se Rim Park, Matthew Thorpe, Dejan Slepčev, and Gustavo K. Rohde. Optimal mass transport: Signal processing and machine-learning applications. *IEEE Signal Processing Magazine*, 34(4):43–59, 2017.

[26] Soheil Kolouri and Gustavo K Rohde. Transport-based single frame super resolution of very low resolution face images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4876–4884, 2015.

[27] Joseph B Kruskal and Myron Wish. *Multidimensional scaling*, volume 11. Sage, 1978.

[28] John A Lee, Michel Verleysen, et al. *Nonlinear dimensionality reduction*, volume 1. Springer, 2007.

[29] Tianyi Lin, Nhat Ho, and Michael I Jordan. On the efficiency of entropic regularized algorithms for optimal transport. *Journal of Machine Learning Research*, 23(137):1–42, 2022.

[30] Yunqian Ma and Yun Fu. *Manifold learning theory and applications*. CRC press, 2011.

[31] Robert J McCann. Polar factorization of maps on riemannian manifolds. *Geometric & Functional Analysis GAFA*, 11(3):589–608, 2001.

[32] Caroline Moosmüller and Alexander Cloninger. Linear optimal transport embedding: Provable wasserstein classification for certain rigid transformations and perturbations. *Information and Inference: A Journal of the IMA*, 12(1):363–389, 2023.

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

[34] 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.

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

[36] Filippo Santambrogio. Optimal transport for applied mathematicians. *Birkäuser, NY*, 55(58-63):94, 2015.

[37] Bernhard Schmitzer. A sparse multiscale algorithm for dense optimal transport. *Journal of Mathematical Imaging and Vision*, 56:238–259, 2016.

[38] Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

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

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

[41] Cédric Villani et al. *Optimal transport: old and new*, volume 338. Springer, 2009.

[42] 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.