Multidimensional Scaling: Infinite Metric Measure Spaces

Henry Adams, Mark Blumstein and Lara Kassab*

Department of Mathematics Colorado State University

Objectives

The objectives of our research project include:

- Defining a notion of MDS on infinite metric measure spaces
- Studying its optimal properties and goodness of fit
- Testing convergence of MDS on metric measure spaces

Multidimensional Scaling

Multidimensional scaling (MDS) is a set of statistical techniques concerned with the problem of constructing a configuration of n points in a Euclidean space using information about the dissimilarities (or distances) between the n objects. The dissimilarities need not be based on Euclidean distances; they can represent many types of dissimilarities between objects. The goal of MDS is to map the objects $x_1, \ldots, x_n$ to configuration (or embedding) points $f(x_1), \ldots, f(x_n)$ in $\mathbb{R}^m$ in such a way that the given dissimilarities $d(x_i, x_j)$ are well-approximated by the Euclidean distance between $f(x_i)$ and $f(x_j)$ [1].

Motivating questions

- If a finite sample $\mathbb{X}_n \subseteq X$ converges to X as we sample more points, then in what sense do the MDS embeddings of these finite samples converge to the MDS embedding of X?
- More generally, if a sequence of metric measure spaces converges to X, then in what sense do the MDS embeddings of these spaces converge to the MDS embedding of X?

MDS on Infinite Metric Measure Spaces and its Convergence

We study multidimensional scaling of infinite metric measure spaces, that is spaces with infinitely many points equipped with some probability measure. Our motivation is to prove convergence properties of MDS of finite metric spaces. Convergence is well-understood when each metric space has the same finite number of points [3], but we are also interested in convergence when the number of points varies (and perhaps tends to infinity). An important application would be understanding how MDS behaves as one samples more and more points from a dataset.

Classical MDS - Infinite MDS

The following table shows a comparison of various elements of classical MDS and infinite MDS.

Elements	**Classical MDS**	**Infinite MDS**
Data	(X_n, d)	(X, d_X, μ)
Distance Representation	$D_{i,j} = d(x_i, x_j), \quad D \in \mathcal{M}_{n\times n}$	$K_D(x,s) = d_X(x,s) \in L^2_{\mu\otimes\mu}(X \times X)$
Linear Operator	$B = -\frac{1}{2} H D^{(2)} H$	$[T_{K_B}\phi](x) = \int K_B(x,s)\phi(s)\mu(\mathrm{d}s)$
Eigenvalues	$\lambda_1, \lambda_2, \ldots \lambda_n$	$\hat{\lambda}_1, \hat{\lambda}_2, \ldots$
Eigenvectors/Eigenfunctions	$v^{(1)}, v^{(2)}, \ldots, v^{(m)} \in \mathbb{R}^n$	$\phi_1(x), \phi_2(x), \ldots \in L^2(X)$
Embedding in $\mathbb{R}^m$ or ℓ^2	$f(x_i) = \left(\sqrt{\lambda_1} v_i^{(1)}, \sqrt{\lambda_2} v_i^{(2)}, \ldots, \sqrt{\lambda_n} v_i^{(m)}\right)$	$f(x) = \left(\sqrt{\hat{\lambda}_1}\phi_1(x), \sqrt{\hat{\lambda}_2}\phi_2(x), \sqrt{\hat{\lambda}_3}\phi_3(x), \ldots\right)$
Strain Minimization	$\sum_{i,j=1}^{n} (b_{i,j} - \hat{b}_{i,j})^2$	$\int\int \left(K_B(x,t) - K_{\hat{B}}(x,t)\right)^2 \mu(\mathrm{d}t)\mu(\mathrm{d}x)$

Example: A Circle

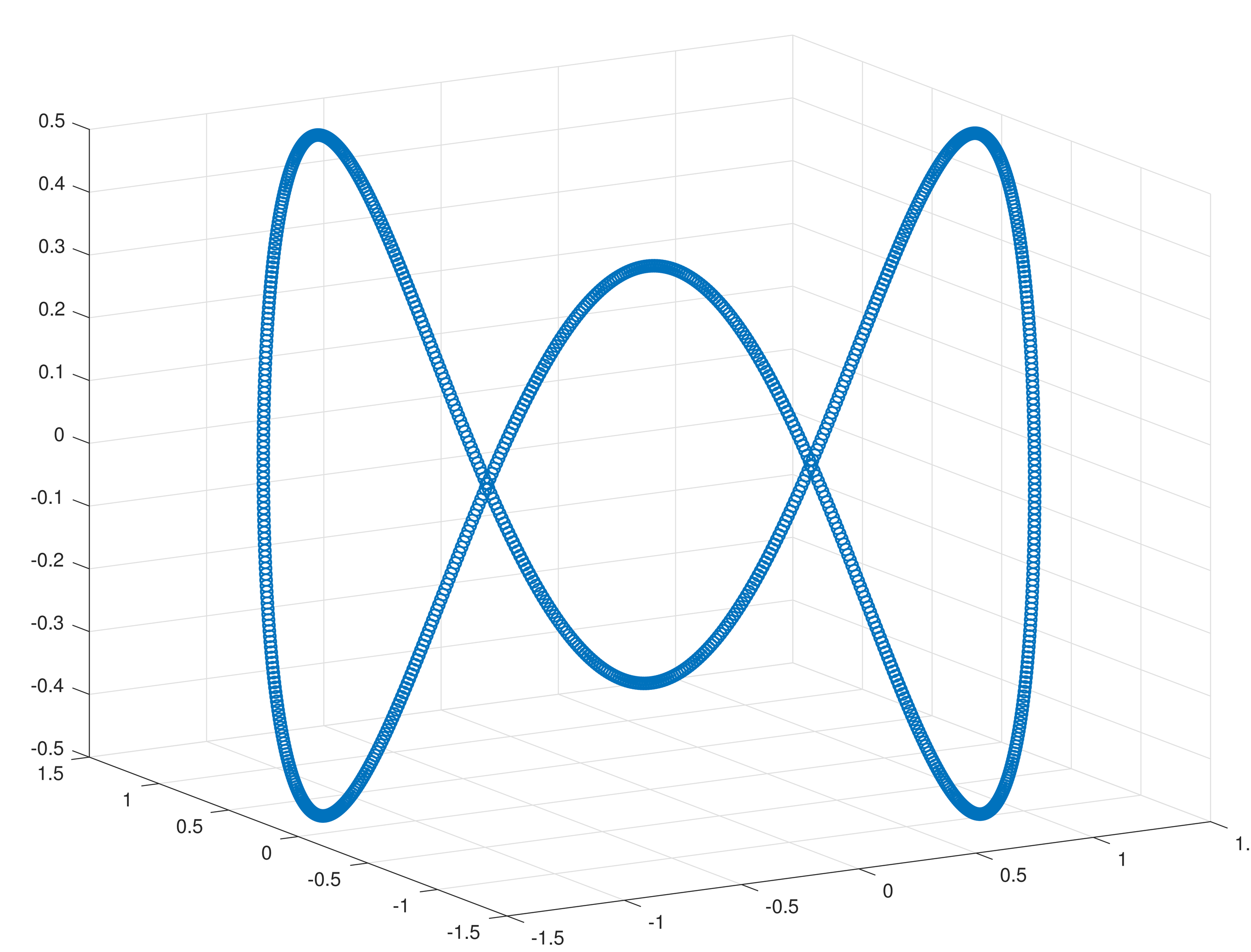

Figure 1: MDS embedding of 1000 evenly spaced points on S^1

References

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

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

[3] 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.

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

Contact Information

- Web: https://larakassab.weebly.com
- Email: lara.kassab@colostate.edu

Colorado State University