Chair of Statistics · School of Business and Economics · Humboldt-Universität zu Berlin

Elastic Full Procrustes Means for Sparse and Irregular Planar Curves

Masters Thesis

in partial fulfillment of the requirements for the degree

M.Sc. Statistics

Manuel Pfeuffer[1]

20th September 2021, Berlin

Advisors: Lisa Maike Steyer, Almond Stöcker

1st Examiner: Prof. Dr. Sonja Greven

2nd Examiner: Prof. Dr. Nadja Klein

[1] `mnl.pfeuffer@gmail.com`, Matriculation number: 577668

Contents

1. Introduction **1**

2. Elastic Full Procrustes Means for Planar Curves **3**

2.1. Equivalence Classes and Shape Invariance 3

2.2. The Elastic Full Procrustes Distance for Planar Curves 6

2.3. The Elastic Full Procrustes Mean for Planar Curves 11

3. Estimation Strategy for Sparse and Irregular Observations **13**

3.1. Efficient Estimation using Hermitian Covariance Smoothing 13

3.2. Estimating the Full Procrustes Mean in a Fixed Basis 13

3.3. Numerical Integration of the Procrustes Fits 15

4. Empirical Applications **16**

4.1. Mean Estimation for Simulated Spirals 16

4.2. Classification of Hand-written Digits 16

4.3. Mean Differences of Tounge Shapes in a Phonetics Study 16

5. Summary **17**

Bibliography **18**

A. Appendix **20**

A.1. Additional Proofs and Derivations 20

A.2. Discussion of Possible Extensions to Closed Curves 21

A.3. Shape-Smoothing Using the Estimated Covariance-Surface 21

B. Supplementary Materials **22**

B.1. Dataset Replication Guide 22

B.2. Implementation Notes 22

1. Introduction

Statistical Shape Analysis (see e.g. Dryden and Mardia 2016) is a branch of statistics concerned with modelling the geometry of objects. Examples might be outlines of bones and organs, handwritten digits, or the folds of a protein. To capture an object's geometrical information, a common approach is the use of *landmarks*, characteristic points on an object, that 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 fixed set of discrete points. The curves are usually themselves represented by functions $\beta : [0,1] \rightarrow \mathbb{R}^k$, which, for example for $k = 2$, 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).

Differences in location, rotation, and size are often not of interest, when analyzing the geometry of objects. 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).

When considering the shapes of curves, one has to additionally take into account effects relating to re-parametrisation, as only the image of any function describing e.g. an object's outline, but not its parametrisation, is indacative of the object's shape.

A prerequisite for many statistical methods is the ability to measure distances between observations. Srivastava, Klassen, et al. 2011 introduced a mathematical framework for analysing the shape of curves, by using their square-root-velocity (SRV) representation and an elastic metric, which is isometric under re-parametrisation. While this SRV framework has been used for the calculation of elastic shape means before, which also include invariance under scaling, rotation and translation, most of these approaches focus on "Riemannian" or "geodesic" mean concepts **[cite]**.

The *Full Procrustes Mean* is a different shape mean concept, which is widely used when working with landmark data, and which has particularly nice properties in two dimensions, when identifying $\mathbb{R}^2$ with $\mathbb{C}$ (see DRYDEN and MARDIA 2016, Chap. 8). When working with planar curves, 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 to functional data and by iteratively applying full Procrustes mean estimation, rotation-alignment and parametrization-alignment, leading to the estimation of *Elastic Full Procrustes Means*. To make use of the nice properties of the Procrustes mean in two dimensions, analysis will be restricted to the case of planar curves. Here, appropriate smoothing techniques for sparse estimation of the complex covariance surfaces, as available in the `R` package `sparseFLMM` (CEDERBAUM, VOLKMANN, and A. STÖCKER 2021), will be used.

The thesis is organized as follows. **[Update this later.]** After covering the relevant background material and deriving an expression for the elastic full Procrustes mean in Section 2. An estimation strategy for the setting of sparse and irregular curves will be proposed in Section 3. The methods will be verified using simulated and empirical datasets in Section 4. Finally, all results will be summarized in section 5. Appendix A and Supplements B offer additional considerations and reproducability guides.

2. Elastic Full Procrustes Means for 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 $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x(t), y(t))^T$, where $x(t)$ and $y(t)$ are the scalar-valued *coordinate functions*. Calculations in two dimensions, 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

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

For reasons that will be discussed in Section 2.2, we furthermore assume the curves to be absolutely continuous or $\beta \in \mathcal{AC}([0,1], \mathbb{C})$. All considerations will be restricted to the case of open curves, with possible extensions to closed curves $\beta \in \mathcal{AC}(\mathbb{S}^1, \mathbb{C})$ discussed in Section A.2 of the appendix.

2.1. Equivalence Classes and Shape Invariance

As mentioned in the introduction, shape is usually defined by its invariance under the transformations of scaling, translation, and rotation. When considering the shape of curves, 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* $\gamma : [0,1] \to [0,1]$ 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-parametrisation are *equivalence relations* with respect to shape, as each action leaves the shape of the curve untouched and only changes the way it 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).

When defining an equivalence class, one has to first consider how the individual transformations act on a planar curve with complex representation $\beta : [0, 1] \rightarrow \mathbb{C}$. This is usually done using the notion of *group actions* and *product groups*, with the later describing multiple transformations acting at once. A brief introduction to group actions may be found in Srivastava and Klassen 2016, Chap. 3.

1. The *translation* group $\mathbb{C}$ acts on β by $(\xi, \beta) \xmapsto{\text{Trl}} \beta + \xi$, for any $\xi \in \mathbb{C}$. We can consider two curves as equivalent with respect to translation $\beta_1 \overset{\text{Trl}}{\backsim} \beta_2$, if there exists a complex scalar $\widetilde{\xi} \in \mathbb{C}$ so that $\beta_1 = \beta_2 + \widetilde{\xi}$. Then, for some function β, the related equivalence class with respect to translation is given by $[\beta]_{\text{Trl}} = \{\beta + \xi \,|\, \xi \in \mathbb{C}\}$.

2. The *scaling* group $\mathbb{R}^+$ acts on β by $(\lambda, \beta) \xmapsto{\text{Scl}} \lambda\beta$, for any $\lambda \in \mathbb{R}^+$. We define $\beta_1 \overset{\text{Scl}}{\backsim} \beta_2$, if there exists a scalar $\widetilde{\lambda} \in \mathbb{R}^+$ so that $\beta_1 = \widetilde{\lambda}\beta_2$. An equivalence class is $[\beta]_{\text{Scl}} = \{\lambda\beta \,|\, \lambda \in \mathbb{R}^+\}$.

3. The *rotation* group $[0, 2\pi]$ acts on β by $(\theta, \beta) \xmapsto{\text{Rot}} e^{i\theta}\beta$, for any $\theta \in [0, 2\pi]$. We define $\beta_1 \overset{\text{Rot}}{\backsim} \beta_2$, if there exists a $\widetilde{\theta} \in [0, 2\pi]$ with $\beta_1 = e^{i\widetilde{\theta}}\beta_2$. An equivalence class is $[\beta]_{\text{Rot}} = \{e^{i\theta}\beta \,|\, \theta \in [0, 2\pi]\}$.

4. The *warping* group Γ acts on β by $(\gamma, \beta) \xmapsto{\text{Wrp}} \beta \circ \gamma$, for any $\gamma \in \Gamma$ with Γ being the set of monotonically increasing and differentiable warping functions. We define $\beta_1 \overset{\text{Wrp}}{\backsim} \beta_2$, if there exists a warping function $\widetilde{\gamma} \in \Gamma$ with $\beta_1 = \beta_2 \circ \widetilde{\gamma}$. An equivalence class is $[\beta]_{\text{Wrp}} = \{\beta \circ \gamma \,|\, \gamma \in \Gamma\}$.

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 $((\lambda, \theta), \beta) \xmapsto{\text{Scl+Rot}} \lambda e^{i\theta}\beta$. These clearly commute, because 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 the joint action of rotation and translation. As the order of translating and rotating or scaling matters, one usually takes the translation to act on the already scaled and rotated curve. The joint action defined using this ordering is usually called an *Euclidean similarity transformation*.

Definition 2.1 (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 2.2 (Shape)**.** The *shape* of an absolutely continous planar curve $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ is given by its equivalence class with respect to all Euclidean similarity transformations and re-parametrisations

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

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

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

where the symbol "$\rtimes$" denotes a semi-direct product, i.e. the translation group acts "after" scaling and rotation. **[This is probably not super clear.]**

2.2. The Elastic Full Procrustes Distance for Planar Curves

Let us now turn to the calculatation of distances between the shapes of curves. As shapes are represented by certain equivalence classes, and are therefore elements of a non-Euclidean quotient space, calculating their distance is not straight-forward. A common approach is to "project" the distance calculation in shape space down into the underlying functional space. For example, consider β_1, $\beta_2 \in \mathbb{L}^2([0,1], \mathbb{C})$ with $[\beta_1]$, $[\beta_2]$ their equivalence classes with respect to all shape-preserving transformations. We might want to calculate their shape-distance as the minimal $\mathbb{L}^2$-distance, when optimizing over all elements of their respective equivalence classes:

$$d([\beta_1],[\beta_2]) = \inf_{\widetilde{\beta}_1 \in [\beta_1], \widetilde{\beta}_2 \in [\beta_2]} d_{\mathbb{L}^2}(\widetilde{\beta}_1, \widetilde{\beta}_2) = \inf_{\widetilde{\beta}_1 \in [\beta_1], \widetilde{\beta}_2 \in [\beta_2]} \|\widetilde{\beta}_1 - \widetilde{\beta}_2\|_{\mathbb{L}^2}.$$

Which is equivalent to optimizing over all shape-preserving transformations:

$$d([\beta_1],[\beta_2]) = \inf_{\lambda_{1,2} \in \mathbb{R}^+, \theta_{1,2} \in [0,2\pi], \xi_{1,2} \in \mathbb{C}, \gamma_{1,2} \in \Gamma} \Big\| \lambda_1 e^{i\theta_1}(\beta_1 \circ \gamma_1) + \xi_1 - \left(\lambda_2 e^{i\theta_2}(\beta_2 \circ \gamma_2) + \xi_2\right) \Big\|_{\mathbb{L}^2}.$$

However, this approach runs into problems, when considering whether all transformations act by isometries on this distance, i.e. whether equally changing the translation, rotation, scaling or re-parametrization of both curves affects their distance.

As it turns out, neither re-parametrization nor scaling are distance preserving when using the $\mathbb{L}^2$-distance: For two equally re-parameterized curves $\widetilde{\beta}_{1,2} = \beta_{1,2} \circ \gamma$, their squared $\mathbb{L}^2$-distance is given by $\|\beta_1 \circ \gamma - \beta_2 \circ \gamma\|^2 = \int_0^1 \|\beta_1(\gamma(t)) - \beta_2(\gamma(t))\|^2 dt = \int_0^1 \|\beta_1(s) - \beta_2(s)\|^2 \frac{1}{\dot{\gamma}(\gamma^{-1}(s))} ds$ with $s = \gamma(t)$. It follows that $\|\widetilde{\beta}_1 - \widetilde{\beta}_2\| \neq \|\beta_1 - \beta_2\|$, as in general $\dot{\gamma}(\gamma^{-1}(s)) \neq 1$. Likewise, it holds for equal re-scaling that $\|\lambda\beta_1 - \lambda\beta_2\| = |\lambda| \, \|\beta_1 - \beta_2\| \neq \|\beta_1 - \beta_2\|$. This has the consequence that the above optimization problem might simply be solved by $\lambda_1, \lambda_2 \to 0$, leading to $d([\beta_1],[\beta_2]) \to 0$ for any two curves β_1, β_2. Furthermore, optimizing over re-parametrization using the $\mathbb{L}^2$-distance has problems relating to the so called *pinching effect* and *inverse-inconsistency*, where the later means that aligning the parametrisation of one curve to another by $\inf_{\gamma \in \Gamma} \|\beta_1 - \beta_2 \circ \gamma\|$ may yield different results than $\inf_{\gamma \in \Gamma} \|\beta_2 - \beta_1 \circ \gamma\|$ (see Srivastava and Klassen 2016, p. 88).

A solution proposed in Srivastava, Klassen, et al. 2011 is to ditch the $\mathbb{L}^2$-metric in favor of an *elastic metric*, which is isometric with respect to re-parametrization. Calculation of this metric, the Fisher-Rao Riemannian metric (Rao 1945), can be greatly simplified by using the *square-root-velocity* (SRV) framework, as the Fisher-Rao metric of two curves can be equivalently calculated as the $\mathbb{L}^2$-distance of their respective SRV curves.

Definition 2.3 (SRV function of a planar curve)**.** The *SRV function* (SRVF) of an absolutely continuous planar curve $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ is given by

$$q(t) = \frac{\dot{\beta}(t)}{\sqrt{||\dot{\beta}(t)||}} \quad \text{for } \dot{\beta}(t) \neq 0, \text{ with } q \in \mathbb{L}^2([0,1], \mathbb{C}),$$

where the original curve β can be re-constructed from its SRVF, up to translation, by $\beta(t) = \beta(0) + \int_0^t q(s)||q(s)||ds$.

As this representation makes use of derivatives, any curve β that has a SRVF must fulfill some kind of differentiability constraint. Here it is enough to consider only curves that are absolutely continuous $\beta \in \mathcal{AC}([0,1], \mathbb{C})$. In particular, this means that the original curves do not have to be smooth but might also be piecewise linear (see Srivastava and Klassen 2016, p. 91). The SRVFs are considered elements of a Hilbert space, which is given by $\mathcal{L}^2([0,1], \mathbb{C})$ equipped with the complex inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $|| \cdot ||$. The complex inner product of $q, q' \in \mathcal{L}^2([0,1], \mathbb{C})$ is defined as

$$\langle q, q' \rangle = \int_0^1 \overline{q(t)} q'(t) dt \, ,$$

with $\overline{z} = \mathrm{Re}(z) - i \, \mathrm{Im}(z)$ denoting the complex conjugate.

As we can always recover the original curve up to translation, the SRV representation holds all relevant information about the shape of a curve. Furthermore, because of the use of derivatives, the SRV representation is invariant under changes in translation of the original curve. As a consequences, all shape-preserving transformations commute on SRV level.

Lemma 2.1. *The actions of the translation, scaling, rotation, and re-parametrization groups commute on SRV level.*

Proof. The SRVF $\widetilde{q}(t)$ of $\widetilde{\beta}(t) = \lambda e^{i\theta}\beta(\gamma(t)) + \xi$ is given by

$$\widetilde{q}(t) = \frac{\lambda e^{i\theta}\dot{\beta}(\gamma(t))\,\dot{\gamma}(t)}{\sqrt{||\lambda e^{i\theta}\dot{\beta}(\gamma(t))\,\dot{\gamma}(t)||}} = \sqrt{\lambda}e^{i\theta}\frac{\dot{\beta}(\gamma(t))}{\sqrt{||\dot{\beta}(\gamma(t))||}}\sqrt{\dot{\gamma}(t)} = \sqrt{\lambda}e^{i\theta}(q\circ\gamma)\sqrt{\dot{\gamma}(t)}.$$

The result is irrespective of the order of applying the transformations. □

Remark. It follows, that the individual transformations translate to SRV level by

$$\text{i.) } (\xi, q) \overset{\text{Trl}}{\longmapsto} q, \quad \text{ii.) } (\lambda, q) \overset{\text{Scl}}{\longmapsto} \sqrt{\lambda}q, \quad \text{iii.) } (\theta, q) \overset{\text{Rot}}{\longmapsto} e^{i\theta}q, \quad \text{iv.) } (\gamma, q) \overset{\text{Wrp}}{\longmapsto} (q\circ\gamma)\sqrt{\dot{\gamma}}.$$

We can use this to identify the shape of a curve as the equivalence class of its respective SRVF modulo scaling, rotation, and warping, where we do not need to consider translation anymore.

Definition 2.4 (Shape (SRV)). The *shape* $[\beta]$ of an absolutely continuous, planar curve $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ can be equivalently represented by the equivalence class $[q]$ of its SRV representation $q = \frac{\dot{\beta}(t)}{\sqrt{|\dot{\beta}(t)|}} \in \mathbb{L}^2([0,1], \mathbb{C})$ modulo scaling, rotation and re-parametrization

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

Going forward, we will now work in the SRV framework and use the elastic metric for distance calculations between shapes. This means, instead of optimizing over the $\mathbb{L}^2$ distance between the original curves, we optimize over the $\mathbb{L}^2$ distance between their respective SRVFs. For two absolutely continuous curves $\beta_1, \beta_2 \in \mathcal{AC}([0,1], \mathbb{C})$ with respective SRV curves $q_1, q_2 \in \mathbb{L}^2([0,1], \mathbb{C})$ we might define the *elastic* distance between their shapes as

$$d([\beta_1], [\beta_2]) = \inf_{\widetilde{q}_1 \in [q_1], \widetilde{q}_2 \in [q_2]} ||\widetilde{q}_1 - \widetilde{q}_2||_{\mathbb{L}^2}$$

or when equivalently optimizing over all possible transformations as

$$d([\beta_1], [\beta_2]) = \inf_{\lambda_{1,2} \in \mathbb{R}^+, \theta_{1,2} \in [0,2\pi], \gamma_{1,2} \in \Gamma} ||\sqrt{\lambda_1}e^{i\theta_1}(q_1\circ\gamma_1)\sqrt{\dot{\gamma}_1} - \sqrt{\lambda_2}e^{i\theta_2}(q_2\circ\gamma_2)\sqrt{\dot{\gamma}_2}||_{\mathbb{L}^2}.$$

We can simplify the optimization, by considering that both rotation and warping act by isometries on the elastic metric, which means in practice it is enough to optimize over the rotation and warping of only one of the curves. For warping, we can reformulate the optimisation as a problem over the relative parametrisation between the curves, as $\inf_{\gamma_1,\gamma_2\in\Gamma}\|q_1\circ\gamma_1 - q_2\circ\gamma_2\| = \inf_{\gamma_1,\gamma_2\in\Gamma}\|q_1 - (q_2\circ(\gamma_2\circ\gamma_1^{-1}))\sqrt{(\gamma_2\circ\dot{\gamma}_1^{-1})}\| = \inf_{\gamma\in\Gamma}\|q_1 - (q_2\circ\gamma)\sqrt{\dot{\gamma}}\|$. And smilarly for rotation, we can optimize over the relative rotation between both curves, as $\inf_{\theta_1,\theta_2\in[0,2\pi]}\|e^{i\theta_1}q_1 - e^{i\theta_2}q_2\| = \inf_{\theta_1,\theta_2\in[0,2\pi]}\|q_1 - e^{i(\theta_2-\theta_1)}q_2\|) = \inf_{\theta\in[0,2\pi]}\|q_1 - e^{i\theta}q_2\|$. Taken together these lead to

$$d([\beta_1],[\beta_2]) = \inf_{\lambda_{1,2}\in\mathbb{R}^+,\,\theta\in[0,2\pi],\,\gamma\in\Gamma}\|\sqrt{\lambda_1}q_1 - \sqrt{\lambda_2}e^{i\theta}(q_2\circ\gamma)\sqrt{\dot{\gamma}}\|\,,$$

which, however, still has the problem of not being isometric with respect to scaling.

A possible solution, mirroring the definition of the *full Procrustes distance* for landmark data, is to work with the normalized representations $z = \frac{q}{\|q\|}$, while only aligning the scaling of one of the curves. This leads to a distance that is isometric with respect to scaling, as for $\widetilde{q}_{1,2} = \lambda q_{1,2}$ it holds that $\|\widetilde{z}_1 - \widetilde{z}_2\| = \|\frac{\lambda q_1}{\|\lambda q_1\|} - \frac{\lambda q_2}{\|\lambda q_2\|}\| \overset{\lambda\in\mathbb{R}^+}{=} \|z_1 - z_2\|$, while also being inverse consistent, as $\inf_{\lambda\in\mathbb{R}^+}\|z_1 - \lambda z_2\|^2 = \inf_{\lambda\in\mathbb{R}^+}\|z_1\|^2 + \lambda^2\|z_2\|^2 - \lambda(\langle z_1, z_2\rangle + \langle z_2, z_1\rangle) = 1 + \lambda^2 - \lambda(\langle z_1, z_2\rangle + \langle z_2, z_1\rangle) = \inf_{\lambda\in\mathbb{R}^+}\|z_2 - \lambda z_1\|^2$. We can finally take everything together, leading to a definition for the *elastic full Procrustes distance*.

Definition 2.5 (Elastic full Procrustes distance)**.** The *elastic full Procrustes distance* between the shapes $[\beta_1]$, $[\beta_2]$ of two continuously differentiable planar curves β_1, $\beta_2 \in \mathcal{AC}([0,1],\mathbb{C})$ is given by

$$d_{EF}([\beta_1],[\beta_2]) = \inf_{\lambda\in\mathbb{R}^+,\,\theta\in[0,2\pi],\,\gamma\in\Gamma}\|z_1 - \lambda e^{i\theta}(z_2\circ\gamma)\sqrt{\dot{\gamma}}\|$$

with normalized SRV representation $z_{1,2} = \frac{q_{1,2}}{\|q_{1,2}\|} \in \mathbb{L}^2([0,1],\mathbb{C})$.

Remark. If the original curve β is of unit length $L[\beta] = \int_0^1|\dot{\beta}(t)|dt = 1$, the SRV curve $q = \frac{\dot{\beta}}{\|\dot{\beta}\|}$ will be automatically normalized, as $\|q\| = \sqrt{\int_0^1 |q(t)|^2\,dt} = \sqrt{\int_0^1 |\dot{\beta}(t)|\,dt} = 1$.

First, we consider the optimisation over rotation and scaling. For a fixed $\gamma\in\Gamma$, the optimisation problem in Definition 2.5 mirrors the optimisation problem of the full Procrustes distance for landmark data, where an explicit solution is known in the planar case (see Dryden and Mardia 2016, Chapter 8).

Lemma 2.2. *The optimisation problem in Definition 2.5 can be reduced to*

$$d_{EF}([\beta_1],[\beta_2]) = \inf_{\gamma\in\Gamma} \sqrt{1 - \langle z_1, (z_2\circ\gamma)\sqrt{\dot{\gamma}}\rangle\langle (z_2\circ\gamma)\sqrt{\dot{\gamma}}, z_1\rangle}$$

Proof. See A.1.1 in the appendix. □

2.3. The Elastic Full Procrustes Mean for Planar Curves

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

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

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

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 $\beta_1, \dots, \beta_n : [0,1] \to \mathbb{C}$, 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.1.1.

Consider a set of planar SRV curves $q_1, \dots, q_n : [0,1] \to \mathbb{C}$ 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. Estimation Strategy for Sparse and Irregular Observations

3.1. Efficient Estimation using Hermitian Covariance Smoothing

Consider the following model for independent curves

$$Y_i(t_{ij}) = \mu(t_{ij}, \mathbf{x}_i) + E_i(t_{ij}) + \epsilon(t_{ij}), \quad j = 1, \dots, D_i,\ i = 1, \dots, n, \tag{3.1}$$

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

3.2. Estimating 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 $b_j(t)$, $j = 1, \ldots, k$ 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. Numerical Integration of the Procrustes Fits

4. Empirical Applications

4.1. Mean Estimation for Simulated Spirals

4.2. Classification of Hand-written Digits

4.3. Mean Differences of Tounge Shapes in a Phonetics Study

5. Summary

Bibliography

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. VOLKMANN, 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.

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.

RAO, C. R. (1945). "Information and accuracy attainable in the estimation of statistical parameters". In: *Bulletin of Calcutta Mathematical Society* 37, pp. 81–91.

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*. arXiv: 2104.11039 [stat.ME].

Stöcker, A and S. Greven (2021). *Functional additive regression on shape and form manifolds of planar curves*. arXiv: 2109.02624 [stat.ME].

Wood, S.N (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.

A. Appendix

A.1. Additional Proofs and Derivations

A.1.1. 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 $\langle \beta_1, \beta_2 \rangle = ae^{i\phi}$:

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

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:

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

A.2. Discussion of Possible Extensions to Closed Curves

A.3. Shape-Smoothing Using the Estimated Covariance-Surface

B. Supplementary Materials

B.1. Dataset Replication Guide

B.2. Implementation Notes

—Discarded—

Math-Basics Recap

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 $G = \mathbb{1}_{n \times n}$.

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:

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

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.

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*.

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 **[...]**

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*

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