CLASSIFICATION OF SHAPES

GREGORY J. MATTHEWS

1. Introduction

1.1. **Statistical Shape Analysis.**

1.2. **Landmarks.**

1.3. **Continuous Curves.** 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 curves

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 $\Gamma := \{\gamma : S \to S$ 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

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

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, 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:

---

[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 $C(t) - C(0) = \int_0^t g(u)du, \forall t \in [0, 1]$.

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

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

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

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

$$(1) \quad d(C_1, C_2) := \inf_{(\gamma, O) \in \Gamma \times SO(2)} \|Q_{C_1} - OQ_{C_2}(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2}$$

$$= \inf_{(\gamma, O) \in \Gamma \times SO(2)} \Big[\int_{\mathbb{S}} \Big\| Q_{C_1}(t) - OQ_{C_2}(\gamma(t))\sqrt{\gamma'(t)} \Big\|_2^2 \, dt \Big]^{1/2}$$

$$= \inf_{(\gamma, O) \in \Gamma \times SO(2)} \|OQ_{C_1}(\gamma)\sqrt{\gamma'} - Q_{C_2}\|_{\mathbb{L}^2}.$$

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

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

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 $Q_C^\star$ is $\int_0^t Q_C^\star(s)ds$. To ensure that the curves are closed we need to impose an additional constraint: $\int_{\mathbb{S}} Q_C^\star(t) = 0$. 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 $C \mapsto Q_C^\star$ is the set

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

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:

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

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:

$$(2) \quad d^\star(C_1, C_2) := \inf_{(\gamma, O) \in \Gamma \times SO(2)} \|Q_{C_1}^\star - OQ_{C_2}^\star(\gamma)\sqrt{\gamma'}\|_{\mathbb{L}^2}$$

$$= \inf_{(\gamma, O) \in \Gamma \times SO(2)} \Big[\int_{\mathbb{S}} \Big\| Q_{C_1}^\star(t) - OQ_{C_2}^\star(\gamma(t))\sqrt{\gamma'(t)} \Big\|_2^2 dt \Big]^{1/2}$$

$$= \inf_{(\gamma, O) \in \Gamma \times SO(2)} \|OQ_{C_1}^\star(\gamma)\sqrt{\gamma'} - Q_{C_2}^\star\|_{\mathbb{L}^2}.$$

When you divide by the norm, the function is of norm one, and hence a point on the unit sphere of radius in L2 function space–in the finite-dimensional vector setup this amounts to keeping just the direction of a vector and discarding all other information.

6. Curve completion and classification

Suppose we are given closed planar curves $C_j^m : \mathbb{S} \to \mathbb{R}^2, j = 1 \ldots, n$ each of which has been observed only on a region $\mathcal{R}_j \subset \mathbb{S}$, where $\mathbb{S}$ is the unit circle in $\mathbb{R}^2$. We assume that the curves are absolutely continuous.

Additionally, a training sample $\{(y_i, C_i^o), 1, \ldots, N\}$ consisting of class labels $y_i \in \{0,1\}$ and $C_i^o : \mathbb{S} \to \mathbb{R}^2$, fully observed on a common domain $\mathbb{S}$, is provided. The set of fully observed curves are elements of the set $\mathcal{C}$; denote by $\mathcal{C}^m$ the set of all partially observed curves.

The problem at hand is to model the shape of and complete each partially observed curve C_j^m 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.

6.1. **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 $\tilde{\beta}(e^{2\pi it}) = e^{2\pi i\beta(t)}$, 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^m and C_i^o can be identified with $[0,1]$ upon unwrapping the circle.

6.2. **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

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

---

[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.

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

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

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:

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

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^* 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).

6.3. Classification.
We outline two approaches for classification of partially observed curves $C_j^m, j = 1, \ldots, N$.

6.3.1. *Combining completion and classification.* The variational formulation for the completion of each curve C_j^m with respect to a fixed curve C can be augmented to address the classification task in the following manner. Using the training sample $\{(y_i, C_i^o), 1, \ldots N\}$ partition the set $\{C_i^o, i = 1, \ldots, N\}$ into $\bigcup_{g \in G}\{C_{ig}^o, i = 1 \ldots, N_g\}$ with $N_1 + \cdots + N_g = N$. Recall that the shape space of a given set of curves is the quotient metric space with the metric d in (2). Such a structure allows us to define the sample Fréchet mean shape curve for a given set of curves. Formally, for each $g \in G$, consider the sample Fréchet functional define on the set $\mathcal{C}$ as

$$F_g : C \to \mathbb{R}, \quad C \ni C \mapsto F_g(C) := \sum_{i=1}^{N_g} d^2(C, C_i^0).$$

The (local) minimizer of F_g is referred to as the Fréchet mean set, since d is a distance between two equivalence classes. We then select a member $\hat{M}_g : \mathbb{S} \to \mathbb{R}^2$ from set this and refer to it as the Fréchet mean of the group $\{C_{ig}^o, i = 1 \ldots, N_g\}$ with $g = 1, \ldots, G$.

The variational classifier based on the augmented optimization problem in (3) is defined by the following rule:

Assign C_j^m to group g^* where $(g^*, \theta_j^*) = \underset{g \in \{1,\ldots,G\} \times \theta_j \in \Theta_j}{\mathrm{argmin}} \; \Phi_{\theta_j}(C_j^m, \hat{M}_g)$.

The classifier assigns C_j^m to the group whose Fréchet mean is closest to C_j^m. The advantage of this approach lies in the fact that completion and classification of partially observed curves are unified under the same metric (distance), and in a certain sense carried out simultaneously.

6.3.2. *Kernel-based classifier.* Consider curves $C_j^m \circ X_j : \mathbb{S} \to \mathbb{R}^2, j = 1\ldots,n$ that have been completed using the variational formulation in (3). For each curve $C_j^m \circ X_j$ a nonparametric estimate of the (conditional) group probabilities $\pi_g := P(y_j = g | C_j^m \circ X_j), g = 1,\ldots,G$ can be constructed using the distance d on the shape space of curves. Using the estimate a curve $C_j^m \circ X_j$ is assigned to the group with the largest probability. We eschew the cumbersome notation involving the completed curves and outline the methodology for a generic curve $C : \mathbb{S} \to \mathbb{R}^2$ to belong to two groups based on the training sample; hence $g = 1, 2$ with labels 0 and 1. Extension to more than two groups is routine.

Assume that training sample $\{(y_i, C_i^0, i = 1\ldots,N\}$ consists of independent realizations of the random element $(\mathbf{y}, \mathbf{C})$ taking values in $\{0, 1\} \times \mathcal{C}$. For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, the random element $\mathbf{C}$ is the mapping $(t, \omega) \mapsto \mathbf{C}(t, \omega)$. Thus $\mathbf{C}$ is an $\mathbb{R}^2$-valued stochastic process $\{\mathbf{C}(t), t \in \mathbb{S}\}$. The kernel-based estimate (see for e.g. Chapter 8 of Ferraty and Vieu [2006]) of the probability of assignment for a curve C to group 1, $\pi(C)$, is given by

$$(4) \qquad \hat{\pi}_N(C) = \frac{\sum_{i=1}^{N} y_i K_{i,h_N}}{\sum_{i=1}^{N} K_{i,h_N}},$$

where $K_{i,h_N} := K(h_N^{-1} d(C, C_i^o))$ for a given kernel K and a positive bandwidth sequence h_N. The distance d is defined on the quotient space of shape curves in (2).

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 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

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

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

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.
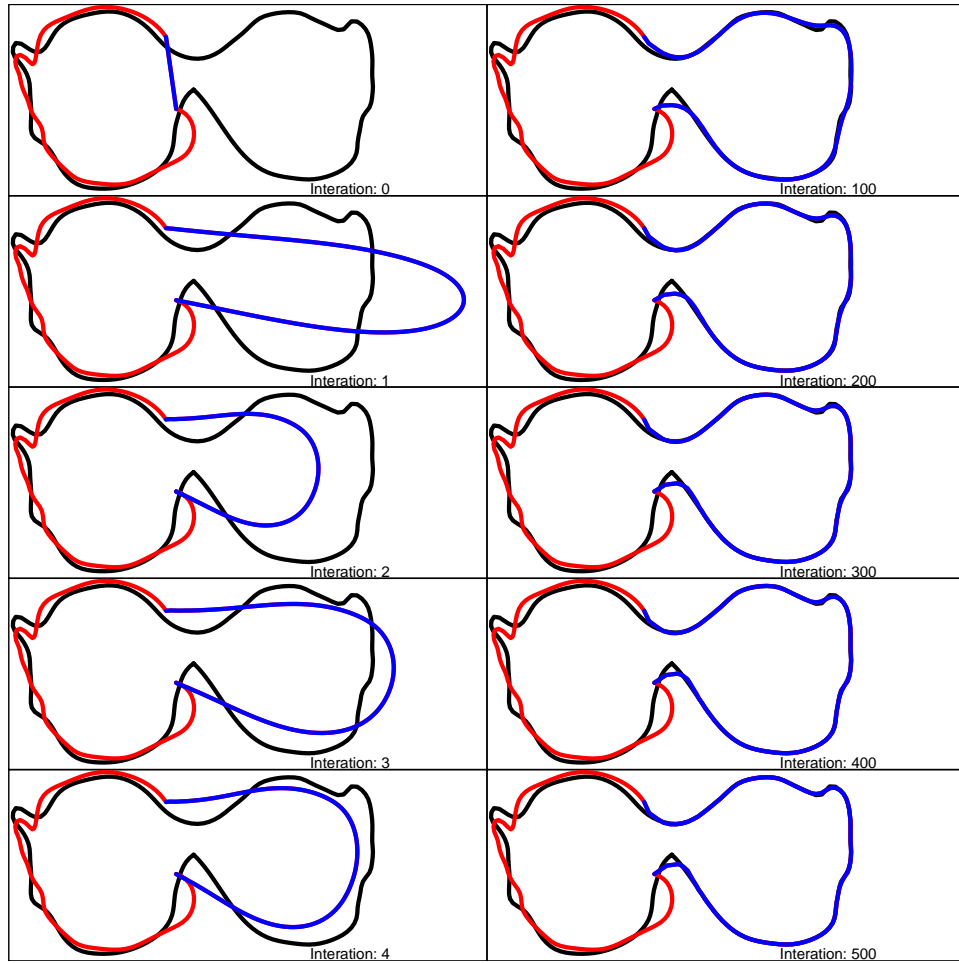
FIGURE 1. Test

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.

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

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

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:

$$\phi(Q_C, h_N) := \mathbb{P}(d(\mathbf{C}, C) < h_N), \quad C \in \mathcal{C}, h_N > 0$$

$$= \mathbb{P}\left(\inf_{(O,\gamma) \in SO(2) \times \Gamma} \|\mathbf{Q_C} - O Q_C(\gamma)\sqrt{|\gamma'|}\|_{\mathbb{L}^2} < h_N \right),$$

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

$$\phi(Q_C, h_N) = \mathbb{P}(d(\mathbf{C}, C) < h_N), \quad C \in \mathcal{C}, h_N > 0$$

$$= \mathbb{P}\left(\inf_{(O,\gamma) \in SO(2) \times \Gamma} \|\mathbf{Q_C} - OQ_C(\gamma)\sqrt{|\gamma'|}\|_{\mathbb{L}^2} < h_N\right)$$

$$= \mathbb{P}\left(\inf_{(O,\gamma) \in SO(2) \times \Gamma} \|\mathbf{Q_C} - OQ_C(\gamma)\sqrt{|\gamma'|}\|_{\mathbb{L}^2}^2 < h_N^2\right)$$

$$= \mathbb{P}\left(\|\mathbf{Q_C} - \tilde{O}Q_C(\tilde{\gamma})\sqrt{|\tilde{\gamma}'|}\|_{\mathbb{L}^2}^2 < h_N^2 \text{ for some } (\tilde{O}, \tilde{\gamma}) \in SO(2) \times \Gamma\right)$$

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 $\{\omega \in \Omega : \|\mathbf{Q_C}(\omega) - Q_C(\omega)\|_{\mathbb{L}^2}^2 < h_N^2\} \subseteq \{\omega \in \Omega : \|\mathbf{Q_C}(\omega) - \tilde{O}Q_C(\omega)(\tilde{\gamma})\sqrt{\tilde{\gamma}'}\|_{\mathbb{L}^2}^2 < h_N^2 \text{ for some } (\tilde{O}, \tilde{\gamma}) \in SO(2) \times \Gamma\}$. 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

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

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 $\mathfrak{C}(C) = \mathfrak{R}(C)$.

---

[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.

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

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

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. $\quad\square$