

# MULTIDIMENSIONAL SCALING ON METRIC MEASURE SPACES

HENRY ADAMS, MARK BLUMSTEIN, AND LARA KASSAB

**ABSTRACT.** Multidimensional scaling (MDS) is a popular technique for mapping a finite metric space into a low-dimensional Euclidean space in a way that best preserves pairwise distances. We overview the theory of classical MDS, along with its optimality properties and goodness of fit. Further, we present a notion of MDS on infinite metric measure spaces that generalizes these optimality properties. As a consequence we can study the MDS embeddings of the geodesic circle  $S^1$  into  $\mathbb{R}^m$  for all  $m$ , and ask questions about the MDS embeddings of the geodesic  $n$ -spheres  $S^n$  into  $\mathbb{R}^m$ . Finally, we address questions on convergence of MDS. For instance, if a sequence of metric measure spaces converges to a fixed metric measure space  $X$ , then in what sense do the MDS embeddings of these spaces converge to the MDS embedding of  $X$ ?

## 1. INTRODUCTION

Given  $n$  objects and a notion of *dissimilarity* between them, the classical multidimensional scaling (MDS) algorithm extracts a configuration of  $n$  points in Euclidean space whose pairwise distances “best” approximate the given dissimilarities. A typical source of dissimilarity data is the distance between high-dimensional objects, in which case MDS serves as a non-linear dimensionality reduction and visualization technique. As such, the MDS algorithm is a popular technique for pattern recognition problems. In this paper, we survey the classical algorithm, and describe an extension to (possibly infinite) metric measure spaces.

The coordinates extracted from an MDS embedding satisfy a least squares optimization problem. While there are several popular choices of MDS loss function (metric or non-metric), we primarily focus on the classical algorithm which minimizes a form of loss function known as *strain*. The classical algorithm is algebraic and not iterative, simple to implement, and guaranteed to discover a configuration which optimizes the strain function. Furthermore, if the input dissimilarities can be realized as distances in a Euclidean space, then classical MDS is guaranteed to recover such a configuration (unique up to translation and orthogonal transformation). However, not all dissimilarity data admits a Euclidean realization. In this case MDS produces a mapping into Euclidean space that distorts the inter-point pairwise distances as little as possible. We make these ideas precise in Section 2.

The classical story is told using finite samples of points, finite dissimilarity matrices, and finite embedding coordinates. Our goal is to extend to an infinite setting, where our input dissimilarity data is replaced by a *metric measure space*: a metric space (with possibly infinitely many points) equipped with some probability measure. This allows us to consider spaces whose points are weighted unequally, along with notions of convergence as more and more points are sampled from an infinite shape.

In more detail, a *metric measure space* is a triple  $(X, d, \mu)$  where  $(X, d)$  is a compact metric space, and  $\mu$  is a Borel probability measure on  $X$ . In Section 4 we generalize the classical MDS algorithm to metric measure spaces, and we show that this generalization minimizes the infinite analogue of strain. As a motivating example, we consider the MDS embedding of the circle with the (non-Euclidean) geodesic metric, and equipped with the uniform measure. By using the properties of circulant matrices, we identify the MDS embeddings of evenly-spaced points from the geodesic circle into  $\mathbb{R}^m$ , for all  $m$ . As the number of points tends to infinity, these embeddings lie along the curve

$$\sqrt{2} \left( \cos \theta, \sin \theta, \frac{1}{3} \cos 3\theta, \frac{1}{3} \sin 3\theta, \frac{1}{5} \cos 5\theta, \frac{1}{5} \sin 5\theta, \dots \right) \in \mathbb{R}^m.$$

As this example illustrates, it is useful to consider the situation where a sequence of metric measure spaces  $X_n$  converges to a fixed metric measure space  $X$  as  $n \rightarrow \infty$ . We survey various notions of convergence in Section 6.

Convergence is well-understood when each metric space has the same finite number of points, for example by Sibson’s perturbation analysis [22]. However, we are also interested in convergence when the number of points varies and is possibly infinite. We survey results of [1, 14] on the convergence of MDS when  $n$  points  $\{x_1, \dots, x_n\}$  are sampled from a metric space according to a probability measure  $\mu$ , in the limit as  $n \rightarrow \infty$ . The law of large numbers describes how the finite measures  $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$  converge

to  $\mu$  as  $n \rightarrow \infty$ . In [13], we reprove these results when instead we are given an *arbitrary* sequence of probability measures  $\mu_n \rightarrow \mu$ . The measures  $\mu_n$  may now be unequally weighted, or have infinite support, for example.

**Organization.** We present an overview on the theory of classical MDS in Section 2. In Section 3, we present necessary background information on operator theory and infinite-dimensional linear algebra. We define a notion of MDS for infinite metric measure spaces in Section 4. In Section 5, we identify the MDS embeddings of the geodesic circle into  $\mathbb{R}^m$ , for all  $m$ , as a motivating example. Lastly, in Section 6, we survey different notions of convergence of MDS.

**Related Work.** The reader is referred to the introduction of [25] and to [10, 12] for some aspects of the history of MDS. There are a variety of papers that study some notion of robustness or convergence of MDS, including [1, 21, 22, 23]. Furthermore, [19, Section 3.3] considers embedding new points in pseudo-Euclidean spaces, [11, Section 3] considers infinite MDS in the case where the underlying space is an interval (equipped with some metric), and [7, Section 6.3] discusses MDS on large numbers of objects.

## 2. CLASSICAL SCALING

Multidimensional scaling (MDS) is a set of statistical techniques concerned with the problem of using information about the dissimilarities between  $n$  objects in order to construct a configuration of  $n$  points in Euclidean space. The input dissimilarities between the objects need not be based on Euclidean distances.

**Definition 2.1.** An  $(n \times n)$  matrix  $\mathbf{D}$  is called a *dissimilarity matrix* if it is symmetric and

$$d_{rr} = 0, \quad \text{with } d_{rs} \geq 0 \quad \text{for all } r \neq s.$$

The first property above is called reflexivity (the dissimilarity between an object and itself is zero), and the second property is called nonnegativity. Symmetry requires that the dissimilarity from object  $r$  to  $s$  is the same as that from  $s$  to  $r$ . Note that there is no need to satisfy the triangle inequality. A dissimilarity matrix  $\mathbf{D}$  is called *Euclidean* if there exists a configuration of points in some Euclidean space whose interpoint distances are given by  $\mathbf{D}$ .

The goal of MDS is to map the objects  $x_1, \dots, x_n$  to a configuration (or embedding) of points  $f(x_1), \dots, f(x_n)$  in  $\mathbb{R}^m$  so that the given dissimilarities  $d(x_i, x_j)$  are well-approximated by the Euclidean distances  $\|f(x_i) - f(x_j)\|_2$ . The different notions of approximation give rise to the different types of MDS.

If the dissimilarity matrix can be realized exactly as the distance matrix of some set of points in  $\mathbb{R}^m$  (i.e. if the dissimilarity matrix is *Euclidean*), then MDS will find such a realization. Furthermore, MDS can be used to identify the minimum Euclidean dimension  $m$  admitting such an isometric embedding. However, some dissimilarity matrices or metric spaces are inherently non-Euclidean (cannot be embedded into  $\mathbb{R}^m$  for any  $m$ ). When a dissimilarity matrix is not Euclidean, then MDS produces a mapping into  $\mathbb{R}^m$  that distorts the interpoint pairwise distances as little as possible. Though we introduce MDS below, the reader is also referred to [2, 8, 12] for more complete introductions.

Classical multidimensional scaling (cMDS) is also known as Principal Coordinates Analysis (PCoA), Torgerson Scaling, or Torgerson–Gower scaling. The cMDS algorithm minimizes a loss function called *strain*, and one of the main advantages of cMDS is that its algorithm is algebraic and not iterative. Therefore, it is simple to implement, and it is guaranteed to discover the optimal configuration in  $\mathbb{R}^m$ . In this section, we describe the cMDS algorithm, and then discuss some of its optimality properties and goodness of fit.

As an illustrative example, we consider ten U.S. cities equipped with the road distance between them, which is a non-Euclidean distance. The classical MDS algorithm produces a two dimensional configuration of points (see Figure 1), where the points represent the different cities. The Euclidean pairwise distances (distances as the crow flies) between the cities in the MDS embedding are the Euclidean distances that best approximate the road distances between them.

Let  $\mathbf{D} = (d_{ij})$  be an  $n \times n$  dissimilarity matrix. Let  $\mathbf{A} = (a_{ij})$ , where  $a_{ij} = -\frac{1}{2}d_{ij}^2$ . Define the matrix  $\mathbf{B}$  to be the double mean-centering of  $\mathbf{A}$ , with entries given by

$$b_{rs} = a_{rs} - \frac{1}{n} \sum_{s=1}^n a_{rs} - \frac{1}{n} \sum_{r=1}^n a_{rs} + \frac{1}{n^2} \sum_{r,s=1}^n a_{rs}. \quad (1)$$

Since  $\mathbf{D}$  is a symmetric matrix, it follows that  $\mathbf{A}$  and  $\mathbf{B}$  are each symmetric, and therefore  $\mathbf{B}$  has  $n$  real eigenvalues.

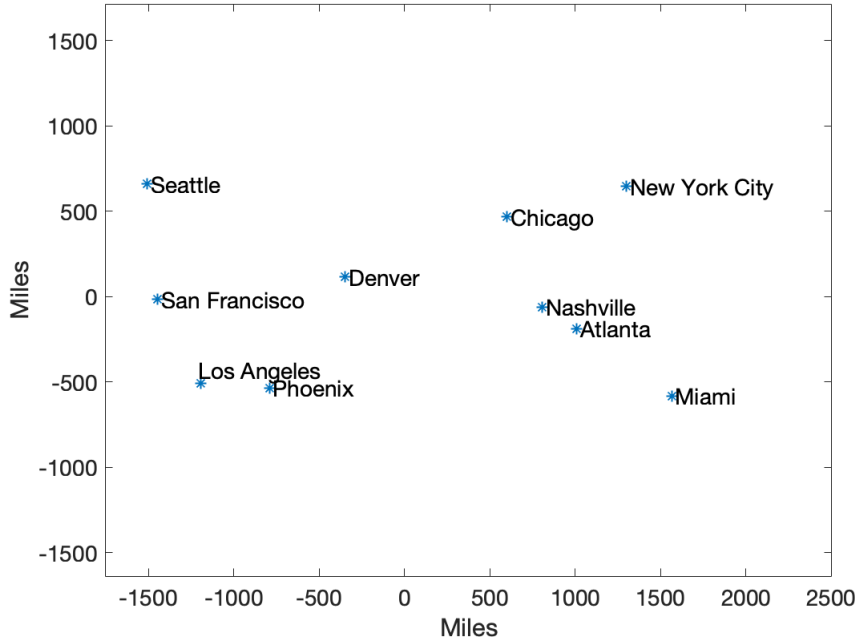


FIGURE 1. Reconstruction of a map of 10 U.S. cities using MDS

Assume for convenience that there are at least  $m$  positive eigenvalues for matrix  $\mathbf{B}$ , where  $m \leq n$ . By the spectral theorem of symmetric matrices, let  $\mathbf{B} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top$ , with  $\mathbf{\Gamma}$  containing unit-length eigenvectors of  $\mathbf{B}$  as its columns, and with the diagonal matrix  $\mathbf{\Lambda}$  containing the eigenvalues of  $\mathbf{B}$  in decreasing order along its diagonal. Let  $\mathbf{\Lambda}_m$  be the  $m \times m$  diagonal matrix of the largest  $m$  eigenvalues sorted in descending order, and let  $\mathbf{\Gamma}_m$  be the  $n \times m$  matrix of the corresponding  $m$  eigenvectors in  $\mathbf{\Gamma}$ . The coordinates of the MDS embedding into  $\mathbb{R}^m$  are then given by the  $n \times m$  matrix  $\mathbf{X} = \mathbf{\Gamma}_m \mathbf{\Lambda}_m^{1/2}$ . More precisely, the MDS embedding consists of the  $n$  points in  $\mathbb{R}^m$  given by the  $n$  rows of  $\mathbf{X}$ . The procedure for classical MDS can be summarized in the following algorithm.

---

**Algorithm 1:** Classical MDS

---

**input:** Dissimilarity matrix  $\mathbf{D}$

**compute:** Compute the matrix  $\mathbf{A} = (a_{ij})$ , where  $a_{ij} = -\frac{1}{2}d_{ij}^2$

**perform :** Perform double-centering to  $\mathbf{A}$ : define  $\mathbf{B}$  by (1)

**compute:** Compute the eigendecomposition of  $\mathbf{B} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top$

**select :** Select the  $m$  largest nonnegative eigenvalues of  $\mathbf{B}$  to obtain  $\mathbf{\Lambda}_m$

**output :** Coordinate matrix given by  $\mathbf{X} = \mathbf{\Gamma}_m \mathbf{\Lambda}_m^{1/2}$

---

We give a small example.

**Example 2.2.** We implement Algorithm 1 on the following  $4 \times 4$  dissimilarity matrix  $\mathbf{D}$ .

$$\mathbf{D} = \begin{pmatrix} 0 & 2 & 2 & 1 \\ 2 & 0 & 2 & 1 \\ 0 & 2 & 2 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \mathbf{A} = -\frac{1}{2} \begin{pmatrix} 0 & 4 & 4 & 1 \\ 4 & 0 & 4 & 1 \\ 0 & 4 & 4 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \mathbf{B} = \frac{1}{16} \begin{pmatrix} 13 & -15 & 5 & -3 \\ -15 & 21 & -7 & 1 \\ 5 & -7 & -3 & 5 \\ -3 & 1 & 5 & -3 \end{pmatrix}$$

The eigenvalues of  $\mathbf{B}$  are  $2.159\dots$ ,  $0.192\dots$ ,  $0$ , and  $-0.602\dots$ , and the MDS embedding of  $\mathbf{D}$  in  $\mathbb{R}^2$  is drawn in Figure 2.

This dissimilarity matrix is not Euclidean. Indeed, label the points  $x_1, x_2, x_3, x_4$  in order of their row/column in  $\mathbf{D}$ . In any isometric embedding in  $\mathbb{R}^n$ , the points  $x_1, x_2, x_3$  would be mapped to an equilateral triangle. The point  $x_4$  would need to get mapped to the midpoint of each edge of this triangle, which is impossible in Euclidean space.

The following fundamental criterion determines algebraically whether a dissimilarity matrix  $\mathbf{D}$  is Euclidean or not.

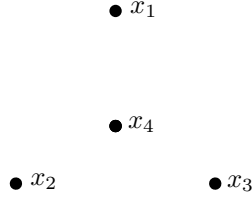


FIGURE 2. MDS embedding of the  $4 \times 4$  distance matrix in Example 2.2.

**Theorem 2.3.** [2, Theorem 14.2.1] *Let  $\mathbf{D}$  be a dissimilarity matrix, and define  $\mathbf{B}$  by equation (1). Then  $\mathbf{D}$  is Euclidean if and only if  $\mathbf{B}$  is a positive semi-definite matrix.*

Moreover, if  $\mathbf{B}$  is positive semi-definite of rank  $m$ , then a perfect realization of the dissimilarities can be found by a collection of points in  $m$ -dimensional Euclidean space.

Let  $\mathbf{D}$  be a dissimilarity matrix, and define  $\mathbf{B}$  via (1). A measure of the goodness of fit of MDS, even in the case when  $\mathbf{D}$  is not Euclidean, can be obtained as follows. If  $\hat{\mathbf{X}}$  is a fitted configuration in  $\mathbb{R}^m$  with centered inner-product matrix  $\hat{\mathbf{B}}$ , then a measure of the discrepancy between  $\mathbf{B}$  and  $\hat{\mathbf{B}}$  is the following strain function [16],

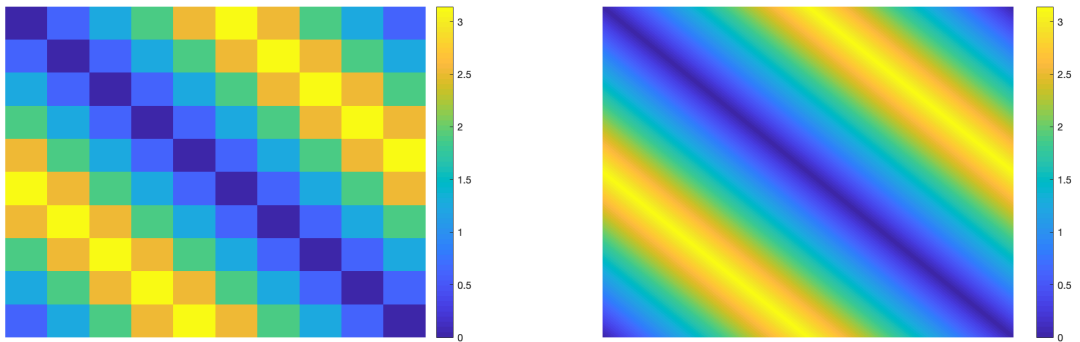
$$\text{tr}((\mathbf{B} - \hat{\mathbf{B}})^2) = \sum_{i,j=1}^n (b_{i,j} - \hat{b}_{i,j})^2. \quad (2)$$

**Theorem 2.4.** [2, Theorem 14.4.2] *Let  $\mathbf{D}$  be a dissimilarity matrix. Then for fixed  $m$ , the strain function in (2) is minimized over all configurations  $\hat{\mathbf{X}}$  in  $m$  dimensions when  $\hat{\mathbf{X}}$  is the classical solution to the MDS problem.*

The reader is referred to [8, Section 2.4] for a summary of a related optimization procedure with a different normalization, due to Sammon [20].

### 3. PRELIMINARIES

We are interested in studying the MDS embeddings of spaces with possibly infinitely many points, and distance matrices aren't enough to store infinitely many pairwise distances. Instead, we use *kernels*, which roughly speaking are distance functions that compute the pairwise distance between any two points in the space. For example, the kernel corresponding to the geodesic distance on a circle is illustrated in Figure 3b.



(A) Distance matrix of ten points.

(B) Distance kernel of infinitely many points.

FIGURE 3. Distance matrix or kernel corresponding to (A) ten equally-spaced points, and to (B) all of the points, on the circle equipped with the geodesic metric.

This section introduces the reader to concepts in infinite-dimensional linear algebra and operator theory used throughout the paper.

**Kernels and Operators.** Let  $X$  be a metric space equipped with a measure  $\mu$ . We denote by  $L^2(X, \mu)$  the set of square-integrable real-valued  $L^2$ -functions with respect to the measure  $\mu$ . We note that  $L^2(X, \mu)$  is furthermore a Hilbert space, after equipping it with the inner product given by  $\langle f, g \rangle = \int_X fg \, d\mu$ .

A *real-valued  $L^2$ -kernel*  $K: X \times X \rightarrow \mathbb{R}$  is a continuous measurable square-integrable function. The kernels that we consider in this paper are symmetric, meaning that  $K(x, s) = K(s, x)$  for all  $x, s \in X$ . A symmetric kernel is *positive semi-definite* if

$$\sum_{i=1}^n \sum_{j=1}^n c_i c_j K(x_i, x_j) \geq 0$$

holds for any  $m \in \mathbb{N}$ , any  $x_1, \dots, x_m \in X$ , and any  $c_1, \dots, c_m \in \mathbb{R}$ . At least in the case when  $X$  is a compact subspace of  $\mathbb{R}^m$  (and probably more generally), a symmetric kernel is positive semi-definite if

$$\int_X \int_X K(x, s) f(x) f(s) \mu(dx) \mu(ds) \geq 0$$

for any  $f \in L^2(X, \mu)$ .

**Definition 3.1** (Hilbert–Schmidt Integral Operator). Let  $(X, \Omega, \mu)$  be a  $\sigma$ -finite measure space, and let  $K$  be an  $L^2$ -kernel on  $X \times X$ . Then the integral operator

$$[T_K \phi](x) = \int_X K(x, s) \phi(s) \mu(ds),$$

which defines a linear mapping from the space  $L^2(X, \mu)$  into itself, is called a *Hilbert–Schmidt integral operator*.

Hilbert–Schmidt integral operators are both continuous (and hence bounded) and compact operators.

**Definition 3.2.** A Hilbert–Schmidt integral operator is a *self-adjoint operator* if  $K(x, y) = K(y, x)$  holds for almost all  $(x, y) \in X \times X$  (with respect to  $\mu \times \mu$ ).

**Definition 3.3.** A bounded self-adjoint operator  $T$  on a Hilbert space  $\mathcal{H}$  is a *positive semi-definite operator* if  $\langle Tx, x \rangle \geq 0$  for any  $x \in \mathcal{H}$ .

It follows that the eigenvalues of a positive semi-definite operator  $A$ , when they exist, are real.

**The Spectral Theorem.** Classical MDS relies on the fact that symmetric matrices are orthogonally diagonalizable with real eigenvalues. Furthermore, positive semi-definite matrices (having nonnegative eigenvalues) may be represented as matrices of Euclidean inner products. The following two theorems give analogues of these results for kernels instead of matrices.

**Theorem 3.4** (Spectral theorem on compact self-adjoint operators). *Let  $\mathcal{H}$  be a Hilbert space, and suppose  $T: \mathcal{H} \rightarrow \mathcal{H}$  is a bounded compact self-adjoint operator. Then  $T$  has at most a countable number of nonzero eigenvalues  $\lambda_n \in \mathbb{R}$ , with a corresponding orthonormal set  $\{e_n\}$  of eigenvectors, such that*

$$T(\cdot) = \sum_{n=1}^{\infty} \lambda_n \langle e_n, \cdot \rangle e_n.$$

*Furthermore, the multiplicity of each nonzero eigenvalue is finite, zero is the only possible accumulation point of  $\{\lambda_n\}$ , and if the set of nonzero eigenvalues is infinite then zero is an accumulation point.*

A fundamental theorem that characterizes positive semi-definite kernels is the Generalized Mercer’s Theorem.

**Theorem 3.5** (Generalized Mercer’s Theorem). [15, Lemma 1] *Let  $X$  be a compact topological Hausdorff space equipped with a finite Borel measure  $\mu$ , and let  $K: X \times X \rightarrow \mathbb{C}$  be a continuous positive semi-definite kernel. Then, there exists a scalar sequence  $\{\lambda_n\} \in \ell_1$  with  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ , and an orthonormal system  $\{\phi_n\}$  of continuous square-integrable functions with respect to  $\mu$ , such that the expansion*

$$K(x, s) = \sum_{n=1}^{\infty} \lambda_n \phi_n(x) \overline{\phi_n(s)}, \quad x, s \in \text{supp}(\mu)$$

*converges uniformly, where  $\text{supp}$  denotes the support of a measure  $\mu$ .*

Therefore, given  $X$  and  $K$  as in Theorem 3.5, the associated Hilbert–Schmidt integral operator

$$[T_K \phi](x) = \int_X K(x, s) \phi(s) \mu(ds)$$

is also positive semi-definite. Moreover, the eigenvalues of  $T_K$  can be arranged in non-increasing order  $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ , indexed according to their algebraic multiplicities, and the orthonormal system  $\{\phi_n\}$  gives the corresponding eigenfunctions of  $T_K$ .

#### 4. MDS OF INFINITE METRIC MEASURE SPACES

Classical multidimensional scaling (cMDS) can be described either as a strain-minimization problem, or as a linear algebra algorithm involving eigenvalues and eigenvectors. Indeed, one of the main theoretical results for cMDS is that the linear algebra algorithm solves the corresponding strain-minimization problem (see Theorem 2.4). In this section, we describe how to generalize both of these formulations to (possibly infinite) metric measure spaces.

This will allow us to discuss the MDS embedding of the circle, for example, without needing to restrict attention to finite subsets thereof.

**Definition 4.1.** A *metric measure space* is a triple  $(X, d, \mu)$  where

- $(X, d)$  is a compact metric space, and
- $\mu$  is a Borel probability measure on  $X$ , i.e.  $\mu(X) = 1$ .

Given a metric space  $(X, d)$ , by a measure on  $X$  we mean a measure on the Borel  $\sigma$ -algebra of  $X$ . When it is clear from the context, the triple  $(X, d, \mu)$  will be denoted by only  $X$ . The reader is referred to [17, 18] for details on metric measure spaces, and for interpretations of these concepts in the context of object matching.

Let  $(X, d, \mu)$  be a metric measure space, with  $d$  a  $L^2$ -function on  $X \times X$ . We say that  $X$  is *Euclidean* if it can be isometrically embedded into  $(\ell^2, \|\cdot\|_2)$ .  $X$  is furthermore Euclidean in the finite-dimensional sense if there is an isometric embedding  $X \rightarrow \mathbb{R}^m$ .

**MDS on Infinite Metric Measure Spaces.** Let  $(X, d, \mu)$  be a metric measure space, where  $d$  is an  $L^2$ -function on  $X \times X$ .

We propose the following MDS method on infinite metric measure spaces:

- (i) From the metric  $d$ , construct the kernel  $K_A: X \times X \rightarrow \mathbb{R}$  defined as

$$K_A(x, s) = -\frac{1}{2}d^2(x, s). \quad (3)$$

- (ii) Obtain the kernel  $K_B: X \times X \rightarrow \mathbb{R}$  via

$$K_B(x, s) = K_A(x, s) - \int_X K_A(w, s) \mu(dw) - \int_X K_A(x, z) \mu(dz) + \int_X \int_X K_A(w, z) \mu(dw) \mu(dz). \quad (4)$$

Assume  $K_B \in L^2(X \times X)$ . Define  $T_{K_B}: L^2(X) \rightarrow L^2(X)$  as

$$[T_{K_B} \phi](x) = \int_X K_B(x, s) \phi(s) \mu(ds). \quad (5)$$

- (iii) Let  $\lambda_1 \geq \lambda_2 \geq \dots$  denote the eigenvalues of  $T_{K_B}$ , with corresponding eigenfunctions  $\phi_1, \phi_2, \dots \in L^2(X)$  forming an orthonormal system in  $L^2(X)$ .

- (iv) Define  $K_{\hat{B}}(x, s) = \sum_{i=1}^{\infty} \hat{\lambda}_i \phi_i(x) \phi_i(s)$ , where  $\hat{\lambda}_i = \lambda_i$  if  $\lambda_i \geq 0$ , and otherwise  $\hat{\lambda}_i = 0$ . Let  $T_{K_{\hat{B}}}: L^2(X) \rightarrow$

$L^2(X)$  be the Hilbert–Schmidt integral operator associated to the kernel  $K_{\hat{B}}$ . The eigenfunctions  $\phi_i$  for  $T_{K_B}$  (with eigenvalues  $\lambda_i$ ) are also the eigenfunctions for  $T_{K_{\hat{B}}}$  (with eigenvalues  $\hat{\lambda}_i$ ). By Mercer’s Theorem (Theorem 3.5),  $K_{\hat{B}}$  converges uniformly.

- (v) Define the MDS embedding of  $X$  into  $\ell^2$  via the map  $f: X \rightarrow \ell^2$  given by

$$f(x) = \left( \sqrt{\hat{\lambda}_1} \phi_1(x), \sqrt{\hat{\lambda}_2} \phi_2(x), \sqrt{\hat{\lambda}_3} \phi_3(x), \dots \right) \quad (6)$$

Similarly, define the MDS embedding of  $X$  into  $\mathbb{R}^m$  via the map  $f_m: X \rightarrow \mathbb{R}^m$  given by

$$f_m(x) = \left( \sqrt{\hat{\lambda}_1} \phi_1(x), \sqrt{\hat{\lambda}_2} \phi_2(x), \dots, \sqrt{\hat{\lambda}_m} \phi_m(x) \right)$$

The procedure for infinite classical MDS can be summarized in the following algorithm.

---

**Algorithm 2:** Classical MDS on Infinite Spaces

---

**input:** Metric measure space  $(X, d, \mu)$   
**compute:** Compute the kernel  $K_A$  by (3)  
**perform :** Perform double-centering to  $K_A$ : define the kernel  $K_B$  by (4)  
**compute:** Compute the eigenvalues and eigenfunctions of the operator  $T_{K_B}$  defined by (5)  
**select** : Select the nonnegative eigenvalues of  $T_{K_B}$  to obtain a new operator  $T_{K_{\hat{B}}}$   
**output** : MDS embedding of  $X$  into  $\ell^2$  via the map  $f: X \rightarrow \ell^2$  given in (6)

---

TABLE 1. A comparison of various aspects of classical and infinite MDS. This table is constructed analogously to that on [27] Principal Component Analysis (PCA) and Functional Principal Component Analysis (FPCA).

Elements	Classical MDS	Infinite MDS
Data	$(X, d)$ with $ X  < \infty$	$(X, d, \mu)$
Distance Representation	$D_{i,j} = d(x_i, x_j), \quad D \in \mathcal{M}_{n \times n}$	$K_D(x, s) = d(x, s) \in L^2_{\mu \otimes \mu}(X \times X)$
Linear Operator	$b_{rs} = a_{rs} - \frac{1}{n} \sum_{s=1}^n a_{rs} - \frac{1}{n} \sum_{r=1}^n a_{rs} + \frac{1}{n^2} \sum_{r,s=1}^n a_{rs}$	$[T_{K_B} \phi](x) = \int_X K_B(x, s) \phi(s) \mu(ds)$
Eigenvalues	$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$	$\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots$
Eigenvectors	$v^{(1)}, v^{(2)}, \dots, v^{(m)} \in \mathbb{R}^n$	$\phi_1(x), \phi_2(x), \dots \in L^2(X)$
Embedding in $\mathbb{R}^m$ or $\ell^2$	$f(x_i) = (\sqrt{\lambda_1} v_1^{(i)}, \sqrt{\lambda_2} v_2^{(i)}, \dots, \sqrt{\lambda_m} v_m^{(i)})$	$f(x) = (\sqrt{\lambda_1} \phi_1(x), \sqrt{\lambda_2} \phi_2(x), \sqrt{\lambda_3} \phi_3(x), \dots)$
Strain Minimization	$\sum_{i,j=1}^n (b_{i,j} - \hat{b}_{i,j})^2$	$\int \int (K_B(x, s) - K_{\hat{B}}(x, s))^2 \mu(dx) \mu(ds)$

**Proposition 4.2.** [13, Proposition 6.3.1.] *The MDS embedding map  $f: X \rightarrow \ell^2$  is continuous.*

The following theorem generalizes Theorem 2.3 to metric measure spaces.

**Theorem 4.3.** [13, Theorem 6.3.3.] *A metric measure space  $(X, d, \mu)$  is Euclidean if and only if  $T_{K_B}$  is a positive semi-definite operator on  $L^2(X, \mu)$ .*

We show that MDS for metric measure spaces minimizes the loss function  $\text{Strain}(f)$ , defined as

$$\text{Strain}(f) = \|T_{K_B} - T_{K_{\hat{B}}}\|_{HS}^2 = \text{Tr}((T_{K_B} - T_{K_{\hat{B}}})^2) = \int \int (K_B(x, s) - K_{\hat{B}}(x, s))^2 \mu(dx) \mu(ds).$$

This result generalizes [2, Theorem 14.4.2], or equivalently [24, Theorem 2], to the infinite case.

**Theorem 4.4.** [13, Theorem 6.4.3.] *Let  $(X, d, \mu)$  be a metric measure space. Then  $\text{Strain}(f)$  is minimized over all maps  $f: X \rightarrow \ell^2$  or  $f: X \rightarrow \mathbb{R}^m$  when  $f$  is the MDS embedding given in Section 4.*

## 5. MDS OF THE CIRCLE

Let  $S^1$  be the unit circle equipped with arc-length distance and the uniform measure  $(\frac{d\theta}{2\pi})$ . Using our definition of MDS as an integral operator, we show that MDS maps  $S^1$  into an infinite dimensional sphere of radius  $\frac{\pi}{2}$  sitting inside  $\ell^2$ . The embedded circle occupies an infinite number of dimensions in

$\ell^2$ , and in fact, the infinite dimensional space is needed—the embedding is better (in the sense of strain minimization) than the MDS embedding into  $\mathbb{R}^m$  for any finite  $m$ .

It is instructive to consider how MDS on finite samples of  $S^1$  converges to the MDS integral operator on the entire circle. We start with the easiest case: let  $S_n^1$  be the sample of  $n$  evenly-spaced points on  $S^1$ .

**Proposition 5.1.** *The classical MDS embedding of  $S_n^1$  lies, up to a rigid motion of  $\mathbb{R}^m$ , on the curve  $\gamma_m: S_n^1 \rightarrow \mathbb{R}^m$  defined by*

$$\gamma_m(\theta) = (a_{1,n} \cos(\theta), a_{1,n} \sin(\theta), a_{3,n} \cos(3\theta), a_{3,n} \sin(3\theta), a_{5,n} \cos(5\theta), a_{5,n} \sin(5\theta), \dots) \in \mathbb{R}^m,$$

where  $\lim_{n \rightarrow \infty} a_{j,n} = \frac{\sqrt{2}}{j}$  (with  $j$  odd).

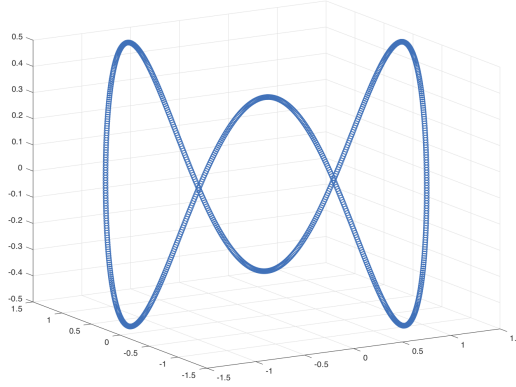


FIGURE 4. MDS embedding of  $S_{1000}^1$ .

Figure 4 shows the MDS configuration in  $\mathbb{R}^3$  of 1000 equally-spaced points on  $S^1$  obtained using the three largest positive eigenvalues.

We sketch the outline of this computation; full details are given in [13]. Let  $\mathbf{D}$  be the arc-length distance matrix for  $S_n^1$ . Following the steps of classical MDS, define  $\mathbf{A} = (a_{ij})$  with  $a_{ij} = -\frac{1}{2}d_{ij}^2$ , and let  $\mathbf{B}$  be the doubly mean-centered version of matrix  $\mathbf{A}$ . A matrix  $\mathbf{M}$  is called *circulant* if cyclically shifting all rows of  $\mathbf{M}$  down by one has the same effect as cyclically shifting all columns of  $\mathbf{M}$  left by one. Both  $\mathbf{D}$  and the double mean centering matrix have this property, and therefore the MDS symmetric matrix  $\mathbf{B}$  is circulant. In coordinates, it has the following form:

$$\mathbf{B} = \begin{pmatrix} b_0 & b_1 & b_2 & \dots & b_3 & b_2 & b_1 \\ b_1 & b_0 & b_1 & \dots & b_4 & b_3 & b_2 \\ b_2 & b_1 & b_0 & \dots & b_5 & b_4 & b_3 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ b_3 & b_4 & b_5 & \dots & b_0 & b_1 & b_2 \\ b_2 & b_3 & b_4 & \dots & b_1 & b_0 & b_1 \\ b_1 & b_2 & b_3 & \dots & b_2 & b_1 & b_0 \end{pmatrix}.$$

For example, if  $\mathbf{D}$  is the distance matrix for  $n = 7$  equally-spaced points on the circle, then we compute

$$\mathbf{D} = \frac{2\pi}{7} \begin{pmatrix} 0 & 1 & 2 & 3 & 3 & 2 & 1 \\ 1 & 0 & 1 & 2 & 3 & 3 & 2 \\ 2 & 1 & 0 & 1 & 2 & 3 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 3 & 2 & 1 & 0 & 1 & 2 \\ 2 & 3 & 3 & 2 & 1 & 0 & 1 \\ 1 & 2 & 3 & 3 & 2 & 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{B} = \frac{2\pi^2}{49} \begin{pmatrix} 4 & 3 & 0 & -5 & -5 & 0 & 3 \\ 3 & 4 & 3 & 0 & -5 & -5 & 0 \\ 0 & 3 & 4 & 3 & 0 & -5 & -5 \\ -5 & 0 & 3 & 4 & 3 & 0 & -5 \\ -5 & -5 & 0 & 3 & 4 & 3 & 0 \\ 0 & -5 & -5 & 0 & 3 & 4 & 3 \\ 3 & 0 & -5 & -5 & 0 & 3 & 4 \end{pmatrix}.$$

The complex eigenvectors of such a matrix are given by the *discrete Fourier modes*, namely  $x_k(n) := (w_n^0, w_n^k, \dots, w_n^{(n-1)k})^\top$  for  $0 \leq k \leq n-1$ , where  $w_n = e^{2\pi i/n}$ . Since the first entry of each vector  $x_k$  is

one, the eigenvalue of  $x_k$  can be computed simply by taking the dot product of the first row of  $\mathbf{B}$  with  $x_k$ . Note that the vector of all ones has eigenvalue zero.

Since  $\mathbf{B}$  is symmetric, each complex eigenvector can be split into its real and imaginary part, which forms two real eigenvectors—this explains the sine and cosine representation of eigenvectors in the proposition. It turns out that the odd Fourier modes have positive eigenvalues, and the even Fourier modes have negative eigenvalues. Since MDS retains coordinates corresponding to positive eigenvalues, we are left with only the odd Fourier modes.

How does this finite MDS computation compare to the MDS integral operator on all of  $S^1$ ? Let  $S^1$  be the unit circle with arc-length distance and uniform measure. If  $\phi_k(x) = e^{ikx}$ , then one may check (use integration by parts) that

$$-\frac{1}{2} \int_{y=x-\pi}^{y=x+\pi} (y-x)^2 e^{iky} \frac{dy}{2\pi} = \frac{1}{k^2} (-1)^{k+1} e^{ikx}.$$

Despite not having performed the double mean centering step to the kernel function, this computation shows that the (complex) eigenfunctions of MDS on  $S^1$  are  $\phi_k(x) = e^{ikx}$  with  $\lambda_k = \frac{1}{k^2} (-1)^{k+1}$ ,  $k \neq 0$ . Indeed, the mean centering step associates the eigenfunction  $\phi_0(x) = 1$  with the eigenvalue 0, and the other Fourier basis functions remain invariant to the double mean centering since they are perpendicular to  $\phi_0$ . Thus, as expected from Proposition 5.1, the MDS embedding  $\gamma$  of  $S^1$  is

$$\gamma(\theta) = \sqrt{2}(\cos \theta, \sin \theta, \frac{1}{3} \cos 3\theta, \frac{1}{3} \sin 3\theta, \frac{1}{5} \cos 5\theta, \frac{1}{5} \sin 5\theta, \dots) \in \ell^2,$$

where the  $\sqrt{2}$  is a normalization factor we picked up moving from a complex to a real eigendecomposition.

A couple of observations:

- (1) Applying the  $\ell^2$  Euclidean distance formula to the image of  $\gamma$  shows that for all  $\theta \in S^1$ ,

$$\|\gamma(\theta)\|_{\ell^2}^2 = 2 \sum_{k \text{ odd}} \frac{1}{k^2} = \frac{\pi^2}{4}.$$

That is, the MDS embedding lies on an infinite-dimensional sphere of radius  $\frac{\pi}{2}$  in  $\ell^2$ .

- (2) The  $\ell^2$  distance between  $\gamma(\theta_1)$  and  $\gamma(\theta_2)$  gives an approximation of the arc-length distance between angles  $\theta_1$  and  $\theta_2$ :

$$(\theta_1 - \theta_2)^2 \approx \|\gamma(\theta_1) - \gamma(\theta_2)\|_{\ell^2}^2 = 4 \sum_{k \text{ odd}} \frac{(1 - \cos(k(\theta_1 - \theta_2)))}{k^2}.$$

We leave it to the reader to verify that the expression above constitutes the odd modes in the Fourier series expansion of the periodic function  $(\theta_1 - \theta_2)^2$ . In fact, the error of MDS comes precisely from the even modes:

$$(\theta_1 - \theta_2)^2 = \|\gamma(\theta_1) - \gamma(\theta_2)\|_{\ell^2}^2 - \left( 4 \sum_{k \text{ even}} \frac{1 - \cos(k(\theta_1 - \theta_2))}{k^2} \right).$$

For this example, the issue of convergence of MDS on finite samples to MDS on the manifold is intuitively clear: the discrete Fourier modes converge (pointwise on the sample points) to the Fourier basis  $\phi(\theta) = e^{ik\theta}$ . However, in general the issue of convergence is not as straightforward. In the next section of the paper we survey results on convergence.

The MDS embeddings of the geodesic circle are closely related to [26], which was written prior to the invention of MDS. In [26, Theorem 1], von Neumann and Schoenberg describe (roughly speaking) which metrics on the circle one can isometrically embed into the Hilbert space  $\ell^2$ . The geodesic metric on the circle is not one of these metrics. However, the MDS embedding of the geodesic circle into  $\ell^2$  must produce a metric on  $S^1$  which is of the form described in [26, Theorem 1]. See also [28, Section 5] and [3, 6, 9].

## 6. CONVERGENCE OF MDS

We saw in the prior section how MDS on an evenly-spaced sample from the geodesic circle generalizes to the MDS integral operator on the entire circle. In this section, we address convergence questions for MDS more generally. Convergence is well-understood when each metric space has the same finite number of points [22], but we are also interested in convergence when the number of points varies and is possibly infinite.

**6.1. Robustness of MDS with Respect to Perturbations.** In a series of papers [21, 22, 23], the authors consider the robustness of multidimensional scaling with respect to perturbations of the underlying dissimilarity or distance matrix, as illustrated in Figure 5. In particular, [22] gives quantitative control over the perturbation of the eigenvalues and vectors determining an MDS embedding in terms of the perturbations of the dissimilarities. These results build upon the fact that if  $\lambda$  and  $v$  are a simple (i.e., non-repeated) eigenvalue and eigenvector of an  $n \times n$  matrix  $\mathbf{B}$ , then one can control the change in  $\lambda$  and  $v$  upon a small symmetric perturbation of the entries in  $\mathbf{B}$ .



FIGURE 5. Perturbation of the given dissimilarities.

Sibson’s perturbation analysis shows that if one has a converging sequence of  $n \times n$  dissimilarity matrices, then the corresponding MDS embeddings of  $n$  points into Euclidean space also converge. In the following sections, we consider the convergence of MDS when the number of points is not fixed. Indeed, we study the convergence of MDS when the number of points is finite but tending to infinity, and alternatively also when the number of points is infinite at each stage in a converging sequence of metric measure spaces.

**6.2. Convergence of MDS by the Law of Large Numbers.** Whereas Sibson’s perturbation analysis was for MDS on a fixed number of points, we now survey results on the convergence of MDS when  $n$  points  $\{x_1, \dots, x_n\}$  are sampled from a metric space according to a probability measure  $\mu$ , in the limit as  $n \rightarrow \infty$ , i.e. when more and more points are sampled. In [1], Bengio et al. study converging measures which are averages of Dirac delta functions, namely  $\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ , with all  $n$  of the random points  $x_i$  weighted equally (see Figure 6). Unsurprisingly, these results rely on the law of large numbers.

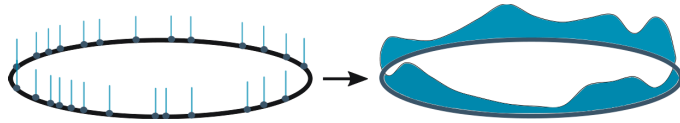


FIGURE 6. Illustration of a notion of convergence of measures discussed in [1].

Consider a data set  $X_n = \{x_1, \dots, x_n\}$  sampled independent and identically distributed (i.i.d.) from an unknown probability measure  $\mu$  on  $X$ . To generalize MDS, Bengio et al. define a corresponding data-dependent kernel that generalizes the mean centering matrix  $B$  (as defined in Section 4). Consequently, they study the convergence of eigenvalues and eigenfunctions of the integral operator associated to the kernel as the number of sampled points increases, and they show the convergence of the MDS embeddings under desirable conditions. They use a fundamental result on the convergence of eigenvalues of this type of integral operator from [14].

**6.3. Convergence of MDS for Arbitrary Measures.** In [13], we reprove the results of the previous section under a different setting which is more general in the sense that we allow for an arbitrary sequence of convergent measures, but which is easier in the sense that this sequence is fixed (i.e. deterministic, not random).

Indeed, let  $X$  be a compact metric space. Suppose  $\mu_n$  is an arbitrary sequence of probability measures on  $X_n$  for all  $n \in \mathbb{N}$ , such that  $\mu_n$  converges to  $\mu$  in total variation as  $n \rightarrow \infty$ . Roughly speaking, this notion of convergence of measures implies the uniform convergence of integrals against bounded measurable functions. For example, a measure  $\mu_n = \sum_{i=1}^n \lambda_i \delta_{x_i}$  in this sequence may again be a sum of Dirac delta functions, although now the weights  $\lambda_i > 0$  (with  $\sum_i \lambda_i = 1$ ) need not be identically equal to  $\frac{1}{n}$  (Figure 7a). Much more generally, the support of any  $\mu_n$  is now allowed to be infinite, as illustrated in Figure 7b. Following [1, 14], we give some first results towards showing that the MDS embeddings of  $(X, d, \mu_n)$  converge to the MDS embedding of  $(X, d, \mu)$  [13]. We similarly define a data-dependent kernel that generalizes the mean centering matrix  $B$  (as defined in Section 4). It is important to note that these kernels depend on the measure on the space. We again show convergence of eigenfunctions and consequently of MDS embeddings.

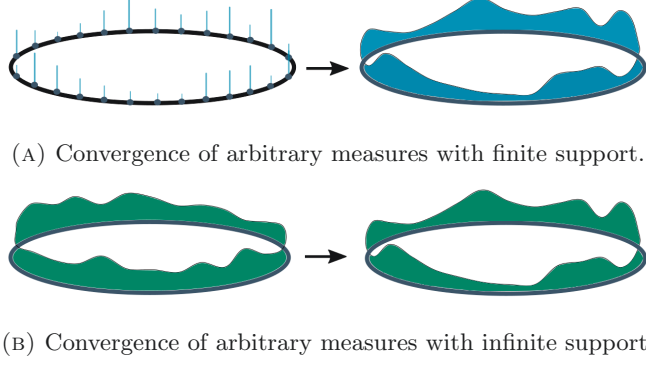


FIGURE 7. Illustration of convergence (in total variation) of arbitrary measures.

**6.4. Convergence of MDS with Respect to Gromov–Wasserstein Distance.** We now consider the more general setting in which  $(X_n, d_n, \mu_n)$  is an arbitrary sequence of metric measure spaces, converging to  $(X, d, \mu)$  in the Gromov–Wasserstein distance, as illustrated in Figure 8a for the finite case and Figure 8b for the infinite case. We remark that  $X_n$  need to no longer equal  $X$ , nor even be a subset of  $X$ . Indeed, the metric  $d_n$  on  $X_n$  is allowed to be different from the metric  $d$  on  $X$ . Sections 6.2 and 6.3 are the particular case when  $(X_n, d_n) = (X, d)$  for all  $n$ , and the measures  $\mu_n$  are converging to  $\mu$ . We now want to consider the case where metric  $d_n$  need no longer be equal to  $d$ .

The Wasserstein (or Kantorovich–Rubinstein) metric is a distance function defined between probability distributions on a given metric space  $X$ . Intuitively, if each distribution is viewed as a unit amount of “dirt” piled on  $X$ , the distance between two distributions is the minimum amount of work required to transform one pile of dirt into the other. More generally, the Gromov–Wasserstein distance between metric measure spaces takes into account not only the variation in measures, but also the variation in metrics between these spaces. Applications of the notion of Gromov–Wasserstein distance arise in shape and data analysis [18].

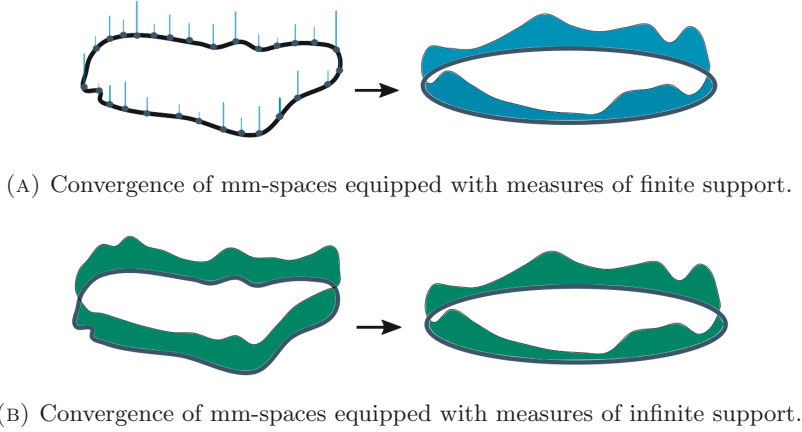


FIGURE 8. Illustration of Gromov–Wasserstein convergence of arbitrary metric measure spaces (mm-spaces).

**Conjecture 6.1.** Let  $(X_n, d_n, \mu_n)$  for  $n \in \mathbb{N}$  be a sequence of metric measure spaces that converges to  $(X, d, \mu)$  in the Gromov–Wasserstein distance. Then the MDS embeddings converge.

**Question 6.2.** Are there other notions of convergence of a sequence of arbitrary (possibly infinite) metric measure spaces  $(X_n, d_n, \mu_n)$  to a limiting metric measure space  $(X, d, \mu)$  that would imply that the MDS embeddings converge in some sense? We remark that one might naturally try to break this into two steps: first analyze which notions of convergence  $(X_n, d_n, \mu_n) \rightarrow (X, d, \mu)$  imply that the corresponding operators converge, and then analyze which notions of convergence on the operators imply that their eigendecompositions and MDS embeddings converge.

## 7. CONCLUSION

MDS is concerned with problem of mapping the objects  $x_1, \dots, x_n$  to a configuration (or embedding) of points  $f(x_1), \dots, f(x_n)$  in  $\mathbb{R}^m$  in such a way that the given dissimilarities  $d_{ij}$  are well-approximated by the Euclidean distances between  $f(x_i)$  and  $f(x_j)$ . We study a notion of MDS on metric measure spaces, which can be simply thought of as spaces of (possibly infinitely many) points equipped with some probability measure. We explain how MDS generalizes to metric measure spaces. Furthermore, we describe in a self-contained fashion an infinite analogue to the classical MDS algorithm. Indeed, classical multidimensional scaling can be described either as a strain-minimization problem, or as a linear algebra algorithm involving eigenvalues and eigenvectors. We describe how to generalize both of these formulations to metric measure spaces. We show that this infinite analogue minimizes a strain function similar to the strain function of classical MDS.

As a motivating example for convergence of MDS, we consider the MDS embeddings of the circle equipped with the (non-Euclidean) geodesic metric. By using the known eigendecomposition of circulant matrices, we identify the MDS embeddings of evenly-spaced points from the geodesic circle into  $\mathbb{R}^m$ , for all  $m$ . Indeed, the MDS embeddings of the geodesic circle are closely related to [26], which was written prior to the invention of MDS.

Lastly, we address convergence questions for MDS. Convergence is understood when each metric space in the sequence has the same finite number of points, or when each metric space has a finite number of points tending to infinity. We are also interested in notions of convergence when each metric space in the sequence has an arbitrary (possibly infinite) number of points. For instance, if a sequence of metric measure spaces converges to a fixed metric measure space  $X$ , then in what sense do the MDS embeddings of these spaces converge to the MDS embedding of  $X$ ?

Several questions remain open. In particular, we would like to have a better understanding of the convergence of MDS under the most unrestrictive assumptions of a sequence of arbitrary (possibly infinite) metric measure spaces converging to a fixed metric measure space. Is there a version that holds under convergence in the Gromov–Wasserstein distance, which that allows for distortion of both the metric and the measure simultaneously (see Conjecture 6.1 and Question 6.2)? Despite all of the work that has been done on MDS by a wide variety of authors, many interesting questions remain open (at least to us). For example, consider the MDS embeddings of the  $n$ -sphere for  $n \geq 2$ .

**Question 7.1.** What are the MDS embeddings of the  $n$ -sphere  $S^n$ , equipped with the geodesic metric, into Euclidean space  $\mathbb{R}^m$ ?

To our knowledge, the MDS embeddings of  $S^n$  into  $\mathbb{R}^m$  are not understood for all positive integers  $m$  except in the case of the circle, when  $n = 1$ . The above question is also interesting, even in the case of the circle, when the  $n$ -sphere is not equipped with the uniform measure. As a specific case, what is the MDS embedding of  $S^1$  into  $\mathbb{R}^m$  when the measure is not uniform on all of  $S^1$ , but instead (for example) uniform with mass  $\frac{2}{3}$  on the northern hemisphere, and uniform with mass  $\frac{1}{3}$  on the southern hemisphere?

We note the work of Blumstein and Kvinge [4], where a finite group representation theoretic perspective on MDS is employed. Adapting these techniques to the analytical setting of compact Lie groups may prove fruitful for the case of infinite MDS on higher dimensional spheres.

We also note the work [5], where the theory of an MDS embedding into pseudo Euclidean space is developed. In this setting, both positive and negative eigenvalues are used to create an embedding. In the example of embedding  $S^1$ , positive and negative eigenvalues occur in a one-to-one fashion. We wonder about the significance of the full spectrum of eigenvalues for the higher dimensional spheres.

## 8. ACKNOWLEDGEMENTS

We would like to thank Bailey Fosdick, Michael Kirby, Henry Kvinge, Facundo Mémoli, Louis Scharf, the students in Michael Kirby’s Spring 2018 class, and the Pattern Analysis Laboratory at Colorado State University for their helpful conversations and support throughout this project.

## REFERENCES

- [1] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-François Paiement, Pascal Vincent, and Marie Ouimet. Learning eigenfunctions links spectral embedding and kernel PCA. *Neural computation*, 16(10):2197–2219, 2004.
- [2] JM Bibby, JT Kent, and KV Mardia. *Multivariate analysis*, 1979.
- [3] Leonard Mascot Blumenthal. *Theory and applications of distance geometry*. Chelsea New York, 1970.
- [4] Mark Blumstein and Henry Kvinge. Letting symmetry guide visualization: Multidimensional scaling on groups. *arXiv preprint arXiv:1812.03362*, 2018.

- [5] Mark Blumstein and Louis Scharf. Pseudo Riemannian multidimensional scaling. 2019.
- [6] Eugène Bogomolny, Oriol Bohigas, and Charles Schmit. Spectral properties of distance matrices. *Journal of Physics A: Mathematical and General*, 36(12):3595, 2003.
- [7] Andreas Buja, Deborah F Swayne, Michael L Littman, Nathaniel Dean, Heike Hofmann, and Lisha Chen. Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics*, 17(2):444–472, 2008.
- [8] Trevor F Cox and Michael AA Cox. *Multidimensional scaling*. CRC press, 2000.
- [9] Jon Dattorro. *Convex optimization & Euclidean distance geometry*. Lulu.com, 2010.
- [10] Jan de Leeuw and Willem Heiser. 13 theory of multidimensional scaling. *Handbook of statistics*, 2:285–316, 1982.
- [11] Persi Diaconis, Sharad Goel, Susan Holmes, et al. Horseshoes in multidimensional scaling and local kernel methods. *The Annals of Applied Statistics*, 2(3):777–807, 2008.
- [12] PJ Groenen and Ingwer Borg. Past, present, and future of multidimensional scaling. *Visualization and verbalization of data*, pages 95–117, 2014.
- [13] Lara Kassab. Multidimensional scaling: Infinite metric measure spaces. Master’s thesis, Colorado State University, 2019. Available at arXiv 1904.07763.
- [14] Vladimir Koltchinskii and Evarist Giné. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):113–167, 2000.
- [15] Thomas Kühn. Eigenvalues of integral operators generated by positive definite Hölder continuous kernels on metric compacta. In *Indagationes Mathematicae (Proceedings)*, volume 90, pages 51–61. Elsevier, 1987.
- [16] Kanti V Mardia. Some properties of classical multi-dimensional scaling. *Communications in Statistics-Theory and Methods*, 7(13):1233–1241, 1978.
- [17] Facundo Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
- [18] Facundo Mémoli. The Gromov–Wasserstein distance: A brief overview. *Axioms*, 3(3):335–341, 2014.
- [19] Elzbieta Pekalska, Pavel Paclik, and Robert PW Duin. A generalized kernel approach to dissimilarity-based classification. *Journal of machine learning research*, 2(Dec):175–211, 2001.
- [20] John W Sammon. A nonlinear mapping for data structure analysis. *IEEE Transactions on computers*, 100(5):401–409, 1969.
- [21] Robin Sibson. Studies in the robustness of multidimensional scaling: Procrustes statistics. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 234–238, 1978.
- [22] Robin Sibson. Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 217–229, 1979.
- [23] Robin Sibson, Adrian Bowyer, and Clive Osmond. Studies in the robustness of multidimensional scaling: Euclidean models and simulation studies. *Journal of Statistical Computation and Simulation*, 13(3-4):273–296, 1981.
- [24] Michael Trosset. Computing distances between convex sets and subsets of the positive semidefinite matrices. Technical report, 1997.
- [25] Michael W Trosset. A new formulation of the nonmetric strain problem in multidimensional scaling. *Journal of Classification*, 15(1):15–35, 1998.
- [26] John Von Neumann and Isaac Jacob Schoenberg. Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*, 50(2):226–251, 1941.
- [27] Wikipedia. Functional principal component analysis. [https://en.wikipedia.org/wiki/Functional\\_principal\\_component\\_analysis](https://en.wikipedia.org/wiki/Functional_principal_component_analysis), 2018. [Online; accessed 28-June-2019].
- [28] WA Wilson. On certain types of continuous transformations of metric spaces. *American Journal of Mathematics*, 57(1):62–68, 1935.

E-mail address: `adams@math.colostate.edu`

E-mail address: `blumstei@math.colostate.edu`

E-mail address: `kassab@math.colostate.edu`