Elastic Full Procrustes Means for Sparse and Irregular Planar Curves

Masters Thesis

Manuel Pfeuffer[1]

12th October 2020, Berlin

Advisors: Lisa Maike Steyer, Almond Stöcker

1st Examiner: Prof. Dr. Sonja Greven

2nd Examiner: Prof. Dr. Nadja Klein

[1] pfeuferm@hu-berlin.de, Matriculation Number: 577668

Contents

1 Introduction

Statistical Shape Analysis is the branch of statistics concerned with modelling the geometry of objects. Examples might be the outlines of bones and organs, a handwritten digit, or the folds of a protein. To capture an object's geometrical information, a common approach is the use of so called "landmarks": characteristic points on an object, which match between and within populations (see Dryden and Mardia 2016, p. 3). However, in recent years an alternative approach has become more popular, where objects are represented using curves. This has the advantage of offering a more flexible representation of an object's geometry, as the analysis is not restricted to a set of discrete points. Usually, the curves are themselves represented by functions $\beta : [0,1] \to \mathbb{R}^k$, which, for example in 2D, might describe the outlines of an object in an image. As each object corresponds to one function, this opens up a connection to the branch of statistics concerned with modelling whole functions: Functional Data Analysis (see e.g. Ramsay 2006).

When analysing the geometry of objects, differences in location, rotation, and size are often not of interest. Instead, the focus lies purely on their differences in *shape*, a widely adapted definition of which was established by Kendall 1977 and which might be formulated in the following way:

Definition 1.1 (Shape)**.** *All geometrical information that remains when location, scale and rotational effects are removed from an object (see* Dryden *and* Mardia *2016, p. 1).*

The functional approach to shape analysis introduces an additional kind of invariance relating to the parametrisation $t \in [0,1]$, as only the image of β but not its parametrisation is indacative of the object's shape.

In statistical shape analysis one is interested in analysing the geometry of objects, like the outlines of bones and organs, a handwritten digit, or the folds of a protein. This geometrical information is usually approximated, by measuring the cordinates of a fixed set of *landmarks*, which are characteristic points on an object that match between and within populations (see Dryden and Mardia 2016, p. 3). As these coordinates depend on the scale, rotation and translation of the object at time of measurement, only the *equivalence class* of the landmark configuration modulo these transformations is indicative of its shape, which will be defined in more detail in section 2.1. A popular type of shape mean that does not depend on the rotation, scaling and translation of the input shapes is the *full Procrustes mean*. Here, the mean is defined as the minimizer of a least squares criterion, using a distance measure that is invariant under the mentioned transformations, which will be discussed in section 2.2.

Instead of approximating the geometry of an object using landmarks, its whole outline might be represented by an open or closed curve $\beta : [0,1] \to \mathbb{R}^d$, $d \in \mathbb{N}$, eliminating the (often subjective) decision of which points to consider as "characteristic". This functional data approach introduces an additional kind of shape invariance relating to the parametrisation t of β, as only the image of β but not its parametrisation is indacative of the objects shape. A functional shape mean, which is invariant with respect to re-parametrisation of the input curves, is called an *elastic* mean and its calculation is greatly simplified by working in the *square-root-velocity* (SRV) framework as introduced in Srivastava, Klassen, et al. 2011. Elastic shape means and the SRV framework will be explained in section 2.3. Even though this functional approach eliminates the subjectivity of choosing a fixed set of landmarks, in practice, a curve will never be fully observed and one has to again work with a set of sampled points along the curve. A particular challenge is working with curves that are sparsely and irregularly sampled, where one has to employ appropriate smoothing techniques to make maximum use of the available (sparse) data. This will be briefly discussed in section 2.4.

The aim of this thesis is to bring all these concepts together in the estimation of "Elastic Full Procrustes Means for Sparse and Irregular Planar Curves". This is a novel approach as, firstly, the literature on elastic shape means has so far mostly focused on the different mean concept of the Riemannian center of mass mean (Srivastava, Klassen, et al. 2011) and secondly, shape means of sparse and irregular shape data have so far only been considered modulo re-parametrisation and translation, excluding rotation and scaling invariance (Steyer, Stöcker, and Greven 2021). The thesis focuses on the special case of 2D (i.e. planar) curves, as it can be shown that the Procrustes mean has particularly nice properties in this setting when working with complex notation. Additionally this thesis will be mainly concerned with the mean estimation for open curves, as the closed curves case is more challenging mathematically. After covering the relevant background material in section 2, an expression for the elastic full Procrustes mean will be derived in section 3. In section 4, an estimation strategy for the setting of sparse and irregular curves is proposed, which will be applied to simulated and empirical datasets in section 5. Finally, in section 6 a possible extension to closed curves will be briefly discussed. All results will be summarized in section 7.

2 Functional and Shape Data Analysis of Planar Curves

Before beginning with a derivation of the elastic full Procrustes mean, it is important to establish a notational and mathematical framework for the treatment of planar shapes. While the restriction to the 2D case might seem a major one, it still covers all shape data extracted from e.g. imagery and is therefore very applicable in practice. Examples of objects that can be analyzed in this way are handwritten digits or symbols, the outlines bones and organs in medical images, or even movement trajectories on a map.

All these objects may be naturally represented as planar curves $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x(t), y(t))^T$, where $x(t)$ and $y(t)$ are the scalar-valud *coordinate functions*. Calculations in 2D, and in particular the derivation of the full Procrustes mean, are greatly simplified by using complex notation. Going forward, we will therefore identify $\mathbb{R}^2$ with $\mathbb{C}$ and always use complex notation when representing a planar curve:

$$\beta : [0,1] \to \mathbb{C}, \quad \beta(t) = x(t) + i\,y(t).$$

2.1 Equivalence Classes and Shape Invariance

The concept of shape is closely related to the concept of invariance under the transformations of scaling, translation and rotation. When considering the shape of a curve, we additionally have to take into account invariance with respect to re-parametrisation. This can be seen by noting that the curves $\beta(t)$ and $\beta(\gamma(t))$, with some *warping function* $\gamma : [0,1] \to [0,1]$ monotonically increasing and differentiable, have the same image and therefore represent the same geometrical object. We can say that the actions of translation, scaling, rotation and re-parametrization are *equivalence relations* with respect to shape, as each action leaves the shape of the curve untouched and only changes the way the shape is represented.

The shape of a curve can then be defined as the respective *equivalence class*, i.e. the set of all possible shape preserving transformations of the curve. As two equivalence classses are neccessarily either disjoint or identical, we can consider two curves as having the same shape, if they are elements of the same equivalence class (see Srivastava and Klassen 2016, p. 40).

Definition 2.1 (Equivalence relation, equivalence class and quotient space)**.** *A relation*

$\backsim$ *on a set* X *is called an* ***equivalence relation*** *if, for all* $x, y, z \in X$*, it has the following properties:*

i. $x \backsim x$ *(reflexivity)*

ii. $x \backsim y \Rightarrow y \backsim x$ *(symmetry)*

iii. $x \backsim y, y \backsim z \Rightarrow x \backsim z$ *(transitivity)*

The ***equivalence class*** $[x]$ *of* $x \in X$ *is given by the set of all* $y \in X$ *so that* $x \backsim y$*. The* ***quotient space*** $X/\backsim$ *of* X *under the relation* $\backsim$ *is the set of all equivalence classes in* X*.*

We now want to define an equivalence relations with respect to shape. With this in mind, let us first consider how the discussed transformations act individually on the set of parametrized planar curves with complex representation $\beta : [0,1] \to \mathbb{C}$:

- The **translation** group $\mathbb{C}$ acts on β by $(\xi, \beta) \mapsto \beta + \xi$, for any $\xi \in \mathbb{C}$. We can consider two curves as equivalent with respect to translation $\beta_1 \backsim \beta_2$, if there exists a complex scalar $\tilde{\xi} \in \mathbb{C}$ so that $\beta_1 = \beta_2 + \tilde{\xi}$. For this relation an equivalence class is $[\beta] = \{\beta + \xi \,|\, \beta : [0,1] \to \mathbb{C},\, \xi \in \mathbb{C}\}$.

- The **scaling** group $\mathbb{R}^+$ acts on β by $(\lambda, \beta) \mapsto \lambda\beta$, for any $\lambda \in \mathbb{R}^+$. We define $\beta_1 \backsim \beta_2$, if there exists a scalar $\tilde{\lambda} \in \mathbb{R}^+$ so that $\beta_1 = \tilde{\lambda}\beta_2$. An equivalence class is $[\beta] = \{\lambda\beta \,|\, \beta : [0,1] \to \mathbb{C},\, \lambda \in \mathbb{R}^+\}$.

- The **rotation** group $[0, 2\pi]$ acts on β by $(\theta, \beta) \mapsto e^{i\theta}\beta$, for any $\theta \in [0, 2\pi]$. We define $\beta_1 \backsim \beta_2$, if there exists a $\tilde{\theta} \in [0, 2\pi]$ with $\beta_1 = e^{i\tilde{\theta}}\beta_2$. An equivalence class is $[\beta] = \{e^{i\theta}\beta \,|\, \beta : [0,1] \to \mathbb{C},\, \theta \in [0, 2\pi]\}$.

- The **re-parametrization** group Γ acts on β by $(\gamma, \beta) \mapsto \beta \circ \gamma$, for any $\gamma \in \Gamma$ with Γ being the set of monotonically increasing and differentiable warping functions. We define $\beta_1 \backsim \beta_2$, if there exists a warping function $\tilde{\gamma} \in \Gamma$ with $\beta_1 = \beta_2 \circ \tilde{\gamma}$. An equivalence class is $[\beta] = \{\beta \circ \gamma \,|\, \beta : [0,1] \to \mathbb{C},\, \gamma \in \Gamma\}$.

In a next step, we can consider how these transformations act in concert and whether they *commute*, that is, whether the order of applying the transformations changes outcomes. Consider for example the actions of the rotation and scaling product group $\mathbb{R}^+ \times [0, 2\pi]$ given by $((\lambda, \theta), \beta) \mapsto \lambda e^{i\theta}\beta$, which clearly commute as $\lambda(e^{i\theta}\beta) = e^{i\theta}(\lambda\beta)$. However, the joint actions of scaling, rotation, and translation do not commute, as $\lambda e^{i\theta}(\beta + \xi) \neq \lambda e^{i\theta}\beta + \xi$. It follows that the order of translating and rotating or scaling matters and when defining such a joint action, one usually takes the translation to act on the already scaled and rotated curve.

Definition 2.2 (Euclidean similarity transformation)**.** *We define an **Euclidean similarity transformation** of a curve $\beta : [0,1] \to \mathbb{C}$ as the joint action of scaling, rotation, and translation by*

$$((\xi, \lambda, \theta), \beta) \mapsto \lambda e^{i\theta}\beta + \xi,$$

with $\xi \in \mathbb{C}$, $\lambda \in \mathbb{R}^+$, and $\theta \in [0, 2\pi]$ (see Dryden *and* Mardia *2016, p. 62).*

With respect to the action of re-parametrization, we can note that it necessarily commutes with all Euclidean similarity transformations, as those only act on the image of β, while the former only acts on the parametrization. Putting everything together, we can finally give a formal definition of the shape of a planar curve.

Definition 2.3 (Shape)**.** *The **shape** of a planar curve $\beta : [0,1] \to \mathbb{C}$ is given by its equivalence class with respect to all Euclidean similarity transformations and re-parametrizations*

$$[\beta] = \left\{ \lambda e^{i\theta}(\beta \circ \gamma) + \xi \,|\, \xi \in \mathbb{C},\, \lambda \in \mathbb{R}^+,\, \theta \in [0, 2\pi],\, \gamma \in \Gamma \right\}.$$

*The **shape space** is then given by the corresponding quotient space*

$$\mathbb{C}^{[0,1]} / \mathbb{C} \rtimes \left(\mathbb{R}^+ \times [0, 2\pi]\right) \times \Gamma = \left\{ [\beta] | \beta \in \mathbb{C}^{[0,1]} \right\},$$

where the symbol "$\rtimes$" denotes a semi-direct product, i.e. that the translation group acts "after" scaling and rotation (for details see Srivastava *and* Klassen *2016, Chapter 3).*

As we are interested in calculating shape means, we now need to find a way to measure simmilarity and dissimilarity between shapes. More formally, this means that we want to measure distances between equivalence classes in shape space. As equivalence classes and quotient spaces are already quite abstract concepts and as the discussed spaces may be quite complex, we need to find a way of simplifying calculations in shape space. The basic idea is to find unique curves as a representation for each equivalence class, so that distance in shape space can be equivalently calculated as the distance between these curves. In short we want to find intersections of the function space, so that there is a one-to-one relationship between elements of that interesection and the squivalence classes in shape space. Consider, for example, the action of the translation group $\mathbb{C}$ with corresponding quotient space $\mathbb{C}^{[0,1]}/\mathbb{C} = \left\{ [\beta] \,|\, \beta \in \mathbb{C}^{[0,1]} \right\}$ and $[\beta] = \{\beta + \xi \,|\, \xi \in \mathbb{C}\}$. For each equivalence class in $\mathbb{C}^{[0,1]}/\mathbb{C}$ there is exactly one element in the space of all centered curves $\mathcal{C}_1 = \left\{ \beta \in \mathbb{C}^{[0,1]} \,|\, \int_0^1 \beta(t)\, dt = 0 \right\}$ belonging to that equivalence class. For the scaling group $\mathbb{R}^+$, when considering only continuously differentiable β, we can similarly form the space of all unit-length curves $\mathcal{C}_2 = \left\{ \beta \in \mathbb{C}^{[0,1]} \,|\, \beta \text{ continuously differentiable, } \int_0^1 |\dot{\beta}(t)|\, dt = 1 \right\}$ and for each equivalence class with respect to scaling there is again exactly one element in $\mathcal{C}_2$ belonging to that class. Both intersections taken together result in the concept of *pre-shape* and the act of removing translation and scaling variability by centering and restricting curves to unit-length is called "quotienting out" (see Srivastava and Klassen 2016, 133 f.).

Definition 2.4 (Pre-shape)**.** *The* ***pre-shape*** *of a countinously differentiable planar curve* $\beta : [0, 1] \to \mathbb{C}$ *is given by its centered unit-length representation*

$$\tilde{\beta} = l_\beta^{-1} \left(\beta - c_\beta \right)$$

with $l_\beta = \int_0^1 |\dot{\beta}(t)|\, dt$ *and* $c_\beta = \int_0^1 \beta(t)\, dt$*. The* ***shape*** *of* β *can then also be defined as*

$$[\beta] = \left\{ e^{i\theta} (\tilde{\beta} \circ \gamma) \,|\, \theta \in [0, 2\pi],\, \gamma \in \Gamma \right\}.$$

We can use the concept of pre-shape to construct an appropriate distance measure for equivalence classes with respect to scaling and translation, by using the usual distance metric for complex functions with the respective pre-shapes.

$$d_{\mathbb{C}^{[0,1]}/\mathbb{C} \rtimes \mathbb{R}^+}([\beta_1], [\beta_2]) = d_{\mathbb{C}^{[0,1]}}(\tilde{\beta}_1, \tilde{\beta}_2)$$

Working with pre-shapes means that we only have to take care of shape equivalence with respect to rotation and re-parametrization going forward. However, both types of transformations are not as easily dealt with as translation and scaling. We will now take a look at a different way of constructing distance measures for quotient spaces, one dealing with rotation and one dealing with re-parametrization.

2.2 The Full Procrustes Mean for Planar Curves

The *full Procrustes distance* is a widely used distance function in classical shape analysis that allows for scale and rotation invariant shape distance calculation. To remove translation and scaling variablity, we chose to restrict our curves to a certain fixed position and scale. This works, because we can simply calculate both for any curve. However, there is no similar way to calculate the "rotation" of a curve and in fact, saying that a curve is "rotated" is only meaningful in comparison to other curves. A solution is to align the rotation of any two curves on a case by case basis, when calculating the distance between them.

Definition 2.5 (Full Procrustes distance). *The* ***full Procrustes distance*** *between the shapes* $[\beta_1]$, $[\beta_2]$ *of two continously differentiable* β_1, β_2 *with* $\beta_i : [0,1] \to \mathbb{C}$ *is given by*

$$d_F([\beta_1], [\beta_2]) = \inf_{\lambda \in \mathbb{R}^+, \theta \in [0, 2\pi]} ||\tilde{\beta}_1 - \lambda e^{i\theta} \tilde{\beta}_2||,$$

where $\tilde{\beta}_{1,2}$ *are the centered, unit-length pre-shapes of* $\beta_{1,2}$.

Definition 2.6 (Full Procrustes mean). *The* ***full Procrustes mean*** *shape for a sample of landmark configurations* X_i *(*$i = 1, \dots, n$*) is then given by the equivalence class [*$\hat{\mu}_F$*] of a landmark configuration that minimizes the sum of squared full Procrustes distances*

$$\hat{\mu}_F = \operatorname*{arginf}_{\mu} \sum_{i=1}^{n} d_F([\mu], [X_i])^2,$$

where μ *is assumed centered and normalized (see* Dryden *and* Mardia *2016, pp. 71, 114).*

[Define an intersection of the quotient space by projections. Proof that this is good. Procrustes Fits.]

2.3 Elastic Means and the Square-Root-Velocity Framework

Instead of working with the original curve β, for calculation of *elastic* means it is advantageous to work with its corresponding *square root velocity curve* given by

$$q : [0,1] \to \mathbb{C}, \quad q(t) = \frac{\dot{\beta}(t)}{\sqrt{||\dot{\beta}(t)||}} \quad \text{for } \dot{\beta}(t) \neq 0$$

where original curve β can be obtained up to translation by back transformation via $\beta(t) = \beta(0) + \int_0^t q(s)||q(s)||ds$. Moreover, if the original curve is of unit length the SRV curve will be automatically normalized:

$$||q|| = \sqrt{\langle q, q \rangle} = \sqrt{\int_0^1 \overline{q(t)} q(t)\, dt} = \sqrt{\int_0^1 |q(t)|^2\, dt} = \sqrt{\int_0^1 |\dot{\beta}(t)|\, dt} = \sqrt{1} = 1.$$

2.4 Functional Data Analysis of Sparse and Irregular Planar Curves

3 The Elastic Full Procrustes Means for Planar Curves

Let β be a continuous planar curve. It can be represented in a parameterized form in $\mathbb{R}^2$ as

$$\beta : [0,1] \to \mathbb{R}^2, \quad \beta(t) = (x(t), y(t)),$$

where x, y are scalar-valued *coordinate functions* of β, parametrized by t. We can equivalently represent a planar curve using complex numbers as

$$\beta : [0,1] \to \mathbb{C}, \quad \beta(t) = x(t) + iy(t),$$

with the added benefit that complex notation often simplifies calculations in the 2D case.

For a set of planar curves $\beta_1, \ldots, \beta_n : [0,1] \to \mathbb{C}$, either centered with $\langle \beta_i, \mathbb{1} \rangle$ or with no relative translation to each other, the *full Procrustes mean* $\hat{\mu}$ is then defined as the curve minimizing the sum of squared *full Procrustes distances* from each β_i to an unknown unit size mean configuration μ, that is

$$\begin{aligned}
\hat{\mu} &= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmin}} \sum_{i=1}^{n} d_F^2(\mu, \beta_i) \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmin}} \sum_{i=1}^{n} 1 - \frac{\langle \mu, \beta_i \rangle \langle \beta_i, \mu \rangle}{\langle \mu, \mu \rangle \langle \beta_i, \beta_i \rangle} \quad \text{s.t. } ||\mu|| = 1
\end{aligned}$$

which we can be further simplified by normalizing $\beta_i := \frac{\beta_i}{||\beta_i||}$ and using $\langle \mu, \mu \rangle = 1$

$$\hat{\mu} = \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \langle \mu, \beta_i \rangle \langle \beta_i, \mu \rangle \quad \text{s.t. } ||\mu|| = 1.$$

The expression for $d_F^2(\mu, \beta_i)$ in the case of planar curves is derived in appendix A.1.

3.1 The Full Procrustes mean

Consider a set of planar SRV curves $q_1, \ldots, q_n : [0,1] \to \mathbb{C}$ of unit length $||q_i|| = 1$ for all i. The *full Procrustes mean* $\hat{\mu}$ is given by

$$\begin{aligned}
\hat{\mu} &= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \langle \mu, q_i \rangle \langle q_i, \mu \rangle \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \int_0^1 \overline{\mu(t)} q_i(t)\, dt \int_0^1 \overline{q_i(s)} \mu(s)\, ds \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu(t)} \underbrace{\left(\sum_{i=1}^{n} q_i(t) \overline{q_i(s)} \right)}_{:= n\hat{C}(s,t)} \mu(s)\, dtds \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \int_0^1 \overline{\mu(t)} \int_0^1 \hat{C}(s,t) \mu(s)\, dsdt \quad \text{s.t. } ||\mu|| = 1
\end{aligned}$$

with the solution given by the eigenfunction corresponding to the largest eigenvector of the complex empirical covariance function $\hat{C}(s,t) = n^{-1} \sum_{i=1}^{n} q_i(t) \overline{q_i(s)}$.

3.2 The Full Procrustes Mean in a fixed basis

To avoid having to sample the estimated covariance surface $\hat{C}(s,t)$ on a large grid when calculating its leading eigenfunction, it might be preferable to calculate this eigenfunction from the vector of basis coefficients directly. After choosing a basis representation $b = (b_1, \dots, b_k)$ with $b_j : \mathbb{R} \to \mathbb{R}$ real-valued basis functions, we want to estimate complex coefficients $\theta_j \in \mathbb{C}$ so that the Full Procrustes mean of SRV curves is given by $\hat{\mu}(t) = \sum_{j=1}^{k} \hat{\theta}_j b_j(t) = b^T \hat{\theta}$:

$$
\begin{aligned}
\hat{\mu} &= \underset{\theta:||b^T\theta||=1}{\operatorname{argmax}} \sum_{i=1}^{n} \langle b^T\theta, q_i \rangle \langle q_i, b^T\theta \rangle \\
&= \underset{\theta:||b^T\theta||=1}{\operatorname{argmax}} \sum_{k,l} \sum_{i=1}^{n} \langle b_k\theta_k, q_i \rangle \langle q_i, b_l\theta_l \rangle \\
&= \underset{\theta:||b^T\theta||=1}{\operatorname{argmax}} \sum_{k,l} \bar{\theta}_k \theta_l \sum_{i=1}^{n} \langle b_k, q_i \rangle \langle q_i, b_l \rangle \\
&= \underset{\theta:||b^T\theta||=1}{\operatorname{argmax}} \, \theta^H S \theta
\end{aligned}
$$

where the matrix $S = \{\sum_{i=1}^{n} \langle b_k, q_i \rangle \langle q_i, b_l \rangle\}_{k,l}$ has to be estimated from the observed SRV curves. We can further simplify S to

$$
\begin{aligned}
S_{kl} &= \sum_{i=1}^{n} \int_0^1 \bar{b}_k(t) q_i(t) dt \int_0^1 \bar{q}_i(s) b_l(s) ds \\
&= \int_0^1 \int_0^1 \bar{b}_k(t) \underbrace{\left(\sum_{i=1}^{n} q_i(t) \bar{q}_i(s) \right)}_{=n\,\hat{C}(s,t)} b_l(s) ds dt \\
&= n \int_0^1 \int_0^1 \bar{b}_k(t) \hat{C}(s,t) b_l(s) ds dt
\end{aligned}
$$

with $\hat{C}(s,t) = \frac{1}{n} \sum_{i=1}^{n} q_i(s) \overline{q_i(t)}$ the sample analogue to the complex population covariance function $C(s,t) = \mathbb{E}[q(s)\overline{q(t)}]$. We may estimate $C(s,t)$ via tensor product splines, so that $\hat{C}(s,t) = \sum_{k,l} \hat{\xi}_{kl} b_k(t) b_l(s)$, where $b_j(t)$, $j = 1, \dots, k$ are the same real valued basis functions as used for the mean and $\hat{\xi}_{kl}$ are the estimated complex coefficients.

We can then further simplify S_{kl}

$$\begin{aligned}
S_{kl} =& n \int_0^1 \int_0^1 b_k(t) \left(\sum_{p,q} \hat{\xi}_{pq} b_q(t) b_p(s) \right) b_l(s) ds dt \\
=& n \sum_{p,q} \hat{\xi}_{pq} \int_0^1 \int_0^1 b_k(t) b_q(t) b_p(s) b_l(s) ds dt \\
=& n \sum_{p,q} \hat{\xi}_{pq} \langle b_k, b_q \rangle \langle b_p, b_l \rangle \\
=& n \sum_{p,q} \hat{\xi}_{pq} g_{kq} g_{pl}
\end{aligned}$$

where g_{ij}, $i, j = 1, \ldots, k$ are the elements of the Gram matrix $G = bb^T$ with $G = \mathbb{I}_k$ in the special case of an orthogonal basis. We can then write the write the matrix S as a function of the estimated coefficient matrix $\hat{\Xi} = (\hat{\xi}_{ij})_{i,j=1,\ldots,k}$:

$$S =n\, G \hat{\Xi} G$$

The full Procrustes mean of SRV curves is then given by the solution to the optimization problem

$$\begin{aligned}
\hat{\mu} &= \underset{\theta}{\operatorname{argmax}}\, n\, \theta^H G \hat{\Xi} G \theta \quad \text{subj. to} \quad ||b^T \theta|| = 1 \\
&= \underset{\theta : ||b^T \theta|| = 1}{\operatorname{argmax}}\ \theta^H G \hat{\Xi} G \theta \quad \text{subj. to} \quad \theta^H G \theta = 1
\end{aligned}$$

One may solve this by using Lagrange optimization with the Langrangian

$$\mathcal{L}(\theta, \lambda) = \theta^H G \hat{\Xi} G \theta - \lambda(\theta^H G \theta - 1)$$

3.3 Estimation of the covariance surface $C(s,t)$

Consider the following model for independent curves

$$Y_i(t_{ij}) = \mu(t_{ij}, \mathbf{x}_i) + E_i(t_{ij}) + \epsilon(t_{ij}), \quad j = 1, \dots, D_i,\ i = 1, \dots, n, \tag{1}$$

[Fast symmetric additive cov smoothing, skew-symmetry, population vs. sample, etc.]

4 Estimation Strategy

5 Empirical Application

6 Outlook

7 Summary

A Derivations and Proofs

A.1 Derivation of the Full Procrustes Distance for Functional Data

Consider two curves $\beta_1, \beta_2 : [0,1] \to \mathbb{C}$ with $\langle \beta_1, \mathbb{1} \rangle = \langle \beta_2, \mathbb{1} \rangle = 0$ where $\mathbb{1}$ is the constant function $\mathbb{1}(t) = 1$ for all $t \in [0,1]$. Then β_1 and β_2 can be considered to be centered as

$$\langle \beta_1, \mathbb{1} \rangle = \int_0^1 \bar{\beta}_1(t)\mathbb{1}(t)dt = \int_0^1 \bar{\beta}_1(t)dt = \int_0^1 (y(t) + ix(t))dt = \underbrace{\int_0^1 y(t)dt}_{\overset{!}{=}0} + i \underbrace{\int_0^1 x(t)dt}_{\overset{!}{=}0} = 0$$

Then the full procrustes distance of β_1, β_2 is given by their minimum distance controlling for translation $\gamma \in \mathbb{C}$, and scaling and rotation $\omega = be^{i\theta} \in \mathbb{C}$:

$$\begin{aligned}
d_F^2 &= \min_{\omega,\gamma \in \mathbb{C}} ||\beta_1 - \gamma\mathbb{1} - \omega\beta_2||^2 \\
&= \min_{\omega,\gamma \in \mathbb{C}} \langle \beta_1 - \gamma\mathbb{1} - \omega\beta_2, \beta_1 - \gamma\mathbb{1} - \omega\beta_2 \rangle \\
&= \min_{\omega,\gamma \in \mathbb{C}} \langle \beta_1 - \omega\beta_2, \beta_1 - \omega\beta_2 \rangle - \underbrace{\langle \beta_1, \gamma\mathbb{1} \rangle}_{=0} - \underbrace{\langle \gamma\mathbb{1}, \beta_1 \rangle}_{=0} + \underbrace{\langle \gamma\mathbb{1}, \omega\beta_2 \rangle}_{=0} + \underbrace{\langle \omega\beta_2, \gamma\mathbb{1} \rangle}_{=0} + \underbrace{\langle \gamma\mathbb{1}, \gamma\mathbb{1} \rangle}_{=||\gamma\mathbb{1}||^2} \\
&\overset{\gamma=0}{=} \min_{\omega \in \mathbb{C}} \langle \beta_1, \beta_1 \rangle + \langle \omega\beta_2, \omega\beta_2 \rangle - \langle \beta_1, \omega\beta_2 \rangle - \langle \omega\beta_2, \beta_1 \rangle \\
&= \min_{\omega \in \mathbb{C}} \langle \beta_1, \beta_1 \rangle + |\omega|^2 \langle \beta_2, \beta_2 \rangle - \omega \langle \beta_1, \beta_2 \rangle - \overline{\omega} \langle \beta_2, \beta_1 \rangle
\end{aligned}$$

To find $\omega \in \mathbb{C}$ that minimizes $||\beta_1 - \omega\beta_2||^2$ we first consider the part of the problem dependent on θ. We need to solve

$$\min_{\omega \in \mathbb{C}} -\omega \langle \beta_1, \beta_2 \rangle - \overline{\omega} \langle \beta_2, \beta_1 \rangle = \max_{\omega \in \mathbb{C}} \omega \langle \beta_1, \beta_2 \rangle + \overline{\omega} \langle \beta_2, \beta_1 \rangle$$

by using $\omega = be^{i\theta}$ and $\langle \beta_1, \beta_2 \rangle = ae^{i\phi}$:

$$\begin{aligned}
\max_{\omega \in \mathbb{C}} \omega \langle \beta_1, \beta_2 \rangle + \overline{\omega} \langle \beta_2, \beta_1 \rangle &= \max_{b \in \mathbb{R}^+, \theta \in [0,2\pi]} be^{i\theta} ae^{i\phi} + be^{-i\theta} ae^{-i\phi} \\
&= \max_{b \in \mathbb{R}^+, \theta \in [0,2\pi]} be^{i\theta} ae^{i\phi} + be^{-i\theta} ae^{-i\phi} \\
&= \max_{b \in \mathbb{R}^+, \theta \in [0,2\pi]} 2ba \cos(\theta + \phi)
\end{aligned}$$

$$\overset{\theta=-\phi}{=} \max_{b\in\mathbb{R}^+} 2ba$$

and using $\theta = -\phi$ the original mimization problem therefore simplifies to

$$\begin{aligned}
d_F^2 &= \min_{b\in\mathbb{R}^+} \langle \beta_1, \beta_1 \rangle + b^2 \langle \beta_2, \beta_2 \rangle - 2ba \\
\frac{\partial d_F^2}{\partial b} &= 2b\langle \beta_2, \beta_2 \rangle - 2a \overset{!}{=} 0 \\
\Rightarrow \quad b &= \frac{a}{\langle \beta_2, \beta_2 \rangle}
\end{aligned}$$

And for the *full Procrustes distance* it follows that

$$d_F^2 = \langle \beta_1, \beta_1 \rangle - \frac{a^2}{\langle \beta_2, \beta_2 \rangle} = \langle \beta_1, \beta_1 \rangle - \frac{\langle \beta_1, \beta_2 \rangle \langle \beta_2, \beta_1 \rangle}{\langle \beta_2, \beta_2 \rangle}$$

As this expression is not symmetric in β_1 and β_2 we can take the curves to be of unit length with $\tilde{\beta}_j = \frac{\beta_j}{||\beta_j||}$, $j = 1, 2$ with $||\beta_j|| = \sqrt{\langle \beta_j, \beta_j \rangle}$, so that $\langle \tilde{\beta}_1, \tilde{\beta}_1 \rangle = \langle \tilde{\beta}_2, \tilde{\beta}_2 \rangle = 1$ and obtain a suitable measure of distance:

$$d_F = \sqrt{1 - \langle \tilde{\beta}_1, \tilde{\beta}_2 \rangle \langle \tilde{\beta}_2, \tilde{\beta}_1 \rangle} = \sqrt{1 - \frac{\langle \beta_1, \beta_2 \rangle \langle \beta_2, \beta_1 \rangle}{\langle \beta_1, \beta_1 \rangle \langle \beta_2, \beta_2 \rangle}}$$

B Supplementary Materials

B.1 Implementation Notes

B.2 Replication Guide

C Math-Basics Recap

C.1 Scalar Products

V n-dimensional vector space with basis $B = (b_1, \ldots, b_n)$, then any scalar product $\langle \cdot, \cdot \rangle$ on V can be expressed using a $(n \times n)$ matrix G, the Gram matrix of the scalar product. Its entries are the scalar products of the basis vectors:

$$G = (g_{ij})_{i,j=1,\ldots,n} \quad \text{with} \quad g_{ij} = \langle b_i, b_j \rangle \quad \text{for} \quad i, j = 1, \ldots, n$$

When vectors $x, y \in V$ are expressed with respect to the basis B as

$$x = \sum_{i=1}^{n} x_i b_i \quad \text{and} \quad y = \sum_{i=1}^{n} y_i b_i$$

the scalar product can be expressed using the Gram matrix, and in the complex case it holds that

$$\langle x, y \rangle = \sum_{i,j=1}^{n} \bar{x}_i y_j \langle b_i, b_j \rangle = \sum_{i,j=1}^{n} \bar{x}_i g_{ij} y_j = x^\dagger G y$$

when $x_i, y_i \in \mathbb{C}$ for $i = 1, \ldots, n$ with $x^\dagger$ indicating the conjugate transpose of $x = (x_1, \ldots, x_n)^T$. If B is an *orthonormal* basis, that is if $\langle b_i, b_j \rangle = \delta_{ij}$, it further holds that $\langle x, y \rangle = x^\dagger y$ as $G = \mathbb{1}_{n \times n}$.

C.2 Functional Scalar Products

This concept can be generalized for vectors in function spaces. Define the scalar product of two functions $f(t), g(t)$ as:

$$\langle f, g \rangle = \int_a^b \bar{f}(t) w(t) g(t) dt$$

with weighting function $w(t)$ and $[a, b]$ depending on the function space. The scalar product has the following properties:

1. $\langle f, g + h \rangle = \langle f, g \rangle + \langle f, h \rangle$
2. $\langle f, g \rangle = \overline{\langle g, f \rangle}$

3. $\langle f, cg\rangle = c\langle f, g\rangle$ or, using (2), $\langle cf, g\rangle = \bar{c}\langle f, g\rangle$ for $c \in \mathbb{C}$

If we have a functional basis $\{\phi_1, \ldots, \phi_n\}$ (and possibly $n \to \infty$) of our function space we can also write the function f as an expansion

$$f = \sum_{i=1}^{n} a_i \phi_i \quad \text{so that} \quad f(t) = \sum_{i=1}^{n} a_i \phi_i(t)$$

Additionally, if we have a *orthogonal* basis, so that $\langle \phi_i, \phi_j\rangle = 0$ for $i \neq j$, we can take the scalar product with ϕ_k from the left

$$\langle \phi_k, f\rangle = \sum_{i=1}^{n} a_i \langle \phi_k, \phi_i\rangle = a_k \langle \phi_k, \phi_k\rangle$$

which yields the coefficients a_k:

$$a_k = \frac{\langle \phi_k, f\rangle}{\langle \phi_k, \phi_k\rangle}$$

For an *orthonormal* basis it holds that $\langle \phi_i, \phi_j\rangle = \delta_{ij}$. Suppose that two functions f, g are expanded in the same orthonormal basis:

$$f = \sum_{i=1}^{n} a_i \phi_i \quad \text{and} \quad g = \sum_{i=1}^{n} b_i \phi_i$$

We can then write the scalar product as:

$$\langle f, g\rangle = \langle \sum_{i=1}^{n} a_i \phi_i, \sum_{i=1}^{n} b_i \phi_i\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} \hat{a}_i b_j \langle \phi_i, \phi_j\rangle = \sum_{i=1}^{n} \bar{a}_i b_i = a^{\dagger} b$$

for coefficient vectors $a, b \in \mathbb{C}^n$. This means that the functional scalar product reduces to a complex dot product. Additionally it holds that for the norm $||\cdot||$ of a function f:

$$||f|| = \langle f, f\rangle^{\frac{1}{2}} = \sqrt{a^{\dagger} a} = \sqrt{\sum_{i=1}^{n} |a_i|^2}$$

D FDA-Basics Recap

As discussed in the last section we can express a function f in its *basis function expansion* using a set of basis functions ϕ_k with $k = 1, \ldots, K$ and a set of coefficients $c_1, \ldots, c_K$

(both possibly $\mathbb{C}$ valued e.g. in the case of $2D$-curves)

$$f = \sum_{k=1}^{K} c_k \phi_k = \boldsymbol{c}'\boldsymbol{\phi}$$

where in the matrix notation $\boldsymbol{c}$ and $\boldsymbol{\phi}$ are the vectors containing the coefficients and basis functions.

When considering a sample of N functions f_i we can write this in matrix notation as

$$\boldsymbol{f} = \boldsymbol{C}\boldsymbol{\phi}$$

where C is a $(N \times K)$ matrix of coefficients and $\boldsymbol{f}$ is a vector containing the N functions.

D.1 Smoothing by Regression

When working with functional data we can usually never observe a function f directly and instead only observe discrete points (x_i, t_i) along the curve, with $f(t_i) = x_i$. As we don't know the exact functional form of f, calculating the scalar products $\langle \phi_k, f \rangle$ and therefore calculating the coefficients c_k of a given basis representation is not possible.

However, we can estimate the basis coefficients using e.g. regression analysis an approach motivated by the error model

$$f(t_i) = \boldsymbol{c}'\boldsymbol{\phi}(\boldsymbol{t_i}) + \epsilon_i$$

If we observe our function n times at $t_1, \ldots, t_n$, we can estimate the coefficients from a least squares problem, where we try to minimize the deviation of the basis expansion from the observed values. Using matrix notation let the vector $\boldsymbol{f}$ contains the observed values $f(t_i)$, $i = 1, \ldots, n$ and $(n \times k)$ matrix $\boldsymbol{\Phi}$ contains the basis function values $\phi_k(t_i)$. Then we have

$$\boldsymbol{f} = \boldsymbol{\Phi}\boldsymbol{c} + \boldsymbol{\epsilon}$$

with the estimate for the coefficient vector $\boldsymbol{c}$ given by

$$\hat{\boldsymbol{c}} = \left(\boldsymbol{\Phi}'\boldsymbol{\Phi}\right)^{-1} \boldsymbol{\Phi}'\boldsymbol{f}.$$

Spline curves fit in this way are often called *regression splines*.

D.2 Common Basis Representations

Piecewise Polynomials (Splines) Splines are defined by their range of validity, the knots, and the order. Their are constructed by dividing the area of observation into subintervals with boundaries at points called *breaks*. Over any subinterval the spline function is a polynomial of fixed degree or order. The term *degree* refers to the highest power in the polynomial while its *order* is one higher than its degree. E.g. a line has degree one but order two because it also has a constant term. **[...]**

Polygonal Basis **[...]**

D.3 Bivariate Functional Data

The analogue of covariance matrices in MVA are covariance surfaces $\sigma(s,t)$ whose values specify the covariance between values $f(s)$ and $f(t)$ over a population of curves. We can write these bivariate functions in a *bivariate basis expansion*

$$r(s,t) = \sum_{k=1}^{K} \sum_{l=1}^{K} b_{k,l} \phi_k(s) \psi_l(t) = \boldsymbol{\phi}(s)' \boldsymbol{B} \boldsymbol{\psi}(t)$$

with a $K \times K$ coefficient matrix B and two sets of basis functions ϕ_k and ψ_l using *Tensor Product Splines*

$$B_{k,l}(s,t) = \phi_k(s)\psi_l(t).$$

References

CEDERBAUM, J., F. SCHEIPL, and S. GREVEN (2018). "Fast Symmetric Additive Covariance Smoothing". In: *Computational Statistics & Data Analysis* 120, pp. 25–41.

DRYDEN, I. L. and K. V. MARDIA (2016). *Statistical Shape Analysis with Applications in R*. 2nd ed. John Wiley and Sons Ltd.

KENDALL, D. G. (1977). "The diffusion of shape". In: *Advances in Applied Probability* 9.3, pp. 428–430. DOI: 10.2307/1426091.

LUDWIG, F. et al. (2013). *Regression. Models, Methods and Applications*. Berlin, Heidelberg: Springer.

RAMSAY (2006). *Functional Data Analysis*. Placeholder.

SRIVASTAVA, A. and E. P. KLASSEN (2016). *Functional and Shape Data Analysis*. New York: Springer.

SRIVASTAVA, A., E. P. KLASSEN, et al. (2011). "Shape Analysis of Elastic Curves in Euclidean Spaces". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 33.7, pp. 1415–1428.

STEYER, L., A. STÖCKER, and S. GREVEN (2021). "Elastic Analysis of Irregularly or Sparsely Sampled Curves". In: Preprint.