INFINITE MULTIDIMENSIONAL SCALING FOR METRIC MEASURE SPACES

ALEXEY KROSHNIN, EUGENE STEPANOV, AND DARIO TREVISAN

Abstract. For a given metric measure space (X, d, μ) we consider finite samples of points, calculate the matrix of distances between them and then reconstruct the points in some finite-dimensional space using the multidimensional scaling (MDS) algorithm with this distance matrix as an input. We show that this procedure gives a natural limit as the number of points in the samples grows to infinity and the density of points approaches the measure μ. This limit can be viewed as "infinite MDS" embedding of the original space, now not anymore into a finite-dimensional space but rather into an infinite-dimensional Hilbert space. We further show that this embedding is stable with respect to the natural convergence of metric measure spaces. However, contrary to what is usually believed in applications, we show that in many cases it does not preserve distances, nor is even bi-Lipschitz, but may provide snowflake (Assouad-type) embeddings of the original space to a Hilbert space (this is, for instance, the case of a sphere and a flat torus equipped with their geodesic distances).

1. Introduction

Let (X, d) be a metric space and suppose that, for a set of points $\{x_1, \dots, x_n\} \subset X$ the distances $d_{ij} := \mathsf{d}(x_i, x_j)$ have been calculated. A common problem in distance geometry is to reconstruct the space X, knowing only the information on these distances. Of course, in general, unless X is finite itself one might only hope to do this in the limit as $n \to \infty$, i.e. with an infinitesimal error as the number of points becomes large. One of the classical algorithms aimed to solve this problem is *multidimensional scaling* (MDS), which is widely used in applications, in particular as a background tool in data science for dimension reduction (it is, for instance, a basic part of Isomap dimension reduction method [20, 21]), for visualization and clustering (see e.g. [10, 24, 11]).

MDS was originally coined for intrinsically Euclidean data, i.e. when $X \subset \mathbb{R}^d$ and the distance $\mathsf{d}(x, y) = |x - y|$ is Euclidean, where it is strongly linked to Principal Component Analysis (PCA), and a rigorous proof that in this case it in fact reconstructs the distances exactly is well known [24]. However, most interest for applications relies on its extension to arbitrary distances or even just for general dissimilarity matrices instead of distance matrices, in which case it can be assimilated to a kernel PCA method [8]. Quite recently, it has been shown in [1] that

Date: July 26, 2022.

The work of the first and second authors on the paper has been carried out within the framework of the HSE University Basic Research Program. The results of Section 5 have been obtained under support of the RSF grant #19-71-30020. The third author was partially supported by GNAMPA-INdAM 2020 project "Problemi di ottimizzazione con vincoli via trasporto ottimo e incertezza" and University of Pisa, Project PRA 2018-49.

exact reconstruction of general, not necessarily Euclidean, metric spaces, cannot be achieved by MDS: for $X = \mathbb{S}^1$ a unit radius circle endowed with its geodesic distance instead of the Euclidean one, and $\{x_i\}_{i=1}^n$ a uniformly spaced grid, MDS yields in the limit as $n \to \infty$ a closed curve, which is very far from being a circle. Infact, a simple computation shows that it is a fractal object, namely, a *snowflake* embedding [22] of a circle in an infinite-dimensional Hilbert space [17], i.e., it becomes an isometric embedding if S^1 is endowed with the geodesic distance raised to some power $\alpha \in (0,1)$ (precisely, for $\alpha = 1/2$). Although this may be unexpected in view of various commonly used applications of MDS, an explanation of this fact may be also traced back to the classical work [23], where all the invariant metrics on the circle that embed isometrically into a Hilbert space are classified, including of course the 1/2-snowflake re-obtained via MDS, which is actually credited to the earlier work [26] (see also the discussion in [14, section 7.3]).

1.1. **Our contribution.** In this paper we show that the above mentioned example is by no means exceptional, namely, in fact MDS for Riemannian manifolds equipped with intrinsic distances quite usually reconstructs snowflake embeddings of such manifolds in an infinite-dimensional Hilbert space. Moreover, we provide natural stability results for MDS maps with respect to their data in terms of natural distances between metric measure spaces. Our results can be summarized as follows:

i) We consider the general MDS on metric measure spaces proposed in [14, 1], which we often refer to as "infinite MDS", as opposed to the usual MDS on a finite subset, and show that, under mild assumptions satisfied on any Riemannian manifold with volume measure, the induced maps are "almost everywhere injective" in a natural sense (Theorem 4.2 and Corollary 4.6), with an explicit inequality.
ii) We prove general stability results for the MDS maps in terms of the Gromov-Kantorovich distance of order 4 (Theorem 5.4), which in particular shows convergence and identifies the limit of the MDS maps obtained by a sequence of families of points $\{x_i\}_{i=1}^n \subset X$, in terms of the limit of their empirical measures $\mu_n := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$, thus settling [1, conjecture 6.1].
iii) We prove that, for every $d \geq 1$, in the case of a sphere $\mathbb{S}^d$ (Proposition 6.1) and a flat torus $(\mathbb{S}^1)^d$ (Proposition 6.3 which follows from a more general result on MDS over product spaces, Proposition 6.2), endowed with their uniform measures, the MDS maps provide snowflake embeddings in infinite dimensional Hilbert spaces, extending the case of $d = 1$ and answering [1, question 7.1].

These results together show that MDS on metric measure spaces behaves in several aspects quite similarly to other embedding methods used in manifold learning, e.g. via eigenfunctions of a Laplace operator or via a heat kernel [3, 9], but with strikingly different features, since contrary to these embeddings, it is not even Lipschitz for quite simple manifolds such as spheres.

We also point out to the readers the very recent preprint [15] dedicated to the same problem, which appeared on the arXiv around the same time as ours and provides similar results. Although the convergence results formulated there are a bit weaker than ours, one can find there some further curious examples of MDS embeddings.

1.2. **Further problems.** We believe that our results contribute to the understanding of the structure of embeddings obtained via MDS on metric measure spaces. Still, several important questions that we describe below remain open.

(1) The most natural question is whether the infinite MDS map is in fact a homeomorphism onto its range, at least for a reasonably large class of metric measure spaces. This reduces to the question whether it is injective (here we are able to prove only a weaker result on "almost everywhere injectivity"); a partial answer is discussed in Remark 4.7, where it is linked to asymptotics of eigenvalues and eigenfunctions of the associated operator, which are however quite far from being easy to obtain even in very explicit examples.
(2) The second question concerns the quantitative description of the limit behavior of (finite) MDS maps in case of randomly sampled points, which is perhaps the most common assumption on the data in applications. Our results may be interpreted in this case as a sort of a qualitative law of large numbers, and more precise information on the order of convergence to infinite MDS maps is quite important to obtain.
(3) The third question is to understand the implications of our results and techniques to more sophisticated algorithms, such as Isomap, the stability of which has also been subject of recent research [7, 5], or to the recently proposed method [17] by the last two authors, together with N. Puchkin and V. Spokoiny, based on a semidefinite programming problem.

1.3. **Structure of the paper.** The paper is structured as follows. In Section 2 we recall some general notation and in Section 3 we focus on MDS, both on a finite sample and on a generic metric measure space, showing in particular some basic properties, including continuity of the embedding map. Section 4 is dedicated to the proof of its almost everywhere injectivity, Section 5 to that of stability with respect to Gromov-Kantovorich convergence. Section 6 deals with the examples of spheres and product spaces, in particular, flat tori.

1.4. **Acknowledgements.** The authors thank two anonymous referees for many constructive comments and in particular for suggesting the use of Vitali spaces and [13] to dispense from the doubling assumption in Remark 4.3.

2. Notation and preliminaries

The metric spaces (X, d) in the sequel will be always assumed separable and complete, and the measures μ will be finite and Borel. For a metric space (X, d) and a set $D \subset X$, we will denote by $\bar{D}$ its closure, by $\mathbf{1}_D$ its characteristic function. The notation $B_r(x) \subset X$ stands for the open ball of X centered at $x \in X$ with radius $r > 0$. For a number $x \in \mathbb{R}$ we denote by x^+ its positive part, i.e. $x^+ := x$ for $x > 0$ and $x^+ := 0$ otherwise.

The Euclidean space $\mathbb{R}^m$ is always assumed to be equipped with the Euclidean norm $|\cdot|$ and by $x \cdot y$ we mean the usual Euclidean scalar product of vectors x and y. If E is a Banach space, then we will denote by $\|\cdot\|_E$ its norm. We let ℓ^p stand for the usual Banach space of p-summable sequences, and $\mathbb{R}^\infty$ stand for the linear space of all real valued sequences (also denoted by $\mathbb{R}^\mathbb{N}$ in some literature), equipped

with its product topology which is metrizable, e.g. by the distance

$$d(x,y) \coloneqq \sum_{n=1}^{\infty} 2^{-n} \frac{|x_n - y_n|}{1 + |x_n - y_n|}$$

which makes it a Polish space. We will always assume $\mathbb{R}^m \subset \ell^p$ and $\mathbb{R}^m \subset \mathbb{R}^\infty$, the restrictions of ℓ^p and $\mathbb{R}^\infty$ to $\mathbb{R}^m$ being the restrictions to first m coordinates. If (X, d) is a metric space with some positive σ-finite measure μ, and E is a Banach space, then $L^p(X, \mu; E)$ stands for the space of Bochner integrable with power p functions $f\colon X \to E$. The notation $L^p(X, \mu; \mathbb{R}^\infty)$ stand for the space of strongly measurable functions $f\colon X \to \mathbb{R}^\infty$ such that $L^p(X, \mu; \mathbb{R}^m)$ for all $m \in \mathbb{N}$. In case $E = \mathbb{R}$ we will omit the reference to E and write just $L^p(X, \mu)$ for the space of integrable with power p real valued functions (or just measurable essentially bounded in case $p = \infty$). By $\operatorname{supp} \mu$ we denote the support of the measure μ.

We write $(\cdot, \cdot)_H$ for the scalar product in a Hilbert space H. For a linear operator T between Hilbert spaces we denote by $\|T\|_{HS}$ its Hilbert-Schmidt norm, and by $\|T\|_1$ its trace class norm. If $\|T\|_1 < \infty$, then T is called trace-class (or nuclear) operator.

3. Multidimensional scaling

3.1. **MDS of a finite sample.** Given a metric space (X, d) and a finite subset $\{x_1, \dots, x_n\} \subset X$, we introduce the matrices

$$\bar{K}_n \coloneqq \left(-\frac{1}{2n} \mathsf{d}^2(x_i, x_j) \right)_{i,j=1,\dots,n}, \tag{3.1}$$

$$\bar{P}_n \coloneqq \mathrm{Id}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^{\mathrm{T}}, \tag{3.2}$$

$$\bar{T}_n \coloneqq \bar{P}_n \bar{K}_n \bar{P}_n, \tag{3.3}$$

where Id_n stands for the identity $n \times n$ matrix, $\mathbf{1}_n$ stands for the column vector $\mathbf{1}_n \coloneqq (1, \dots, 1)^{\mathrm{T}}$, T stands for the matrix transpose. Then the classical MDS consists in minimizing the strain function defined by

$$\mathrm{strain}_n(y_1, \dots, y_n) \coloneqq \sum_{i,j=1}^{n} \left((\bar{T}_n)_{i,j} - y_i \cdot y_j \right)^2 \tag{3.4}$$

among all $\{y_1, \dots, y_n\} \subset \mathcal{H}$, where $\mathcal{H}$ is some chosen finite-dimensional Euclidean space [1, 14]. Note that one can easily rewrite

$$\mathrm{strain}_n(y_1, \dots, y_n) = \left\| \bar{T}_n - G_y \right\|_{HS}^2,$$

where the positive symmetric $n \times n$ matrix G_y is the Gram matrix of $y_1, \dots, y_n$. Then minimization with respect to $y_1, \dots, y_n$ is equivalent to minimization over positive-semidefinite matrices G_y. In particular, one possible MDS maps from the finite set $\{x_1, \dots, x_n\}$ to $\mathbb{R}^n$ has the form

$$\bar{M}_n(x_i) = \left(\sqrt{\lambda_k^+(\bar{T}_n)} [v_k^+(\bar{T}_n)]_i \right)_{k=1,\dots,n}, \tag{3.5}$$

where $\lambda_k^+(\bar{T}_n)$ are positive eigenvalues of $\bar{T}_n$ (counting multiplicity) and $v_k^+(\bar{T}_n) \in \mathbb{R}^n$ are corresponding orthonormal eigenvectors.

3.2. **MDS of a metric measure space.** Let μ be a Borel probability measure on (X, d) with finite 4th moment, i.e.

$$\int_X \mathsf{d}^4(x_0, y)\, d\mu(y) < \infty \tag{3.6}$$

for some $x_0 \in X$. Define now

$$\mathsf{k}(x, y) := -\frac{1}{2}\mathsf{d}^2(x, y),$$
$$(Kf)(x) := \int_X \mathsf{k}(x, y) f(y)\, d\mu(y).$$

Clearly, under condition (3.6) one has that $\mathsf{k} \in L^2(X \times X, \mu \otimes \mu)$ and hence the operator $K\colon L^2(X, \mu) \to L^2(X, \mu)$ is a linear self-adjoint Hilbert–Schmidt (and hence compact) operator with

$$\|K\|_{HS}^2 = \frac{1}{4}\int_{X\times X} \mathsf{d}^4(x, y)\, d\mu \otimes \mu(x, y) < \infty$$

because

$$\mathsf{d}^2(x, y) \le 8\left(\mathsf{d}^2(x, x_0) + \mathsf{d}^2(x_0, y)\right).$$

Consider the projector operator $P := \mathrm{Id} - \mathbf{1} \otimes \mathbf{1}$ to the orthogonal complement of constant functions in $L^2(X, \mu)$, where Id stands for the identity operator in $L^2(X, \mu)$, and denote

$$T := PKP.$$

Clearly T is also an integral operator

$$\begin{aligned}(Tf)(x) &= \int_X \mathsf{k}_T(x, y) f(y)\, d\mu(y), \quad \text{where}\\ \mathsf{k}_T(x, y) &:= \mathsf{k}(x, y) - \int_X \mathsf{k}(x, y')\, d\mu(y') - \int_X \mathsf{k}(x', y)\, d\mu(x')\\ &\quad + \int_X\int_X \mathsf{k}(x', y')\, d\mu(x')\, d\mu(y').\end{aligned} \tag{3.7}$$

Moreover, T is also a self-adjoint Hilbert–Schmidt operator. It is immediate to see that zero is an eigenvalue of T whose eigenspace contains constant functions.

We will call in analogy with (3.5) the *infinite MDS embedding* map from X to $\mathbb{R}^\infty$ the map defined by

$$M(x) := \left(\sqrt{\lambda_j^+(T)}\, u_j^+(T)(x)\right)_{j\ge 1}, \tag{3.8}$$

where $\lambda_1^+(T) \ge \lambda_2^+(T) \ge \cdots > 0$ are positive eigenvalues of T (counting multiplicity), and $u_j^+(T) \in L^2(X, \mu)$ are corresponding orthonormal eigenfunctions. If the set of positive eigenvalues of T contains $J < \infty$ elements, we set $(M(x))_j := 0$ for $j > J$. In Proposition 3.2 we prove that at least one eigenvalue of T is positive, and thus infinite MDS embedding does not reduce to a singleton. Note that M is only defined once all the eigenfuctions $u_j^+(T)$ are chosen; a different choice of eigenfunctions gives a different map M.

It is conjectured in [14] that this map is a limit of MDS embeddings of finite samples. In what follows we rigorously formulate and prove this conjecture.

3.3. **Basic properties.** We start with the following easy calculation showing that eigenfunctions of T are locally Lipschitz continuous.

Lemma 3.1. *Let $u \in L^2(X,\mu)$, $\|u\|_{L^2(X,\mu)} = 1$ be an eigenfunction of T with eigenvalue $\lambda \neq 0$, i.e., $Tu = \lambda u$. Then for any $x_0, x, x' \in X$, one has*

$$|u(x) - u(x')| \leq \frac{\mathsf{d}(x,x')}{|\lambda|}\left[\mathsf{d}(x_0,x) + \mathsf{d}(x_0,x') + 2\left(\int_X \mathsf{d}^2(x_0,y)\,d\mu(y)\right)^{1/2}\right].$$

In particular, if $\operatorname{diam} X < \infty$, then

$$|u(x) - u(x')| \leq \frac{4\operatorname{diam} X}{|\lambda|}\mathsf{d}(x,x').$$

Proof. One has

$$\begin{aligned}
|\lambda|\,|u(x) - u(x')| &\leq \int_X |(\mathsf{k}_T(x,y) - \mathsf{k}_T(x',y))u(y)|\,d\mu(y)\\
&\leq \int_X |\mathsf{k}(x,y) - \mathsf{k}(x',y)| \cdot |u(y)|\,d\mu(y)\\
&\qquad + \left|\int_X \mathsf{k}(x,y)\,d\mu(y) - \int_X \mathsf{k}(x',y)\,d\mu(y)\right| \int_X |u(y)|\,d\mu(y)\\
&\leq \int_X |\mathsf{k}(x,y) - \mathsf{k}(x',y)| \cdot |u(y)|\,d\mu(y) + \int_X |\mathsf{k}(x,y) - \mathsf{k}(x',y)|\,d\mu(y)\\
&\leq 2\left(\int_X (\mathsf{k}(x,y) - \mathsf{k}(x',y))^2\,d\mu(y)\right)^{1/2}\\
&= \left(\int_X \left(\mathsf{d}^2(x,y) - \mathsf{d}^2(x',y)\right)^2 d\mu(y)\right)^{1/2}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\int_X \left(\mathsf{d}^2(x,y) - \mathsf{d}^2(x',y)\right)^2 d\mu(y) &\leq \int_X \mathsf{d}^2(x,x')\left(\mathsf{d}(x,y) + \mathsf{d}(x',y)\right)^2 d\mu(y)\\
&\leq \mathsf{d}^2(x,x')\int_X \left(2\mathsf{d}(x_0,y) + \mathsf{d}(x_0,x) + \mathsf{d}(x_0,x')\right)^2 d\mu(y)\\
&= \mathsf{d}^2(x,x')\,\|2\mathsf{d}(x_0,\cdot) + \mathsf{d}(x_0,x) + \mathsf{d}(x_0,x')\|^2_{L^2(X,\mu)},
\end{aligned}$$

thus

$$\begin{aligned}
|\lambda|\,|u(x) - u(x')| &\leq \mathsf{d}(x,x')\,\|2\mathsf{d}(x_0,\cdot) + \mathsf{d}(x_0,x) + \mathsf{d}(x_0,x')\|_{L^2(X,\mu)}\\
&\leq \mathsf{d}(x,x')\left(\mathsf{d}(x_0,x) + \mathsf{d}(x_0,x') + 2\,\|\mathsf{d}(x_0,\cdot)\|_{L^2(X,\mu)}\right),
\end{aligned}$$

since $\mu(X) = 1$ is a probability. The thesis follows. □

We now show that at least one positive eigenvalue always exists, thus the infinite MDS map M is never trivial, and, moreover, M is continuous when viewed as a map between X and $\mathbb{R}^\infty$.

Proposition 3.2. *Under condition (3.6) the operator T has at least one positive eigenvalue. One has that $M\colon X \to \mathbb{R}^\infty$ is continuous and, moreover, $M \in L^2(X,\mu;\ell^2)$ when $\sum_j \lambda_j^+(T) < +\infty$, which happens, for instance, when T is a trace-class (i.e. nuclear) operator.*

Remark 3.3. In particular examples one might know the asymptotics of the eigenvalues of T. For instance, if $X = \mathbb{S}^d$ is the d-dimensional sphere equipped with its intrinsic distance d and the surface measure μ, then (A.5) gives the power asymptotics for positive eigenvalues of T, which implies according to the above Proposition 3.2 that $M \in L^2(X, \mu; \ell^2)$. The same holds if instead X is just a subset of a sphere not containing any couple of antipodal points (d and μ being the same), because in this case the kernel of T is smooth and hence the eigenvalues vanish quicker than any power (and this is of course not specific to a sphere, but rather holds for X subset of a smooth Riemannian manifold with diameter $\operatorname{diam} X$ smaller than the injectivity radius of the latter, d the Riemannian distance and μ a volume measure).

Proof. If T has only negative eigenvalues, i.e. $-T$ is positive definite, then by [25, Satz 1] one has

$$\begin{aligned} 0 < \sum_j (-\lambda_j) = \operatorname{trace}(-T) &:= -\int_X \mathsf{k}_T(x,x)\, d\mu(x) \\ &= -\int_X \mathsf{k}(x,x)\, d\mu(x) + \int_X \int_X \mathsf{k}(x,y)\, d\mu(x)\, d\mu(y) \\ &= -\frac{1}{2} \int_X \int_X \mathsf{d}^2(x,y)\, d\mu(x)\, d\mu(y) < 0, \end{aligned}$$

a contradiction proving the existence of at least one positive eigenvalue of T.

Continuity of $M\colon X \to \mathbb{R}^\infty$ follows from Lemma 3.1. Moreover, under additional assumption that

$$\sum_{j=1}^{\infty} \lambda_j^+(T) < \infty,$$

for every $s \in \ell^2$ one has the relationship

$$(s, M(x))_{\ell^2} = \sum_j s_j \sqrt{\lambda_j^+(T)} u_j^+(T)(x),$$

and the series

$$\sum_j \lambda_j^+(T)(u_j^+(T))^2(x)$$

is convergent for μ-a.e. $x \in X$, because

$$\int_X \sum_{j=1}^{m} \lambda_j^+(T)(u_j^+(T))^2(x)\, d\mu(x) = \sum_{j=1}^{m} \lambda_j^+(T) < +\infty.$$

Thus one has

$$\begin{aligned} \lim_{m\to\infty} \left| \sum_{j>m} s_j \sqrt{\lambda_j^+(T)} u_j^+(T)(x) \right| &\le \lim_{m\to\infty} \left(\sum_{j>m} s_j^2 \right)^{1/2} \left(\sum_{j>m} \lambda_j^+(T)(u_j^+(T))^2(x) \right)^{1/2} \\ &= 0, \end{aligned}$$

and therefore,

$$(s, M(x))_{\ell^2} = \lim_{m\to\infty} \sum_{j=1}^{m} s_j \sqrt{\lambda_j^+(T)} u_j^+(T)(x)$$

is well-defined, for μ-a.e. $x \in X$, hence the function $s \in \ell^2 \mapsto (s, M(x))_{\ell^2} \in \mathbb{R}$ is measurable as a μ-a.e. limit of a sequence of continuous functions. This means that $M\colon X \to \ell^2$ is weakly (hence also strongly) measurable, and finally the relationship

$$\int_X |M(x)|^2 \, d\mu(x) = \sum_j \lambda_j^+(T) < \infty$$

shows that in this case $M \in L^2(X, \mu; \ell^2)$. □

Calculation of the infinite MDS embedding map for a space homogeneous under the action of some group by measure preserving isometries may be simplified with the help of the following statement.

Proposition 3.4. *Let (X, d, μ) be homogeneous, i.e. there is a group G acting transitively on X by isometries preserving μ. Then $\mathbf{1}$ is an eigenfunction of K, hence T and K share all eigenfunctions and eigenvalues except for the eigenvalue corresponding to $\mathbf{1}$.*

Proof. Fix $x_0 \in X$ and take an arbitrary $x \in X$. There exists $g \in G$ such that $x_0 = gx$. Then

$$\begin{aligned}(K\mathbf{1})(x) &= \int_X \mathsf{k}(x, y) \, d\mu(y) = \int_X \mathsf{k}(gx, gy) \, d\mu(y) \\ &= \int_X \mathsf{k}(x_0, y) \, d(g_\# \mu)(y) = \int_X \mathsf{k}(x_0, y) \, d\mu(y) = (K\mathbf{1})(x_0),\end{aligned}$$

that is, $\mathbf{1}$ is an eigenfunction of K corresponding to the eigenvalue $\lambda \coloneqq (K\mathbf{1})(x_0)$. □

4. Injectivity almost everywhere

We let $\{\lambda_i\}_{i=1}^\infty$ denote all the eigenvalues of the operator T defined over $L^2(X, \mu)$ by (3.7) and by $u_i \in L^2(x, \mu)$ the eigenfunctions corresponding to λ_i, chosen so as to form an orthonormal basis $\{u_i\}_{i=1}^\infty$ in $L^2(X, \mu)$. One has

$$\sum_{i=1}^\infty \lambda_i u_i \otimes u_i = \mathsf{k}_T$$

the convergence being understood in the sense of $L^2(X \times X, \mu \otimes \mu)$, and even more, since both k_T and all u_i are continuous if $\lambda_i \neq 0$, then in fact

$$\sum_{i=1}^\infty \lambda_i u_i(x) u_i(\cdot) = k_T(x, \cdot)$$

in the sense of $L^2(X, \mu)$ for every $x \in X$.

Lemma 4.1. *Let T be defined over $L^2(X, \mu)$ by (3.7), and assume that T is a trace-class (or nuclear) operator, i.e.*

$$\|T\|_1 \coloneqq \sum_i |\lambda_i| < +\infty. \tag{4.1}$$

Define the function

$$f \coloneqq \sum_{i=1}^\infty |\lambda_i| \, u_i^2 \in L^1(X, \mu). \tag{4.2}$$

If $x \in X$ is a Lebesgue point of f, then

$$\sum_{i=1}^{\infty} \lambda_i u_i^2(x) = k_T(x,x).$$

Proof. Fix $\varepsilon > 0$. Take $n \in \mathbb{N}$ such that

$$\sum_{i=n}^{\infty} |\lambda_i| \, u_i^2(x) < \varepsilon.$$

Now notice that

$$\begin{aligned}
\sum_{i=1}^{\infty} \lambda_i \left(\fint_{B_r(x)} u_i \, d\mu \right)^2 &= \sum_{i=1}^{\infty} \lambda_i \fint_{B_r(x)\times B_r(x)} u_i \otimes u_i \, d\mu \otimes \mu \\
&= \fint_{B_r(x)\times B_r(x)} \left(\sum_{i=1}^{\infty} \lambda_i u_i \otimes u_i \right) d\mu \otimes \mu \\
&= \fint_{B_r(x)\times B_r(x)} \mathsf{k}_T \, d\mu \otimes \mu \to \mathsf{k}_T(x,x) \quad \text{as} \quad r \to 0.
\end{aligned}$$

Furthermore, since all functions u_i are continuous,

$$\sum_{i=1}^{n-1} \lambda_i \left(\fint_{B_r(x)} u_i \, d\mu \right)^2 \to \sum_{i=1}^{n-1} \lambda_i u_i^2(x) \quad \text{as } r \to 0,$$

and since x is a Lebesgue point of f,

$$\begin{aligned}
\left| \sum_{i=n}^{\infty} \lambda_i \left(\fint_{B_r(x)} u_i \, d\mu \right)^2 \right| &\le \sum_{i=n}^{\infty} |\lambda_i| \fint_{B_r(x)} u_i^2 \, d\mu \\
&= \fint_{B_r(x)} f \, d\mu - \sum_{i=1}^{n-1} |\lambda_i| \fint_{B_r(x)} u_i^2 \, d\mu \\
&\to f(x) - \sum_{i=1}^{n-1} |\lambda_i| \, u_i^2(x) = \sum_{i=n}^{\infty} |\lambda_i| \, u_i^2(x) < \varepsilon \quad \text{as} \quad r \to 0.
\end{aligned}$$

Combining the above results, we conclude. □

Theorem 4.2. *Assume that (4.1) holds and let μ-a.e. point $x \in X$ be a Lebesgue point of f. Then*

$$\sum_{i=1}^{\infty} \lambda_i \left(u_i(x) - u_i(y) \right)^2 = \mathsf{d}^2(x,y) \quad \textit{for } \mu \otimes \mu\textit{-a.e. } (x,y). \tag{4.3}$$

Remark 4.3. The assumption on f from the above theorem is fulfilled, e.g., if μ is doubling, i.e. there is a constant $C > 0$ such that for any $x \in X$ and $r > 0$ one has

$$\mu\left(B_{2r}(x)\right) \le C\mu\left(B_r(x)\right)$$

(see [4, theorem 5.2.3] or alternatively [12, corollary 2.9.9, theorem 2.8.17])). In fact, it holds also as well on infinitesimally doubling spaces, also called Vitali spaces, see [13, theorem 3.4.3], a class that includes any smooth Riemannian manifold.

Remark 4.4. It is also worth remarking that under condition (4.1) one has

$$\sum_{i=1}^{\infty} \lambda_i = \operatorname{trace}(T) := \int_X \mathsf{k}_T(x,x)\, d\mu(x) = \frac{1}{2}\int_X \int_X \mathsf{d}^2(x,y)\, d\mu(x) d\mu(y).$$

Proof. From Lemma 4.1 we get

$$\begin{aligned}\sum_{i=1}^{\infty} \lambda_i \left(u_i(x) - u_i(y)\right)^2 &= \sum_{i=1}^{\infty} \lambda_i u_i^2(x) + \sum_{i=1}^{\infty} \lambda_i u_i^2(y) - 2\sum_{i=1}^{\infty} \lambda_i u_i(x) u_i(y) \\ &= \mathsf{k}_T(x,x) + \mathsf{k}_T(y,y) - 2\mathsf{k}_T(x,y)\end{aligned}$$

for $\mu \otimes \mu$-a.e. (x,y). The identity

$$\mathsf{k}_T(x,x) + \mathsf{k}_T(y,y) - 2\mathsf{k}_T(x,y) = \mathsf{k}(x,x) + \mathsf{k}(y,y) - 2\mathsf{k}(x,y) = \mathsf{d}^2(x,y)$$

then concludes the proof. □

Remark 4.5. If under conditions of Theorem 4.2 one has additionally that T is positive semidefinite (i.e. $\lambda_i \geq 0$ for all $i \in \mathbb{N}$), then in view of Mercer's theorem one has that (4.3) holds for all $(x,y) \in \operatorname{supp}\mu \times \operatorname{supp}\mu$ rather than just for $\mu \otimes \mu$-a.e. (x,y). In other words, M is an isometric embedding of $\operatorname{supp}\mu \subset X$ into ℓ^2.

Now we define a new map

$$M^-(x) := \left(\sqrt{\lambda_j^-(T)}\, u_j^-(T)(x)\right)_{j\geq 1},$$

where $\lambda_j^-(T)$ and $u_j^-(T)$ are the absolute values of negative eigenvalues of T and the respective orthonormal eigenfunctions (with the same convention as with M if these are a finite number). In the same way as M, one has $M^- \in L^2(X,\mu;\ell^2)$. Theorem 4.2 ensures that for $\mu \otimes \mu$-a.e. (x,y)

$$\|M(x) - M(y)\|_2^2 - \left\|M^-(x) - M^-(y)\right\|_2^2 = \mathsf{d}^2(x,y). \tag{4.4}$$

As an important corollary we have the following statement.

Corollary 4.6. *Under conditions of Theorem 4.2 one has*

$$\|M(x) - M(y)\|_2^2 \geq \mathsf{d}^2(x,y) \tag{4.5}$$

for $\mu \otimes \mu$-a.e. (x,y).

Moreover, the map $x \mapsto N(x) := (M(x), M^-(x)) \in \mathcal{K}$ considered as a map between X and the Krein space $\mathcal{K} := \ell^2 \times \ell^2$ with the indefinite inner product

$$((f_1,g_1),(f_2,g_2))_{\mathcal{K}} := (f_1,f_2) - (g_1,g_2)$$

and the corresponding pseudo-norm

$$\|(f,g)\|_{\mathcal{K}}^2 := \|f\|_2^2 - \|g\|_2^2,$$

is an a.e. isometry in the sense that

$$\|N(x) - N(y)\|_{\mathcal{K}}^2 = \mathsf{d}^2(x,y) \tag{4.6}$$

for $\mu \otimes \mu$-a.e. $(x,y) \in X \times X$.

Remark 4.7. Under conditions of Theorem 4.2 and hence also of Corollary 4.6 one has $M \in L^2(X, \mu; \ell^2)$. If one knows additionally that

$$\lim_m \sup_{x \in X} \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))^2(x) = 0, \tag{4.7}$$

then more can be said, namely, (4.5) holds for all $(x, y) \in X \times X$ rather than just for $\mu \otimes \mu$-a.e., i.e. in particular M is injective. When X is compact, this also implies that $M \colon X \to \mathbb{R}^\infty$ is a homeomorphism onto its image since it is a continuous map. In fact, by (4.7) for every $\varepsilon > 0$ there is an $m \in \mathbb{N}$ such that

$$\|M(x) - M(y)\|_2^2 \leq |M^m(x) - M^m(y)|^2 + \varepsilon$$

for all $(x, y) \in X \times X$, where $M^m \colon X \to \mathbb{R}^m$ stands for the restriction of M to its first m components, because

$$\begin{aligned} 0 \leq \|M(x) - M(y)\|_2^2 &- |M^m(x) - M^m(y)|^2 \\ &= \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))^2(x) + \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))^2(y) \\ &\qquad - 2 \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))(x)(u_j^+(T))(y) \\ &\leq 2 \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))^2(x) + 2 \sum_{j \geq m} \lambda_j^+(T)(u_j^+(T))^2(y). \end{aligned}$$

Thus by (4.5) one would have

$$\mathsf{d}^2(x, y) \leq |M^m(x) - M^m(y)|^2 + \varepsilon$$

for $\mu \otimes \mu$-a.e. $(x, y) \in X \times X$, and minding that M^m is continuous by Lemma 3.1, in fact, also for all $(x, y) \in X \times X$. Letting then $\varepsilon \to 0$ we get the desired claim.

The condition (4.7) is satisfied, for instance, when one has

$$\left\|u_j^+(T)\right\|_{L^\infty(X,\mu)} \leq C\lambda_j^{-\alpha} \quad \text{and} \quad \sum_{j=1}^{\infty} (\lambda_j^+(T))^{1-2\alpha} < +\infty$$

for some $C > 0$ and $\alpha \in [0, 1/2)$. Another interesting particular case is when the range of M is some $\mathbb{R}^m$, i.e. $M = (M^m, 0, \dots)$, or the operator T has only $m \in \mathbb{N}$ positive eigenvalues. Combining (4.5) with Lemma 3.1 one has that in the latter case

$$\mathsf{d}(x, y) \leq |M(x) - M(y)| \leq C\mathsf{d}(x, y)$$

for all $(x, y) \in X \times X$, i.e. in other words M provides a bi-Lipschitz embedding of X into $\mathbb{R}^m$.

Let us finally mention that Corollary 4.6 actually holds for any continuous symmetric kernel k, not necessarily the negative squared distance, with $\mathsf{d}^2(\cdot, \cdot)$ in all assertions replaced with $-2\mathsf{k}(\cdot, \cdot)$.

5. Stability of MDS

5.1. **Stability under Gromov–Kantorovich convergence.** Let (X_n, d_n, μ_n) be a sequence of separable complete metric measure spaces. Recall that the Gromov–Kantorovich distance[1] of order p is defined as [16, definition 5.7]

$$GW_p^p(X, X_n) \coloneqq \inf_{\gamma \in \Pi(\mu,\mu_n)} \int_{(X\times X_n)^2} |\mathsf{d}(x,y) - \mathsf{d}_n(x',y')|^p \, d\gamma(x,x')\, d\gamma(y,y'), \tag{5.1}$$

where $\Pi(\mu, \mu_n)$ denote the set of transport plans between μ and μ_n, i.e., joint probability densities on $X \times X_n$ with marginals respectively given by μ, μ_n. Assume $\lim_n GW_4(X, X_n) = 0$. We are going to show that in this case

$$\lim_n GW_2\big(M^m(X), M_n^m(X_n)\big) = 0$$

for every $m \in \mathbb{N}$, where M^m (resp. M_n^m) is any MDS map of (X, d, μ) (resp. $(X_n, \mathsf{d}_n, \mu_n)$) into $\mathbb{R}^m$.

Denote by $\gamma_n \in \Pi(\mu, \mu_n)$ any optimal transport plan in (5.1). According to [2, lemma 5.3.4] there exists a probability measure $\boldsymbol{\gamma}$ on the product space

$$X^\infty \coloneqq X \times \prod_n X_n,$$

such that its projection on any pair $X \times X_n$ is equal to γ_n. Note that X^∞ is still Polish as a countable product of Polish spaces, e.g. with the metric

$$\mathsf{d}_\infty(\boldsymbol{x}, \boldsymbol{y}) \coloneqq \mathsf{d}(x,y) + \sum_{n=1}^\infty 2^{-n} \frac{\mathsf{d}_n(x_n, y_n)}{1 + \mathsf{d}_n(x_n, y_n)}.$$

Embeddings of $L^2(X, \mu)$ and $L^2(X_n, \mu_n)$ to $L^2(X^\infty, \boldsymbol{\gamma})$ also allow us to naturally extend the operators K, K_n, T, T_n on this space, with K, T vanishing on $\big(L^2(X,\mu)\big)^\perp$, and K_n, T_n vanishing on $\big(L^2(X_n,\mu_n)\big)^\perp$, respectively. Note that K is induced by the kernel $\mathsf{k}(\boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{2}\mathsf{d}^2(x,y)$ and K_n is induced by the kernel $\mathsf{k}_n(\boldsymbol{x}, \boldsymbol{y}) = -\frac{1}{2}\mathsf{d}_n^2(x_n, y_n)$. Moreover, it is straightforward to check that

$$T = P_{\gamma_n} K P_{\gamma_n}, \quad T_n = P_{\gamma_n} K_n P_{\gamma_n}$$

where $P_{\gamma_n} \coloneqq \mathrm{Id} - \mathbf{1} \otimes \mathbf{1}$ is an orthogonal projector in $L^2(X^\infty, \boldsymbol{\gamma})$. Clearly, this embedding does not change any eigenfunctions of T and T_n corresponding to non-zero eigenvalues, thus we can equivalently define the MDS maps M^m and M_n^m using these embedded operators. The next lemma ensures that $\|T_n - T\|_{HS} \to 0$ as $GW_4(X, X_n) \to 0$.

Lemma 5.1. *Setting $C_X = \|\mathsf{d}(\cdot,\cdot)\|_{L^4(X\times X, \mu\otimes\mu)}$, one has*

$$\|T_n - T\|_{HS} \le \|K_n - K\|_{HS} \le C_X GW_4(X, X_n) + \frac{1}{2} GW_4^2(X, X_n).$$

[1] It is nowadays customary to call these distances and the respective convergence Gromov–Wasserstein, but we would like to avoid this historically incorrect naming.

Proof. Clearly, it is enough to consider the restrictions of the operators to $L^2(X \times X_n, \gamma_n)$, with $\gamma \in \Pi(\mu, \mu_n)$:

$$\begin{aligned}
\|T_n - T\|_{HS} &\leq \|K_n - K\|_{HS} = \|\mathsf{k}_n - \mathsf{k}\|_{L^2((X\times X_n)^2, \gamma_n \otimes \gamma_n)} \\
&= \frac{1}{2}\left(\int_{(X\times X_n)^2} (\mathsf{d}_n^2(x_n, y_n) - \mathsf{d}^2(x,y))^2 d\gamma_n(x, x_n) d\gamma_n(y, y_n)\right)^{1/2} \\
&\leq \left(\int_{(X\times X_n)^2} \mathsf{d}(x,y)\left(\mathsf{d}_n(x_n, y_n) - \mathsf{d}(x,y)\right) d\gamma_n(x, x_n) d\gamma_n(y, y_n)\right)^{1/2} \\
&\quad + \frac{1}{2}\left(\int_{(X\times X_n)^2} \left(\mathsf{d}_n(x_n, y_n) - \mathsf{d}(x,y)\right)^2 d\gamma_n(x, x_n) d\gamma_n(y, y_n)\right)^{1/2} \\
&\leq \left(\int_{X\times X} \mathsf{d}(x,y)^4 d\mu \otimes \mu(x,y)\right)^{1/4} GW_4(X, X_n) + \frac{1}{2} GW_4^2(X, X_n). \qquad \square
\end{aligned}$$

Now let us recall some stability results for compact self-adjoint operators on a Hilbert space (see [18]).

Proposition 5.2 (Kato 1987)**.** *For any self-adjoint compact operators A, B on a Hilbert space $\mathcal{H}$ there are enumerations $\{\alpha_i\}_i$, $\{\beta_i\}_i$ of their eigenvalues (counting multiplicity) such that*

$$\sum_{i=1}^{\infty} |\alpha_i - \beta_i|^2 \leq \|A - B\|_{HS}$$

and, with possibly another order of eigenvalues,

$$\sup_i |\alpha_i - \beta_i| \leq \|A - B\|_{HS}.$$

Proposition 5.3. *Let A and B be compact self-adjoint operators on a Hilbert space $\mathcal{H}$. Take $k \in \mathbb{N}$ such that $\lambda_k(A) \neq 0$ and define $r_k := \frac{1}{2} d\big(\lambda_k(A), \sigma(A) \setminus \{\lambda_k(A)\}\big)$. If $\|A - B\|_{HS} \leq \frac{r_k}{2}$, then*

$$\|P_A - P_B\|_{HS} \leq \frac{2}{r_k} \|A - B\|_{HS},$$

where P_A and P_B are the orthogonal projectors onto $\operatorname{span}\{u_i(A) : \lambda_i(A) = \lambda_k(A)\}$ *and* $\operatorname{span}\{u_i(B) : |\lambda_i(B) - \lambda_k(A)| \leq r_k\}$*, respectively. Moreover, the dimensions of their ranges coincide.*

Proof. The result follows from theorem 20 in [18] (recall that for self-adjoint operators spectral projectors considered in that theorem are orthogonal). Indeed, take as a curve $\Gamma \subset \mathbb{C}$ enclosing $\lambda_k(A)$ the circle of radius r_k centered at $\lambda_k(A)$, so

$$\delta := \min_{\lambda \in \sigma(A), z \in \Gamma} |z - \lambda| = r_k$$

and $\ell(\Gamma) = 2\pi\delta$. Then

$$\|P_A - P_B\| \leq \frac{\ell(\Gamma)}{2\pi\delta} \frac{\|A - B\|}{\delta - \|A - B\|} \leq \frac{2\|A - B\|}{\delta}$$

and the dimensions of their ranges coincide. $\square$

Now we are ready to prove the main result of this section.

Theorem 5.4. *Let* $\lim_n GW_4(X, X_n) = 0$. *If* $m \in \mathbb{N}$ *is such that* $\lambda_m^+(T) > \lambda_{m+1}^+(T)$, *then*

$$\min_{Q \in O(m)} \|M^m - QM_n^m\|_{L^2(X^\infty, \boldsymbol{\gamma})} \to 0$$

as $n \to \infty$, *where* $O(m)$ *is the orthogonal group on* $\mathbb{R}^m$. *Moreover, if* $\deg \lambda_j^+(T) = 1$ *for all* $1 \le j \le m$, *then for every* n *there exists some diagonal matrix* $D_n \in \mathbb{R}^{m \times m}$ *with diagonal elements in* $\{-1, 1\}$, *such that*

$$\lim_{n \to \infty} \|M^m - D_n M_n^m\|_{L^2(X^\infty, \boldsymbol{\gamma})} = 0.$$

Proof. Fix some $\lambda_j^+(T) > 0$ and consider the projectors P_T and P_{T_n} coming from Proposition 5.3. Let $P_T = \sum_{i \in I} u_i(T) \otimes u_i(T)$ for a suitable index set I of size $d = \dim P_T$. For simplicity we can assume without loss of generality that $I = \{1, \dots, d\}$. Lemma 5.1 implies that $\|T_n - T\|_{HS} \to 0$. Then Propositions 5.2 and 5.3 yield that $P_{T_n} \to P_T$, and for n large enough $P_{T_n} = \sum_{i \in I} u_i(T_n) \otimes u_i(T_n)$. Since

$$\|(P_T - P_{T_n}) u_i(T_n)\| = \|P_T u_i(T_n) - u_i(T_n)\| \le \|P_T - P_{T_n}\| \to 0,$$

$\{u_i(T_n)\}_{i \in I}$ converges, up to a subsequence, to an orthonormal basis of $\mathrm{range}(T) = \mathrm{span}\{u_i(T) : i \in I\}$. Since all orthonormal bases are equivalent up to an orthogonal transformation, we obtain that in fact there are $Q^n \in O(d)$ such that for all $i = 1, \dots, d$,

$$\sum_{j=1}^d Q_{ij}^n u_j(T_n) \to u_i(T) \quad \text{in} \quad L^2(X^\infty, \boldsymbol{\gamma}).$$

Combining the above result with the convergence of the spectra (Proposition 5.2), we conclude that

$$\min_{Q \in O(m)} \|M^m - QM_n^m\|^2_{L^2(X^\infty, \boldsymbol{\gamma})} \to 0. \qquad \square$$

Remark 5.5. We require $\lambda_m^+(T) > \lambda_{m+1}^+(T)$ to ensure that the MDS map M^m is uniquely defined (up to an orthogonal transformation).

Corollary 5.6. *Under the assumptions of Theorem 5.4 for every* $m \in \mathbb{N}$ *one has*

$$GW_2\big(M^m(X), M_n^m(X_n)\big) \to 0$$

as $n \to \infty$, *where* $M^m(X)$ *stands for the space* $(M^m(X), \mathsf{d}_m, M^m_\# \mu)$ *and* $M_n^m(X_n)$ *stands for* $(M_n^m(X_n), \mathsf{d}_m, (M_n^m)_\# \mu_n)$, d_m *standing for the usual Euclidean distance in* $\mathbb{R}^m$.

Proof. We bound the Gromov–Kantorovich distance between the MDS images with the distance between the maps themselves in $L^2(X^\infty, \boldsymbol{\gamma})$: for any $Q \in O(m)$,

$$
\begin{aligned}
&GW_2^2\big(M^m(X), M_n^m(X_n)\big) \\
&\quad\le \int\limits_{(X\times X_n)^2} \left(|M^m(x)-M^m(y)| - |M_n^m(x')-M_n^m(y')|\right)^2 d\gamma_n(x,x')\, d\gamma_n(y,y') \\
&\quad= \int\limits_{(X\times X_n)^2} \left(|M^m(x)-M^m(y)| - |QM_n^m(x')-QM_n^m(y')|\right)^2 d\gamma_n(x,x')\, d\gamma_n(y,y') \\
&\quad\le \int\limits_{(X\times X_n)^2} \left(|M^m(x)-QM_n^m(x')| + |M^m(y)-QM_n^m(y')|\right)^2 d\gamma_n(x,x')\, d\gamma_n(y,y') \\
&\quad\le 4 \int\limits_{X\times X_n} |M^m(x)-QM_n^m(x')|^2\, d\gamma_n(x,x') \\
&\quad= 4\,\|M^m - QM_n^m\|^2_{L^2(X\times X_n,\gamma_n)} = 4\,\|M^m - QM_n^m\|^2_{L^2(X^\infty,\boldsymbol{\gamma})}\,.
\end{aligned}
$$

Using Theorem 5.4, we conclude the proof. □

5.2. **Stability under Kantorovich convergence.** In this section we consider the case where μ and μ_n are defined on the same space X, and μ_n converge to μ in the Kantorovich distance W_4.

5.2.1. *Embedding of a finite sample to $L^2(X,\mu)$.* Assume again we are given points $x_1,\dots,x_n \in X$. Define the empirical measure $\mu_n = \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$. Suppose that an optimal transport plan from μ to μ_n with a cost function $\mathsf{d}^4(\cdot,\cdot)$ gives us a partition of X into disjoint sets $V_1,\dots,V_n$, such that any $x \in V_i$ is transported to $t_n(x) := x_i$, i.e. there exists a solution to the Monge optimal transportation problem (which is the case e.g., if μ has no atoms). Define the operator

$$
A_n\colon L^2(X,\mu) \to \mathbb{R}^n, \qquad A_n f := \sqrt{n}\left(\int_{V_i} f\, d\mu\right)_{i=1,\dots,n}.
$$

It is easy to see that the adjoint operator $A_n^*\colon \mathbb{R}^n \to L^2(X,\mu)$ is given by

$$
A_n^* x = \sqrt{n}\sum_{i=1}^n x_i \mathbf{1}_{V_i}.
$$

In particular, $A_n A_n^* = \mathrm{Id}_n$ since $\mu(V_i) = \frac{1}{n}$, hence A_n^* is an isometric embedding of $\mathbb{R}^n$ into $L^2(X,\mu)$. Furthermore, $A_n\mathbf{1} = \frac{1}{\sqrt{n}}\mathbf{1}_n$, $A_n^*\mathbf{1}_n = \sqrt{n}\mathbf{1}$, and

$$
A_n(\mathbf{1}\otimes\mathbf{1}) = \frac{1}{\sqrt{n}}\mathbf{1}_n\otimes\mathbf{1}, \quad \mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} A_n = \mathbf{1}_n \otimes (A_n^*\mathbf{1}_n) = \sqrt{n}\mathbf{1}_n\otimes\mathbf{1}.
$$

Thus,

$$
A_n P = A_n(\mathrm{Id} - \mathbf{1}\otimes\mathbf{1}) = A_n - \frac{1}{\sqrt{n}}\mathbf{1}_n\otimes\mathbf{1} = A_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^{\mathrm{T}} A_n = \bar{P}_n A_n.
$$

Now we consider the operators

$$
K_n := A_n^*\bar{K}_n A_n \quad\text{and}\quad T_n := A_n^*\bar{T}_n A_n = P K_n P,
$$

where $\bar{K}_n$ and $\bar{T}_n$ are defined by (3.1) and (3.3) respectively. The eigenvalues of T_n are $\lambda_1(\bar{T}_n),\dots,\lambda_n(\bar{T}_n)$, all the remaining eigenvalues being zero and the first n

eigenfunctions are given by $u_j(T_n) = A_n^* v_j(\bar{T}_n)$, where $v_j(\bar{T}_n)$ is an eigenvector of $\bar{T}_n$. Indeed,

$$T_n u_j(T_n) = nA_n^* \bar{T}_n A_n A_n^* v_j(\bar{T}_n) = A_n^* \bar{T}_n v_j(\bar{T}_n) = \lambda_j(\bar{T}_n) A_n^* v_j(\bar{T}_n) = \lambda_j(\bar{T}_n) u_j(T_n).$$

Note that $\|u_j(T_n)\|_{L^2(X,\mu)} = \left|v_j(\bar{T}_n)\right|$, this allows us to define an empirical MDS map μ-a.e. by the formula

$$M_n(x) \coloneqq \left(\sqrt{\lambda_j^+(T_n)} u_j^+(T_n)(x)\right)_{j\in\mathbb{N}}. \tag{5.2}$$

Clearly, the metric measure spaces obtained by $\bar{M}_n$ and M_n from $\{x_1, \dots, x_n\}$ and X, respectively, are the same.

5.2.2. *Convergence.* One can see that K_n is also an integral operator, that is,

$$K_n f(x) = \int_X k_n(x,y) f(y)\, d\mu(x),$$

where $k_n(x,y) \coloneqq k(t_n(x), t_n(y))$. The following lemma controls $\|T_n - T\|_{HS}$ with $W_4(\mu_n, \mu)$ similarly to Lemma 5.1.

Lemma 5.7. *One has, setting* $C_X = \|\mathsf{d}(\cdot,\cdot)\|_{L^4(X\times X, \mu\otimes\mu)}$,

$$\|T_n - T\|_{HS} \le 2C_X W_4(\mu_n, \mu) + 2W_4^2(\mu_n, \mu).$$

Proof. First of all,

$$\begin{aligned}\|T_n - T\|_{HS} &\le \|K_n - K\|_{HS} = \|k_n - k\|_{L^2(X\times X, \mu\otimes\mu)} \\ &= \frac{1}{2}\left(\int_{X\times X} \left(\mathsf{d}^2(t_n(x), t_n(y)) - \mathsf{d}^2(x,y)\right)^2 d\mu\otimes\mu(x,y)\right)^{1/2}.\end{aligned}$$

Note that

$$\begin{aligned}&\left|\mathsf{d}^2(t_n(x), t_n(y)) - \mathsf{d}^2(x,y)\right| \\ &\qquad \le 2\mathsf{d}(x,y)\left|\mathsf{d}(t_n(x), t_n(y)) - \mathsf{d}(x,y)\right| + \left|\mathsf{d}(t_n(x), t_n(y)) - \mathsf{d}(x,y)\right|^2\end{aligned}$$

and

$$\begin{aligned}&\left(\int_{X\times X} \left(\mathsf{d}(t_n(x), t_n(y)) - \mathsf{d}(x,y)\right)^4 d\mu\otimes\mu(x,y)\right)^{1/4} \\ &\qquad \le \left(\int_{X\times X} \left(\mathsf{d}(x, t_n(x)) + \mathsf{d}(y, t_n(y))\right)^4 d\mu\otimes\mu(x,y)\right)^{1/4} \\ &\qquad \le 2\left(\int_X \mathsf{d}(x, t_n(x))^4 d\mu(x)\right)^{1/4} = 2W_4(\mu_n, \mu).\end{aligned}$$

Therefore, we have by the triangle and Hölder inequalities that

$$\begin{aligned}\Big(\int_{X\times X}\big(\mathsf{d}^2(t_n(x),t_n(y))-\mathsf{d}^2(x,y)\big)^2\,d\mu\otimes\mu(x,y)\Big)^{1/2}&\\ \le 2\Big(\int_{X\times X}(\mathsf{d}(x,y)[\mathsf{d}(t_n(x),t_n(y))-\mathsf{d}(x,y)])^2\,d\mu\otimes\mu(x,y)\Big)^{1/2}&\\ +\Big(\int_{X\times X}(\mathsf{d}(t_n(x),t_n(y))-\mathsf{d}(x,y))^4\,d\mu\otimes\mu(x,y)\Big)^{1/2}&\\ \le 4\Big(\int_{X\times X}\mathsf{d}(x,y)^4 d\mu\otimes\mu(x,y)\Big)^{1/4}W_4(\mu_n,\mu)+4W_4^2(\mu_n,\mu).&\end{aligned}$$

□

In the same way as in Section 5.1 we now obtain the following statement.

Theorem 5.8. *Let* $\lim_n W_4(\mu_n,\mu)=0$. *If* $\lambda_m^+(T)>\lambda_{m+1}^+(T)$, *then (up to an orthogonal transformation of* $\mathbb{R}^m$*)*

$$M_n^m\to M^m \text{ in } L^2(X,\mu;\mathbb{R}^m)$$

as $n\to\infty$. *In particular,* M_n *converge to* M *in measure* μ *with respect to the product topology on* $\mathbb{R}^\infty$ *(which is metrizable, so one can define the convergence in measure).*

Note that to obtain the strong convergence in ℓ^2 we need to have some uniform (in n) bounds on the tails $\sum_{i=j}^\infty\lambda_i^+(T_n)$ and $\sum_{i=j}^\infty\lambda_i^+(T)$. It is not clear how can we get it, even if the operator T is trace-class.

Another way to obtain a strong convergence is to modify slightly the definition of the infinite MDS as for instance proposed in the following statement.

Proposition 5.9. *Suppose that* $\operatorname{diam}(X)<\infty$, $p\ge 4$, *and set*

$$\tilde M(x):=\left(\sqrt{\lambda_j^+(T)}\frac{u_j^+(T)(x)}{\|u_j^+(T)\|_{L^p(X,\mu)}}\right)_{j\ge1},$$

$$\tilde M_n(x):=\left(\sqrt{\lambda_j^+(T_n)}\frac{u_j^+(T_n)(x)}{\|u_j^+(T_n)\|_{L^p(X,\mu)}}\right)_{j=1,\dots,n},$$

i.e. the maps of the same form as M *and* M_n *respectively, but with the eigenfunctions normalized in* $L^p(X,\mu)$ *instead of* $L^2(X,\mu)$. *One has then* $\tilde M\in L^p(X,\mu;\ell^p)$. *Moreover, there are orthogonal matrices* Q_n *(with different dimensions) such that*

$$\left\|\tilde M-Q_n\tilde M_n\right\|_{L^p(X,\mu;\ell^p)}\to 0 \tag{5.3}$$

as $n\to\infty$. *Here* Q_nx *acts on the first* $\dim Q_n$ *components of* $x\in\ell^p$, *leaving all the other components unchanged.*

Remark 5.10. The condition on boundedness of X (i.e. $\operatorname{diam}(X)<\infty$) is only taken to ensure that $\|u_j^+(T)\|_{L^p(X,\mu)}<+\infty$ for all j. Of course it can be weakened to just an appropriate integrability of the kernel k_T (or, equivalently, k) with respect to μ.

Proof. If $\operatorname{diam}(X) < \infty$, then by Lemma 3.1 every $u_j^+(T)$ is bounded (by C/λ_j for some $C > 0$ depending only on $\operatorname{diam} X$) so that $\left\|u_j^+(T)\right\|_{L^p(X,\mu)} < +\infty$ and hence in view of Proposition 3.2 the map $\tilde{M}$ is correctly defined. We assume $u_j^+(T)$ to be already normalized in $L^p(X,\mu)$, so that

$$\tilde{M}(x) := \left(\sqrt{\lambda_j^+(T)}u_j^+(T)(x)\right)_{j\geq 1}.$$

Denoting by $\langle\cdot,\cdot\rangle$ the duality pairing between ℓ^p and $\ell^{p'}$, for every $s \in (\ell^p)' = \ell^{p'}$, $p' = \frac{p}{p-1}$, one has

$$\langle\tilde{M}(x), s\rangle = \sum_j s_j\sqrt{\lambda_j^+(T)}u_j^+(T)(x),$$

one has

$$|\langle\tilde{M}(x), s\rangle| \leq \|s\|_{\ell^{p'}} \sum_j (\lambda_j^+(T))^{p/2}|u_j^+(T)|^p(x).$$

Let us set $q := \frac{p}{2} \geq 2$ and denote by $\|A\|_{S_q}$ the q-Schatten norm of a self-adjoint operator A, i.e.

$$\|A\|_{S_q} := \left(\sum_j |\lambda_j(A)|^q\right)^{1/q}.$$

Then by [19, theorem 1], since T is self-adjoint, one has

$$\begin{aligned}\|T\|_{S_q} &\leq \left(\int_X \|\mathsf{k}_T(\cdot,x)\|^q_{L^{q'}(X,\mu)}\,d\mu(x)\right)^{1/q}\\ &\leq 4\left(\int_X \|\mathsf{k}(\cdot,x)\|^q_{L^{q'}(X,\mu)}\,d\mu(x)\right)^{1/q} \leq 2\operatorname{diam}^2(X).\end{aligned}$$

Hence, the series

$$\sum_j (\lambda_j^+(T))^{p/2}|u_j^+(T)|^p(x)$$

is convergent for μ-a.e. $x \in X$, because

$$\int_X \sum_j (\lambda_j^+(T))^{p/2}|u_j^+(T)|^p(x) = \sum_j (\lambda_j^+(T))^{p/2} < +\infty.$$

Thus

$$\langle\tilde{M}(x), s\rangle = \lim_m \sum_{j=1}^m s_j\sqrt{\lambda_j^+(T)}u_j^+(T)(x)$$

for μ-a.e. $x \in X$, and hence the function $s \in (\ell^p)' \mapsto \langle\tilde{M}(x), s\rangle \in \mathbb{R}$ is measurable as a μ-a.e. limit of a sequence of continuous functions, implying that $\tilde{M}\colon X \to \ell^p$ is weakly (hence also strongly) measurable. The equality

$$\int_X \left\|\tilde{M}(x)\right\|^p_{\ell^p}\,d\mu(x) = \sum_j (\lambda_j^+(T))^{p/2} \leq \|T\|^{p/2}_{S_{p/2}} < \infty$$

shows that $\tilde{M} \in L^p(X,\mu;\ell^p)$.

Using again [19, theorem 1], the Lyapunov inequality (or, equivalently, just Hölder inequality), and the fact that $q' \leq 2 \leq q$ we obtain that

$$\begin{aligned}\|T_n - T\|_{S_q} &\leq \left(\int_X \|\mathsf{k}_T(\cdot,x) - \mathsf{k}_{T_n}(\cdot,x)\|^q_{L^{q'}(X,\mu)}\, d\mu(x)\right)^{1/q} \\ &\leq \|\mathsf{k}_T - \mathsf{k}_{T_n}\|_{L^q(X\times X,\mu\otimes\mu)}\,.\end{aligned}$$

Since

$$\begin{aligned}|\mathsf{k}(x,y) - \mathsf{k}_n(x,y)| &= \frac{\mathsf{d}(x,y) + \mathsf{d}(t_n(x),t_n(y))}{2}|\mathsf{d}(x,y) - \mathsf{d}(t_n(x),t_n(y))| \\ &\leq \operatorname{diam}(X)\left[\mathsf{d}(x,t_n(x)) + \mathsf{d}(y,t_n(y))\right],\end{aligned}$$

it is easy to see that

$$\begin{aligned}\|\mathsf{k}_T - \mathsf{k}_{T_n}\|_{L^q(X\times X,\mu\otimes\mu)} &\leq 4\,\|\mathsf{k} - \mathsf{k}_n\|_{L^q(X\times X,\mu\otimes\mu)} \\ &\leq 8\operatorname{diam}(X)\left(\int_X \mathsf{d}(x,t_n(x))^q d\mu(x)\right)^{1/q} \\ &\leq 8\operatorname{diam}(X)^{2-2/q} W_2^{2/q}(\mu,\mu_n) \to 0.\end{aligned}$$

Recall that if $v_n \in L^\infty(X,\mu)$ are uniformly bounded and $v_n \to v$ in $L^2(X,\mu)$, then they also converge in $L^p(X,\mu)$. Thus in a similar way to Theorem 5.8, using the uniform bounds on $u_j^+(T)$, we get that

$$\lim_n \min_{Q\in O(m)} \left\|\tilde{M}^m - Q\tilde{M}_n^m\right\|_{L^p(X,\mu;\ell^p)} = 0 \tag{5.4}$$

for every $m \in \mathbb{N}$ such that $\lambda_m^+(T) > \lambda_{m+1}^+(T)$. Now note that

$$\left\|\tilde{M}^m - \tilde{M}\right\|^p_{L^p(X,\mu;\ell^p)} = \sum_{j=m+1}^\infty \left(\lambda_j^+(T)\right)^{p/2} \to 0 \tag{5.5}$$

as $m \to \infty$, and extending a $Q \in O(m)$ to an operator (still denoted by Q) over ℓ^p, we get

$$\begin{aligned}\lim_{n\to\infty} \left\|Q\tilde{M}_n^m - Q\tilde{M}_n\right\|^p_{L^p(X,\mu;\ell^p)} &= \lim_{n\to\infty} \left\|\tilde{M}_n^m - \tilde{M}_n\right\|^p_{L^p(X,\mu;\ell^p)} \\ &= \lim_{n\to\infty} \sum_{j=m+1}^\infty \left(\lambda_j^+(T_n)\right)^{p/2} \\ &= \sum_{j=m+1}^\infty \left(\lambda_j^+(T)\right)^{p/2}\end{aligned}$$

by Proposition 5.2, since $\|T_n - T\|_{S_{p/2}} \to 0$. Hence,

$$\lim_{n\to\infty} \left\|Q\tilde{M}_n^m - Q\tilde{M}_n\right\|_{L^p(X,\mu;\ell^p)} \to 0 \tag{5.6}$$

as $m \to \infty$. Combining (5.4) with (5.5) and (5.6), we get (5.3), concluding the proof. □

6. MDS for sample spaces

In this section we show by means of several examples that MDS for manifolds equipped with intrinsic distances might produce a snowflake embedding of such manifolds in an infinite-dimensional Hilbert space, as, for instance, in the case of finite-dimensional spheres or flat tori.

6.1. **MDS embedding of spheres.** Consider the d-dimensional sphere $\mathbb{S}^d$ endowed with the inner metric $\mathsf{d}(x,y) = \arccos(x \cdot y)$ and a normalized surface measure $\mu := \frac{\sigma}{\sigma(\mathbb{S}^d)}$. Along with this space we consider its "snowflake", i.e. the same sphere but with the distance $\mathsf{d}^{1/2}$ and the same measure. We let M and $M_{1/2}$ stand for the respective infinite MDS embeddings of $\mathbb{S}^d$ into ℓ^2. The following statement is valid.

Proposition 6.1. *One has*

$$\|M(x) - M(y)\|_2^2 = \pi \left\|M_{1/2}(x) - M_{1/2}(y)\right\|_2^2 = \pi \mathsf{d}(x,y). \tag{6.1}$$

Proof. Define the kernel

$$\mathsf{k}(x,y) = -\frac{1}{2}\mathsf{d}^2(x,y) = -\frac{1}{2}\arccos^2(x \cdot y)$$

over $\mathbb{S}^d \times \mathbb{S}^d$, the corresponding integral operator K on $L^2(X,\mu)$

$$K[f](x) = \int_{\mathbb{S}^d} \mathsf{k}(x,y) f(y)\, d\mu(y) \tag{6.2}$$

and $T = PKP$, where P is the orthogonal projector onto the orthogonal complement of constant functions $\operatorname{span}^\perp(\mathbf{1})$ in $L^2(\mathbb{S}^d,\mu)$ defined by the formula

$$Pf := f - \mathbf{1}(\mathbf{1}, f)_{L^2(\mathbb{S}^d,\mu)}.$$

We justify below (see (6.6)) that the kernel k enjoys the explicit series expansion

$$\mathsf{k}(x,y) = \sum_{n=0}^{\infty} a_n (x \cdot y)^n, \quad \text{where } a_n = \begin{cases} -\frac{\pi^2}{8}, & n = 0, \\ \frac{\pi(2j)!}{(2j+1)2^{2j+1}(j!)^2}, & n = 2j+1, \\ -\frac{4^j (j!)^2}{2(j+1)(2j+1)}, & n = 2j+2. \end{cases} \tag{6.3}$$

Then by the theorem from [6, section 2] any spherical harmonic of degree j is an eigenfunction of K with eigenvalue

$$\lambda_j(K) = \frac{\Gamma(d/2)}{2^{j+1}} \sum_{s=0}^{\infty} a_{2s+j} \frac{(2s+j)!}{(2s)!} \cdot \frac{\Gamma(s+1/2)}{\Gamma(s+j+(d+1)/2)}. \tag{6.4}$$

In particular, $\lambda_0(K)$ corresponds to a constant function, thus any other eigenfunction u_j of K has zero mean. Since $\mathbb{S}^d$ is homogeneous under the action of the orthogonal group $O(d+1)$, and μ is invariant under this action, then by Proposition 3.4 the operators T and K share the same set of eigenfunctions and eigenvalues except for $\lambda_0(T) = 0$. We are interested only in positive eigenvalues, i.e. with odd indices. It is worth noting that according to (A.1) one has

$$\sum_n \lambda_n^+(T) = \sum_n \lambda_n^+(K) < +\infty,$$

and hence by Proposition 3.2 one has $M \in L^2(\mathbb{S}^d, \mu; \ell^2)$.

For the snowflake of $\mathbb{S}^d$ (endowed with the distance $\mathsf{d}^{1/2}$) we define the kernel

$$\mathsf{k}_{1/2}(x,y) := -\frac{1}{2}\arccos(x\cdot y)$$

over $\mathbb{S}^d \times \mathbb{S}^d$ and operators

$$K_{1/2}[f](x) := \int_{\mathbb{S}^d} \mathsf{k}_{1/2}(x,y) f(y)\, d\mu(y), \quad T_{1/2} := PK_{1/2}P. \tag{6.5}$$

in $L^2(X,\mu)$. The Taylor series expansion of arccos function gives

$$\mathsf{k}_{1/2}(x,y) = \sum_{n=0}^{\infty} b_n (x\cdot y)^n, \quad \text{where}$$
$$b_n = \begin{cases} -\frac{\pi}{4}, & n=0, \\ \frac{(2j)!}{(2j+1)2^{2j+1}(j!)^2}, & n = 2j+1, \\ 0, & \text{otherwise.} \end{cases}$$

The formula (6.3) for a_n can be either obtained from Taylor series expansion of the square of arccos function, or just from the above expression for b_n through the simple calculation

$$\begin{aligned} \sum_{n=0}^{\infty}(-2a_n)(-1)^n x^n &= \arccos^2(-x) = (\pi - \arccos x)^2 \\ &= \pi^2 - 2\pi \arccos x + \arccos^2 x \\ &= \pi^2 - 2\pi \sum_{n=0}^{\infty}(-2b_n)x^n + \sum_{n=0}^{\infty}(-2a_n)x^n, \end{aligned}$$

which implies

$$\sum_{n=0}^{\infty}(-2a_n)((-1)^n - 1)x^n = \pi^2 - 2\pi \sum_{n=0}^{\infty}(-2b_n)x^n \tag{6.6}$$

for all $x \in \mathbb{R}$. In particular, for $n \neq 0$ odd one has $a_n = \pi b_n$, while for $n \neq 0$ even one has $b_n = 0$. In short, for $n \geq 1$ we have $\pi b_n = (a_n)^+$. Using again results from [6] we immediately get that any spherical harmonic of degree j is an eigenfunction of $K_{1/2}$ with the eigenvalue

$$\lambda_j(K_{1/2}) = \frac{\Gamma(d/2)}{2^{j+1}} \sum_{s=0}^{\infty} b_{2s+j} \frac{(2s+j)!}{(2s)!} \cdot \frac{\Gamma(s+1/2)}{\Gamma(s+j+(d+1)/2)}.$$

The latter clearly is positive for $j \neq 0$, if and only if j is odd, in which case $\pi\lambda_j(K_{1/2}) = \lambda_j(K)$. By the same reason as the operators T and K, also the operators $T_{1/2}$ and $K_{1/2}$ share the same set of eigenfunctions and eigenvalues except for $\lambda_0(T_{1/2}) = 0$. Hence

$$T_+ = \pi T_{1/2}, \tag{6.7}$$

where T_+ is the "positive part" of T, i.e.

$$T_+ v := \sum_j \lambda_j^+(T)(v, u_j^+(T))_{L^2(\mathbb{S}^d,\mu)} u_j^+(T).$$

Therefore, $T_{1/2}$ is positive semidefinite and trace-class, hence an MDS embedding of $(\mathbb{S}^d, \mathsf{d}^{1/2})$ into ℓ^2 is an isometric in view of Remark 4.5. Hence the second equality of (6.1) holds. The first equality of (6.1) is true in view of (6.7). □

6.2. **MDS of product spaces.** Let $(X_i, \mathsf{d}_i, \mu_i)$ be a metric space with probability measure μ_i, $i = 1, 2$. Then define a metric on $X := X_1 \times X_2$ by

$$\mathsf{d}(x, y) := \sqrt{\mathsf{d}_1^2(x_1, y_1) + \mathsf{d}_2^2(x_2, y_2)}, \quad \text{where } x = (x_1, x_2),\ y = (y_1, y_2),$$

and endow this space with the product measure $\mu := \mu_1 \otimes \mu_2$.

Proposition 6.2. *Denote by $M^{(i)} \colon X^{(i)} \to \ell^2$, the infinite MDS map for each space $(X_i, \mathsf{d}_i, \mu_i)$, $i = 1, 2$. Then the infinite MDS map M for the space (X, d, μ) can be represented as*

$$\begin{aligned} M \colon X &\to \ell^2 \times \ell^2, \\ x &\mapsto M(x) = (M^{(1)}(x_1), M^{(2)}(x_2)), \end{aligned}$$

the space $\ell^2 \times \ell^2$ being equipped with the natural Hilbert space structure, and

$$\|M(x) - M(y)\|_2^2 = \left\|M^{(1)}(x_1) - M^{(1)}(y_1)\right\|_2^2 + \left\|M^{(2)}(x_2) - M^{(2)}(y_2)\right\|_2^2.$$

Proof. Let $\mathsf{k}^{(i)}$, $\mathsf{k}_T^{(i)}$, $K^{(i)}$, and $T^{(i)}$ denote the corresponding kernels and operators on X_i. Consider the kernel

$$\mathsf{k}(x, y) := -\frac{1}{2}\mathsf{d}^2(x, y) = \mathsf{k}^{(1)}(x_1, y_1) + \mathsf{k}^{(2)}(x_2, y_2)$$

over X and the respective operators K and T. Note that

$$\begin{aligned} \int_X \mathsf{k}(x, y')\, d\mu(y') &= \int_{X_1 \times X_2} \left(\mathsf{k}^{(1)}(x_1, y_1') + \mathsf{k}^{(1)}(x_2, y_2')\right) d\mu_1(y_1') \otimes \mu_2(y_2') \\ &= \int_{X_1} \mathsf{k}^{(1)}(x_1, y_1')\, d\mu_1(y_1') + \int_{X_2} \mathsf{k}^{(2)}(x_2, y_2')\, d\mu_2(y_2'), \end{aligned}$$

thus the kernel of T

$$\begin{aligned} \mathsf{k}_T(x, y) &= \mathsf{k}(x, y) - \int_X \mathsf{k}(x, y')\, d\mu(y') - \int_X \mathsf{k}(x', y)\, d\mu(x') \\ &\quad + \int_X \int_X \mathsf{k}(x', y')\, d\mu(x')\, d\mu(y') \\ &= \mathsf{k}_T^{(1)}(x_1, y_1) + \mathsf{k}_T^{(2)}(x_2, y_2). \end{aligned}$$

To study the eigenfunctions of T recall that $\{u_i \otimes v_j\}_{i,j}$ is an orthonormal basis in $L^2(X, \mu)$, where u_i is an eigenfunction of $T^{(1)}$ with eigenvalue ν_i, v_j is an

eigenfunction of $T^{(2)}$ with eigenvalue η_j, and $u_i \otimes v_j := u_i(x_1)v_j(x_2)$. Then

$$\begin{aligned}
(Tu_i \otimes v_j)(x) &= \int_X \left(\mathsf{k}_T^{(1)}(x_1,y_1) + \mathsf{k}_T^{(2)}(x_2,y_2)\right) u_i(y_1)v_j(y_2)\, d\mu_1(y_1) \otimes \mu_2(y_2) \\
&= \int_{X_1} \mathsf{k}_T^{(1)}(x_1,y_1)u_i(y_1)\, d\mu_1(y_1) \int_{X_2} v_j(y_2)\, d\mu_2(y_2) \\
&\quad + \int_{X_1} u_i(y_1)\, d\mu_1(y_1) \int_{X_2} \mathsf{k}_T^{(2)}(x_2,y_2)v_j(y_2)\, d\mu_2(y_2) \\
&= \left(T^{(1)}u_i\right)(x_1) \cdot (v_j, \mathbf{1})_{L^2(X_2,\mu_2)} + \left(T^{(2)}v_j\right)(x_2) \cdot (u_i, \mathbf{1})_{L^2(X_1,\mu_1)} \\
&= \nu_i u_i(x_1)(v_j, \mathbf{1})_{L^2(X_2,\mu_2)} + \eta_j v_j(x_2)(u_i, \mathbf{1})_{L^2(X_1,\mu_1)}.
\end{aligned}$$

Note that constant functions belong to $\ker T^{(1)}$ and $\ker T^{(2)}$, thus other eigenfunctions have zero mean. Hence, $T(u_i \otimes v_j) \neq 0$ only if $v_j = \mathbf{1}$, $\nu_i \neq 0$, or $u_i = \mathbf{1}$, $\eta_j \neq 0$. In the former case we have

$$T(u_i \otimes v_j) = \nu_i u_i \otimes \mathbf{1} = \nu_i(u_i \otimes v_j),$$

and in the latter case

$$T(u_i \otimes v_j) = \eta_j \mathbf{1} \otimes v_j = \eta_j(u_i \otimes v_j).$$

Therefore, $\{u_i \otimes v_j\}_{i,j}$ is a basis of eigenfunctions of T, with only eigenfunctions outside $\ker T$ having the form $\{u_i \otimes \mathbf{1}\}_i$ or $\{\mathbf{1} \otimes v_j\}_j$. We immediately obtain that one can represent the MDS map M of (X, d, μ) as claimed. □

An immediate corollary for the case of the flat torus $\mathbb{S}^1 \times \mathbb{S}^1$ is as follows.

Proposition 6.3. *The infinite MDS map $M\colon \mathbb{S}^1 \times \mathbb{S}^1 \to \ell^2$ for the flat torus $\mathbb{S}^1 \times \mathbb{S}^1$ equipped with the locally Euclidean (flat) metric and the volume measure satisfies*

$$\|M((x_1,x_2)) - M((y_1,y_2))\|_2^2 = \pi\left(\mathsf{d}_{\mathbb{S}^1}(x_1,y_1) + \mathsf{d}_{\mathbb{S}^1}(x_2,y_2)\right),$$

where $\mathsf{d}_{\mathbb{S}^1}$ stands for the spherical distance (i.e. the geodesic distance), i.e. provides a bi-Hölder embedding of the flat torus in a Hilbert space.

Proof. Combine Proposition 6.2 with Proposition 6.1 (with $d := 1$). □

Of course, a similar result can be easily obtained for a flat torus of any dimension.

Appendix A. Estimates of eigenvalues on a sphere

We show that for the positive eigenvalues of the operator K defined by (6.2), i.e. those with odd indices, one has

$$\lambda_{2n+1}(K) = \Theta_d\left(n^{-d-1}\right), \tag{A.1}$$

where we use Knuth's big Theta notation: $f(\cdot) = \Theta(g(\cdot))$ means that $f(\cdot) = O(g(\cdot))$ and $g(\cdot) = O(f(\cdot))$ hold simultaneously. Here and below the subscript d means that we consider d fixed, so the constant in $O(\cdot)$ can depend on it, and all the asymptotic notation, unless otherwise stated, refers to $n \to \infty$. To this aim note that by (6.4) one has

$$\lambda_n^+(K) = \lambda_{2n+1}(K) = \Gamma(d/2)\sum_{s=0}^{\infty} \theta_n(s), \tag{A.2}$$

where

$$\begin{aligned}
\theta_n(s) &= \frac{1}{2^{2n+2}} a_{2s+2n+1} \frac{(2s+2n+1)!}{(2s)!} \cdot \frac{\Gamma(s+1/2)}{\Gamma(s+2n+(d+3)/2)} \\
&= \frac{1}{2^{2n+2}} \cdot \frac{\pi(2(s+n))!}{(2(s+n)+1)2^{2(s+n)+1}((s+n)!)^2} \cdot \frac{(2(s+n)+1)!}{(2s)!} \cdot \\
&\quad \cdot \frac{\Gamma(s+1/2)}{\Gamma(s+2n+(d+3)/2)} \qquad \text{(writing } a_{2s+2n+1} \text{ from (6.3))} \\
&= 2^{2s-3} \left(\frac{(2(s+n))!\sqrt{\pi}}{2^{2(s+n)}(s+n)!} \right)^2 \frac{1}{(2s)!} \cdot \frac{(2s)!\sqrt{\pi}}{2^{2s}s!} \cdot \frac{1}{\Gamma(s+2n+(d+3)/2)} \\
&= \frac{\sqrt{\pi}}{8} \frac{\Gamma^2(s+n+1/2)}{\Gamma(s+2n+(d+3)/2)s!}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\alpha_n(s) &:= \frac{\theta_n(s+1)}{\theta_n(s)} = \frac{(s+n+1/2)^2}{(s+1)(s+2n+(d+3)/2)} \\
&= 1 - \frac{2(d+3)(s+1)-(2n-1)^2}{2(s+1)(2s+4n+d+3)} \\
&= 1 - \frac{1}{2s+4n+d+3} \left((d+3) - \frac{(2n-1)^2}{2(s+1)} \right).
\end{aligned}$$

From the above relationship we obtain that for

$$s^* := \frac{(2n-1)^2}{2(d+3)} - 1$$

one has that $\alpha_n(s^*) = 1$, $\alpha_n(s) \le 1$ when $s \ge s^*$, and $\alpha_n(s) \ge 1$ when $s \le s^*$. Thus we get

$$s_n := \operatorname{argmax}_{s\in\mathbb{N}} \theta_n(s) = \lceil s^* \rceil = \left\lceil \frac{(2n-1)^2}{2(d+3)} - 1 \right\rceil = \Theta_d(n^2). \tag{A.3}$$

Note that

$$\sum_{s=s_n}^{\infty} \theta_n(s) \le \sum_{s=0}^{\infty} \theta_n(s) \le s_n \theta_n(s_n) + \sum_{s=s_n}^{\infty} \theta_n(s). \tag{A.4}$$

Stirling's formula yields

$$\Gamma(x) = \Theta\left(\frac{1}{\sqrt{x}} \left(\frac{x}{e}\right)^x \right),$$

for large $x > 0$, and thus

$$\begin{aligned}
\theta_n(s) &= \Theta\left(\frac{\left(\frac{s+n-1/2}{e}\right)^{2(s+n)}}{\left(\frac{s+2n+(d+1)/2}{e}\right)^{s+2n+d/2+1} \left(\frac{s}{e}\right)^{s+1/2}} \right) \\
&= \Theta\left(\frac{(s+n-1/2)^{2s+2n}}{(s+2n+(d+1)/2)^{s+2n+d/2+1}\, s^{s+1/2}} \right) \\
&= \Theta\left(\left(\frac{e}{s}\right)^{(d+3)/2} \left(1 + \frac{n-1/2}{s}\right)^{2s+2n} \left(1 + \frac{2n+(d+1)/2}{s}\right)^{-(s+2n+d/2+1)} \right).
\end{aligned}$$

Since $\ln(1+x) = x - O(x^2)$ as $x \to 0$, we obtain that

$$(2s+2n)\ln\left(1+\frac{n-1/2}{s}\right) = 2n-1+O\left(\frac{n^2}{s}+\frac{n^3}{s^2}\right),$$

as well as

$$\left(s+2n+\frac{d}{2}+1\right)\ln\left(1+\frac{2n+(d+1)/2}{s}\right) = 2n+\frac{d+1}{2}+O\left(\frac{(n+d)^2}{s}+\frac{(n+d)^3}{s^2}\right).$$

Hence for any $s \geq s_n = \Theta_d(n^2)$ one has

$$\ln\theta_n(s) = -\frac{d+3}{2}(\ln s - 1) - 1 - \frac{d+1}{2} + O_d\left(\frac{n^2}{s}\right) = -\frac{d+3}{2}\ln s + O_d(1),$$

and thus $\theta_n(s) = \Theta_d\left(s^{-(d+3)/2}\right)$. Therefore,

$$\sum_{s=s_n}^{\infty}\theta_n(s) = \Theta_d\left(\sum_{s=s_n}^{\infty} s^{-(d+3)/2}\right) = \Theta_d\left(s_n^{-(d+1)/2}\right).$$

Moreover,

$$s_n\theta_n(s_n) = \Theta_d\left(s_n^{-(d+1)/2}\right).$$

Finally, substituting the latter estimate into (A.4), from (A.2) we conclude that

$$\lambda_{2n+1}(K) = \Theta_d\left(\sum_{s=0}^{\infty}\theta_n(s)\right) = \Theta_d\left(s_n^{-(d+1)/2}\right) = \Theta_d\left(n^{-d-1}\right), \tag{A.5}$$

where in the latter equility we used again $s_n = \Theta_d(n^2)$. This is in fact (A.1) as claimed.

References

[1] H. Adams, M. Blumstein, and L. Kassab. Multidimensional scaling on metric measure spaces. *Rocky Mountain Journal of Mathematics*, 50(2):397–413, 2020.

[2] L. Ambrosio, N. Gigli, and G. Savaré. *Gradient Flows: In Metric Spaces and in the Space of Probability Measures*. Springer Science & Business Media, 2008.

[3] L. Ambrosio, S. Honda, J.W. Portegies, and D. Tewodrose. Embedding of $\mathrm{RCD}^*(K,N)$ spaces in L^2 via eigenfunctions. *J. Funct. Anal.*, 280(10):Paper No. 108968, 72, 2021.

[4] L. Ambrosio and P. Tilli. *Topics on analysis in metric spaces*, volume 25 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2004.

[5] E. Arias-Castro, A. Javanmard, and B. Pelletier. Perturbation bounds for procrustes, classical scaling, and trilateration, with applications to manifold learning. *Journal of Machine Learning Research*, 21:15–1, 2020.

[6] D. Azevedo and V.A. Menegatto. Eigenvalues of dot-product kernels on the sphere. *Proceeding Series of the Brazilian Society of Computational and Applied Mathematics*, 3(1), 2015.

[7] M. Balasubramanian and E.L. Schwartz. The isomap algorithm and topological stability. *Science*, 295(5552):7–7, 2002.

[8] Y. Bengio, O. Delalleau, N. Le Roux, J-F. Paiement, P. Vincent, and M. Ouimet. Learning eigenfunctions links spectral embedding and kernel pca. *Neural computation*, 16(10):2197–2219, 2004.

[9] P. Bérard, G. Besson, and S. Gallot. Embedding Riemannian manifolds by their heat kernel. *Geom. Funct. Anal.*, 4(4):373–398, 1994.

[10] I. Borg and P.J.F. Groenen. *Modern multidimensional scaling: Theory and applications*. Springer Science & Business Media, 2005.

[11] I. Chesser, T. Francis, M. De Graef, and E.A. Holm. Learning the grain boundary manifold: tools for visualizing and fitting grain boundary properties. *Acta Materialia*, 195:209–218, 2020.

[12] H. Federer. *Geometric Measure Theory*. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen. Springer, 1969.
[13] J. Heinonen, P. Koskela, N. Shanmugalingam, and J.T. Tyson. *Sobolev spaces on metric measure spaces*. Number 27. Cambridge university press, 2015.
[14] L. Kassab. Multidimensional scaling: Infinite metric measure spaces. *arXiv preprint arXiv:1904.07763*, 2019. Masters thesis.
[15] S. Lim and F. Memoli. Classical mds on metric measure spaces. *arXiv preprint arXiv:2201.09385*, 2022.
[16] F. Mémoli. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11(4):417–487, 2011.
[17] N. Puchkin,, V. Spokoiny, E. Stepanov, and D. Trevisan. Reconstruction of manifold embeddings into euclidean spaces via intrinsic distances, 2020. arxiv 2012.13770.
[18] L. Rosasco, M. Belkin, and E. De Vito. On learning with integral operators. *Journal of Machine Learning Research*, 11(2), 2010.
[19] B. Russo. On the Hausdorff–Young theorem for integral operators. *Pacific Journal of Mathematics*, 68(1):241–253, 1977.
[20] J. Tenenbaum. Mapping a manifold of perceptual observations. *Advances in neural information processing systems*, 10, 1997.
[21] J. Tenenbaum, V. de Silva, and J.C. Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.
[22] J.T. Tyson and J.-M. Wu. Characterizations of snowflake metric spaces. *Ann. Acad. Sci. Fenn. Math*, 30(2):313–336, 2005.
[23] J. von Neumann and I.J. Schoenberg. Fourier integrals and metric geometry. *Transactions of the American Mathematical Society*, 50(2):226–251, 1941.
[24] J. Wang. *Geometric structure of high-dimensional data and dimensionality reduction*. Springer, 2012.
[25] J. Weidmann. Integraloperatoren der Spurklasse. *Math. Ann.*, 163:340–345, 1966.
[26] W.A. Wilson. On certain types of continuous transformations of metric spaces. *American Journal of Mathematics*, 57(1):62–68, 1935.

HSE University and Institute for Information Transmission Problems, Moscow
Email address: akroshnin@hse.ru

St.Petersburg Branch of the Steklov Mathematical Institute of the Russian Academy of Sciences, Fontanka 27, 191023 St.Petersburg, Russia and HSE University, Moscow and Department of Mathematical Physics, Faculty of Mathematics and Mechanics, St. Petersburg State University, St.Petersburg, Russia
Email address: stepanov.eugene@gmail.com

Dario Trevisan, Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, I-56127, Pisa
Email address: dario.trevisan@unipi.it