Jose Perea

Assistant Professor, CMSE; Department of Mathematics
joperea@msu.edu | 517.355.1600 | 619 Red Cedar Rd., Room C212

BIO SKETCH

Jose Perea joined the Michigan State University faculty in 2015 as an assistant professor, with joint appointments in the Department of Computational Mathematics, Science and Engineering, and the Department of Mathematics.

Jose Perea is an active researcher in the area of computational topology and topological data analysis. Broadly speaking, his work entails applications and adaptations of ideas from algebraic and geometric topology to the study of high-dimensional and complex data. He recently joined the Michigan State University faculty as an assistant professor in the Department of Computational Mathematics, Science and Engineering.

Prior to arriving at Michigan State, Perea held a post-doctoral position as a Visiting Assistant Professor in the Department of Mathematics at Duke University (2011–2015). He was a member of the Institute for Mathematics and Applications at the University of Minnesota, during the annual thematic program on Scientific and Engineering Applications of Algebraic Topology, in the spring of 2014. Dr. Perea received his B.S. in Mathematics from Universidad del Valle in 2006 (valedictorian and summa cum laude) under the supervision of Gonzalo Garcia, and a Ph.D. in Mathematics from Stanford University in 2011 under the supervision of Gunnar Carlsson.

RESEARCH INTERESTS

Computational geometric and algebraic topology; data analysis; machine learning; computer vision; computational biology

CURRENT RESEARCH FOCUS

Suppose you have a large but finite subset S of RD, with D (potentially much) bigger than 5.

Is it "connected" (as you squint)? If not, how many clusters are there? Holes? Voids? Does it accumulate around a familiar object? In short, what is the shape of S?

Questions like these try to identify the topology of the most likely space underlying the point cloud S. That is, they describe what we call the topology of the data. The collection of techniques one uses for this purpose, often adapted from algebraic and geometric topology, is what falls under the banner of Topological Data Analysis (TDA).

The goal of TDA is twofold: on the one hand it looks for new insights, and on the other, it augments classic data analytic methods by making them shape-aware. For instance, a data set which sits naturally on a torus (the surface of a doughnut) embedded in a high dimensional Euclidean space, should be regarded—e.g., for density estimation or model fitting—as sampled from the torus. Not only does this reduce the intrinsic dimensionality of the problem, but indicates the existence of underlying periodic processes—which get lost in the Euclidean picture.

It also suggests polar coordinates, instead of linear ones, as a more natural representation. This way of thinking has profound implications in, for example, how one produces and updates models for data, on how features for machine/deep learning purposes are extracted and on how density/parameter estimation is carried out. In what follows, we will see how some

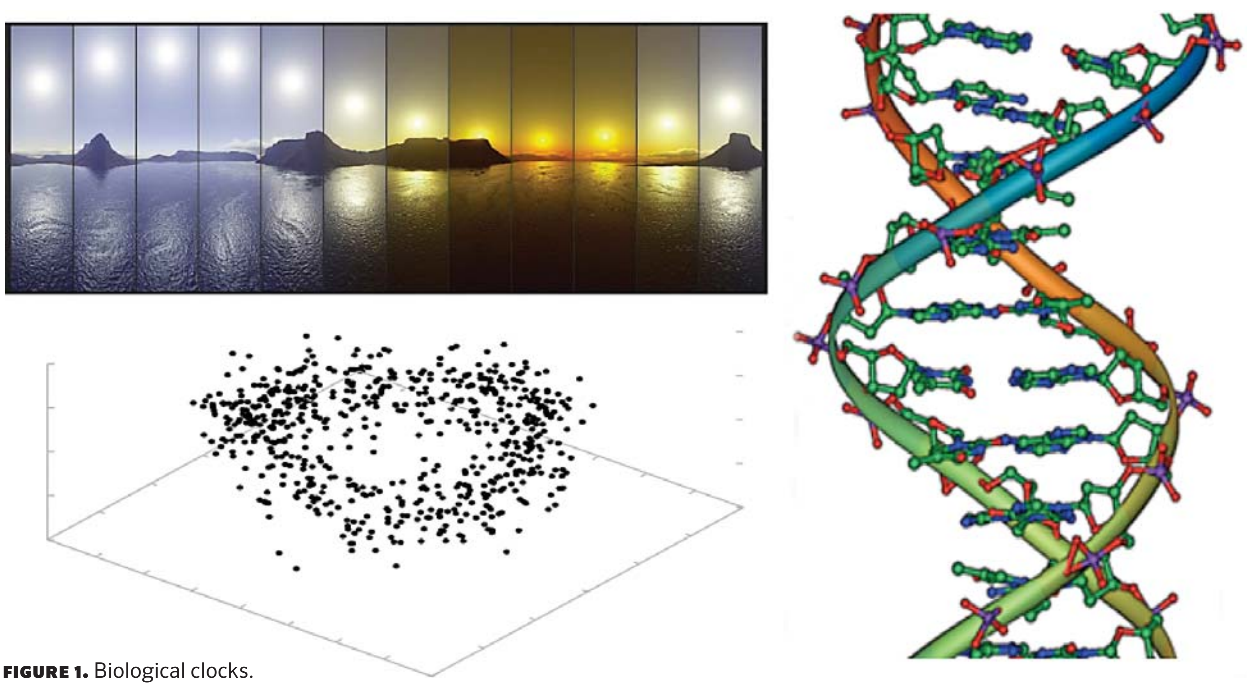
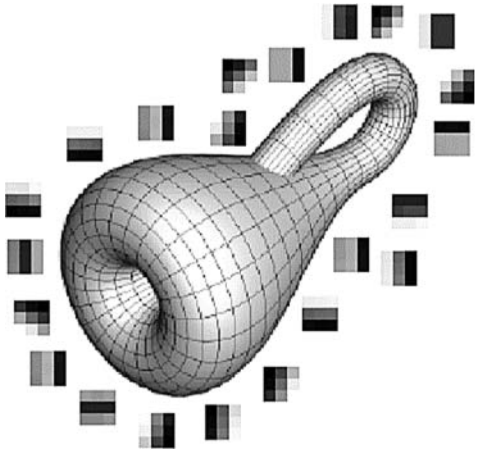


FIGURE 1. Biological clocks.



of these methodologies are applied to real-world data, and the vision for the work that will come out of the Applied Topology Group at Michigan State University.

Spaces of micro-images. One of the early surprises out of TDA was the work of Carlsson et al. (2008). They show that the majority of relevant small patches from natural images, accumulate with high-density around a surface with the shape of a Klein bottle.

Using this geometric model in conjunction with a compact Fourier-like representation, we introduced (Perea and Carlsson, 2014) a novel invariant for images. This representation was applied to the problem of texture classification, with results on par with the most successful methods in computer vision.

We know now that several other collections of features from images, both in patch-space and in the scale-space framework of Koenderink (1984), have descriptions in terms of simple topological spaces. We expect to identify the topology and useful parametrizations for these spaces, so that they can be used in image classification and restoration. The idea being that estimating the distribution of features from images (e.g., in order to produce the most likely patch on a corrupted area) should be done in their natural parametrizing space.

Sliding windows and persistence. Time series are ubiquitous across disciplines, and hence are an extremely important category of data. In (Perea and Harer, 2015) we investigate the sliding window (or time-delay) embedding used in dynamical systems—see for example (Takens, 1981)—from a geometric and topological perspective.

This work was the first theoretical analysis of how persistent homology, a fundamental tool in TDA, behaves when applied to time delay embeddings. Also in (Perea et al., 2015) it led to the SW1PerS algorithm for detection of periodic gene expression time series, outperforming some of the most popular methods in the literature. For instance, in data from the Yeast cell cycle, it identities genes not preferred by other algorithms, hence not previously reported as periodic, but found in other relevant experiments.

This same framework is currently being developed and deployed in music classification, business cycle analysis, chaos quantification and mechanical chatter detection. These are all projects which fit in the program of transforming data to TDA ready formats for further analysis.

Projective coordinates. Dimensionality reduction is a fundamental tool in data science. Not only does it allow one to produce visualizations for exploratory analyses, but it also makes tractable many complex learning problems with scarce data.

We have recently introduced (Perea, 2015) a methodology for dimensionality reduction which is informed by the topology of the data. In short, we construct maps to the relevant projective spaces which best capture the underlying topological features.

REFERENCES

G. Carlsson, T. Ishkhanov, V. De Silva, A. Zomorodian, "On the local behavior of spaces of natural images," *International Journal of Computer Vision*, 76(1), 1–12, 2008.

J.J. Koenderink, "The structure of images," *Biological Cybernetics*, 50(5), 363–370, 1984.

J.A. Perea, "Projective coordinates for the analysis of data," in preparation, 2015.

J.A. Perea, G. Carlsson, "A klein-bottle-based dictionary for texture representation," *International Journal of Computer Vision*, 107(1), 75–97, 2014.

J.A. Perea, J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, 15(3), 799–838. 2015.

J.A. Perea, A. Deckard, S.B. Hasse, J. Harer, "Sw1pers: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data," *BMC Bioinformatics*, 16(1), 257, 2015.

F. Takens, "Detecting strange attractors in turbulence," in *Dynamical Systems and Turbulence*, Warwick 1980, pp. 366–381. Springer, 1981.

RECENT PUBLICATIONS

J.A. Perea, A. Deckard, S.B. Haase, J. Harer, "SW1PerS: Sliding windows and 1-persistence scoring; discovering periodicity in gene expression time series data," *BMC Bioinformatics*, 16(1), 257 (2015).

J.A. Perea, J. Harer, "Sliding windows and persistence: An application of topological methods to signal analysis," *Foundations of Computational Mathematics*, 15(3), 799–838 (2015).

J.A. Perea, G. Carlsson, "A Klein-bottle-based dictionary for texture representation," *International Journal of Computer Vision*, 107(1), 75–97 (2014).