Multidimensional Scaling: Infinite Metric Measure Spaces

Lara Kassab - Master's Thesis Defense
February 25, 2019

Advisor: Dr. Henry Adams
Committee: Dr. Michael Kirby, Dr. Bailey Fosdick

Multidimensional Scaling

- Multidimensional scaling (MDS) is a set of statistical techniques concerned with the problem of constructing a configuration of n points in Euclidean space using information about the dissimilarities between the n objects.

Purpose of Multidimensional Scaling

- MDS mainly serves as a visualization technique for proximity data, the input of MDS, which is usually represented in the form of an $n \times n$ dissimilarity matrix.
- The choice of the embedding dimension m is arbitrary in principle, but low in practice $m = 1, 2,$ or 3.

Some Applications of MDS

- MDS was invented for the analysis of proximity data which arise in the following areas:
  - Social sciences, behavioral sciences, psychometrics
  - Archeology
  - Chemistry (molecular conformation)
  - Graph layout techniques
  - Classification problems
  - Dimension reduction
  - Machine learning (Isomap, kernel PCA $\cdots$)
- Similarities can represent for instance:
  - People's ratings of similarities between objects
  - The percent agreement between judges
  - The number of times a subjects fails to discriminate between stimuli etc.

Visualization of MDS

Consider the following dissimilarity matrix, $D_1 = \begin{pmatrix} 0 & 6 & 8 \\ 6 & 0 & 10 \\ 8 & 10 & 0 \end{pmatrix}$.

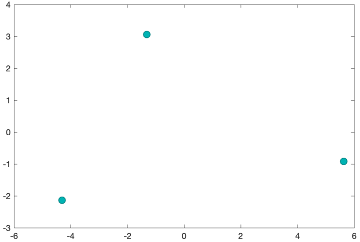


Figure: MDS embedding of D_1 into $\mathbb{R}^2$.

Configuration Points: $(-1.3163, 3.0624)$, $(-4.3046, -2.1404)$ and $(5.6209, -0.9220)$.

Visualization of MDS

Consider the following dissimilarity matrix,

$$D_2 = \begin{pmatrix} 0 & 1 & 1 & \sqrt{2} & 1 \\ 1 & 0 & \sqrt{2} & 1 & 1 \\ 1 & \sqrt{2} & 0 & 1 & 1 \\ \sqrt{2} & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{pmatrix}.$$

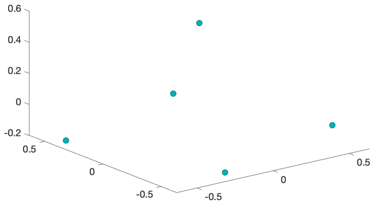


Figure: MDS embedding of D_2 into $\mathbb{R}^3$.

Visualization of MDS

Consider the following dissimilarity matrix, $D_3 = \begin{pmatrix} 0 & 2 & 2 & 1 \\ 2 & 0 & 2 & 1 \\ 2 & 2 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$.

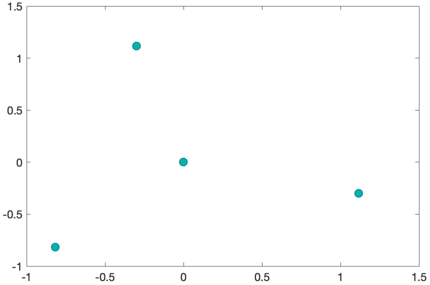


Figure: MDS embedding of D_3 into $\mathbb{R}^2$.

Types of Multidimensional Scaling

There are several types of MDS, and they differ mostly in the loss function they minimize. In general, there are two dichotomies:

- Kruskal-Shepard **distance scaling** versus classical Torgerson-Gower **inner-product scaling**.
- **Metric** scaling versus **nonmetric** scaling.

Common Loss Functions

A Stress Function:

$$\text{Stress}(f) = \sqrt{\frac{\sum_{i,j} \left(d_{ij} - \hat{d}_{ij}\right)^2}{scale}}.$$

A Strain Function:

$$\text{Strain}(f) = \sum_{i,j} (b_{ij} - \langle f(x_i), f(x_j)\rangle)^2.$$

Motivation Behind Our Work

- We address questions on convergence of MDS: 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?

MDS of evenly-spaced points on the circle equipped with the geodesic metric:

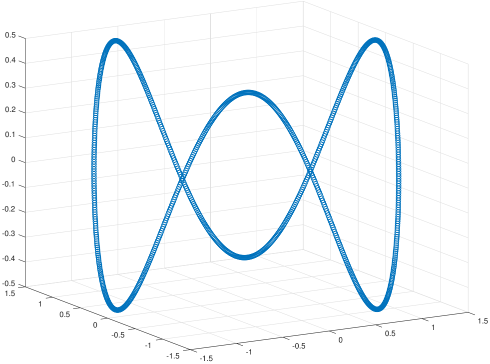


Figure: MDS embedding of S^1_{1000}.

MDS of evenly spaced points on a Circle

Proposition

The classical MDS embedding of S^1_n lies, up to a rigid motion of $\mathbb{R}^m$, on the curve $\gamma_m \colon S^1 \to \mathbb{R}^m$ defined by

$$\gamma_m(\theta) = (a_1(n)\cos(\theta), a_1(n)\sin(\theta), a_3(n)\cos(3\theta), a_3(n)\sin(3\theta), \ldots) \in \mathbb{R}^m,$$

where $\lim_{n\to\infty} a_j(n) = \frac{\sqrt{2}}{j}$ (with j odd).

The MDS embeddings of the geodesic circle are closely related to [6].

Motivation Behind Our Work

- Convergence is well-understood when each metric space has the same finite number of points, and also fairly well-understood when each metric space has a finite number of points tending to infinity.

- An important example is the behavior of MDS as one samples more and more points from a dataset.

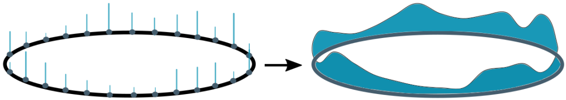


Figure: Convergence of arbitrary measures with finite support.

- We are also interested in convergence when the metric measure spaces in the sequence perhaps have an infinite number of points.

- In order to prove such results, we first need to define the MDS embedding of an infinite metric measure space X, and study its optimal properties and goodness of fit.

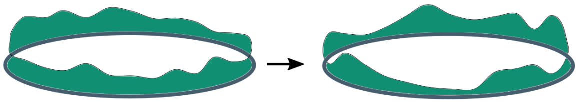


Figure: Convergence of arbitrary measures with infinite support.

Classical MDS: Algorithm

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

Let $\mathbf{D} = (d_{ij})$ be a $n \times n$ distance matrix.

1. Compute the matrix $\mathbf{A} = (a_{ij})$, where $a_{ij} = -\frac{1}{2} d_{ij}^2$.
2. Apply double centering to $\mathbf{A}$. Define $\mathbf{B} = \mathbf{HAH}$, where $\mathbf{H} = \mathbf{I} - n^{-1}\mathbf{1}\mathbf{1}^\top$.
3. Compute the eigendecomposition of $\mathbf{B} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^\top$.
4. Let $\mathbf{\Lambda}_m$ be the matrix of the largest m eigenvalues sorted in descending order, and let $\mathbf{\Gamma}_m$ be the matrix of the corresponding m eigenvectors. Then, the coordinate matrix of classical MDS is given by $\mathbf{X} = \mathbf{\Gamma}_m \mathbf{\Lambda}_m^{1/2}$.

Classical MDS: Theory

Theorem

[2, Theorem 14.2.1] Let $\mathbf{D}$ be a dissimilarity matrix. Then $\mathbf{D}$ is Euclidean if and only if $\mathbf{B}$ is a positive semi-definite matrix.

Theorem

[2, Theorem 14.4.1] Let $\mathbf{D}$ be a Euclidean distance matrix corresponding to a configuration $\mathbf{X}$ in $\mathbb{R}^m$, and fix k $(1 \leq k \leq m)$. Then amongst all projections $\mathbf{XL_1}$ of $\mathbf{X}$ onto k-dimensional subspaces of $\mathbb{R}^m$, the quantity $\sum_{r,s=1}^{n} (d_{rs}^2 - \hat{d}_{rs}^2)$ is minimized when $\mathbf{X}$ is projected onto its principal coordinates in k dimensions.

Classical MDS: Optimality Property

When $\mathbf{D}$ is not necessarily Euclidean, it is more convenient to work with the matrix $\mathbf{B} = \mathbf{HAH}$. 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:

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

Theorem

[2, Theorem 14.4.2] Let $\mathbf{D}$ be a dissimilarity matrix (not necessarily Euclidean). Then for fixed m, (1) is minimized over all configurations $\hat{\mathbf{X}}$ in m dimensions when $\hat{\mathbf{X}}$ is the classical solution to the MDS problem.

Metric Measure Space

Definition

A *metric measure space* (mm-space) is a triple (X, d_X, μ_X) where

- (X, d_X) is a compact metric space.
- μ_X is a Borel probability measure on X, i.e. $\mu_X(X) = 1$.

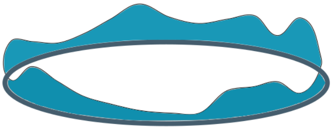


Figure: An illustration of a metric measure space.

Euclidean Metric Measure Spaces

Definition

A metric space (X, d_X) is said to be *Euclidean* if (X, d_X) can be isometrically embedded into $(\ell^2, \|\cdot\|_2)$. That is, (X, d_X) is Euclidean if there exists an isometric embedding $f\colon X \to \ell^2$, meaning $\forall x, s \in X$, we have that $d_X(x, s) = d_{\ell^2}(f(x), f(s))$.

Furthermore, we call a metric measure space (X, d_X, μ_X) *Euclidean* if its underlying metric space (X, d_X) is.

Indeed, $(\hat{X}, d_{\hat{X}})$ could be finite dimensional, i.e., $\hat{X} \subseteq \mathbb{R}^m$ and $d_{\hat{X}}$ is the Euclidean metric on $\mathbb{R}^m$.

Square-Integrable Functions

We denote by $L^2(X,\mu)$ the set of square integrable L^2-functions with respect to the measure μ. We note that $L^2(X,\mu)$ is furthermore a Hilbert space, after equipping it with the inner product given by

$$\langle f, g\rangle = \int_X fg \; d\mu.$$

Definition (Roughly Speaking)

A measurable function f on $X \times X$ is said to be *square-integrable* if

$$\int_X \int_X |f(x,s)|^2 \; \mu(dx)\mu(ds) < \infty.$$

We denote by $L^2_{\mu\otimes\mu}(X \times X)$ the set of square integrable functions with respect to the measure $\mu \otimes \mu$.

Kernels

In this context, a *real-valued L^2-kernel* $K\colon X \times X \to \mathbb{R}$ is a continuous measurable square-integrable function i.e. $K \in L^2_{\mu\otimes\mu}(X \times X)$.

Definition

A kernel K is *symmetric* (or *complex symmetric* or *Hermitian*) if

$$K(x,s) = \overline{K(s,x)} \quad \text{for all } x, s \in X,$$

where the overline denotes the complex conjuguate.

Most of the kernels that we define in our work are symmetric.

Symmetric Kernels

Definition

A symmetric function $K : X \times X \to \mathbb{R}$ is called a *positive semi-definite (p.s.d.) kernel* on X 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 $x_1, \ldots, x_m \in X$, and any $c_1, \ldots, c_m \in \mathbb{R}$.

Hilbert–Schmidt Integral Operator

Definition (Hilbert–Schmidt Integral Operator)

Let (X, Ω, μ) be a σ-finite measure space, and let $K \in L^2_{\mu\otimes\mu}(X \times X)$. Then the integral operator

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

defines a linear mapping acting from the space $L^2(X,\mu)$ into itself.

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

Spectral theorem on compact self-adjoint operators

Theorem (Spectral theorem on compact self-adjoint operators)

Let $\mathcal{H}$ be a not necessarily separable Hilbert space, and suppose $T \in \mathcal{B}(\mathcal{H})$ is 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

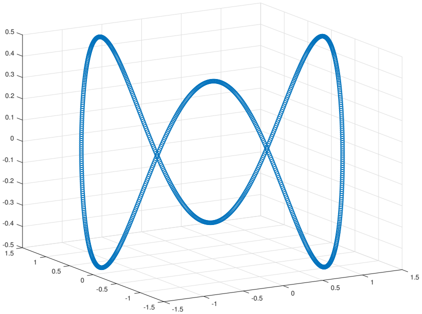
$$T(\cdot) = \sum_n \lambda_n \langle e_n, \cdot \rangle \; e_n.$$

An important consequence of the spectral theorem, is the **Generalized Mercer's theorem**.

MDS on Infinite metric measure spaces

Let (X, d, μ) be a bounded metric measure space, where d is a real-valued L^2-function on $X \times X$ with respect to the measure $\mu \otimes \mu$. We propose the following MDS method on infinite metric measure spaces:

1. From the metric d, construct the kernel $K_A \colon X \times X \to \mathbb{R}$ defined as $K_A(x, s) = -\frac{1}{2} d^2(x, s)$.



2. Obtain the kernel $K_B \colon X \times X \to \mathbb{R}$ defined as

$$K_B(x,s) = K_A(x,s) - \int_X K_A(w,s)\mu(\mathrm{d}w) - \int_X K_A(x,z)\mu(\mathrm{d}z) + \int_{X\times X} K_A(w,z)\mu(\mathrm{d}w \times \mathrm{d}z).$$

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

$$[T_{K_B}\phi](x) = \int_X K_B(x,s)\phi(s)\mu(\mathrm{d}s).$$

3. Let $\lambda_1 \geq \lambda_2 \geq \dots$ denote the eigenvalues of T_{K_B} with corresponding eigenfunctions $\phi_1, \phi_2, \dots$, where the $\phi_i \in L^2(X)$ are real-valued functions. Indeed, $\{\phi_i\}_{i \in \mathbb{N}}$ forms an orthonormal system of $L^2(X)$.

4. Define $K_{\hat{B}}(x,s) = \sum_{i=1}^{\infty} \hat{\lambda}_i \phi_i(x)\phi_i(s)$, where

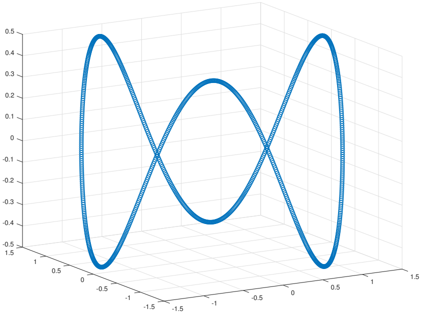
$$\hat{\lambda}_i = \begin{cases} \lambda_i & \text{if } \lambda_i \geq 0, \\ 0 & \text{if } \lambda_i < 0. \end{cases}$$

Define $T_{K_{\hat{B}}} : L^2(X) \to L^2(X)$ to be the Hilbert–Schmidt integral operator associated to the kernel $K_{\hat{B}}$. Note that the eigenfunctions ϕ_i for T_{K_B} (with eigenvalues λ_i) are also the eigenfunctions for $T_{K_{\hat{B}}}$ (with eigenvalues $\hat{\lambda}_i$).

5. Define the MDS embedding of X into ℓ^2 via the map $f\colon X \to \ell^2$ given by

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

for all $x \in X$.



MDS on Infinite metric measure spaces

Proposition

The MDS embedding map $f\colon X \to \ell^2$ defined by

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

is a continuous map.

Proposition

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

Optimality Property

Definition

Define the Strain function of f as follows

$$\begin{aligned}\mathrm{Strain}(f) &= \|T_{K_B} - T_{K_{\hat{B}}}\|_{HS}^2 = \mathrm{Tr}((T_{K_B} - T_{K_{\hat{B}}})^2) \\ &= \int\int \left(K_B(x,t) - K_{\hat{B}}(x,t)\right)^2 \mu(dt)\mu(dx).\end{aligned}$$

Theorem

Let (X, d, μ) be a bounded (and possibly non-Euclidean) metric measure space. Then $\mathrm{Strain}(f)$ is minimized over all maps $f\colon X \to \ell^2$ or $f\colon X \to \mathbb{R}^m$ when f is the MDS embedding.

Convergence of MDS for Arbitrary Measures:

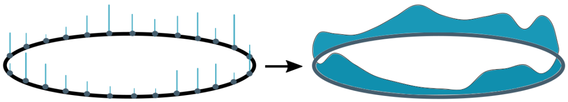


Figure: Convergence of arbitrary measures with finite support.

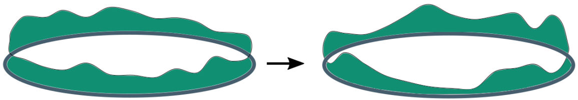


Figure: Convergence of arbitrary measures with infinite support.

Convergence of MDS with Respect to Gromov–Wasserstein Distance:

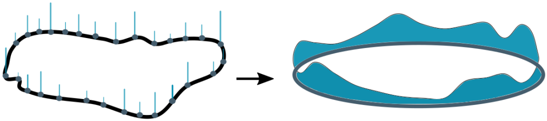


Figure: Convergence of mm-spaces equipped with measures of finite support.

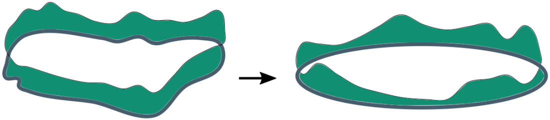


Figure: Convergence of mm-spaces equipped with measures of infinite support.

Robustness of MDS with Respect to Perturbations:

In a series of papers, Sibson and his collaborators consider the robustness of multidimensional scaling with respect to perturbations of the underlying distance or dissimilarity matrix.

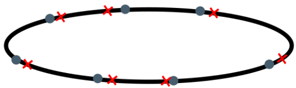


Figure: Perturbation of the given dissimilarities.

Sibson's perturbation analysis shows that if one is has a converging sequence of $n \times n$ dissimilarity matrices, then the corresponding MDS embeddings of n points into Euclidean space also converge.

Convergence of MDS by the Law of Large Numbers [1]:

Suppose we are given the data set $X_n = \{x_1, \ldots, x_n\}$ with $x_i \in \mathbb{R}^k$ sampled independent and identically distributed (i.i.d.) from an unknown probability measure μ on X.

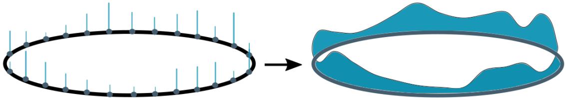


Figure: Convergence of arbitrary measures with finite support.

Data-Dependent Kernel:

$$K(x,y) = \frac{1}{2}\Big(-d(x,y)^2 + \int_X d(w,y)^2 \mu(\mathrm{d}w) + \int_X d(x,z)^2 \mu(\mathrm{d}z) - \int_{X\times X} d(w,z)^2 \mu(\mathrm{d}w \times \mathrm{d}z)\Big)$$

Associated Operator:

Define $T_K\colon L^2(X) \to L^2(X)$ as

$$[T_K f](x) = \int K(x,s) f(s) \mu(\mathrm{d}s).$$

Convergence of MDS

Theorem

[3, Theorem 3.1] The ordered spectrum of T_{K_n} converges to the ordered spectrum of T_K as $n \to \infty$ with respect to the ℓ^2-distance, namely

$$\ell^2(\lambda(T_{K_n}), \lambda(T_K)) \to 0 \quad \text{a.s.}$$

Convergence of MDS

Theorem

[1, Proposition 2] If K_n converges uniformly in its arguments and in probability, with the eigendecomposition of the Gram matrix converging, and if the eigenfunctions $\phi_{k,n}(x)$ of T_{K_n} associated with non-zero eigenvalues converge uniformly in probability, then their limit are the corresponding eigenfunctions of T_K.

Convergence of MDS

Definition (Total-variation convergence of measures)

Let $(X, \mathcal{F})$ be a measurable space. The total variation distance between two (positive) measures μ and ν is then given by

$$\|\mu - \nu\|_{\mathsf{TV}} = \sup_f \left\{ \int_X f \, d\mu - \int_X f \, d\nu \right\}.$$

Indeed, convergence of measures in total-variation implies convergence of integrals against bounded measurable functions, and the convergence is uniform over all functions bounded by any fixed constant.

Convergence of MDS for Finite Measures:

Proposition

Suppose $\mu_n = \frac{1}{n} \sum_{x \in X_n} \delta_x$ *converges to* μ *in total variation. If the eigenfunctions* $\phi_{k,n}$ *of* T_{K_n} *converge uniformly to* $\phi_{k,\infty}$ *as* $n \to \infty$*, then their limit are the corresponding eigenfunctions of* T_K*.*

Convergence of MDS for Arbitrary Measures:

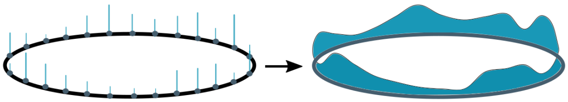


Figure: Convergence of arbitrary measures with finite support.

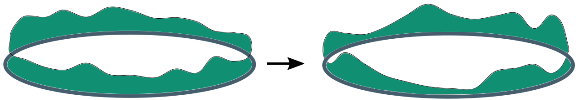


Figure: Convergence of arbitrary measures with infinite support.

Convergence of MDS

Proposition

Suppose μ_n converges to μ in total variation. If the eigenvalues $\lambda_{k,n}$ of T_{K_n} converge to λ_k, and if their corresponding eigenfunctions $\phi_{k,n}$ of T_{K_n} converge uniformly to $\phi_{k,\infty}$ as $n \to \infty$, then the $\phi_{k,\infty}$ are eigenfunctions of T_K with eigenvalue λ_k.

Convergence of MDS

Conjecture

Suppose we have the convergence of measures $\mu_n \to \mu$ in total variation. The ordered spectrum of T_{K_n} converges to the ordered spectrum of T_K as $n \to \infty$ with respect to the ℓ^2–distance,

$$\ell^2(\lambda(T_{K_n}), \lambda(T_K)) \to 0.$$

Convergence of MDS with Respect to Gromov–Wasserstein Distance:

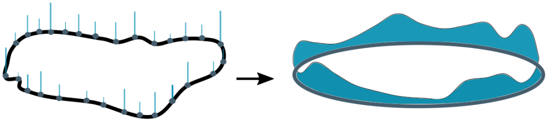


Figure: Convergence of mm-spaces equipped with measures of finite support.

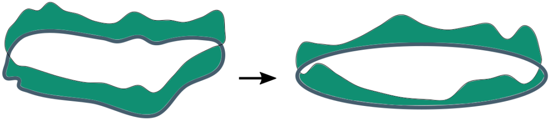


Figure: Convergence of mm-spaces equipped with measures of infinite support.

Convergence of MDS

Conjecture

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

References

[1] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-Franois Paiement, Pascal Vin-cent, and Marie Ouimet. Learning eigenfunctions links spectral embedding and kernel PCA. *Neural computation*, 16(10):21972219, 2004.

[2] JM Bibby, JT Kent, and KV Mardia. Multivariate analysis, 1979.

[3] Vladimir Koltchinskii, Evarist Giné, et al. Random matrix approximation of spectra of integral operators. *Bernoulli*, 6(1):113-167, 2000.

[4] Facundo Mémoli. Gromov-Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417-487, 2011.

[5] Robin Sibson. Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the Royal Statistical Society, Series B*, 217-229, 1979.

[6] J Von Neumann and IJ Schoenberg. Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*, 50(2):226-251, 1941.