Received: 16 November 2022 | Revised: 30 January 2023 | Accepted: 17 February 2023

DOI: 10.1002/sta4.551

ORIGINAL ARTICLE

WILEY

Dimensionality reduction using elastic measures

J. Derek Tucker[1] | Matthew T. Martinez[2] | Jose M. Laborde[3]

[1]Statistical Sciences, Sandia National Laboratories, Albuquerque, New Mexico, USA

[2]Boeing, Albuquerque, New Mexico, USA

[3]Moffitt Cancer Center, Tampa, Florida, USA

Correspondence
J. Derek Tucker, Statistical Sciences, Sandia National Laboratories, NM, USA.
Email: jdtuck@sandia.gov

Present address
J. Derek Tucker, P.O. Box 5800 MS 0829, Albuquerque, NM 87185, USA.

Funding information
U.S. Department of Energy

With the recent surge in big data analytics for hyperdimensional data, there is a renewed interest in dimensionality reduction techniques. In order for these methods to improve performance gains and understanding of the underlying data, a proper metric needs to be identified. This step is often overlooked, and metrics are typically chosen without consideration of the underlying geometry of the data. In this paper, we present a method for incorporating elastic metrics into the t-distributed stochastic neighbour embedding (t-SNE) and Uniform Manifold Approximation and Projection (UMAP). We apply our method to functional data, which is uniquely characterized by rotations, parameterization and scale. If these properties are ignored, they can lead to incorrect analysis and poor classification performance. Through our method, we demonstrate improved performance on shape identification tasks for three benchmark data sets (MPEG-7, Car data set and Plane data set of Thankoor), where we achieve 0.77, 0.95 and 1.00 F1 score, respectively.

KEYWORDS
dimension reduction, functional data analysis, shape analysis

1 | INTRODUCTION

With the ever increasing number of observations and expanding number of variables in modern datasets, the use of dimensionality reduction as a first step for training machine learning models has gained a renewed interest. Methods such as the t-distributed stochastic neighbour embedding (t-SNE) (van der Maaten & Hinton, 2008) and the Uniform Manifold Approximation and Projection (UMAP) (McInnes et al., 2020) have enabled improved classification performance over methods such as principal component analysis (PCA) and factor analysis. This is due to their ability to accurately embed complicated data structures into small-dimensional spaces by computing the distances between observations in higher-dimensional space. The choice of distance function is typically left to the user, where the default metric is the Euclidean distance, which is deficient in higher-dimensional space. It has been shown that the UMAP is sensitive to the choice of distance metric and can effect the results significantly (McInnes et al., 2020).

When computing the distances between observations, a metric which is symmetric, isometric and obeys the triangle inequality is preferred. These properties lead to pairwise distance computations that are independent of input order and can be degenerate for some data types (Srivastava et al., 2011; Tucker et al., 2013). For example, the Euclidean distance is not a proper distance on functional data in $\mathbb{R}^1$ under phase variation (Tucker et al., 2013), the space of open and closed curves (Srivastava & Klassen, 2016) and any trajectory on a Riemmanian manifold (Su et al., 2014).

In this paper, we demonstrate that selecting a distance metric that is consistent with the data topology drastically improves the performance of both t-SNE and UMAP on functional data. We apply this method to a clustering of curves tasks in $\mathbb{R}^n$, where we choose our distance metric such that it measures the *shape* of these curves by accounting for rotation, scale and parameterization. We utilize the elastic shape framework (Srivastava & Klassen, 2016), which yields two metrics that properly measure distance in both shape and amplitude and provides a distance from one shape to another in sampling variability. When applied to the space of open curves, a Riemmanian manifold, we are able to obtain an initialization point for dimensionality reduction that is consistent with the properties of the data. We also apply this framework to any trajectory on a Riemmanian manifold using the square-root velocity transformation. Additionally, we can change out the cost-function in t-SNE from

Kullabck–Liebler divergence to a proper distance on the space of PDFs which provides performance gains depending on the balance of precision and recall desired.

This paper is organized as follows. In Section 2, we provide an overview of the elastic shape analysis (ESA) framework. In Section 3, we review both t-SNE and UMAP and provide modification in using a proper distance for functional data and a modification to the cost-function to t-SNE. Section 4 provides results on two data sets, one being a set of shapes from the MPEG7, shapes of cars extracted from video, and fighter jet shapes, and finally, we provide a discussion and concluding remarks in Section 5.

2 | ESA

ESA is a collection of techniques for registering and analysing functional data, using the process of phase-amplitude separation. From the separated components, statistical analysis can be performed on the individuals phase and amplitude components (Kurtek et al., 2011; Srivastava et al., 2011; Tucker et al., 2013). In this section, we provide a review of ESA and refer the reader to Kurtek et al. (2011), Srivastava et al. (2011), Srivastava and Klassen (2016) and Tucker et al. (2013) for a complete overview.

Phase and amplitude represent two orthogonal components of a function's shape, which are properties of a function that remain invariant to shape-preserving transformations (rotation, translation, scaling and phase) (Dryden & Mardia, 2016; Srivastava & Klassen, 2016). The phase component represents the *horizontal* variability within trajectories or reparameterization, where as the amplitude component represents the *vertical* variability independent of phase. In order to analyse these components, the need to be separated through alignment, Srivastava et al. (2011) demonstrated that the Fisher–Rao metric is a proper distance and provides this separation.

To compute the alignment using the Fisher–Rao metric, the square-root velocity function (SRVF) (Srivastava et al., 2011) is used for registration of functions in R^n. For real-valued curves, the SRVF performs a bijective mapping of the real-valued curve f to its normalized gradient $f'/\sqrt{|f'|}$. The registration of two real-valued curves is performed by elastically deforming the domain of one function such that the L^2 distance between the SRVFs of the two curves is minimized. The amount of elastic deformation required is measured by the phase distance (Section 2.2), while the residual distance between the SRVF, postregistration (rotation and scaling), defines the amplitude distance between them (Section 2.1). Together, they are known as elastic distances. By registering SRVFs instead of trajectories directly, the phase and amplitude distances become proper metrics and are invariant to shape-preserving transformations.

2.1 | Amplitude distance for R^n valued functions

Let $F_{R^n} = \{f : [0,1] \mapsto \mathbb{R}^n, f \text{ differentiable}\}$ be the class of differentiable trajectories on $[0,1]$ mapping to R^n with $n \geq 2$.[1] In higher-dimensional Euclidean space ($n \geq 2$), the scale, rotation and phase of a trajectory need to be accounted for in order to isolate its shape. Scale variability is removed by standardizing each trajectory to have unit length, that is, each trajectory is divided by the L^2 norm of its gradient:

$$f(t) \mapsto f(t)/||f'|| \; \forall t \in [0,1],$$

where $||f'|| = \sqrt{\int_0^1 f'(t)^2 dt}$.

Rotation variability is accounted for using the space of rotation matrices, the special orthogonal group $SO(n)$, and is defined as the group of orthogonal matrices with a determinant of one. For trajectories in R^2, $SO(2)$ defines the rotations around a point and in R^3 $SO(3)$ defines rotations around a line. The action of $SO(n)$ on a trajectory f is denoted as $O(f)$ and is defined pointwise as follows:

$$O(f) = \{Of(t) : \forall t \in [0,1]\},$$

where $Of(t)$ represents a standard matrix multiplication of the $n \times n$ matrix O and the $n \times 1$ vector $f(t)$. See Vladimir (1994) for more details on the orthogonal groups and their properties. The optimal rotation matrix in $SO(n)$ is found alongside the optimal reparameterization when computing the amplitude distance.

Phase variability in R^n is represented using the space $\Gamma = \{\gamma : I \to I | \{\gamma, \gamma^{-1}\} \in C^1(I), \dot{\gamma} > 0\}$ of all positive slope diffeomorphisms of the unit interval. Together with the operation of composition, Γ is a group and is defined with respect to the domain $I = [0,1]$. The properties of the diffeomorphisms are what allows the bending and stretching described above and why we use the term *elastic*. The reparameterization of a curve f by a warping function $\gamma \in \Gamma$ is denoted as the operation $f \circ \gamma$.

[1] We are focusing here on $n \geq 2$, and similar metrics are available for $n = 1$.

We desire a distance that is invariant to simultaneous reparameterizations, which allows us to use the SRVF as the following transformation on trajectories $f \in F_R^n$: The SRVF is defined as the following transformation on trajectories $f \in F_R^n$:

$$q_f(t) = \frac{f'(t)}{\sqrt{|f'(t)|}}, \tag{1}$$

where $|f'(t)|$ is the absolute value of f' at t. This transformation maps trajectories onto the L^2 ball of radius one and is a bijective mapping, up to an additive constant (Srivastava & Klassen, 2016). The reparmeterization of $q_f(t)$ is then $(q_f(t),\gamma) = (q_f(t) \circ \gamma)\sqrt{\dot{\gamma}}$.

Since the space of SRVFs is the L^2 ball, then the norm on the SRVFs is the arc length distance between points on the sphere:

$$d(q_f, q_g) = \arccos \int_0^1 \langle q_f(t), q_g(t) \rangle dt, \tag{2}$$

where $q_f = SRVF(f)$ and $q_g = SRVF(g)$ for two trajectories $f, g \in F_{R^n}$ and $\langle q_f(t), q_g(t) \rangle$ denotes the inner product of the vectors $q_f(t), q_g(t)$. To convert this into an amplitude distance, we need to place f and g in phase and rotation with each other. This optimization is summarized as the following amplitude distance on R^n valued trajectories (Srivastava et al., 2011):

$$(f,g) = \inf_{\gamma \in \Gamma, O \in SO(n)} \arccos \int_0^1 \langle q_f(t), q_{O(g \circ \gamma)}(t) \rangle dt, \tag{3}$$

where q_f and $q_{O(g \circ \gamma)}$ denote the SRVF's of f and $O(g \circ \gamma)$, respectively.

2.2 | Phase space

We will define the phase space using diffeomorphic mapping. The phase space of the unit interval $[0,1]$ as defined above is

$$\Gamma = \{\gamma : [0,1] \mapsto [0,1] \,|\, \gamma(0) = 0, \gamma(1) = 1, \gamma \text{ is a diffeomorphism}\}.$$

This diffeomorphic constraint gives rise to the notion of elasticity because the elements of Γ, that is, phase functions, can only smoothly stretch and contract portions of the unit interval so that it maps back to itself. Phase is generally thought of as the representation of a trajectory because any trajectory with domain $[0,1]$ can be *warped* by a phase function to appear differently. The amplitude will be taken to be those features of a trajectory that remain unchanged under any possible warping.

2.2.1 | Phase distance

To define phase distance, we use the optimal γ that defines the amplitude distance. The phase space Γ is a nonlinear manifold with no known geometry so we use the SRVF to map Γ to a known geometry (Tucker et al., 2013). Phase functions are positive for all $t \in [0,1]$ and $||q_\gamma|| = \int_0^1 \sqrt{\gamma'(t)}^2 dt = 1$, so the SRVF maps Γ onto the positive orthant of a unit Hilbert Sphere. Thus, the phase distance is defined as

$$d_p(\gamma_1, \gamma_2) = \arccos \int_0^1 \langle \psi_1, \psi_2 \rangle dt, \tag{4}$$

where $\psi = \sqrt{\dot{\gamma}}$ and $\langle \cdot, \cdot \rangle$ is the inner product between the two vectors (Srivastava & Klassen, 2016). The metric $d_p(\gamma_1, \gamma_2)$ is essentially measuring the amount of elastic deformation needed to compare the shapes of f_1 and f_2, when $\gamma_1 = id$ is the identity warping.

3 | DIMENSION REDUCTION

In dimensionality reduction, we are interested in finding a low-dimensional embedding space, Y, for the high-dimensional space, X, such that Y is as similar to X as possible and that the representation of the reduced space facilitates better data analysis (e.g. classification with good class separation). For the dimensionality reduction to work, we want to define the similarities between two objects x_i and x_j in X and their mapping to Y.

Recently, two methods have become highly utilized in the literature for performing dimensionality reduction; t-SNE and UMAP. In the following sections, we will give a brief overview of the two methods.

3.1 | t-SNE

A t-SNE (van der Maaten & Hinton, 2008), embedding defines probabilities proportional to the similarity of the objects x_i and x_j and is defined as

$$p_{j|i} = \frac{\exp(-d(x_i,x_j)^2/2\sigma_i^2)}{\sum_{k\neq i}\exp(-d(x_i,x_k)^2/2\sigma_i^2)},$$

where $p_{i|i} = 0$ and the $\sum_j p_{j|i} = 1 \forall i, d(x_i,x_j)$ is the Euclidean distance between the two objects (as defined in van der Maaten & Hinton, (2008)). The similarity between the two objects is the conditional probability that x_i is similar to x_j in proportion to the probability density under a Gaussian distribution centred at x_i.

The total probability or similarity is defined as

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N},$$

where t-SNE aims to learn a d-dimensional mapping, Y, that reflects the similarities p_{ij}. The similarities between y_i and y_j are defined as follows:

$$q_{ij} = \frac{(1+d(y_i,y_j)^2)^{-1}}{\sum_k\sum_{l\neq k}(1+d(y_i,y_j)^2)^{-1}},$$

for which the student t-distribution measures these similarities in the mapping space.

The location of the points (y_i) in the lower-dimensional space is determined by minimizing the Kullback–Leibler divergence of the distribution P from the distribution Q, that is,

$$KL(P||Q) = \sum_{i\neq j} p_{ij}\log\left(\frac{p_{ij}}{q_{ij}}\right).$$

This minimization is performed using gradient descent and is the mapping between X and Y.

3.2 | UMAP

UMAP approximates a manifold by constructing a fuzzy simplicial set representation, which is performed on high-dimensional data (X) and on a low-dimensional representation ($Y \in \mathbb{R}^d$). The selected representation is the one that optimizes the cross entropy between the high and low-dimensional spaces.

UMAP is similar to the approach of t-SNE, which constructs a probability distribution over pairs of high-dimensional objects. In t-SNE, similar objects have a high probability of being picked, whereas dissimilar points have an extremely small probability of being picked. t-SNE defines a similar probability distribution over the points in the low-dimensional map and minimizes the Kullback–Leibler divergence between the two distributions.

In UMAP, the high-dimensional similarities, $v_{j|i}$, are local fuzzy simplicial set memberships based on smooth nearest neighbour (NN) distances $v_{j|i}$ from $x_i \in X$ to one of its k distinct NNs $x_j \in X$

$$v_{j|i} = \exp[(-d(x_i,x_j) + \rho_i)/\sigma_i],$$

where $d(x_i,x_j)$ is the Euclidean distance on the learned manifold. The parameter ρ_i is the distance to NN, and σ_i is a normalizing constant. The symmetrization is used to produce an undirected graph structure representing the one-dimensional frame of the fuzzy simplicial set and is carried out by fuzzy set union using the probabilistic t-conorm

$$v_{ij} = (v_{j|i} + v_{i|j}) - v_{j|i}v_{i|j}.$$

The graph defined by the v_{ij} is then embedded into a low-dimensional space Y, where the dimension is prescribed as a parameter. The low-dimensional similarities between the projections y_i and y_j, of x_i and x_j, into Y via the initial embedding are given by

$$w_{ij} = \left(1 + a||y_i - y_j||_2^{2b}\right)^{-1},$$

where a and b are user defined and a gradient descent procedure is used to find them. The defaults for UMAP are $a \approx 1.929$ and $b \approx 0.7915$ (McInnes et al., 2020).

The cost function that is optimized to find the embedding is

$$C_{UMAP} = \sum_{i \neq j} v_{ij} \log\left(\frac{v_{ij}}{w_{ij}}\right) + (1 - v_{ij}) \log\left(\frac{1 - v_{ij}}{1 - w_{ij}}\right),$$

which penalizes discrepancies in the relative distributions of similarities in X and Y.

3.3 | Elastic modifications

To perform dimensionality reduction, it is critical that the distance, $d(x_i, x_j)$, is appropriate for the underlying data. In this work, we perform dimensionality reduction on curves that lie in $\mathbb{R}^n$. We then can utilize the distance defined in Equation (3) for measuring the distance between two open curves, where the distance is symmetric, isometric and obeys the triangle inequality, unlike the Euclidean distance. Additionally, we perform clustering on the phase variability of the data using the phase distance in Equation (4).

The cost function for t-SNE uses the Kullback–Leibler divergence which is not a metric in the space of probability density functions (PDFs); however, different divergences have been studied for t-SNE (Bunte et al., 2012; Im et al., 2018). Depending on the choice of divergence, a different balance is sought between precision and recall of the embedding. For example, in Im et al. (2018), the authors showed that the Hellinger distance, an extrinsic metric on the transformed space of PDFs, balances precision and recalls while penalizing small embeddings and preserving neighbourhood sizes. For a complete review of the effect of various diverges the reader is referred to Im et al. (2018). The use of an intrinsic metric instead of an extrinsic metric should be done as it is not an approximation. The Fisher–Rao metric is an intrinsic on the space of PDFs and can be used in favour of the Hellinger or Kullback–Leibler.

We give a brief review of the geometry of the space of PDFs but refer the reader to Kurtek and Bharath (2015) for a more in-depth review. We focus here on univariate densities on $[0,1]$, but the theory can easily be extended for all finite dimensional densities. Let $\mathcal{P} = \left\{p : [0,1] \to \mathbb{R}_{\geq 0} | \int_0^1 p(x)dx = 1\right\}$ as the space of all PDFs, and we define the tangent space $T_p(\mathcal{P}) = \left\{d_p : [0,1] \to \mathbb{R} | \int_0^1 d_p(x)p(x)dx = 0\right\}$. The inner product on this space, known as the Fisher–Rao Riemmanian metric (Kass & Vos, 1997; Rao, 1945), is given by the following:

$$\langle\langle d_{p_1}, d_{p_2}\rangle\rangle_p = \int_0^1 d_{p_1}(x) d_{p_2}(x) \frac{1}{p(x)} dx.$$

Čencov (1982) showed that this metric is invariant to reparameterizations. However, the metric is difficult to compute, requiring numerical methods and large computations. Bhattacharya (1943) proposed a convenient square-root transformation with simplifies the computation of the metric and is the transformed space where the Hellinger distance is an extrinsic metric. Define the transformation $\phi(p) = \sqrt{p}$ as the square-root density (SRD) of PDF p, which is similar to SRVF described previously. The corresponding space of SRDs is $\Psi = \left\{\sqrt{p} : [0,1] \to \mathbb{R}_{\geq 0} | \int_0^1 p(x)dx = 1\right\}$ which is the positive orthant of the Hilbert sphere. Since the geometry is known in this transformed space, we can define the geodesic distance between two PDFs $p_1, p_2 \in \mathcal{P}$ under the Fisher–Rao metric using the SRDS $\sqrt{p_1}, \sqrt{p_2} \in \Psi$ is the great circle defined as follows:

$$d_{FR}(p_1, p_2) = \cos^{-1}\left(\int_0^1 \sqrt{p_1(x)}\sqrt{p_2(x)}dx\right). \tag{5}$$

Therefore, we can replace the Kullback–Leibler divergence in t-SNE with a proper metric on the space of PDFs using Equation (5) similar to divergences done in Im et al. (2018) and Bunte et al. (2012). The use of an intrinsic metric is more natural than an extrinsic method as it is the exact distance between two PDFs and is invariant to reparameterizations of the PDFs. We will denote this modification to t-SNE as *Fisher–Rao t-SNE* or frt-SNE and will compare t-SNE, UMAP, and Fisher–Rao t-SNE using the elastic metrics on the space of curves in $\mathbb{R}^n$.

4 | RESULTS

In this section, we provide results on two data set. The first data set will be analysing a set of shapes where we have 65 shape classes from the MPEG-7 Core Experiment CE-Shape-1 Test Set(Jeannin & Bober, 1999). This data set is used for benchmarking shape matching algorithms and includes binary images grouped into categories by their content not their appearance or shape. The second is the fighter plane shapes and the car shapes from Thakoor et al. (2007) and includes part of the MPEG-7 data set.

4.1 | MPEG7

In this section, we classify real-world curve data taken from the MPEG-7 database (Jeannin & Bober, 1999). The full database has 1300 shape samples, 20 shapes for each class and 65 shape classes. An example of some of the shapes from the MPEG-7 bases is presented in Figure 1. Note the differences in rotations and scales of each of the shapes which complicates the analysis.

Using the shape distance in Definition 3, we can compute the pairwise distance matrix between all shapes in the data set. Figure 2a provides the distance matrix using the elastic distance, and for comparison, we computed the distance metric using the Euclidean distance and is provided in Figure 2b. The euclidean distance computed does remove rotation or scale variation. It should be noted this is **not** the default distance in the original implementation of t-SNE and UMAP. We will utilize this Euclidean distance for all results in this section. There is a definite difference in structure, and visually, there is a noticeable large amount of clustering in the elastic distance where the Euclidean distance is only showing roughly three large groups.

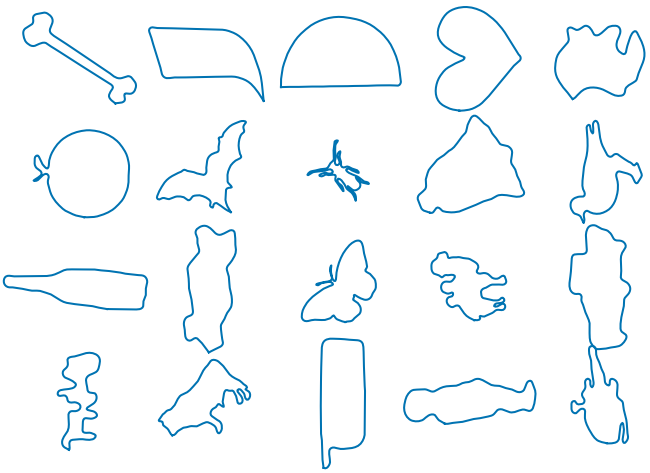
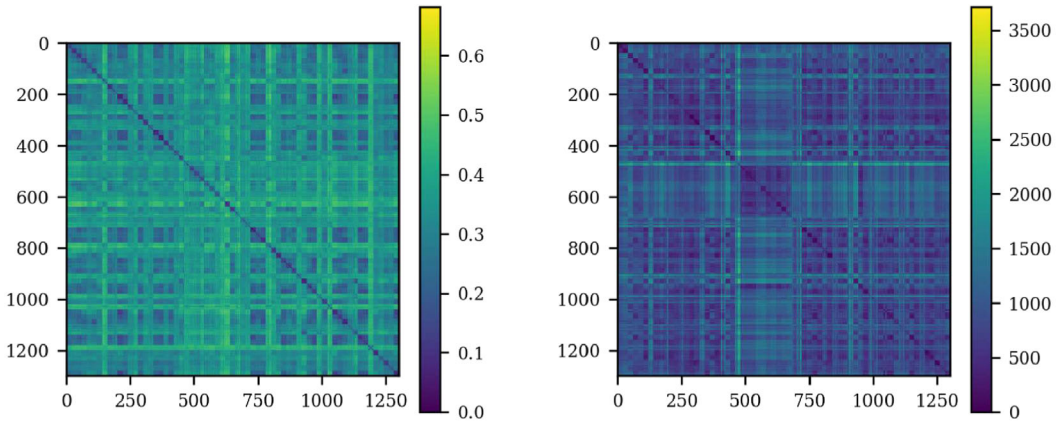


FIGURE 1 Example shapes from the MPEG-7 shape database.



(a) Elastic Distance (b) Euclidean Distance

FIGURE 2 (a, b) MPEG7 distance matrices.

We computed t-SNE, frt-SNE and UMAP projecting the data onto two components using the elastic distance and the Euclidean distance for comparison. Figure 3 provides the t-SNE results with panel (a) using the elastic distance and panel (b) using the Euclidean distance. There is a distinct clustering of each of the classes and separability of the classes versus the Euclidean distance. Figure 4 provides the Fisher–Rao t-SNE results with panel (a) using the elastic distance and panel (b) using the Euclidean distance. Comparing with standard t-SNE using the Kullback–Leibler divergence, we see a slightly better separation of classes, and as mentioned in above, we see that neighbourhood sizes are more preserved. Figure 5 provides the UMAP results with panel (a) using the elastic distance and panel (b) using the Euclidean distance. Again, we see a distinct clustering of the classes and better separability using a proper metric for the space of open curves.

To add a more complete picture to the embedding, we trained a random forest (RF) classifier (Breiman, 2001) using fivefold cross-validation. Since t-SNE was not originally defined for out-of-sample data, we will focus on UMAP. In the following section, for the full data analysis, we will utilize parametric t-SNE (Maaten, 2009; Policar et al., 2023) which allows for out-of-sample data. Figure 6 provides the average F1 score across folds with the standard deviation shaded, using the elastic metric and the Euclidean metric. Overall, for each class, the F1 score is larger for each class and maintains on average over 0.8 while on average of the Euclidean is 0.6. Figure 7 provides the confidence matrix for the elastic distance

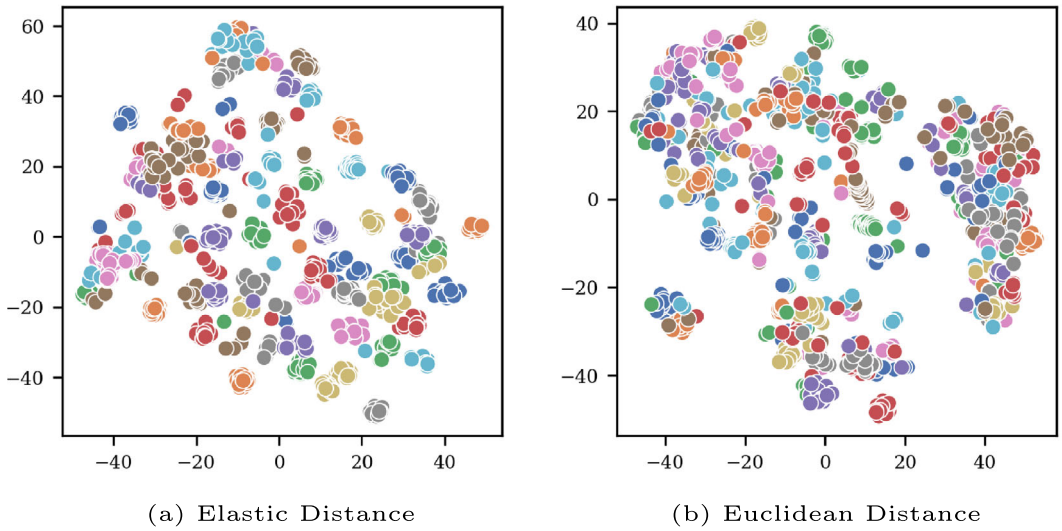


FIGURE 3 (a, b) MPEG7 t-distributed stochastic neighbour embedding (t-SNE) projection into two dimensions.

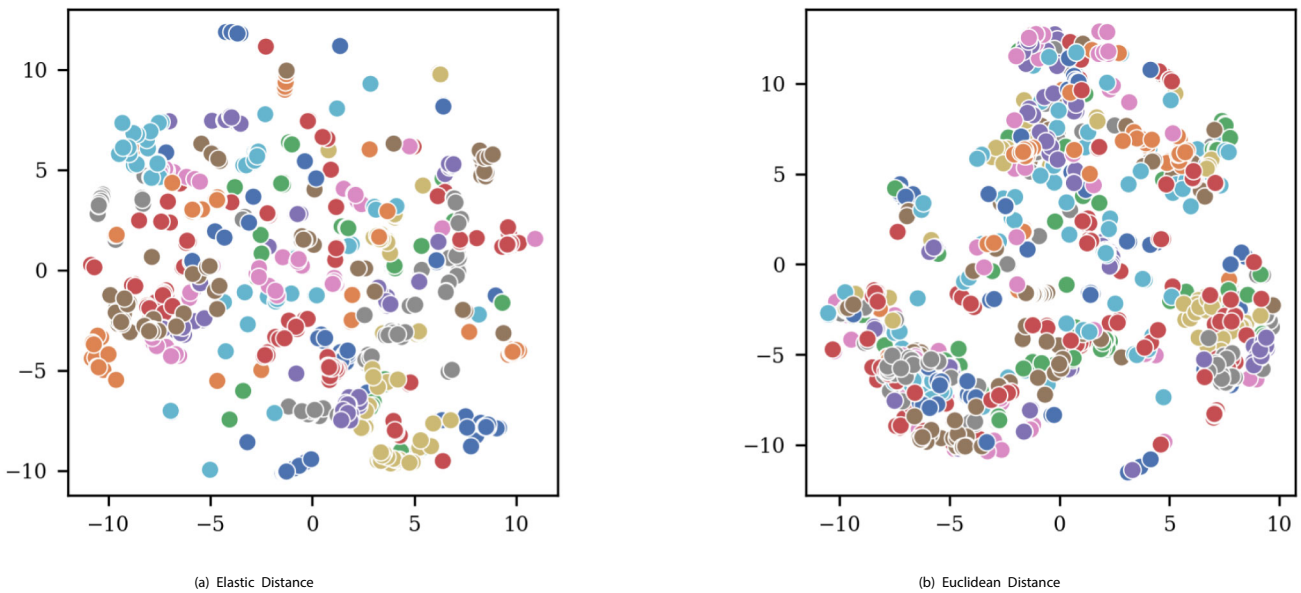


FIGURE 4 (a, b) MPEG7 Fisher–Rao t-distributed stochastic neighbour embedding (t-SNE) projection into two dimensions.

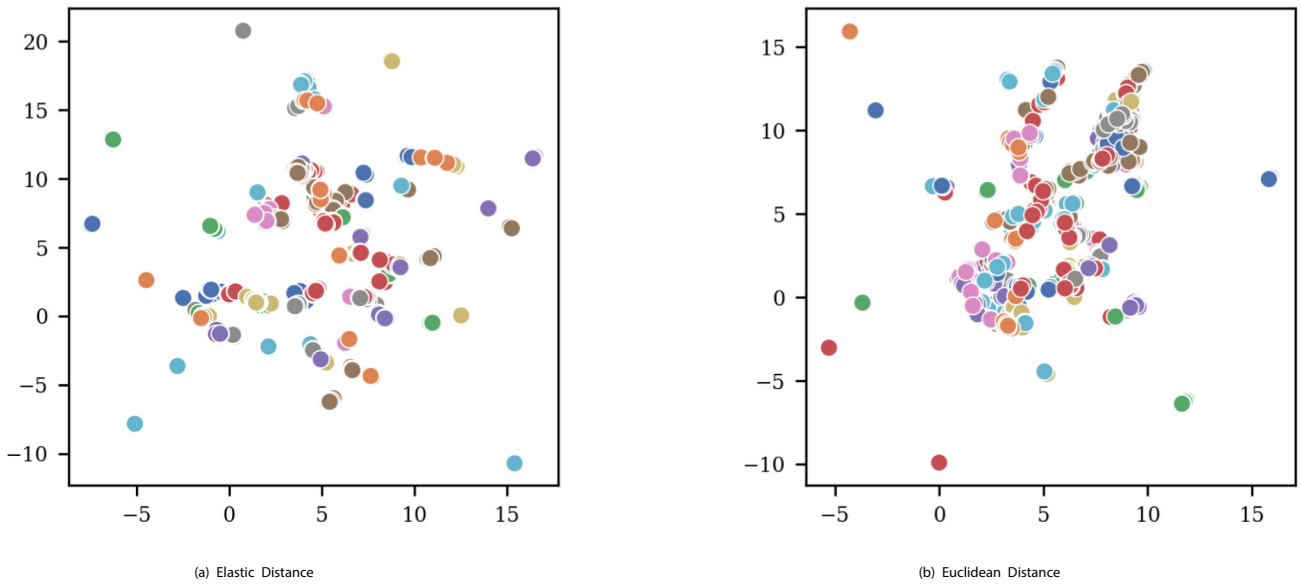


FIGURE 5 (a, b) MPEG7 Uniform Manifold Approximation and Projection (UMAP) projection into two dimensions.

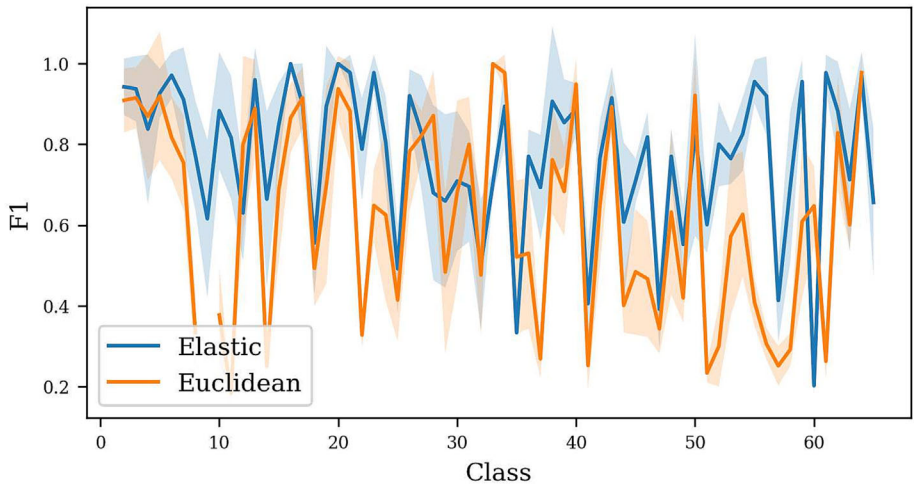


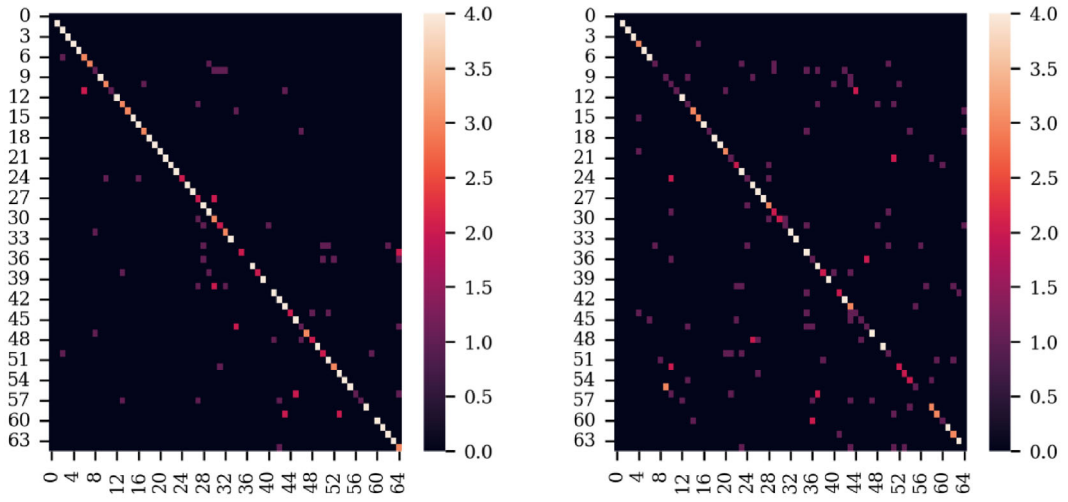
FIGURE 6 Average F1 score MPEG7

in panel (a) and the Euclidean metric in panel (b). As was noted from the F1 score, the elastic distance for most of the classes does extremely well with few classes being misspecified. The Euclidean distance does have trouble with a few of the classes with more than a few miss classifications.

4.2 | Shape database

In this section, we look at the overall classification performance utilizing UMAP and parametric t-SNE as a dimensionality reduction tool and then performing classification using a RF classifier. Again, we will compare the results against using the standard Euclidean metric in both dimension reduction techniques. For this study, we will utilize the fighter jet shapes, car shapes and the full MPEG7 database described and analysed in the previous section.

The fighter airplane shape database includes Mirage, Eurofighter, F-14, Harrier, F-22 and F-15 shapes. Since the F-14 has two possible shapes, one when its wings are closed and another when its wings are opened, the total number of shape classes is seven, where there are 30 shapes for each class for a total of 210 examples. The database was created from digital pictures of die-cast replica of the airplanes and were segmented to extract the contour.



(a) Elastic Distance (b) Euclidean Distance

FIGURE 7 (a, b) MPEG7 confidence matrices.

The car shapes were generated using the method described in Thakoor and Gao (2005) where contours of vehicles were extracted from video clips. There are four classes in the group: sedan, pickup, minivan and SUV, and each has 30 samples for each class. The data set shows larger within-class variation, as shapes of vehicles of different makes and models vary and some contours are distorted due to the shadow also being extracted. This data set contains more real-world variability in applying shape classification metrics.

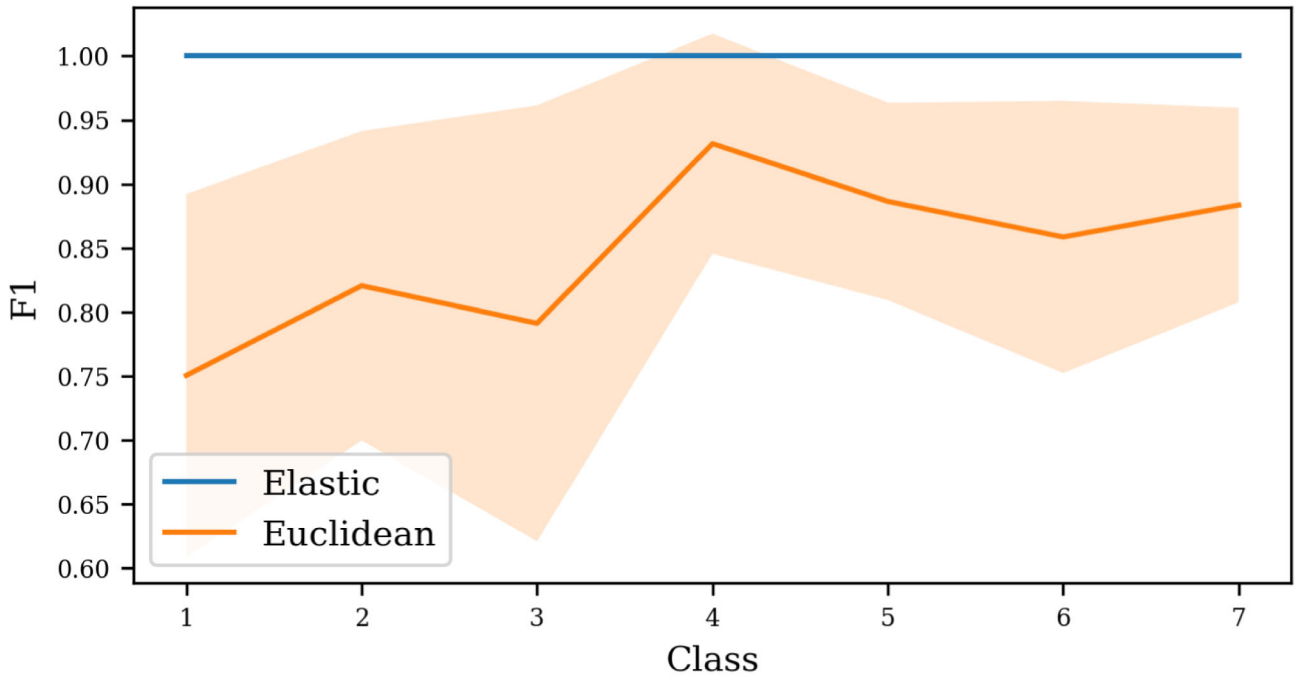
Figure 8 presents the F1 score for each class for the plane data set in panel (a) and the car data set in panel (b). A RF classifier was trained from the UMAP embedding using the designated distance. The shaded region shows the variability across the fivefold cross-validation. In both data sets, the elastic metric vastly outperforms the the Euclidean metric for all of the classes and shows less variability across the folds. The increase in performance is directly related to utilizing a proper distance and a metric that is invariant to scale, rotation and reparameterization of the shape. Being invariant to these actions is critical in any shape analysis/recognition task.

Table 1 presents a look at the class average F1 score and class averaged Matthew's correlation coefficient (MCC) for all three data sets: planes, cars and MPEG7. The average was taken across classes and across folds from the fivefold cross-validation of a RF classifier with the designated metric and dimensionality reduction technique. Panel (a) presents the metrics utilizing UMAP, panel (b) presents the metrics utilizing t-SNE and panel (c) presents the metrics using Fisher–Rao t-SNE. For UMAP, the elastic method outperforms the Euclidean metric for both the *F*-score and MCC, and for the easier planes, data set was perfect. For t-SNE, we achieve similar results where the elastic metrics out perform the Euclidean metric. It is the same story over improvement in using the metrics when for Fisher–Rao t-SNE. We see an improvement across all data sets when using the Fisher–Rao metric as the cost-function for t-SNE. The performance overall is degraded using t-SNE over UMAP as UMAP computes a more separable lower-dimensional representation. Additionally, as has been well-known across the literature, t-SNE is more computationally expensive as the number of examples in the data set increases. The one benefit of t-SNE over UMAP is the statistical interpretation which can be used for uncertainty quantification.

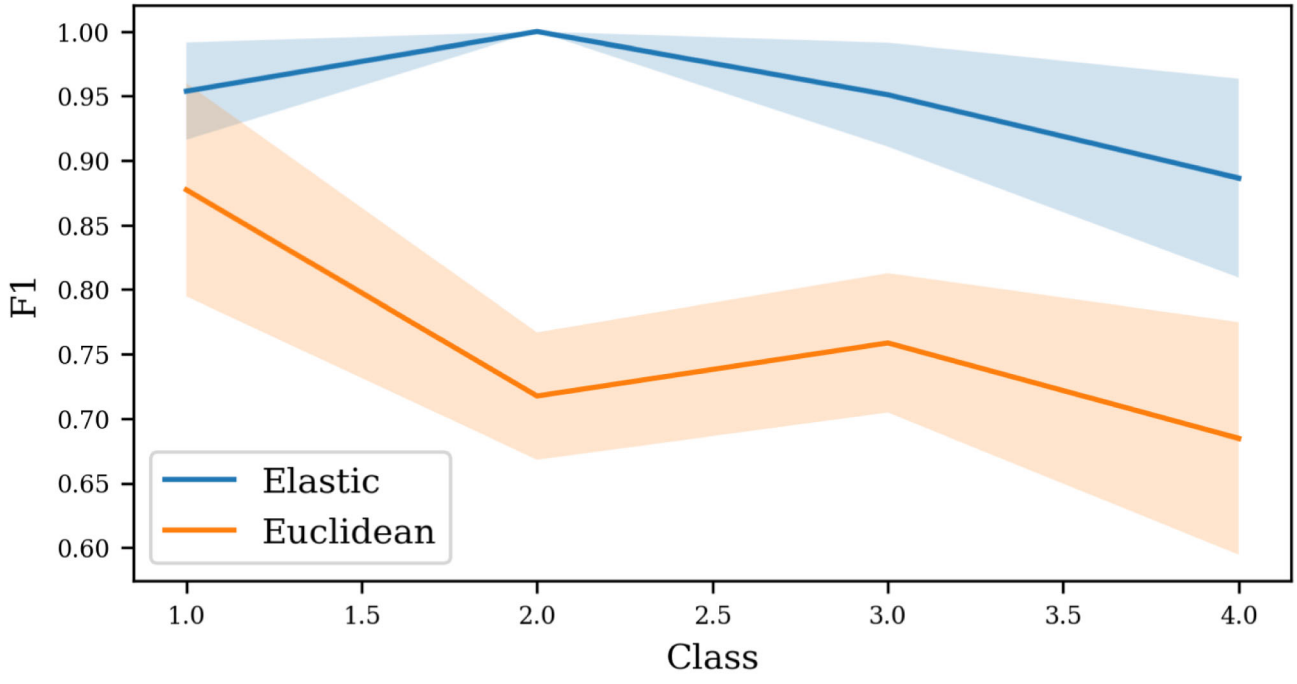
Lastly, to examine performance across multiple classifiers, we compared the performance of the RF against the gradient boosted trees (GBT) (Friedman, 2001) and K-nearest neighbours (KNN) classifiers (Cover & Hart, 1967). The purpose of this comparison is to demonstrate that the method is independent of the classifier and to show that the power is in choosing the proper metric for the geometry of the data under consideration. Table 2 presents the class averaged F1 score and MCC for the three classifiers for the MPEG-7 data set. For all three classifiers under both UMAP and t-SNE, we see an improvement in classification performance when using the elastic metric over the Euclidean. Across the classifiers, UMAP has the most consistent performance, while t-SNE obtains the highest under KNN, the performance under the other classifiers is quite reduced. The increase in performance between the two metrics is attributed to picking a metric that is a *proper* distance for the data being classified.

5 | DISCUSSION

We have proposed a new flexible elastic metric for functional data utilizing state-of-the art dimensionality reduction methods. We have demonstrated its advantages over the current methods using current shape analysis benchmarking data set. Unlike the current cross-sectional metrics,



(a) Planes



(b) Cars

FIGURE 8 (a, b) F1 score for each class for the plane and car data sets using Uniform Manifold Approximation and Projection (UMAP) and random forest.

the elastic method accurately estimates the mean of the underlying data generating mechanism and handles rotation and parameterization as a nuisance accordingly. It is imperative to choose a metric that is respective of the underlying variability of the data and is respective of the geometry. Doing so will give advantages in multiple machine learning tasks. We have presented the metric for open and closed shapes in $\mathbb{R}^n$, but equivalent metrics exists for functions in $\mathbb{R}^1$ (Tucker et al., 2013), surfaces (Kurtek et al., 2017) and images (Xie & Srivastava, 2016).

TABLE 1 Class averaged metrics for shape data set.

	Elastic		Euclidean	
Data set	***F*-score**	**MCC**	***F*-score**	**MCC**
(a) UMAP				
Planes	1.000	1.000	0.8571	0.8426
Cars	0.9478	0.9367	0.7786	0.6997
MPEG-7	0.7694	0.7587	0.6300	0.5917
(b) t-SNE				
Planes	0.9554	0.9400	0.74526	0.7234
Cars	0.8130	0.7662	0.7230	0.6455
MPEG-7	0.5833	0.4614	0.5546	0.4113
(c) Fisher–Rao t-SNE				
Planes	0.9904	0.9891	0.8148	0.7261
Cars	0.8240	0.7962	0.7571	0.6845
MPEG-7	0.5990	0.5282	0.5321	0.4322

Abbreviations: MCC, Matthew's correlation coefficient; t-SNE, t-distributed stochastic neighbour embedding; UMAP, Uniform Manifold Approximation and Projection.

TABLE 2 Class averaged metrics for MPEG-7 across multiple classifiers.

	Elastic		Euclidean	
Data set	***F*-score**	**MCC**	***F*-score**	**MCC**
(a) UMAP				
RF	0.7694	0.7587	0.6314	0.5987
GBT	0.6926	0.6209	0.6300	0.4911
KNN	0.7942	0.7813	0.7011	0.6540
(b) t-SNE				
RF	0.5990	0.5282	0.5104	0.4134
GBT	0.5409	0.3432	0.4799	0.2432
KNN	0.8480	0.8352	0.7103	0.6542
(c) Fisher–Rao t-SNE				
RF	0.5833	0.4614	0.5742	0.4324
GBT	0.4402	0.3133	0.4762	0.3546
KNN	0.6212	0.5706	0.5642	0.4724

Abbreviations: GBT, gradient boosted trees; KNN, K-nearest neighbours; MCC, Matthew's correlation coefficient; RF, random forest; t-SNE, t-distributed stochastic neighbour embedding; UMAP, Uniform Manifold Approximation and Projection.

ACKNOWLEDGEMENTS

This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the US Department of Energy or the US Government. This work was supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the US Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

ORCID

J. Derek Tucker https://orcid.org/0000-0001-8844-2169

REFERENCES

Bhattacharya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society*, *35*, 99–109.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.
Bunte, K., Haase, S., Biehl, M., & Villmann, T. (2012). Stochastic neighbor embedding (SNE) for dimension reduction and visualization using arbitrary divergences. *Neurocomputing*, *90*, 23–45.
Čencov, N. N. (1982). *Statistical decision rules and optimal inferences*. AMS.
Cover, T., & Hart, P. (1967). Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, *13*(1), 21–27.
Dryden, I. L., & Mardia, K. V. (2016). *Statistical shape analysis with application in R*. Wiley.
Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232.
Im, D. J., Verma, N., & Branson, K. (2018). Stochastic neighbor embedding under f-divergences. Journal Article. arXiv:1811.01247 [cs.LG]. Retrieved from https://dx.doi.org/10.48550/arxiv.1811.01247
Jeannin, S., & Bober, M. (1999). Shape data for the MPEG-7 core experiment ce-shape-1. https://www.dabi.temple.edu/shape/MPEG7/dataset.html
Kass, R. E., & Vos, P. W. (1997). *Geometric foundations of asymptotic inference*. John Wiley & Sons, Inc.
Kurtek, S., & Bharath, K. (2015). Bayesian sensitivity analysis with the Fisher–Rao metric [Journal Article. *Biometrika*, *102*(3), 601–616. https://doi.org/10.1093/biomet/asv026
Kurtek, S., Laga, H., Srivastava, A., & Jermyn, I. H. (2017). *Elastic shape analysis of three-dimensional objects*. Springer International Publishing.
Kurtek, S., Srivastava, A., & Wu, W. (2011). Signal estimation under random time-warpings and nonlinear signal alignment.
Maaten, L. V. D. (2009). Learning a parametric embedding by preserving local structure. *Artificial intelligence and statistics*, 26–41.
McInnes, L., Healy, J., & Melville, J. (2020). UMAP: Uniform manifold approximation and projection for dimension reduction. arXiv:1802.03426v3 [stat.ML].
Policar, P. G., Stražar, M., & Zupan, B. (2023). Embedding to reference t-SNE space addresses batch effects in single-cell classification. *Machine Learning*, *112*, 721–740. https://doi.org/10.1007/s10994-021-06043-1
Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters. *Bulletin of Calcutta Mathematical Society*, *37*, 81–91.
Srivastava, A., Klassen, E., Joshi, S. H., & Jermyn, I. H. (2011). Shape analysis of elastic curves in Euclidean spaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *33*(7), 1415–1428.
Srivastava, A., & Klassen, E. P. (2016). *Functional and shape data analysis*. Springer.
Srivastava, A., Wu, W., Kurtek, S., Klassen, E., & Marron, J. S. (2011). Registration of functional data using Fisher-Rao metric. arXiv:1103.3817v2 [math.ST]. Retrieved from http://arxiv.org/abs/1103.3817v2
Su, J., Kurtek, S., Klassen, E., & Srivastava, A. (2014). Statistical analysis of trajectories on Riemannian manifolds: Bird migration, hurricane tracking and video surveillance. *The Annals of Applied Statistics*, *8*(1), 530–552.
Thakoor, N., & Gao, J. (2005). Automatic video object shape extraction and its classification with camera in motion. *Proc. IEEE Int. Conf. Image Processing*, *3*, 437–440.
Thakoor, N., Gao, J., & Jung, S. (2007). Hidden Markov model-based weighted likelihood discriminant for 2-D shape classification. *IEEE Transactions on Image Processing*, *16*(11), 2707–2719.
Tucker, J. D., Wu, W., & Srivastava, A. (2013). Generative models for functional data using phase and amplitude separation. *Computational Statistics and Data Analysis*, *61*, 50–66.
van der Maaten, L. J. P., & Hinton, G. E. (2008). Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.
Vladimir, P. L. (1994). Encyclopedia of mathematics. In *(chap. Orthogonal group)*. Springer-Netherlands.
Xie, Q., & Srivastava, A. (2016). Image registration using phase amplitude separation. *Variational Methods*, 1–25.

How to cite this article: Tucker, J. D., Martinez, M. T., & Laborde, J. M. (2023). Dimensionality reduction using elastic measures. *Stat*, *12*(1), e551. https://doi.org/10.1002/sta4.551