

Elastic Full Procrustes Means for Sparse and Irregular Planar Curves

Masters Thesis

MANUEL PFEUFFER¹

12th October 2020, Berlin

Advisors: Lisa Maike Steyer, Almond Stöcker

1st Examiner: Prof. Dr. Sonja Greven

2nd Examiner: Prof. Dr. Nadja Klein

¹pfeufferm@hu-berlin.de, Matriculation Number: 577668

Contents

| | | |
|----------|--|-----------|
| 1 | Introduction | 3 |
| 2 | Functional and Shape Data Analysis of Planar Curves | 5 |
| 2.1 | Equivalence Classes and Shape Invariance | 5 |
| 2.2 | The SRV Framework the Elastic Distance | 8 |
| 2.3 | The Full Procrustes Mean for Planar Curves | 8 |
| 2.4 | Functional Data Analysis of Sparse and Irregular Planar Curves | 10 |
| 3 | The Elastic Full Procrustes Means for Planar Curves | 11 |
| 3.1 | The Full Procrustes mean | 12 |
| 3.2 | The Full Procrustes Mean in a fixed basis | 13 |
| 3.3 | Estimation of the covariance surface $C(s, t)$ | 15 |
| 4 | Estimation Strategy | 15 |
| 5 | Empirical Application | 15 |
| 6 | Outlook | 15 |
| 7 | Summary | 15 |
| A | Derivations and Proofs | 16 |
| A.1 | Derivation of the Full Procrustes Distance for Functional Data | 16 |
| B | Supplementary Materials | 18 |
| B.1 | Implementation Notes | 18 |
| B.2 | Replication Guide | 18 |
| C | Math-Basics Recap | 19 |
| C.1 | Scalar Products | 19 |
| C.2 | Functional Scalar Products | 19 |
| D | FDA-Basics Recap | 20 |
| D.1 | Smoothing by Regression | 21 |

| | |
|--|-----------|
| D.2 Common Basis Representations | 22 |
| D.3 Bivariate Functional Data | 22 |
| References | 23 |

1 Introduction

Statistical Shape Analysis (see e.g. DRYDEN and MARDIA 2016) 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 gained in popularity, where objects are represented using curves. This has the advantage of a more flexible representation of an object's geometry, as the analysis is not restricted to a set of discrete points along the object. Usually, the curves are themselves represented by functions $\beta : [0, 1] \rightarrow \mathbb{R}^k$, which, for example in 2D, might describe the outlines of an object in an image. As each object corresponds to one observation, this opens up a connection to the branch of statistics concerned with observations that are whole functions: Functional Data Analysis (see e.g. RAMSAY and SILVERMAN 2005).

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).*

Additionally, the functional approach to shape analysis necessitates invariance relating to parametrisation, as only the image of any function describing e.g. an object's outline, but not its parametrisation, is indicative of the object's shape.

When analysing object data, a prerequisite for many statistical methods is the ability to measure the similarity and dissimilarity of observations in the form of a distance function. SRIVASTAVA, KLASSEN, et al. 2011 introduced a mathematical framework for the analysis of shapes as curves using their square-root-velocity (SRV) representation and a so called "elastic" distance, which is invariant under re-parametrisation. While this SRV framework has been used for the calculation of elastic shape means, which also

include invariance under scaling, rotation and translation, most of these approaches focus on “Riemannian” or “geodesic” mean concepts [cite examples here? Srivastava?].

Another popular mean concept from landmark shape analysis is the *Full Procrustes Mean*, which has particularly nice properties in two dimensions, when identifying \mathbb{R}^2 with \mathbb{C} (see DRYDEN and MARDIA 2016, Chap. 8), as its calculation can be shown to be related to an eigenfunction problem of the complex covariance surface of the observed curves. This offers an advantage when working in the challenging setting of sparsely and irregularly sampled curves, as appropriate smoothing techniques for estimation of covariance surfaces in this setting are already known. Here, in particular, CEDERBAUM, SCHEIPL, and GREVEN 2018 offers a method for efficient covariance smoothing in the sparse setting.

The aim of this thesis is to extend existing methods for elastic mean estimation of sparse and irregularly sampled curves, as proposed by STEYER, A. STÖCKER, and GREVEN 2021 and implemented in the R package *elasdics* (STEYER 2021), to also include invariance with respect to rotation and scaling. The later will be achieved by generalizing the concept of the *Full Procrustes Mean* from landmark data to functional data [I should be doing more, actually], leading to the estimation of shape means in the form of *Elastic Full Procrustes Means*. To make use of the nice properties of the Procrustes mean, analysis will be restricted to the case of planar curves. Here, appropriate smoothing techniques for efficient estimation of the complex covariance surfaces from sparse data, as available in the R package *sparseFLMM* (CEDERBAUM, VOLKMANN, and A. STÖCKER 2021), will be used.

The thesis is organized as follows. [Organize this at the end.] 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

As a starting point, 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. The outline of a 2D object may be naturally represented by a planar curves $\beta : [0, 1] \rightarrow \mathbb{R}^2$ with $\beta(t) = (x(t), y(t))^T$, where $x(t)$ and $y(t)$ are the scalar-valued *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

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

which we furthermore consider as absolutely continuous and integrable. **[Maybe better define spaces here? Hilbert space? Inner product? Or do that later on SRV level?]**

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 re-parametrisation or *warping function* $\gamma : [0, 1] \rightarrow [0, 1]$ 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-parametrisation 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 classes are necessarily 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).

We now want to define an equivalence relation with respect to shape. This is usually done using the notion of *group actions* and *product groups*, to which a brief introduction may be found in SRIVASTAVA and KLASSEN 2016, Chap. 3. Let us first consider how the discussed transformations act individually on the set of parametrized planar curves with complex representation $\beta : [0, 1] \rightarrow \mathbb{C}$:

- 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 $\beta_1 \stackrel{\text{Trl}}{\sim} \beta_2$, if there exists a complex scalar $\tilde{\xi} \in \mathbb{C}$ so that $\beta_1 = \beta_2 + \tilde{\xi}$. Then, for some function β , the related equivalence class with respect to translation is given by $[\beta]_{\text{Trl}} = \{\beta + \xi \mid \xi \in \mathbb{C}\}$.
- The **scaling** group \mathbb{R}^+ acts on β by $(\lambda, \beta) \mapsto \lambda\beta$, for any $\lambda \in \mathbb{R}^+$. We define $\beta_1 \stackrel{\text{Scl}}{\sim} \beta_2$, if there exists a scalar $\tilde{\lambda} \in \mathbb{R}^+$ so that $\beta_1 = \tilde{\lambda}\beta_2$. An equivalence class is $[\beta]_{\text{Scl}} = \{\lambda\beta \mid \lambda \in \mathbb{R}^+\}$.
- 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 $\beta_1 \stackrel{\text{Rot}}{\sim} \beta_2$, if there exists a $\tilde{\theta} \in [0, 2\pi]$ with $\beta_1 = e^{i\tilde{\theta}}\beta_2$. An equivalence class is $[\beta]_{\text{Rot}} = \{e^{i\theta}\beta \mid \theta \in [0, 2\pi]\}$.
- The **warping** 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 $\beta_1 \stackrel{\text{Wrp}}{\sim} \beta_2$, if there exists a warping function $\tilde{\gamma} \in \Gamma$ with $\beta_1 = \beta_2 \circ \tilde{\gamma}$. An equivalence class is $[\beta]_{\text{Wrp}} = \{\beta \circ \gamma \mid \gamma \in \Gamma\}$.

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 $((\lambda, \theta), \beta) \mapsto \lambda e^{i\theta}\beta$. These clearly commute, as the order of applying rotation or scaling do not make a difference, as $\lambda(e^{i\theta}\beta) = e^{i\theta}(\lambda\beta)$. However, the joint actions of scaling and translation do not commute, as $\lambda(\beta + \xi) \neq \lambda\beta + \xi$, with the same holding for rotation. 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. [Maybe better define rot/scaling on the centered and retranslated curve, as in Stöcker and Greven 2021?]

Definition 2.1 (Euclidean similarity transformation). We define an *Euclidean similarity transformation* of a curve $\beta : [0, 1] \rightarrow \mathbb{C}$ 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 2.2 (Shape). The *shape* of a planar curve $\beta : [0, 1] \rightarrow \mathbb{C}$ is given by its equivalence class with respect to all Euclidean similarity transformations and re-parametrizations

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

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

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

where the symbol “ \rtimes ” denotes a semi-direct product, i.e. that the translation group acts “after” scaling and rotation [Note: this notation is crap.] (for details see SRIVASTAVA and KLASSEN 2016, Chapter 3).

[This doesn’t lead well to the next subsection.] As we are ultimately interested in calculating shape means, we now need to find a way to measure similarity and dissimilarity between equivalence classes in the form of a distance function. 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 representations in curve space. In short we want to find intersections of the function space, so that there is a one-to-one relationship between elements of that intersection and the equivalence classes in shape space.

2.2 The SRV Framework the Elastic Distance

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

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

where the original curve β is assumed absolutely continuous. It can be obtained from its SRV representation up to translation by the back transformation $\beta(t) = \beta(0) + \int_0^t q(s) \|q(s)\| ds$.

Moreover, if the original curve is of unit length the SRV curve will be automatically normalized **[Keep this for later]**:

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

2.3 The Full Procrustes Mean for Planar Curves

Consider, for example, the action of the translation group \mathbb{C} with corresponding quotient space $\mathbb{C}^{[0,1]}/\mathbb{C} = \{[\beta] \mid \beta \in \mathbb{C}^{[0,1]}\}$ and $[\beta] = \{\beta + \xi \mid \xi \in \mathbb{C}\}$. For each equivalence class in $\mathbb{C}^{[0,1]}/\mathbb{C}$ there is exactly one element in the space of all centered curves $\mathcal{C}_1 = \{\beta \in \mathbb{C}^{[0,1]} \mid \int_0^1 \beta(t) dt = 0\}$ 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 $\mathcal{C}_2 = \{\beta \in \mathbb{C}^{[0,1]} \mid \beta \text{ continuously differentiable, } \int_0^1 |\dot{\beta}(t)| dt = 1\}$ 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 2.3 (Pre-shape). *The **pre-shape** of a continuously differentiable planar curve $\beta : [0, 1] \rightarrow \mathbb{C}$ is given by its centered unit-length representation*

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

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

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

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.

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

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.

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 variability, 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 2.4 (Full Procrustes distance). *The **full Procrustes distance** between the shapes $[\beta_1], [\beta_2]$ of two continuously differentiable β_1, β_2 with $\beta_i : [0, 1] \rightarrow \mathbb{C}$ 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 2.5 (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.]

2.4 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

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

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

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

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

For a set of planar curves $\beta_1, \dots, \beta_n : [0, 1] \rightarrow \mathbb{C}$, 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

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

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

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

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 $q_1, \dots, q_n : [0, 1] \rightarrow \mathbb{C}$ of unit length $\|q_i\| = 1$ for all i . The *full Procrustes mean* $\hat{\mu}$ is given by

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

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 $b_j : \mathbb{R} \rightarrow \mathbb{R}$ 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 $\hat{\mu}(t) = \sum_{j=1}^k \hat{\theta}_j b_j(t) = b^T \hat{\theta}$:

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

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 $\hat{C}(s, t) = \sum_{k,l} \hat{\xi}_{kl} b_k(t) \overline{b_l(s)}$, 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} , $i, j = 1, \dots, k$ 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 matrix S as a function of the estimated coefficient matrix $\hat{\Xi} = (\hat{\xi}_{ij})_{i,j=1,\dots,k}$:

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

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

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

[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 $\beta_1, \beta_2 : [0, 1] \rightarrow \mathbb{C}$ 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

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

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 $\omega = be^{i\theta} \in \mathbb{C}$:

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

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

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

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

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

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

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

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

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 $B = (b_1, \dots, b_n)$, 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:

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

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 $i = 1, \dots, n$ with x^\dagger indicating the conjugate transpose of $x = (x_1, \dots, x_n)^T$. 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 $\{\phi_1, \dots, \phi_n\}$ (and possibly $n \rightarrow \infty$) 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:

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

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 $k = 1, \dots, K$ and a set of coefficients c_1, \dots, c_K

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

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

where in the matrix notation \mathbf{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

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

where \mathbf{C} is a $(N \times K)$ matrix of coefficients and \mathbf{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

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

If we observe our function n times at t_1, \dots, 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 \mathbf{f} contains the observed values $f(t_i)$, $i = 1, \dots, n$ and $(n \times k)$ matrix $\boldsymbol{\Phi}$ contains the basis function values $\phi_k(t_i)$. Then we have

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

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

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

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. They 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*

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

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

- ANDERSON, C. R. (1997). *Object recognition using statistical shape analysis*. PhD thesis. University of Leeds.
- CEDERBAUM, J., F. SCHEIPL, and S. GREVEN (2018). “Fast Symmetric Additive Covariance Smoothing”. In: *Computational Statistics & Data Analysis* 120, pp. 25–41.
- CEDERBAUM, J., A. VOLKMAN, and A. STÖCKER (2021). *sparseFLMM: Functional Linear Mixed Models for Irregularly or Sparsely Sampled Data*. R package version 0.4.0. URL: <https://CRAN.R-project.org/package=sparseFLMM>.
- DRYDEN, I. L. (2019). *shapes package*. Contributed package, Version 1.2.5. R Foundation for Statistical Computing. Vienna, Austria. URL: <http://www.R-project.org>.
- 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, J. and B. W. SILVERMAN (2005). *Functional Data Analysis*. Springer series in statistics.
- 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. (2021). *elasdics: Elastic Analysis of Sparse, Dense and Irregular Curves*. R package version 0.1.1. URL: <https://CRAN.R-project.org/package=elasdics>.
- STEYER, L., A. STÖCKER, and S. GREVEN (2021). “Elastic analysis of irregularly or sparsely sampled curves”. In: Preprint.
- STÖCKER, A and S. GREVEN (2021). “Functional additive regression on shape and form manifolds of planar curves”. In: Preprint.
- WOOD, S.N (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.