CLASSIFICATION OF SHAPES

GREGORY J. MATTHEWS ET. AL.

1. Introduction

Statistical analysis of shape plays an important role in many areas of science such as biology (O'Higgins [1989], O'Higgins and Dryden [1993], Goo [1989]), chemistry (Dryden and Melville [2007], Czogiel and Brignell [2011]), medicine (Bookstein [1996], Brignell [2010]), bioinformatics (Green and Mardia [2006]), genetics (Horgan and Fenton [1992]), geology (Lohman [1983]) and anthropology Brophy [2014](Matthews 2017). Formal techniques for statistical shape analysis have been developed to extend classical statistical methods to shapes, such as computation of a mean shape and shape variability, extensions of Hotelling's T^2-test for inference about mean shapes, and principal components analysis of shapes (PCA), to name just a few Dryden and Mardia [1998]. Statistical shape analysis techniques generally assume that the entire shape is observed. However, in practice, there are instances where the shapes of interest are only partially observed. For example, the shape of the occlusal surface of isolated fossil teeth from the Family Bovidae is useful to biological anthropologists for identifying the taxa of specimens, which in turn is used to reconstruct paleoenvirnoments. In order to classify these isolated fossil tooth remains, scientists rely on, among other factors, the shape of the outline of the occlusal surface of the tooth to make accurate taxonomic classifications. Currently, only complete (or nearly complete) teeth are generally considered in classification analyses, however, there are numerous specimens of fractured fossil Bovid teeth that are not considered as strongly or ignored entirely in the analysis process because the full shape is not observed. There are countless examples in the missing data literature where dropping observations with missing data (i.e. complete case analysis) has been demonstrated to produce biased results, therefore, it is scientifically beneficial to have methods for including partial observations in an analysis of interest, in this example the classification of fossilized teeth from the family Bovidae.

1.1. Statistical Shape Analysis.

1.2. Landmarks.

Much of the early work in shape analysis was based on defining shapes by a series of landmarks (Bookstein [1986], Mardia and Dryden [1989], Kendall [1984], Le and Kendall [1993]) where each point

Date: June 26, 2019.

represents a meaningful location on a shape and there is a natural correspondence between points across shapes in a sample. This allows for a natural correspondence of points across shapes and greatly simplifies analysis in many cases, such as computing a mean shape or variability of a collection of shapes. The use of landmark data also allows for a natural methods for the alignment of shapes to one another using a method called Procrustes analysis Green [1952], Goodall and Mardia [1991], which aligns shapes subject to translations, rotations, and scaling. Additionally, shapes can be defined by unlabeled points where there is some true labeling, but it is unobserved. This setting is more complicated than labeled landmarks since the correspondence between points across shapes is unknown. In many situations, however, there are no reasonable choices for landmarks on the shape. Later,the concept of sliding landmarks Green, Bookstein [1997] was proposed. Sliding landmarks are the same as landmarks except that they are allowed to "slide" along the contour of the curve when being aligned with a reference set of landmarks.

1.3. **Continuous Curves1.** Rather than describing a closed shape with a discrete set of points, a function can be used:

$$\beta : \mathbb{S} \to \mathbb{R}^n$$

where $\mathbb{S}$ is the unit sphere and guarantees that the function has the same start and end point (i.e. is a closed curve). The function β is a closed curve but scalings, translations, rotations and reparameterizations of this function will usually result in different functions, which means that this framework does not offer invariance across these transformations. Rather than using this function directly, the first derivative could be used, which results in a translation invariant representation.

This representation is the foundation of the square root velocity function (SRVF) (CITE) framework, which allows invariance across tranlsation, rotation, (optional) scaling, and reparameterization. Frameworks for representing shapes that have invariance over reparameterization are said to be *elastic*.

To define the SRVF start by considering an absolutely continuous function $\beta : \mathbb{S} \to \mathbb{R}^n$. The SRVF of this function $q : \mathbb{S} \to \mathbb{R}^n$ is defined as the following:

where $\dot{\beta}(t)$ is the first derivative of the function $\beta(t)$.

A full review of other methods for shape analysis can be found in CITE(FSDA)

Sliding landmarks by Bookstein.

Section 1.2.4 in FSU disserttion is a good description of this.

Shape analysis in general.

Partial Shape matching.

2. Missing Data

Missing data problems arise in many statistical settings due to survey non-response, designed missingness in surveys (CITE), measurement error, and latent classes, to name a few examples. There are numerous methods proposed in the literaure describing methods for handling missing data in a statistically principled manner. For example, (Demspter Laird and Rubin) describes the EM algorithm for finding maximum likelihood estimates in the presence of missing data. Another method, of particular interest in this setting, for handling mising data is multiple imputation (CITE Rubin and Little, Schafer). The idea behind multiple imputation is to blah blah blah.....

EM algorithm (Depster Laird and Rubin) Multiple Imputation (Rubin and Little)

For a more thorough review of the missing data literature, see cite(HortonAndKleinman2007) (Add more).

3. Missing Data and Shapes

3.1. **13.1 - Incomplete Data.** Papers mentioned:

- Albers and Gower (2010) - Missing data in procrustes
- Mitchelson (2013) - handling occluded landmarks.
- Gunz et. al. (2009) - missing data in studies of biological evolution of bone surfaces

4. Shapes of closed curves1

Our interest is in the completion of shapes of partially observed closed curves, and their classification. This first requires us to adopt a suitable representation for the shape of a fully observed curve. We adopt a parametric representation of a closed curve by representing as an absolutely continuous function[1] $C : \mathbb{S} \to \mathbb{R}^2$, thus automatically ensuring that the curve is closed. The notion of a shape of such a curve requires invariances to transformations that represent nuisance information. Specifically, if $\Gamma := \{\gamma : S \to S \text{ is an orientation-preserving diffeomorphism}\}$ and $SO(2)$ is the rotation group in $\mathbb{R}^3$, the shape of a parameterized curve $C : D \to \mathbb{R}^2$ is defined to be the equivalence class

$$[C] := \Big\{\sigma OC(\gamma(t)) + a, \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma > 0\Big\}.$$

Thus $[C]$ is the set of all possible curves that can be obtained through a translation $(C + a)$, rotation (OC), scale change (σC), a reparametrization $(C(\gamma))$, or a combination of the transformations, of the curve C. In words,

[1]Upon identification of $\mathbb{S}$ with $[0, 1] \cong \mathbb{R}/2\pi\mathbb{Z}$, a curve $C : [0, 1] \to \mathbb{R}^2$ is absolutely continuous if and only if there exists an integrable function $g : [0, 1] \to \mathbb{R}^2$ such that $C(t) - C(0) = \int_0^t g(u)du, \forall t \in [0, 1]$.

the shape of a curve C is what is left once variations due scale, translation, rotation and reparametrisation has been accounted for.

In this instance, it is also of interest to preserve the size and shape of the curve. This simply requires removing any scale changes from the definition of the equivalence class. This size-and-shape of a parameterized $C : D \to \mathbb{R}^2$ is defined to be the equivalence class:

$$[C^\star] := \Big\{\sigma OC(\gamma(t)) + a, \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \Big\}.$$

Thus $[C^\star]$ is the set of all possible curves that can be obtained through a translation $(C + a)$, rotation (OC), a reparameterization $(C(\gamma))$, or any combination of the transofrmations of curve C, notably with scaling not included as one of the possible transformations.

A key ingredient in several classification methods (e.g. linear/quadratic discriminant analysis; kernel-based methods) for functional data is the notion of similarity or distance between curves. A popular choice is the distance induced by the $\mathbb{L}^2$ norm of a Hilbert space of square-integrable functions. However, it is well known that the $\mathbb{L}^2$ is unsuitable for comparing curves in the presence of parameterisation variability [Kurtek et al., 2012, Srivastava and Klassen, 2016]. To this end we employ a suitable representation (transformation) of a curve C that allows us to compute distances easily while accounting for the necessary invariances.

4.1. **Notation.** A closed planar curve C is always an absolutely continuous mapping $C : \mathbb{S} \to \mathbb{R}^2$, and the set of planar closed curves will be denoted by $\mathcal{C}$. The set $\mathcal{C}$ is equipped with the norm $\|x\|_{\mathbb{L}^2} := [\int_{\mathbb{S}} \|x(t)\|_2 dt]^{1/2}$ where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^2$. $SO(2)$ is the special orthogonal group of rotation matrices of $\mathbb{R}^2$, and Γ is group of orientation-preserving diffeomorphisms of $\mathbb{S}$. The set Γ_I will denote the group $\{\gamma : [0,1] \to [0,1], \gamma' > 0, \gamma(0) = 0, \gamma(1) = 1\}$.

4.2. **Square-root velocity transform.** For a detailed introduction to the transform, its properties and advantages we refer to the reader to Chapter 6 of the book by Srivastava and Klassen [2016]. Here we briefly outline the important concepts required for our purposes.

For an absolutely continuous curve $C : \mathbb{S} \to \mathbb{R}^2$ consider the transformation

$$C \mapsto \frac{C'}{\|C'\|_{\mathbb{L}^2}} =: Q_C,$$

where C' is the (vector) derivative of $C(t)$ with respect to t, and $\|\cdot\|$ is the usual Euclidean norm in $\mathbb{R}^2$.

The unique (up to translations) inverse of a square-root transformed curve Q_C is $\int_o^t Q_C(s)\|Q_C(s)\|_2 ds$. To ensure that the curves are closed we need to impose an additional constraint: $\int_{\mathbb{S}} \|C'(t)\|_2 dt = \int_{\mathbb{S}} Q_C(t)\|Q_C(t)\|_2 = 0$.

For a curve C, by taking its derivative and dividing by its length $\|C'\|_{\mathbb{L}^2}$, the transform accounts for translation and scale variabilities. Thus the image

of the set of absolutely continuous closed curves with fixed lengths l, $\{C : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|C'(t)\|_2 dt = l\}$, under the square-root transform map $C \mapsto Q_C$ is the set

$$\mathcal{Q} := \Big\{ Q_C : \int_{\mathbb{S}} \|Q_C\|_2^2 = 1, \int_{\mathbb{S}} Q_C(t)\|Q_C(t)\|_2 dt = 0 \Big\},$$

since $\|Q_C\|_{\mathbb{L}^2} = \|C'\|_{\mathbb{L}^2}^{1/2}$. Thus the set $\mathcal{Q}$ is a subset of $\mathbb{L}^2(\mathbb{S}, \mathbb{R}^2) := \{Q : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|Q(t)\|_2^2 < \infty\}$. It is referred to as the pre-shape space corresponding to the curves, since variations due to rotation and parameterization are yet to be accounted for. It is not a linear space and is a manifold [Srivastava and Klassen, 2016].

Before defining the shape space, we discuss the actions of groups Γ and $SO(2)$ on the set $\mathcal{Q}$. The set Γ of reparameterisations (or warp maps) of $\mathbb{S}$ is group with group action given by composition. Its action on $\mathcal{Q}$ is defined by $(Q_C, \gamma) \mapsto Q_C(\gamma)\sqrt{|\gamma'|}$, where γ' is the derivative of γ (see Chapters 5 and 6 Srivastava and Klassen [2016] for more details). The derivative of $\gamma : \mathbb{S} \to \mathbb{S}$ needs to be viewed as a derivative of $\gamma : [0, 1] \to [0, 1]$ based on the identification $\mathbb{S} \cong \mathbb{R}/2\pi\mathbb{Z}$, and hence $|z|$ is just the absolute value of the real number z. The action of the rotation group $SO(2)$ is defined in the usual way as the map $SO(2) \times \mathcal{Q} \to \mathcal{Q}$ with $(O, Q_C) \mapsto \{OQ_C(t) : t \in \mathbb{S}\}$.

Two important ramifications of the described framework, motivating its use in our work for analyzing shapes of curves, are the following. Under the square-root velocity framework:

1. The actions of $SO(2)$ and Γ on $\mathcal{Q}$ commute, i.e. they can be applied to a curve in any order. This ensures that their combined action is given by the product group $\Gamma \times SO(2)$.
2. The action of $\Gamma \times SO(2)$ on $\mathcal{Q}$ is by isometries: Given two curves C_1 and C_2, we have $\|OQ_{C_1}(\gamma)\sqrt{\gamma'} - OQ_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} = \|Q_{C_1} - Q_{C_2}\|_{\mathbb{L}^2}$, for every $(\gamma, O) \in \Gamma \times SO(2)$. This ensures that if two square-root velocity transformed curves are rotated and reparameterized the same way, their distance remains unchanged.

Starting with a curve C we can now define its shape to be the equivalence class or its orbit of its corresponding square-root transform:

$$[Q_C] = \text{closure}\{OQ_C(\gamma)\sqrt{\gamma'} : (\gamma, O) \in \Gamma \times SO(2)\},$$

where the closure is with respect to the norm $\|\cdot\|_{\mathbb{L}^2}$ on Q. The shape space consequently is defined as $\mathcal{Q}_s := \{[Q_C] : Q_C \in \mathcal{Q}\}$. Property (2) ensures that the metric induced by the norm $\|\cdot\|_{\mathbb{L}^2}$ on $\mathcal{Q}$ descends onto a metric d on the shape space (quotient space) $\mathcal{Q}_s$ in a natural way. Given two curves

C_1 and C_2, the shape distance between them is defined as

$$
\begin{aligned}
(1) \quad d(C_1, C_2) &:= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \|Q_{C_1} - OQ_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} \\
&= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \Big[\int_{\mathbb{S}} \Big\|Q_{C_1}(t) - OQ_{C_2}(\gamma(t))\sqrt{\gamma'(t)}\Big\|_2^2 dt\Big]^{1/2} \\
&= \inf_{(\gamma,O)\in\Gamma\times SO(2)} \|OQ_{C_1}(\gamma)\sqrt{\gamma'} - Q_{C_2}\|_{\mathbb{L}^2}.
\end{aligned}
$$

The symmetry with respect to the action either on Q_{C_1} or Q_{C_2} is an attractive feature and will be used profitably in the sequel.

5. Preserving scale

For an absolutely continuous curve $C : \mathbb{S} \to \mathbb{R}^2$ consider the transformation

$$C \mapsto C' =: Q^\star_C = Q_C\|C'\|_{\mathbb{L}^2}$$

where C' is the (vector) derivative of $C(t)$ with respect to t.

In this case, the unique (again, up to translations) inverse of this curve $Q^\star_C$ is $\int_0^t Q^\star_C(s)ds$. To ensure that the curves are closed we need to impose an additional constraint: $\int_{\mathbb{S}} Q^\star_C(t) = 0$. Thus the image of the set of absolutely continuous closed curves , $\{C : \mathbb{S} \to \mathbb{R}^2 : \int_{\mathbb{S}} \|C'(t)\|_2 dt = l\}$, under the scale preserving transform map $C \mapsto Q^\star_C$ is the set

$$\mathcal{Q} := \Big\{Q^\star_C : \int_{\mathbb{S}} Q^\star_C(t)dt = 0\Big\},$$

Starting with a curve C we can now define its size and shape to be the equivalence class or its orbit of its corresponding velocity transform:

$$[Q^\star_C] = \text{closure}\{OQ^\star_C(\gamma)\sqrt{\gamma'} : (\gamma, O) \in \Gamma \times SO(2)\},$$

where the closure is with respect to the norm $\|\cdot\|_{\mathbb{L}^2}$ on $Q^\star$.

In the square-root velocity framework, dividing the velocity function by its norm gives us a function that is of norm one. This implies that it is a point on the unit sphere. In the finite-dimensional vecotr set-up this amounts to keeping just the direction of the vectors and discarding all other information and the distance between two curves can then be viewed as the angle between these two vectors. However, when preserving scale this defintion of distance will need to be modified as both the angle AND magnitude of the vectors are both of interest.

Therefore, a modified defintion of distance that preserves scaling between two curves will be defined:

$$
\begin{aligned}
(2)\quad d^{\star}(C_1, C_2) &:= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \|Q^{\star}_{C_1} - OQ^{\star}_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2} \\
&= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \Big[\int_{\mathbb{S}} \Big\|Q^{\star}_{C_1}(t) - OQ^{\star}_{C_2}(\gamma(t))\sqrt{\gamma'(t)}\Big\|_2^2\, dt\Big]^{1/2} \\
&= \inf_{(\gamma, O)\in\Gamma\times SO(2)} \|OQ^{\star}_{C_1}(\gamma)\sqrt{\gamma'} - Q^{\star}_{C_2}\|_{\mathbb{L}^2}.
\end{aligned}
$$

6. Missingness Function

Now consider a function, R, defined over the set $\mathbb{S}$, where $R(t) = 1$ if the shape function is observed at $t \in \mathbb{S}$ and $R(t) = 0$ otherwise.

Specifically,

$$R : \mathbb{S} \to \{0, 1\}^2$$

In $\mathbb{R}^2$, for a given point t we have $C(t) = (x_t, y_t)$ and the defintion of the function R allows these two be missing individually (e.g. x_t could be observed and y_t is missing or y_t could be observed and x_t is missing), in our setting both the x_t and y_t values are either both observed or both missing and so specifically in this case the function is of the following form:

$$R : \mathbb{S} \to \{\{0, 0\}, \{1, 1\}\}$$

Now for a specific curve C in the equivalence class $[C]$ (or $[C^{\star}]$), we can define the function R over $\mathbb{S}$. However, for every reparameterization of C in the equivalence class $[C]$ (or $[C^{\star}]$), there is a corresponding reparameterization of R. Also, note that $R^{-1}(1)$ is the subset of domain $\mathbb{S}$ such that the coordinates of the shape are observed for every point in this set. Further, here we only the case where $R^{-1}(1)$ be a connected subset of $\mathbb{S}$ is considered, though in general, this does not have to be the case. This prevents us from considering partially observed shapes with multiple "gaps" (though this is an interesting idea for future work).

For a given shape, there are many possible reparameterizations of the missingness function, and an equivalence class can be defined for the missingness function across reparametertizations as follows:

$$[\mathcal{R}] = \{R(\gamma(t)), \gamma \in \Gamma\}$$

.

Note that for every curve C and corresponding R, in order to maintain the correct relationship, if C is reparameterized by $\gamma \in \Gamma$ then R must be reparameterized using the same γ that acted on C.

Given $\mathcal{R}$, we can break $[C^{\star}]$ into the observed and missing part of the size and shape as follows:

$$[C^{\star}]_{obs} = \{OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2 | R(\gamma(t)) = 1\}$$

and

$$[C^\star]_{mis} = \{OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2 | R(\gamma(t)) = 0\}$$

.

Likewise for $[C]$ we can do the same thing, but also consider scaling:

$$[C]_{obs} = \{\sigma OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma 0 | R(\gamma(t)) = 1\}$$

and

$$[C]_{mis} = \{\sigma OC(\gamma(t)) + a, R(\gamma(t)), \gamma \in \Gamma, O \in SO(2), a \in \mathbb{R}^2, \sigma 0 | R(\gamma(t)) = 0\}$$

.

6.1. **Completing the curve.** Suppose we are given closed planar curves $C_j^p : \mathbb{S} \to \mathbb{R}^2, j = 1 \ldots, n_p$ each of which has been observed only on a region $\{t : R(\gamma(t) = 1\} \subset \mathbb{S}$, where $\mathbb{S}$ is the unit circle in $\mathbb{R}^2$. We assume that the curves are absolutely continuous. Additionally, a training sample $\{(y_i, C_i^f), 1, \ldots, n_f\}$ consisting of class labels $y_i \in \{0, 1, \cdots G\}$ and $C_i^f : \mathbb{S} \to \mathbb{R}^2$, fully observed on a common domain $\mathbb{S}$, is provided. The set of fully observed curves are elements of the set $\mathcal{C}^f$; denote by $\mathcal{C}^p$ the set of all partially observed curves.

The problem at hand is to model the shape (and size) of and complete each partially observed curve C_j^p and assign it to one of G groups. We consider two approaches: one based on a variational formulation, and the other using kernel-based classifier. Before describing our approaches, some comments on the set Γ are in order.

6.2. **The set of reparameterizations Γ.** Elements of the group Γ of diffeomorphisms of $\mathbb{S}$ can be viewed in the following manner. The unit circle $\mathbb{S}$ can be identified with the quotient group $\mathbb{R}/2\pi\mathbb{Z} \cong [0, 1]$. Through this identification, every continuous mapping $\beta : \mathbb{R} \to \mathbb{R}$ induces a continuous mapping of $\mathbb{S}$ onto itself such that $\beta(t + 1) = \beta(t) + 1$ for all $t \in \mathbb{R}$. If β is monotone increasing, we say that the induced map on $\mathbb{S}$ is orientation-preserving (based on a choice of clockwise or anti-clockwise orientation).

Consider now the set

$$\Gamma_{\mathbb{R}} := \{\beta : \mathbb{R} \to \mathbb{R} : \beta(t + 1) = \beta(t) + 1, \text{ continuous and increasing}\}$$

.

Each member β of $W_{\mathbb{R}}$ induces a warp map $\tilde{\beta} : \mathbb{S} \to \mathbb{S}$ with $\tilde{\beta}(e^{2\pi i t}) = e^{2\pi i \beta(t)}$, where β is referred to as the lift of $\tilde{\beta}$. This β satisfies $\beta(t+1) = \beta(t)+1$ for all $t \in [0, 1]$, and consequently we have, for $t \in [0, 1]$, $\beta(t) = \gamma(t) + c$, where γ is a warp map of $[0, 1]$ and $c \in (0, 1]$ (through the identification of $[0, 1]$ with $\mathbb{R}/2\pi\mathbb{Z}$). This procedure can be viewed as one that produces a warp map of $\mathbb{S}$ by 'unwrapping' $\mathbb{S}$ at a chosen point s and generating a warp map of $[0, 1]$. If $\Gamma_I := \{\gamma : [0, 1] \to [0, 1] : \gamma' > 0, \gamma(0) = 0, \gamma(1) = 1\}$ is

the group of diffeomorphisms of $[0,1]$, the map $\Gamma \mapsto \mathbb{S} \times \Gamma_I$ is a bijection [2]. We will hence employ the product group $\mathbb{S} \times \Gamma_I$ in place of Γ. This ensures that the domain of each curve C_j^p and C_i^f can be identified with $[0,1]$ upon unwrapping the circle.

6.3. **Curve Completion1.** Out goal is to complete $[C]_{obs}$ or $([C^\star]_{obs})$ by randomly drawing from the distribution on the missing part of the shape conditional on the observed part. Specifically, for a fixed translation, rotation, scaling, and parameterization, we seek to approximate the distribution of $C_{mis}|C_{obs}$. If we $X_j \sim C_{mis}|C_{obs}$ to be a random draw from this distribution, that is a $X_j \in \mathcal{X}_j$ is an open curve that is a plausible completion of the curve C_{obs} and $\mathcal{X}_j$ is defined as

$$\mathcal{X}_j := \{X : \gamma^{-1}(R^{-1}(t)) \to \mathbb{R}^2\}$$
$$t \in \gamma^{-1}(R^{-1}(1))$$

6.4. **Curve completion.** The observed region $\mathcal{R}_j$ associated with a partially observed curve C_j^m is the subinterval $[0, t_j]$ with $t_j < 1$ for all $j = 1, \ldots, N$. Suppose that $C_j^m(0) = \mathbf{a}_j := (a_{1j}, a_{2j})^T$ and $C_j^m(t_j) = \mathbf{b}_j := (b_{1j}, b_{2j})^T$. Then the set of curves comprising the missing segment of curve C_j^m is

$$\mathcal{X}_j := \{X : [t_j, 1] \to \mathbb{R}^2 : X(t_j) = \mathbf{b}, X(1) = \mathbf{a}\}.$$

For a partially observed $C_j^m : [0, t_j]$ and an $X \in \mathcal{X}_j$ with $j = 1, \ldots, n$, define its completion to be the concatenated closed curve

$$C_j \circ X(t) := C_j^m(t)\mathbb{I}_{t\in[0,t_j]} + X(t)\mathbb{I}_{(t_j,1]}.$$

Denote by $Q_{C_j^m \circ X}$ the square-root transform of $C_j^m \circ X$. For each $j = 1, \ldots, n$, let $\Theta_j := \mathbb{S} \times \Gamma_I \times SO(2) \times \mathcal{X}_j$, and recall that $\mathcal{C}^o$ and $\mathcal{C}^m$ denote the sets of fully and partially observed curves, respectively. For a fixed $C \in \mathcal{C}$, for each $j = 1, \ldots, n$ define the cost functional $\Phi_{\theta_j} : \mathcal{C}^m \times \mathcal{C} \to \mathbb{R}$ by

$$\begin{aligned}\Phi_{\theta_j}(C_j^m, C) &:= d^2(C_j^m \circ X_j, OC(\gamma)), \quad \theta_j \in \Theta_j \\ &= \inf_{(s,\gamma,O)\in\mathbb{S}\times\Gamma_I\times SO(2)} \|Q_{C_j^m\circ X_j}, OQ_C(\gamma)\sqrt{\gamma'}\|^2_{\mathbb{L}^2}.\end{aligned}$$

The optimal shape completion of a partially observed curve $Q_{C_j^m}, j = 1, \ldots, n$ is $Q_{C_j^m \circ X_j^*}$, where X_j^* is obtained from the solution set of:

$$\theta_j^* := (s_j^*, \gamma_j^*, O_j^*, X_j^*) = \operatorname*{argmin}_{\theta_j\in\Theta_j} \Phi_{\theta_j}(C_j^m, C). \tag{3}$$

In the expression above s_j^* corresponds to the optimal point at which $\mathbb{S}$ was unwrapped in order to identify Γ with $\mathbb{S} \times \Gamma_I$; $\gamma_j^* : [0,1] \to [0,1]$ and O_j^*

[2]Technically, this is not a bijection since for a $\gamma \in \Gamma$, the corresponding $\beta \in \Gamma_I$ has a jump discontinuity at the point $t_c \in [0,1]$ where $\beta(t_c) + c = 1$. This can be circumvented by assuming that the members of Γ and Γ_I are absolutely continuous (as opposed to diffeomorphims); then the map between to the sets is bijective a.e.

represents the optimal reparameterization of the curve $C_j^m \circ X_j^*$. The use of a valid distance on the quotient shape space $\mathcal{Q}_s$ allows us to apply the shape transformations on $Q_{C_i}, i = 1 \ldots, n$ in the variational problem with introducing any arbitrariness. The (product) group structure of $\mathbb{S} \times \Gamma_I \times SO(2)$, and its action on $\mathcal{Q}$, ensures that if the transformations were to have been applied to $Q_{C_j^m \circ X^*}$, the resulting solution set would instead contain the corresponding group inverses. (INSERT ILLUSTRATIVE FIGURE).

7. Results

Describe the simulation study.

GRADY'S DESCRIPTION: Required scripts and data are loaded into the R environment from external files. R package fdasrvf is used for requisite shape distance functions, namely for resampling points around a curve and finding the distance between two closed shapes.

A loop begins for both the individual teeth and their left and right halves. The two halves are generated at this step via finding the two points which are separated with the smallest Euclidean distance, provided that these two points are roughly twenty indices away in either direction, ensuring that teeth will be split roughly at the midpoint, where the X would be on a figure-eight.

Teeth halves are stored as "partial shape" alongside an indicator for whether it is the first or second half of the tooth in question, and the complete shape list is amended to not have the tooth in question present so that it would not be matched with itself.

Using the partial shape distance function, distances between the given half of the given tooth and the remaining completed teeth are taken. From the K smallest distances, a sample of size M with replacement is taken. For each of unique teeth in M, an imputed version of the tooth is then created: First, the partial shape is closed via a series of tightly-spaced points connecting the starting and ending point. Both the partial and complete shape are then resampled with splinesfor the partial shape, the straight segment connecting the start and end is done separately from this, in order to avoid spline-based issues where the shape is predicted to curve inwards in a manner not reflective of the true shape.

The partial and whole shape are then compared in order to find the portion of the complete shape which the observed partial shape most adequately corresponds to. This is done by comparing a series of starting segments of the completed shape to the observed partial shape. For each of the subsets, the partial complete shape is scaled and centered, then a best rotation onto the observed partial shape is made. From this, an error term is found, and the segment with the minimum error term is selected as where the closed partial shape will then be overlaid. From here, the straight line segment from earlier is then gradually morphed so that it more and more adequately

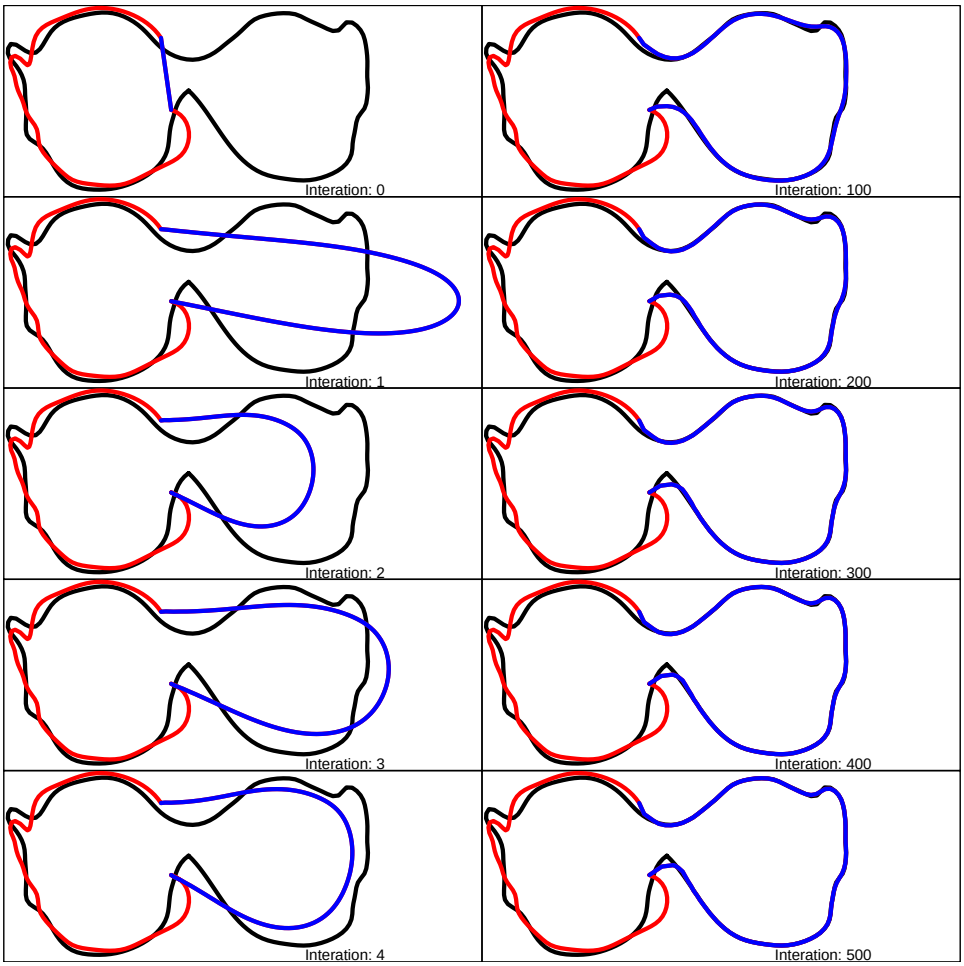

FIGURE 1. Test

fits the curvature of the complete shape. After NUMBER iterations, a final version of the imputed partial tooth is obtained.

Classification is done using KNN, where the distances between the imputed tooth and the whole teeth are measured, and the most common class designation amongst the TEN ? teeth with the smallest distance measure is used to predict the class designation of the imputed tooth. A total of M of these predictions is then obtained based on the sampling from the K smallest distances earlier. Further, KNN classification based off the earlier, partial distances is used a means of assessing the performance of classifying based upon imputation.

7.1. **Completing the curves.**

References

Growth curve models for correlated traingular shapes, 1989. Interface Foundation, Fairfax Station.

Bookstein. Size and shape spaces for landmark data in two dimensions (with discussion). *Statistical Science*, 1:181–242, 1986.

F.L. Bookstein. Biometrics, biomathematics and the morphometric synthesis. *Bulletin of Mathematical Biology*, 58(2):313–365, 1996.

F.L. Bookstein. Landmark methods for forms without landmarks: morphometrics of group differences in outline shape. *Medical Image Analysis*, 1 (3):225–243, 1997.

Dryde I.L. Gattone S.A. et. al. Brignell, C.J. Surface shape analysis, with an application to brain surface asymmetry in schizophrenia. *Biostatistics*, 11:609–630, 2010.

de Ruiter D.J. Athreya S. DeWitt T.J. Brophy, J.K. Quantitative morphological analysis of bovid teeth and its implications for paleoenvironmental reconstructions in south africa. *Journal of Archaeological Science*, 41: 376–388, 2014.

G. Changy and A. Roche. Adaptive estimation in the functional nonparametric regression model. *Journal of Multivariate Analysis*, 146:105–118, 2016.

Dryden I.L. Czogiel, I. and C.J. Brignell. Bayesian matching of unlabeled marked point sets using random fields, with application to molecular alignment. *Annals of Applied Statistics*, 5:2603–2629, 2011.

Hirst J.D. Dryden, I.L. and J.L. Melville. Statistical analysis of unlabeled points sets: comparing molecules in chemoinformatics. *Biometrics*, 63(1): 237–251, 2007.

I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis*. John Wiley & Sons, Chichester, 1998.

F. Ferraty and P. Vieu. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer Series in Statistics, 2006.

C.R. Goodall and K.V. Mardia. A geometrical derivation of the shape density. *Advances in Applied Probability*, 23:496–514, 1991.

B.F. Green. The othogonal approximation of an oblique structure in factor analysis. *Psychometrika*, 17:492–440, 1952.

P.J. Green and K.V. Mardia. Bayesian alignment using hierarchical models, with applications in protein bioinformatics. *Biometrika*, 93(2):235–254, 2006.

W.D.K. Green. The thin-plate spline and images with curving features. In Gill C.A. Dryden I.L. Mardia, K.V., editor, *Image Fusion and Shape Variability*.

Creasey A. Horgan, G.W. and B. Fenton. Superimposing two-dimensional gels to study genetics variation in malaria parasites. *Electrophoresis*, 13: 871–875, 1992.

D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

S. Kurtek, A. Srivastava, E. Klassen, and Z. Ding. Statistical modeling of curves using shapes and related features. *Journal of the American Statistical Association*, 107(499):1152–1165, 2012.

S. Lahiri, D. Robinson, and E. Klassen. Precise matching of PL curves in $\mathbb{R}^n$ in the Square Root Velocity framework. *Geometry, Imaging and Computing*, pages 133–186, 2015.

H. Le and D.G. Kendall. The riemannian structure of euclidean shape spaces: A novel environment for statistics. *The Annals of Statistics*, 31 (3), 1993.

W. V. Li and Q. M. Shao. Gaussian process: inequalities, small ball probabilities and applicaitons. *Stochastic process: theory and methods*, 19: 533–597, 2001.

G.P. Lohman. Eigenshape analysis of micro fossils: a general morphometric procedure for describing changes in shape. *Mathematical Geology*, 15: 659–672, 1983.

K.V. Mardia and I.L. Dryden. Shape distributions for landmark data. *Advances in Applied Probability*, 21:742–755, 1989.

A. Mas. Lower bound in regression for functional data by representation of small ball probabilities. *Electronic Journal of Statistics*, 6:1745–1178, 2012.

P. O'Higgins. *A morphometric study of cranial shape in Hominoidea*. PhD thesis, University of Leeds, 1989.

P. O'Higgins and I.L. Dryden. Sexual dimorphism in hominoids: further studies of craniofacial shape differences in *pan*, *gorilla*, *pongo*. *Journal of Human Evolution*, 25:183–205, 1993.

A. Srivastava and E. P. Klassen. *Functional and Shape Data Analysis*. Springer-Verlag, New York, 2016.

8. Appendix

The asymptotic properties of the estimator $\hat{\pi}_N(C)$ are intimately related to the (shifted) small-ball probability of the process $\mathbf{C}$ under the metric d:

$$\begin{aligned}\phi(Q_C, h_N) &:= \mathbb{P}(d(\mathbf{C}, C) < h_N), \quad C \in \mathcal{C}, h_N > 0 \\ &= \mathbb{P}\left(\inf_{(O,\gamma)\in SO(2)\times\Gamma} \|\mathbf{Q_C} - OQ_C(\gamma)\sqrt{|\gamma'|}\|_{\mathbb{L}^2} < h_N \right),\end{aligned}$$

where $\mathbf{Q_C}$ is the (pathwise) square-root transform of the random curve $\mathbf{C}$. For a detailed account of small-ball and shifted small-ball probabilities of processes, and their role in kernel-based estimators involving functional data, see Ferraty and Vieu [2006], Changy and Roche [2016], Mas [2012], Li and Shao [2001] and references therein. To the best of our knowledge results regarding small-ball probabilities are available only for processes with values in linear function spaces (e.g. Hilbert space with $\mathbb{L}^2$ norm or Banach

space with supremum norm). The process $\mathbf{Q_C} = \frac{\mathbf{C}'}{\|\mathbf{C}'\|_{\mathbb{L}^2}}$ takes values in an infinite-dimensional manifold (pre-shape space) $\mathcal{Q} \subset \mathbb{L}^2(\mathbb{S}, \mathbb{R}^2)$, defined earlier. Moreover, d is on the quotient shape space, and is defined between orbits of $\mathbf{Q_C}$ and Q_C (with respect to the action of Γ and $\mathbb{S}$).

We can view the class probability π (conditional expectation of $\mathbf{y}$ given $\mathbf{C}$) as a map from $\mathcal{C}$ to $[0,1]$. We make the following assumptions.

A1. The kernel K is supported on $[0,1]$ and bounded away from 0 and 1.
A2. The bandwidth $h_N \to 0$ as $N \to \infty$.
A3. $\phi(C, h_N) > 0$ for every $h_N > 0$ with $N\phi(C, h_N) \to \infty$ as $N \to \infty$.
A4. The conditional probability $\pi : \mathcal{C} \to [0,1]$ is α-Lipschitz, i.e. there exists a $\lambda > 0$ such that for every $\tilde{C} \in \mathcal{C}$, $|\pi(C) - \pi(\tilde{C})| \leq \lambda \|C - \tilde{C}\|^\alpha$.

The following result relates the behaviour of $\phi(Q_C, h_N)$ to the small-ball probability $\phi(C, h_N) = \mathbb{P}(\|\mathbf{C} - C\| < h_N)$ of the process $\mathbf{C}$ (taking values in a linear space), and establishes consistency and rate of convergence of the estimate $\hat{\pi}_N$, as $N \to \infty$.

Theorem 1. *Under assumptions A1-A3, $\phi(Q_C, h_N) > 0$ for every $h_N > 0$ and $N\phi(Q_C, h_N) \to \infty$ as $N \to \infty$. As a consequence, $N \to \infty$:*

1. *$\hat{\pi}_N$ converges in probability to π;*
2. *$\hat{\pi}_N - \pi = O_{\mathbb{P}}(h_N^\beta)$.*

Proof. The key argument is to demonstrate that $\phi(Q_C, h_N) > 0$ and $N\phi(Q_C, h_N) \to \infty$ under assumptions A1-A3. Proofs of consistency and rate of convergence follow using almost identical arguments as in the proofs of Theorems 6.1 (p. 63) and 8.2 (p. 123) of Ferraty and Vieu [2006], and are omitted.

The shifted small-ball probability satisfies

$$\begin{aligned}
\phi(Q_C, h_N) &= \mathbb{P}(d(\mathbf{C}, C) < h_N), \quad C \in \mathcal{C}, h_N > 0 \\
&= \mathbb{P}\left(\inf_{(O,\gamma) \in SO(2) \times \Gamma} \|\mathbf{Q_C} - O Q_C(\gamma) \sqrt{|\gamma'|}\|_{\mathbb{L}^2} < h_N \right) \\
&= \mathbb{P}\left(\inf_{(O,\gamma) \in SO(2) \times \Gamma} \|\mathbf{Q_C} - O Q_C(\gamma) \sqrt{|\gamma'|}\|_{\mathbb{L}^2}^2 < h_N^2 \right) \\
&= \mathbb{P}\left(\|\mathbf{Q_C} - \tilde{O} Q_C(\tilde{\gamma}) \sqrt{|\tilde{\gamma}'|}\|_{\mathbb{L}^2}^2 < h_N^2 \text{ for some } (\tilde{O}, \tilde{\gamma}) \in SO(2) \times \Gamma \right)
\end{aligned}$$

under the assumption that the (unique) infimum is attained at $(\tilde{O}, \tilde{\gamma}) \in SO(2) \times \Gamma$. The infimum will be attained if the orbits of the elements of $\mathcal{Q}$ are closed under the action of the product group $SO(2) \times \Gamma$. While the orbit under $SO(2)$ is closed, the same isn't generally true for Γ. A technical adjustment in the definition of Γ rectifies this; for details see Lahiri et al. [2015].[3] Observe that $\{\omega \in \Omega : \|\mathbf{Q_C}(\omega) - Q_C(\omega)\|_{\mathbb{L}^2}^2 < h_N^2\} \subseteq \{\omega \in \Omega :$

[3]The group Γ needs to extended to the semi-group $\tilde{\Gamma}$ that allows for derivatives to be 0 at some points. We could then alter the action of $\tilde{\Gamma}$ on $\mathcal{Q}$ to be just the composition $Q_C(\gamma), \gamma \in \tilde{\Gamma}$, and the arguments in the proof remain valid.

$\|\mathbf{Q_C}(\omega) - \tilde{O}Q_C(\omega)(\tilde{\gamma})\sqrt{\dot{\tilde{\gamma}}}\|^2_{\mathbb{L}^2} < h_N^2$ for some $(\tilde{O}, \tilde{\gamma}) \in SO(2) \times \Gamma\}$. This can be seen by noting that the relationship is trivially true if $\tilde{\gamma}$ is the identity map in Γ. Thus we have that

$$\phi(Q_C, h_N) \geq \mathbb{P}\left(\|\mathbf{Q_C} - Q_C\|^2_{\mathbb{L}^2} < h_N^2\right).$$

Consider the square-root map $\mathfrak{C} : \mathcal{C} \to \mathcal{Q}$, $\mathfrak{C}(C) = Q_C$. The map is a bijection between the two spaces [Srivastava and Klassen, 2016]. It is however a complicated map between two Hilbert spaces. Instead of directly dealing with the map in order to relate $\phi(Q_C, \cdot)$ to $\phi(C, \cdot)$, we adopt the following strategy.

Denote by $\mathbb{S}^\infty$ the unit sphere in $\mathbb{L}^2(\mathbb{S}, \mathbb{R}^2)$. Note that the pre-shape space $\mathcal{Q} = \left\{Q_C : \int_{\mathbb{S}} \|Q_C\|_2^2 = 1, \int_{\mathbb{S}} Q_C(t)\|Q_C(t)\|_2 dt = 0\right\}$ is proper subset of $\mathbb{S}^\infty$. Thus we can view $\mathbf{C}(C) = \mathbf{Q_C}$ and $\mathfrak{C}(C) = Q_C$ as random elements taking values in $\mathbb{S}^\infty$. Consider the radial map in a Hilbert space $\mathfrak{R} : \mathcal{C} \to \mathbb{S}^\infty$ given by

$$\mathfrak{R}C = \begin{cases} C & \text{if } \|C\| \leq 1; \\ \frac{C}{\|C\|} & \text{if } \|C\| > 1. \end{cases}$$

The map $\mathfrak{R}$ is the unique metric projection of $\mathcal{C}$ onto $\mathbb{S}^\infty$ and is 1-Lipschitz and nonexpansive, that is,

$$\|\mathfrak{R}C_1 - \mathfrak{R}C_2\| \leq \|C_1 - C_2\|, \quad C_1, C_2 \in \mathcal{C}.$$

Since the image of $\mathfrak{C}$ (i.e. $\mathcal{Q}$) is contained in the image of $\mathfrak{R}$ (i.e. $\mathbb{S}^\infty$), and the noting that $\mathfrak{C}$ is bijective and $\mathfrak{R}$ is the unique projection on $\mathbb{S}^\infty$, for $C \in \mathcal{C}$ we necessarily have $\mathfrak{C}(C) = \mathfrak{R}(C)$.

From the definitions of the maps $\mathfrak{C}$ and $\mathfrak{R}$, from equation (8) we have,

$$\begin{aligned} \phi(Q_C, h_N) &\geq \mathbb{P}\left(\|\mathfrak{C}(\mathbf{C}) - \mathfrak{C}(C)\|^2_{\mathbb{L}^2} < h_N^2\right) \\ &= \mathbb{P}\left(\|\mathfrak{R}(\mathbf{C}) - \mathfrak{C}(C)\|^2_{\mathbb{L}^2} < h_N^2\right) \\ &\geq \mathbb{P}\left(\|\mathbf{C} - C\|^2_{\mathbb{L}^2} < h_N^2\right) \\ &= \phi(C, h_N), \end{aligned}$$

since for a fixed N, $\{\omega \in \Omega : \|\mathbf{C}(\omega) - C\|_{\mathbb{L}^2} < h_N\} \subseteq \{\omega \in \Omega : \|\mathfrak{R}(\mathbf{C}(\omega)) - \mathfrak{R}(C)\|_{\mathbb{L}^2} < h_N|\}$ with $\mathfrak{R}$ being 1-Lipschitz. This completes the proof. □