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]

17th November, 2021, Berlin

Advisors: Lisa 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 6

2.1. Equivalence Classes and Shape 6

2.2. The Elastic Full Procrustes Distance for Planar Curves 9

2.3. The Elastic Full Procrustes Mean for Planar Curves 17

3. Estimation Strategy for Sparse and Irregular Observations 22

3.1. Discrete Treatment of SRV Curves 22

3.2. Efficient Estimation using Hermitian Covariance Smoothing 24

3.3. Estimating the Elastic Full Procrustes Mean in a Fixed Basis 25

3.4. Numerical Integration of the Procrustes Fits 27

4. Empirical Applications 28

4.1. Mean Estimation for Simulated Spirals 28

4.2. Classification of Hand-written Digits 28

4.3. Mean Differences of Tounge Shapes in a Phonetics Study 28

5. Summary 29

Bibliography 30

A. Appendix 32

A.1. Proofs and Derivations 32

A.2. Discussion of Possible Extensions to Closed Curves 33

A.3. Shape-Smoothing Using the Estimated Covariance-Surface 33

B. Supplementary Materials 34

B.1. Dataset Replication Guide 34

B.2. Implementation Notes 34

Todo list

Write abit more here. Citations! 13
Citation, write more here! 14
Write iterative procedure as algorithm! Basically as in Srivastava, Klassen, et al. 2011. 16
Nummerierung der Equations überarbeiten 17
Figure: Example original Covariance surface (real part) 19
Figure: Example SRV Covariance surface (real part 19
Figure: Example original Covariance surface (imaginary part) 19
Figure: Example SRV Covariance surface (imaginary part 19
Create figures for everything. 22
Figure: Sparsely observed digit 3. 24
Figure: Sparse SRV of digit 3. 24
Rewrite proof for on curve level. 32

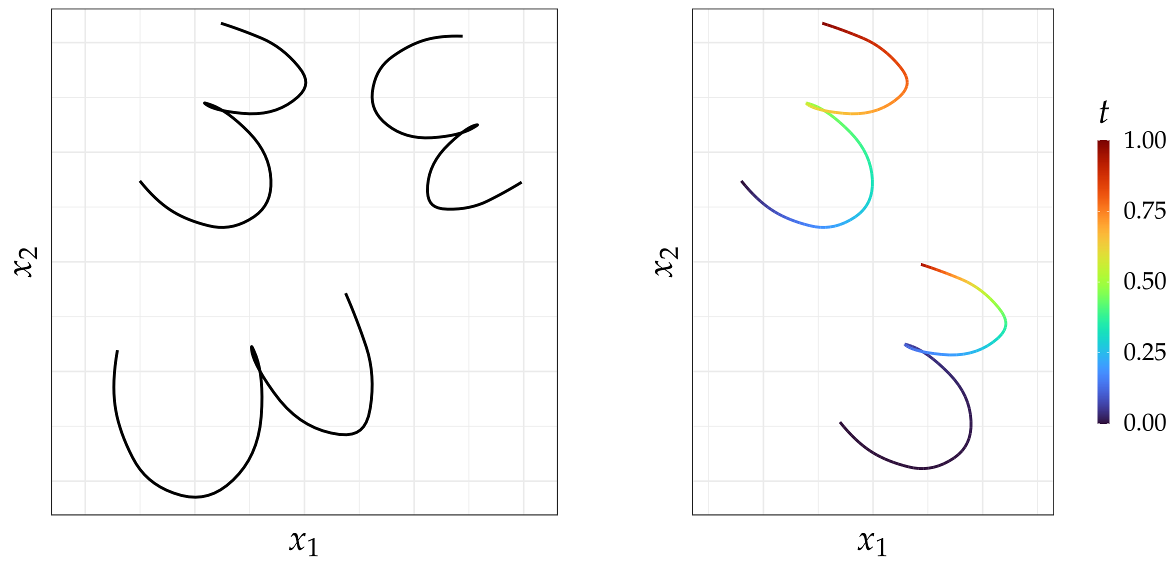
1. Introduction

sec:1

Statistical Shape Analysis (see e.g. Dryden and Mardia 2016) is the branch of statistics concerned with modelling geometrical information. Such information might come in the form of outlines of bones or organs in a medical image, traced points along a handwritten digit, or data on the folding structure of a protein. This data is commonly captured using *landmarks*, which are characteristic points on the objects of interest that "match between and within populations"(Dryden and Mardia 2016, p. 3). As an example, we might geometrically compare a set of mouse vertebrae by comparing the coordinates of prominent points along the bone outlines, which are common between all mouse vertebrae. More formally, we could say that each mouse vertebra's geometrical information is then represented by a landmark configuration $X \in \mathbb{R}^{k \times d}$, which is the stacked matrix of the k d-dimensional landmark coordinates, allowing for a multivariate treatment of shape or geometrical form.

A more flexible approach might be to treat e.g. the outline of an object as a whole, represented in the form of a continous curve $\beta : [0,1] \to \mathbb{R}^d$. Landmarks have the drawback that there is no clear way of choosing which points to include in the configuration, leaving the decision up to the subjectivity of the researcher. Furthermore, using landmarks leads to an inherently discrete treatment of the available data, which means modes of variation that lie in between landmarks may not be picked up by the analysis. By using curves, the analysis is not restricted to a fixed set of k discrete points, but instead uses all the available information. At the same time, the subjectivity in choosing the landmarks is eliminated. As each object then 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 analyzing 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



(a) The same digit with randomized rotation, scaling and translation applied three times.

fig:1-eucl

fig:1-warp

(b) Original (left) and re-parameterized digit (right; $t \mapsto t^5$). Color indicates the value of the parametrization $t \in [0,1]$ at point $\beta(t)$.

Figure 1.1.: Several representations of the same shape. Data: `digits3.dat` from the `shapes` package (Dryden 2019) for the R programming language (R Core Team 2021) with smoothing applied using methods discussed in Appendix A.3. Original dataset collected by Anderson 1997.

ig:1-shape

formulated in the following way: "[A]ll the geometrical information that remains when location, scale and rotational effects are removed from an object"(Dryden and Mardia 2016, p. 1). This is illustrated in Figure 1.1a, where the same shape of a handwritten digit '3' is plotted in three different orientations and sizes. When considering the shape of a curve $\beta : [0,1] \to \mathbb{R}^d$, one has to additionally take into account effects relating to the parametrization $t \in [0,1]$. As illustrated in Figure 1.1b, curves $\beta(t)$ and $\beta(\gamma(t))$, with some re-parametrization or *warping function* $\gamma : [0,1] \to [0,1]$ monotonically increasing and differentiable, have the same image and therefore represent the same shape as well.

A pre-requisite for any statistical analysis of shape is the ability to calculate a distance between and to estimate a mean from observations, in a fashion that does not depend on location, rotation, scale and/or parametrization of the input. In this thesis two established approaches to shape analysis will be combined: Firstly, the *full Procrustes distance* and *mean* are widely used for translation-, rotation-, and scaling-invariant analysis of landmark data (see e.g. Dryden and Mardia 2016, Chap. 4, 6). Secondly, Srivastava, Klassen, et al. 2011 introduced a mathematical framework for the *elastic*

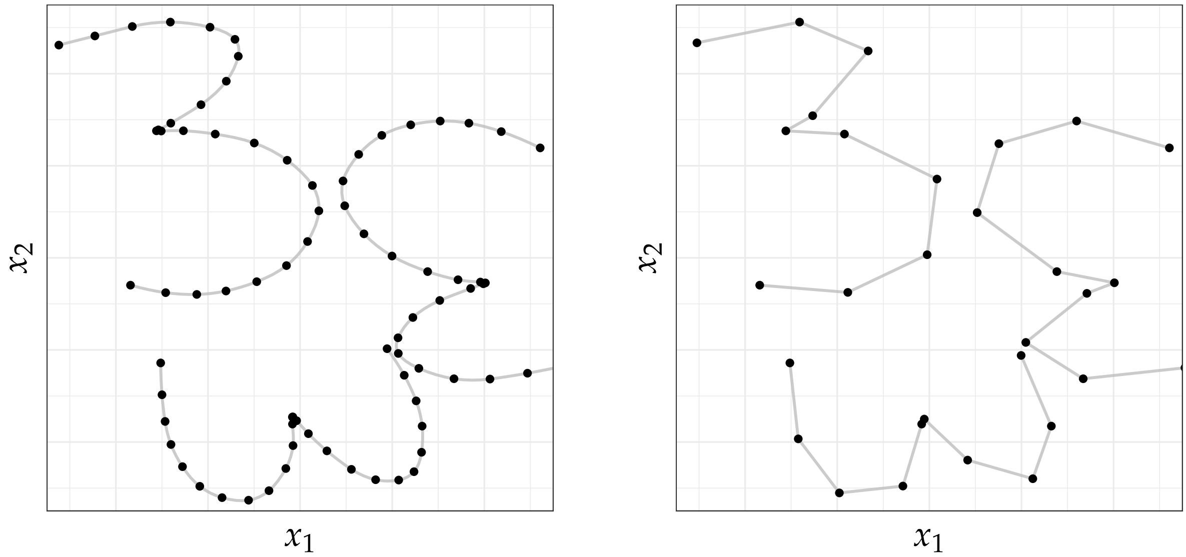


Figure 1.2.: Dense (left) and sparse (right) observations of the same three digits. Data: see Figure 1.1, with the smooth curves sampled on a dense (left) and sparse (right) grid.

g:1-sparse

(re-parametrization invariant) shape analysis of curves, by using their square-root-velocity (SRV) representations. Taken together, both approaches allow for analysing curves in a fashion that is invariant to all four shape-preserving transformations, leading to an *elastic full Procrustes distance* and *mean*. As the full Procrustes mean has particularly nice properties in two dimensions, when identifying $\mathbb{R}^2$ with $\mathbb{C}$ (see Dryden and Mardia 2016, Chap. 8), this thesis will be restricted to the case of planar curves.

While we are interested in modelling a (planar) object's geometrical information as a continous curve $\beta : [0,1] \to \mathbb{R}^2$, the curve itself is usually only observed as a discrete set of points $\beta(t_1), \beta(t_2), \dots, \beta(t_m)$. As shown on the left side of Figure 1.2 this is no problem when the number m of observed points is high and the whole length of the curve is densly observed, as we can easily interpolate $\beta(t)$ for any $t \in [0,1]$. However, in cases where β is only observed over a small number of points (right side) and where the density and position of observed points may even vary between different curves—a setting known as *sparse* and *irregular*—more sophisticated smoothing techniques have to be applied. While the SRV framework has been combined with a Procrustes distance before, to estimate elastic shape means that also include invariance under scaling, rotation and translation (see Srivastava, Klassen, et al. 2011), these approaches

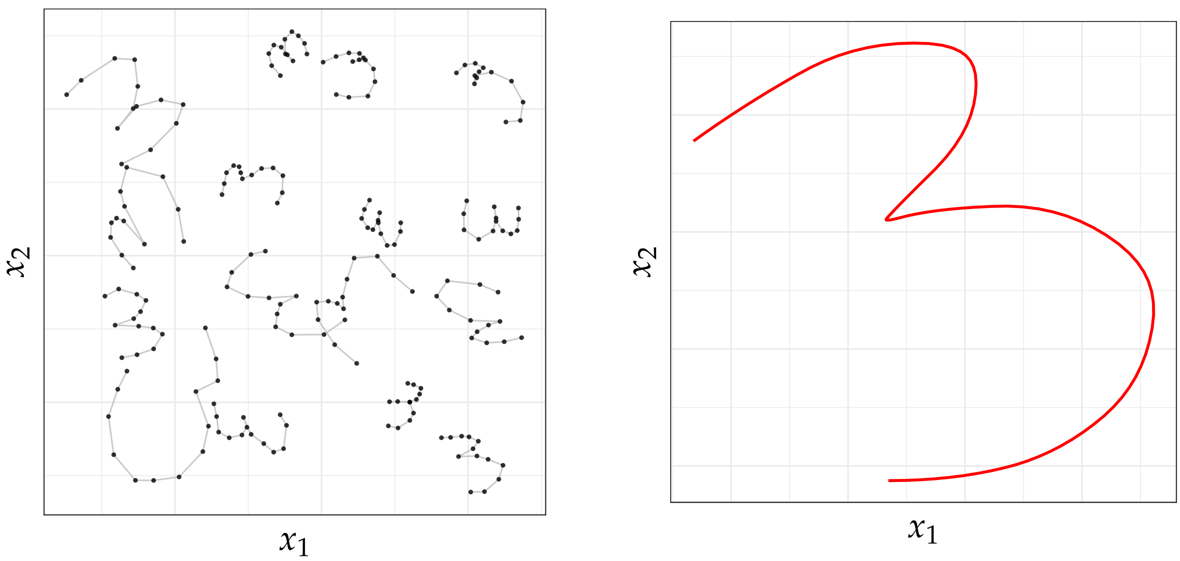


Figure 1.3.: Elastic full Procrustes mean function (right) estimated from sparse and irregular observations (left) using 13 knots and linear p-splines with a 2nd order penalty on SRV level. Data: Original (un-smoothed) `digits3.dat` with additional random rotation, scaling and translation applied.

fig:1-mean

have mostly focused on *Riemannian* or *geodesic* mean concepts and are not specially designed with sparse or irregular observations in mind. On the other hand, as will be shown, the estimation of the *elastic full Procrustes mean* in two dimensions is related to an eigenfunction problem over the complex covariance surface of the observed curves. This offers an advantage, as we can then make use of established smoothing techniques for the estimation of covariance surfaces in the sparse and irregular setting. Here, in particular Cederbaum, Scheipl, and Greven 2018 offers a method for efficient covariance smoothing using *tensor product p-splines* (see e.g. Fahrmeier et al. 2013, Chap. 8.2).

The aim of this thesis, as illustrated in Figure 1.3, 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 package `elasdics` (Steyer 2021) for the R programming language (R Core Team 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. Here, techniques for hermitian smoothing 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. In Section 2 relevant background material is reviewed and the elastic full Procrustes mean is derived, in the case where curves $\beta : [0,1] \to \mathbb{R}^2$ are fully observed. In Section 3 an estimation procedure for the setting of sparse and irregularly observed curves $\beta(t_1), \ldots, \beta(t_m)$ is proposed, concluding the theoretical part of this thesis. In Section 4 the proposed methods will be verified and applied using simulated and empirical datasets. Finally, all results will be summarized in Section 5. Appendix A and Supplements B offer additional considerations and some reproducability guidelines.

2. Elastic Full Procrustes Means for Planar Curves

sec:2

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 curve $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x_1(t), x_2(t))^T$, where $x_1(t)$ and $x_2(t)$ 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. We will therefore identify $\mathbb{R}^2$ with $\mathbb{C}$, as shown in Figure 2.1, and always use complex notation when representing a planar curve:

$$\beta : [0,1] \to \mathbb{C}, \quad \beta(t) = x_1(t) + i\, x_2(t).$$

We will assume the curves to be absolutely continuous, denoted as $\beta \in \mathcal{AC}([0,1],\mathbb{C})$, guaranteeing us that $\beta(t)$ has an integrable derivative. This is important when working in the square-root-velocity (SRV) framework, as will be discussed in Section 2.2. 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

ec:2-shape

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 (see Figure 1.1b). We can say that the actions of translation, scaling, rotation, and

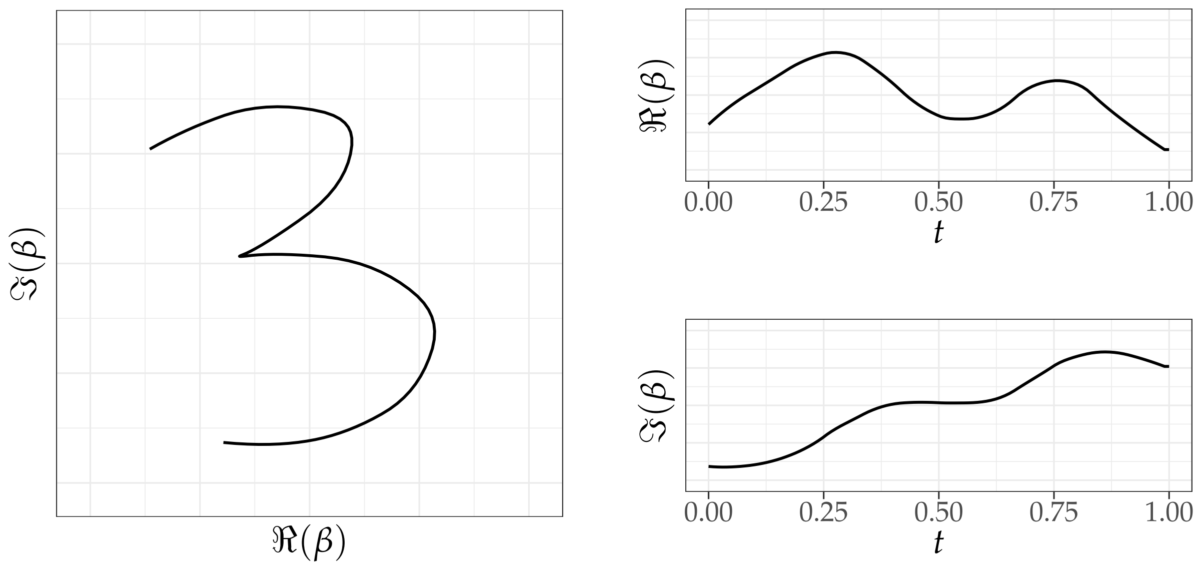


Figure 2.1.: Example of a planar curve (left) with respective coordinate functions (right) using complex notation. Data: see Figure 1.1.

ig:2-curve

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 each individual transformation acts on a planar curve $\beta : [0,1] \to \mathbb{C}$. This is usually done using the notion of *group actions* and *product groups*, with the later desciribing 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 $(\xi, \beta) \overset{\text{Trl}}{\longmapsto} \beta + \xi$, 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 $(\lambda, \beta) \overset{\text{Scl}}{\longmapsto} \lambda\beta$, 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*, i.e. 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 $((\lambda, \theta), \beta) \xmapsto{\text{Scl+Rot}} \lambda e^{i\theta}\beta$, which clearly commutes as $\lambda(e^{i\theta}\beta) = e^{i\theta}(\lambda\beta)$. On the other hand, 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 actions 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 called an *Euclidean similarity transformation* with $((\xi, \lambda, \theta), \beta) \xmapsto{\text{Eucl}} \lambda e^{i\theta}\beta + \xi$ (see Dryden and Mardia 2016, p. 62). Considering the action of *warping* or 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 give a formal definition of the shape of a planar curve as the following equivalence class:

Definition 2.1 (Shape)**.** The *shape* of an absolutely continous planar curve $\beta \in \mathcal{AC}([0, 1], \mathbb{C})$ is given by its equivalence class $[\beta]$ 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* $\mathcal{S}$ is then given by $\mathcal{S} = \{[\beta] \,|\, \beta \in \mathcal{AC}([0, 1], \mathbb{C})\}$.

2.2. The Elastic Full Procrustes Distance for Planar Curves

sec:2-dist

Let us now turn to the construction of an appropriate *shape distance* $d([\beta_1], [\beta_2])$ for two curves β_1, β_2. As the shapes $[\beta_1]$ and $[\beta_2]$ are elements of a non-Euclidean quotient space (the shape space $\mathcal{S}$), calculating a distance between them is already not straight-forward. A common approach is to map each equivalence class $[\beta]$ to a suitable representative, so that the distance calculation in shape space can be identified with a (much simpler) distance calculation over the representatives in the underlying functional space.

To illustrate this, let us first discuss each type of shape-preserving transformation individually, starting with the Euclidean similarity transformations. Consider two equivalence classes with respect to translation $[\beta_1]_{\mathrm{Trl}}, [\beta_2]_{\mathrm{Trl}}$. They might be uniquely mapped to their centered elements $\widetilde{\beta}_i^{\mathrm{Trl}} = \beta_i - \overline{\beta}_i \in [\beta_i]_{\mathrm{Trl}}$ for $i = 1, 2$. We can then define an appropriate distance that is invariant under translation as $d_{\mathrm{Trl}}([\beta_1]_{\mathrm{Trl}}, [\beta_2]_{\mathrm{Trl}}) = \|\widetilde{\beta}_1^{\mathrm{Trl}} - \widetilde{\beta}_2^{\mathrm{Trl}}\|$. Similarly, a distance that is invariant under scaling might be defined over the normalized elements $\widetilde{\beta}_i^{\mathrm{Scl}} = \frac{\beta_i}{\|\beta_i\|} \in [\beta_i]_{\mathrm{Scl}}$ for $i = 1, 2$, as $d_{\mathrm{Scl}}([\beta_1]_{\mathrm{Scl}}, [\beta_2]_{\mathrm{Scl}}) = \|\widetilde{\beta}_1^{\mathrm{Scl}} - \widetilde{\beta}_i^{\mathrm{Scl}}\|$. When considering invariance under rotation, we can first note that no "standardization" procedure compareable to normalizing and centering exists for the case of rotation. Instead of mapping $[\beta]_{\mathrm{Rot}}$ to a fixed representative, we therefore have to identify an appropriate representative on a case-by-case basis. This can be achieved by defining the distance as the minimal distance $d_{\mathrm{Rot}}([\beta_1]_{\mathrm{Rot}}, [\beta_2]_{\mathrm{Rot}}) = \min_{\widetilde{\beta}_2^{\mathrm{Rot}} \in [\beta_2]_{\mathrm{Rot}}} \|\beta_1 - \widetilde{\beta}_2^{\mathrm{Rot}}\| = \min_{\theta \in [0, 2\pi]} \|\beta_1 - e^{i\theta}\beta_2\|$, when keeping one curve fixed and rotationally aligning the other curve (compare e.g Stöcker and Greven 2021).

The Full Procrustes Distance

The three approaches can be combined to formulate the two *Procrustes* distances, which are invariant under all Euclidean similarity transforms. The *partial Procrustes distance* is defined as the minimizing distance $d_{PP}([\beta_1]_{\mathrm{Eucl}}, [\beta_2]_{\mathrm{Eucl}}) = \min_{\theta \in [0, 2\pi]} \|\widetilde{\beta}_1 - e^{i\theta}\widetilde{\beta}_2\|$, when rotationally aligning the centered and normalized curves $\widetilde{\beta}_i = \frac{\beta_i - \overline{\beta}_i}{\|\beta_i - \overline{\beta}_i\|}$, $i = 1, 2$. On the other hand, the *full Procrustes distance* (see Def. 2.2) includes an additional alignment over scaling, leading to a slightly different geometrical interpretation (see

Dryden and Mardia 2016, pp. 77–78). In this thesis we will only consider the full Procrustes distance, although no distance definition is inherently better than the other. In the context of mean estimation for sparse and irregular curves, the full Procrustes distance might be slightly more suitable, as the additional scaling alignment offers more flexibility in a setting where calculating a norm $\|\beta\| = \int_0^1 \|\beta(t)\|\, dt$ may already present a challange. Note that in Def. 2.2 the optimization over scaling $\lambda \in \mathbb{R}$ and rotation $\theta \in [0, 2\pi]$ was combined into a single optimization over *rotation and scaling* $\omega = \lambda e^{i\theta} \in \mathbb{C}$. Furthermore, Fig. 2.4 shows an example of two curves that where aligned by minimizing their full Procrustes distance. Curves aligned in such a way are called *Procrustes fits*.

def:2-fpdist **Definition 2.2** (Full Procrustes distance)**.** The *full Procrustes distance* for two equivalence classes $[\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}}$ is defined as

$$d_{FP}([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}}) = \min_{\omega \in \mathbb{C}} \|\widetilde{\beta}_1 - \omega \widetilde{\beta}_2\| \tag{2.1}$$ eq:2-fpdist-c

with centered and normalized representatives $\widetilde{\beta}_i = \frac{\beta_i - \overline{\beta}_i}{\|\beta_i - \overline{\beta}_i\|}$.

m:2-fpdist **Lemma 2.1.** *Let $\beta_1, \beta_2 : [0,1] \to \mathbb{C}$ be two planar curves with corresponding equivalence classes $[\beta_1]_{Eucl}, [\beta_2]_{Eucl}$ with respect to Euclidean similarity transforms and let $\widetilde{\beta}_i = \frac{\beta_i - \overline{\beta}_i}{\|\beta_i - \overline{\beta}_i\|}$.*

i.) *The full Procrustes distance between $[\beta_1]_{Eucl}$ and $[\beta_2]_{Eucl}$ is given by*

$$d_{FP}([\beta_1]_{Eucl}, [\beta_2]_{Eucl}) = \sqrt{1 - \langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle} \tag{2.2}$$

ii.) *The optimal rotation and scaling alignment of $\widetilde{\beta}_2$ onto $\widetilde{\beta}_1$ is given by $\omega^{opt} = \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle$. The aligned curve $\widetilde{\beta}_2^P = \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle \cdot \widetilde{\beta}_2$ is then called the* Procrustes fit *of $\widetilde{\beta}_2$ onto $\widetilde{\beta}_1$.*

Proof. See Appendix A.1.1. □

The Elastic Distance

When considering warping, we would like to do something similar to rotation in trying to find an optimal warping alignment between two curves β_1, β_2 by optimizing over

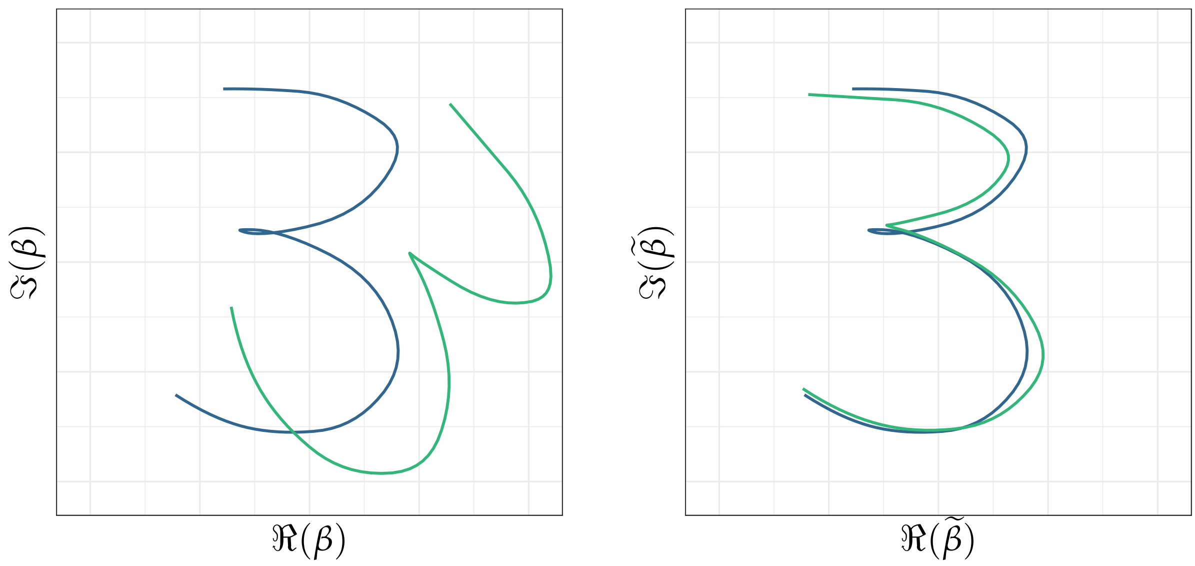


Figure 2.2.: Procrustes fit (right; normalized and centered) of two example curves (left). The Procrustes fit of β_2 (green) onto β_1 (blue) is given by $\widetilde{\beta}_2^P = \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle \widetilde{\beta}_2$. Data: See Figure 1.1.

fig:2-pfit

their distance $\inf_{\gamma \in \Gamma} \|\beta_1 - (\beta_2 \circ \gamma)\|$, where Γ is the space of warping functions. A usual choice would be to optimize this over the $\mathbb{L}^2$-distance between the curves, but in this case the result would not define a proper distance. 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, pp. 88–90).

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 warping. 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. As this SRV representation makes use of derivatives, any curve β that has a SRV curve must fulfill some kind of differentiability constraint. Here it is enough to consider only curves that are absolutely continuous $\beta \in \mathcal{AC}([0,1], \mathbb{C})$, which in particular means that the original curves do not have to be smooth but might also be piecewise linear (see Srivastava and Klassen 2016, p. 91). Note that because of the use of derivatives, any elastic analysis of curves will automatically be translation invariant as well. See

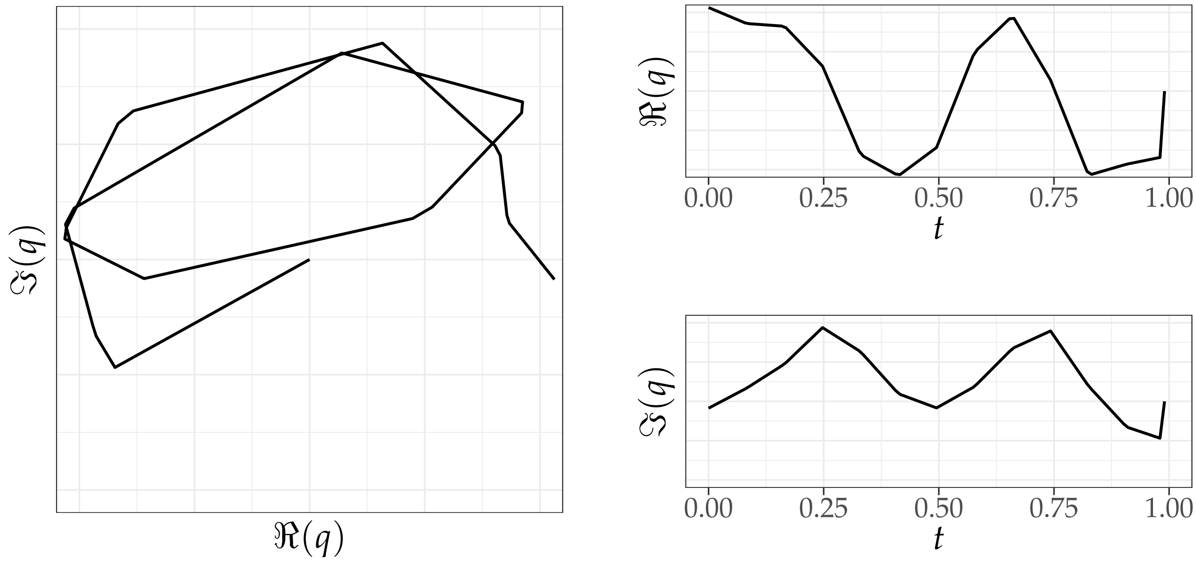


Figure 2.3.: SRV function (left) of the planar curve in Figure 2.1 with respective SRV coordinate functions (right). Note that the polygon-like look of the SRV curve is an artifact of the (linear) smoothing applied to the original data on SRV level (see Appendix A.3). Data: see Figure 1.1.

fig:2-srv

Figure 2.3 for an example SRV curve of a digit '3'.

def:2-eldist

Definition 2.3 (Elastic distance (Srivastava, Klassen, et al. 2011))**.** The *elastic distance* between equivalence classes $[\beta_1]_{\text{Wrp+Trl}}$, $[\beta_2]_{\text{Wrp+Trl}}$ is defined as

$$d_E([\beta_1]_{\text{Wrp+Trl}}, [\beta_2]_{\text{Wrp+Trl}}) = \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \circ \gamma) \cdot \sqrt{\dot{\gamma}}\|_{\mathbb{L}^2} \tag{2.3}$$

eq:2-eldist-c

with the respective *square-root-velocity* (SRV) representations $q_i \in \mathbb{L}^2([0,1], \mathbb{C})$ given by

$$q_i(t) = \begin{cases} \frac{\dot{\beta}_i(t)}{\sqrt{\|\dot{\beta}_i(t)\|}} & \text{for} \quad \dot{\beta}_i(t) \neq 0, \\ 0 & \text{for} \quad \dot{\beta}_i(t) = 0, \end{cases} \tag{2.4}$$

where $\beta_i \in \mathcal{AC}([0,1], \mathbb{C})$ and $\dot{\beta}_i(t) = \frac{\partial \beta_i(t)}{\partial t}$ for $i = 1, 2$.

Unlike the optimization over rotation in the definition of the full Procrustes distance (see Eq. 2.1), no analytical solution exists for the optimization over warping (see Eq. 2.3) in the definition of the elastic distance. Instead, Eq. 2.3 is usually solved numerically, by minimizing a cost function $H[\gamma] = \int_0^1 \|q_1(t) - q_2(\gamma(t)) \sqrt{\dot{\gamma}(t)}\| \, dt$ using a dynamic programming algorithm (see e.g. Srivastava and Klassen 2016, p. 152) or gradient based methods. In this thesis we will use the gradient based methods laid out in Steyer, A. Stöcker, and Greven 2021 to calculate optimal warping alignment between curves in the setting of sparse and irregularly sampled curves.

The Elastic Full Procrustes Distance

When the original curves β are absolutely continuous, the SRV curves are always ensured to be $\mathbb{L}^2$-integrable. As a consequence, we can re-construct the original curve β up to translation from its respective SRV curve q by integration $\beta(t) = \beta(0) + \int_0^t q(s)\|q(s)\|ds$. Because the translation of the original curve is usually not of interest from the point of shape analysis, the SRV curve holds all relevant information about the shape of β. This means, in particular, that instead of analysing the shape of β, we can equivalently analyse the shape of q. The shape preserving transformations on original curve level translate to SRV curve level by actions laid out in Lem. 2.2.

Write abit more here. Citations!

m:2-transf

Lemma 2.2. *The actions of the translation, scaling, rotation, and re-parametrization groups commute on SRV level. Furthermore, the individual transformations translate to SRV level by*

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

(see e.g. Srivastava *and* Klassen *2016, p. 142).*

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. □

We can note that the SRV curves are invariant under translation of the original curves, that the rotation is preserved on the SRV level and that scaling translates to SRV level by $\sqrt{\cdot}$. It is noteworthy that warping the original curve changes the image of the SRV curve.

Going forward, we will work in the SRV framework and combine the elastic distance with the full Procrustes distance. While the full Procrustes distance (see Def. 2.2) was defined over the normalized and centered curves, the SRV curves are already translation invariant so additional centering is not necessary. We will therefore define the *elastic full Procrustes distance* minimal distance, when aligning the scaling, rotation, and warping of the normalized SRV curves $\widetilde{q} = \frac{q}{\|q\|}$. Note that when 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 normalized, as $\|q\| = \sqrt{\int_0^1 |q(t)|^2\, dt} = \sqrt{\int_0^1 |\dot{\beta}(t)|\, dt} = \sqrt{L[\beta]}$.

def:dist **Definition 2.4** (Elastic full Procrustes distance)**.** The *elastic full Procrustes distance* between 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

$$d_{EF}([\beta_1], [\beta_2]) = \inf_{\omega \in \mathbb{C}, \gamma \in \Gamma} \quad \|\widetilde{q}_1 - \lambda e^{i\theta} (\widetilde{q}_2 \circ \gamma) \sqrt{\dot{\gamma}}\|$$

with normalized SRV representation $\widetilde{q}_i = \frac{q_i}{\|q_i\|} \in \mathbb{L}^2([0,1], \mathbb{C})$, where q_i is the SRV representation of β_i, for $i = 1, 2$.

To calculate the distance, we need to solve the joint optimization problem over $\omega \times \Gamma$. For a fixed $\gamma \in \Gamma$, the optimization problem in Definition 2.4 mirrors the full Procrustes distance for landmark data, where an explicit solution is known in the planar case (see Dryden and Mardia 2016, Chapter 8). Likewise, for fixed rotation $\theta \in [0, 2\pi]$ and scaling $\lambda \in \mathbb{R}^+$, there are known optimization techniques dealing with re-parametrisation.

$$(\lambda^*, \theta^*, \gamma^*) = \underset{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi], \gamma \in \Gamma}{\operatorname{argmin}} \quad \|z_1 - \lambda e^{i\theta} (z_2 \circ \gamma) \sqrt{\dot{\gamma}}\|$$

Here, using one parameter $\omega = \lambda e^{i\theta} \in \mathbb{C}$ for rotation and scaling can simplify notation. The rotation and scaling parameters can always be recovered by $\lambda = |\omega|$ and $\theta = \arg(\omega)$.

$$(\omega^*, \gamma^*) = \underset{\omega \in \mathbb{C}, \gamma \in \Gamma}{\operatorname{argmin}} \|z_1 - \omega(z_2 \circ \gamma) \sqrt{\dot{\gamma}}\| \tag{2.5}$$ eq:argmin

The usual strategy is to optimize over the sets of parameters individually and then to iterate through both solutions. Let us first consider the optimization with respect to

Citation, write more here!

rotation and scaling. For fixed $\gamma \in \Gamma$ with $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$ we can write Eq. 2.5 as

$$\omega^* = \underset{\omega \in \mathbb{C}}{\operatorname{argmin}} \|z_1 - \omega \widetilde{z}_2\|, \tag{2.6}$$ eq:argmin_

which can be solved analytically.

lem:dist **Lemma 2.3.** i.) *For a fixed $\gamma \in \Gamma$, the optimal scaling and rotation solving Eq. 2.6 is*

$$\omega^* = \langle \widetilde{z}_2, z_1 \rangle = \langle (z_2 \circ \gamma)\sqrt{\dot{\gamma}}, z_1 \rangle$$

ii.) *The optimization problem in Definition 2.4 can then 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 **??** in the appendix. □

Remark. For fixed $\gamma \in \Gamma$, we can use the first part of Lemma 2.3 to calculate the optimal rotation and scaling alignment of $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$ onto z_1. The alignment $\widetilde{z}_2^P = \langle \widetilde{z}_2, z_1 \rangle \widetilde{z}_2$ is called the *Procrustes fit* of $\widetilde{z}_2$ onto z_1. The second part of Lemma 2.3 will be useful for mean calculation in the next section.

Given the current $\omega \in \mathbb{C}$ with $z_2^P = \omega z_2$, the optimization with respect to reparametrization can be written as

$$\gamma^* = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \|z_1 - (z_2^P \circ \gamma)\sqrt{\dot{\gamma}}\|. \tag{2.7}$$ eq:argmin_

This is a well known problem and usually solved numerically by minimizing a cost function $H[\gamma] = \int_0^1 \|z_1(t) - z_2^P(\gamma(t))\sqrt{\dot{\gamma}(t)}\| \, dt$ using a dynamic programming algorithm (DPA) or gradient based methods (see Srivastava, Klassen, et al. 2011). In this thesis we will use the methods laid out in Steyer, A. Stöcker, and Greven 2021, for solving Eq. 2.7 in the setting of sparse and irregularly sampled curves.

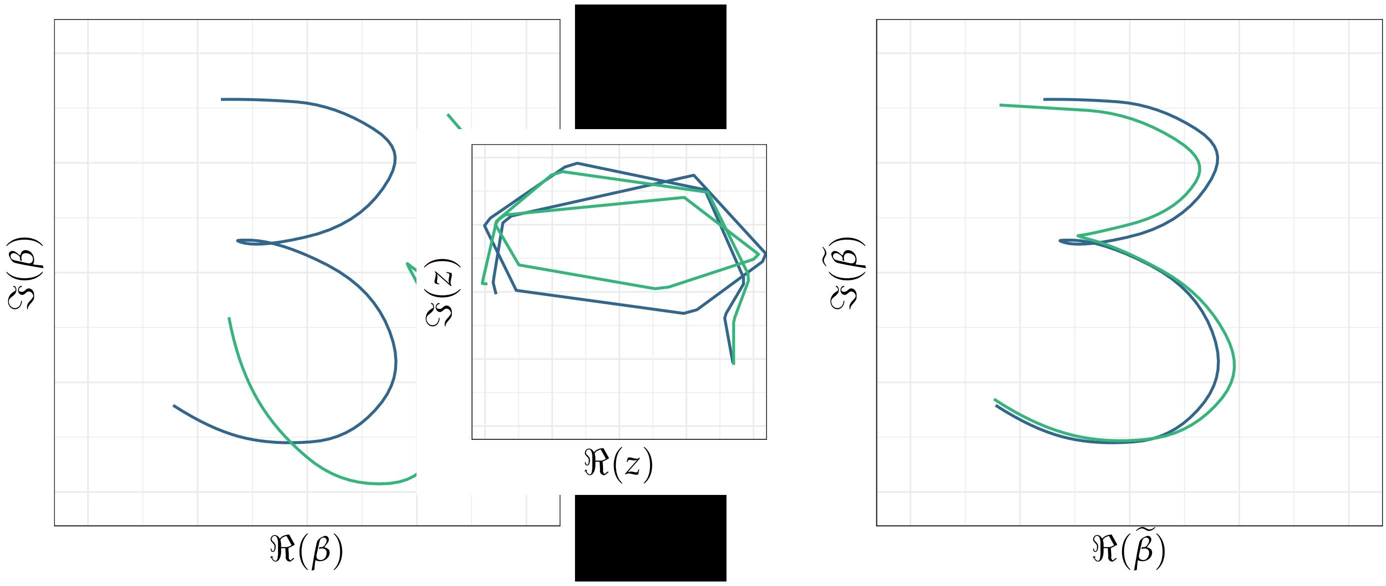


fig:2-pfit

Figure 2.4.: Procrustes fit of two example curves (left) on normalized SRV (middle) and original curve level (right). On normalized SRV level, the Procrustes fit of z_2 (green) onto z_1 (blue) is calculated as $\omega^* z_2$, where $\omega^* = \langle z_2, z_1 \rangle$ is the optimal scaling and rotation alignment. On original curve level, the Procrustes fit can then be constructed as Data: see Figure 1.1.

Anmerkung für Lisa Ich habe mich noch gefragt ob man anstatt Iteration über Eq. 2.6 und Eq. 2.7 die Distanz in Lemma 2.3 ii.) auch direkter optimieren kann.

$$\begin{aligned}
\gamma^* &= \underset{\gamma \in \Gamma}{\operatorname{argmin}} \sqrt{1 - \langle z_1, \widetilde{z}_2 \rangle \langle \widetilde{z}_2, z_1 \rangle} \\
&= \underset{\gamma \in \Gamma}{\operatorname{argmax}} \langle z_1, \widetilde{z}_2 \rangle \langle \widetilde{z}_2, z_1 \rangle \\
&= \underset{\gamma \in \Gamma}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{z_1(t)}\, \underbrace{\widetilde{z}_2(t) \overline{\widetilde{z}_2(s)}}_{:= \widetilde{C}(s,t)} z_1(s)\, dtds \\
&= \underset{\gamma \in \Gamma}{\operatorname{argmax}} \langle \widetilde{C} z_1, z_1 \rangle
\end{aligned}$$

Mit $\widetilde{C}(s,t)$ der Kovarianz-Funktion von $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$. Habe das jetzt nicht weiter verfolgt (weil Zeit), aber vielleicht ist das ganz interessant? Man kann das glaube ich auch umschreiben mit $\check{z}_1 = (z_1 \circ \gamma^{-1})\sqrt{\dot{\gamma}^{-1}}$ als $\operatorname{argmax}_{\gamma \in \Gamma} \langle C\, \check{z}_1, \check{z}_1 \rangle$.

Write iterative procedure as algorithm! Basically as in Srivastava, Klassen, et al. 2011.

2.3. The Elastic Full Procrustes Mean for Planar Curves

sec:2-mean

Nummerierung der Equations überarbeiten

We now want to calculate shape means for a random sample of planar curves. Again, we assume all curves to be absolutely continous $\beta_i \in \mathcal{AC}([0,1], \mathbb{C})$ with corresponding SRV curves $q_i \in \mathbb{L}^2([0,1], \mathbb{C})$, $i = 1, \ldots, N$. As we want to take into account shape invariance in the mean calculation, we cannot simply use sums or integrals to calculate a sample mean shape. Instead, we can use a more general concept for mean calculation, where the mean is defined as a minimizer over the sum of squared distances to each observation, for choice of a sensible distance in the space of interest. If the resulting mean is a global minimum, it is usually called a "sample Fréchet mean" (Fréchet 1948), if it is a local minimum a "sample Karcher mean" (Karcher 1977) (see Dryden and Mardia 2016, p. 111).

def:mean

Definition 2.5 (Sample elastic full Procrustes mean)**.** For a set of curves $\beta_i \in \mathcal{AC}([0,1], \mathbb{C})$, $i = 1, \ldots, N$, their *sample elastic full Procrustes mean* is given by the minimizing shape $[\hat{\mu}]$ with

$$[\hat{\mu}] = \underset{[\mu] \in \mathcal{S}}{\operatorname{arginf}} \sum_{i=1}^{N} d_{EF}([\mu], [\beta_i])^2 ,$$

where $\mathcal{S} = \{[\beta] : \beta \in \mathcal{AC}([0,1], \mathbb{C})\}$ is the shape space.

Instead of working with equivalence classes as in Def. 2.5, it is often simpler to work with a specific element $\hat{\mu} \in [\hat{\mu}]$, that acts as a "representation" of the sample mean shape. One possibility is to use a representation that is of unit-length and starts at the origin, so that $L[\hat{\mu}] = \int_0^1 \|\dot{\hat{\mu}}(t)\| \, dt = 1$ and $\hat{\mu}(0) = 0$. This is an attractive choice when working in the SRV framework, as we do not have to worry about reconstructing translation when calculating $\hat{\mu}$ from its respective SRV curve $\hat{\mu}_q$ by $\hat{\mu}(t) = \hat{\mu}(0) + \int_0^t \hat{\mu}_q(s) \|\hat{\mu}_q(s)\| \, ds$. Another possibility would be to use a unit-length representation that is centered, so that $|\int_0^1 \hat{\mu}(t) \, dt| = 0$, which may be achieved by setting $\hat{\mu}(0) = \int_0^1 \int_0^t \hat{\mu}_q(s) \|\hat{\mu}_q(s)\| \, ds \, dt$ when reconstructing $\hat{\mu}$ from $\hat{\mu}_q$. From the point of shape analysis, the choice of representation does not make a difference, as both mean curves are elements of $[\hat{\mu}]$ and therefore have the same shape. However, the distinction becomes important when the estimated mean curve $\hat{\mu}$ is used in concert with other curves, for example when visualizing multiple curves or when comparing multiple class mean shapes, as those do not typically share the same center or starting point. Differences between both representations will be explored using empirical data of tounge shapes in Section **??**.

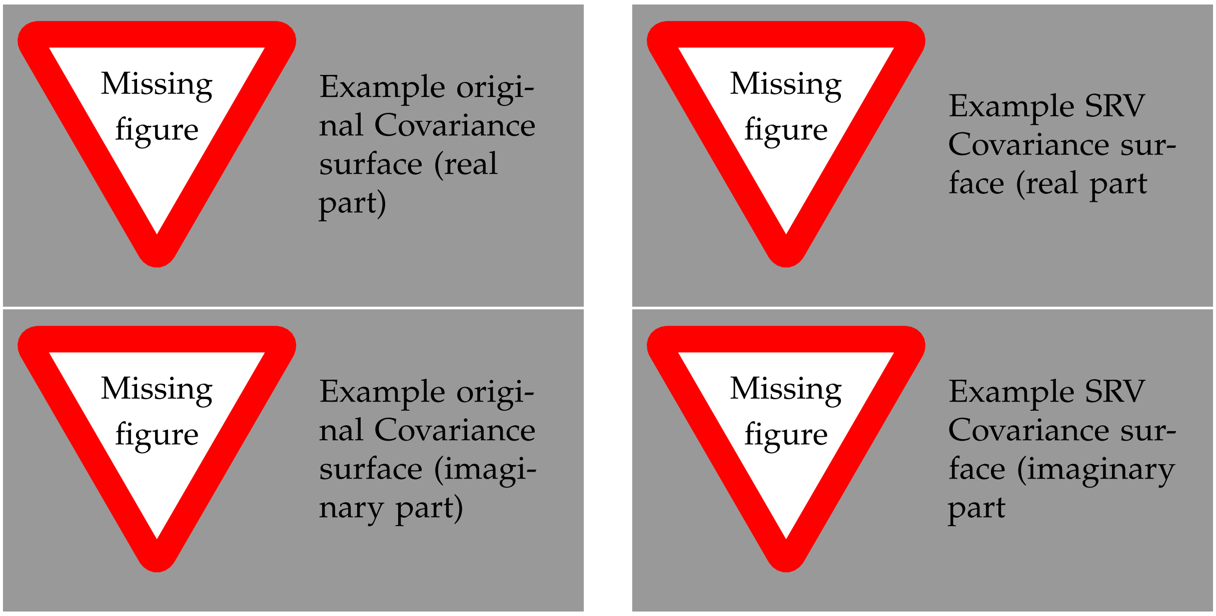
Note that we can always construct $\hat{\mu}$ (and therefore $[\hat{\mu}]$) from $\hat{\mu}_q$ by integration. Turning back to the actual mean calculation, we can therefore use the Def. 2.4 to reformulate the optimization problem in Def. 2.5 on SRV level, where we now optimize over possible normalized SRV representations μ_q of the mean shape $[\mu]$.

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmin}} \sum_{i=1}^{N} \left(\inf_{\omega_i \in \mathbb{C}, \gamma_i \in \Gamma} \|\mu_q - \omega_i (z_i \circ \gamma_i) \sqrt{\dot{\gamma}_i}\| \right)^2$$

Here, $z_i = \frac{q_i}{\|q_i\|}$ are the observed normalized SRV curves, ω_i the rotation and scaling alignment to μ_q and γ_i the re-parametrisation alignment to μ_q. We can further simplify this by solving the optimization over ω_i using Lemma 2.3 ii.

$$\begin{aligned}
\hat{\mu}_q &= \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmin}} \sum_{i=1}^{N} \inf_{\gamma_i \in \Gamma} \Big(1 - \langle \mu_q, (z_i \circ \gamma_i) \sqrt{\dot{\gamma}_i} \rangle \langle (z_i \circ \gamma_i) \sqrt{\dot{\gamma}_i}, \mu_q \rangle \Big) \\
\hat{\mu}_q &= \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \sup_{\gamma_i \in \Gamma} \langle \mu_q, (z_i \circ \gamma_i) \sqrt{\dot{\gamma}_i} \rangle \langle (z_i \circ \gamma_i) \sqrt{\dot{\gamma}_i}, \mu_q \rangle \\
\hat{\mu}_q &= \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \sup_{\gamma_i \in \Gamma} \big\langle \mu_q, (\gamma_i, z_i) \big\rangle \big\langle (\gamma_i, z_i), \mu_q \big\rangle
\end{aligned}$$

We end up with a two step optimization problem consisting of an outer optimization over μ_q and an inner optimization over the set $\{\gamma_i\}_{i=1,\dots,N}$. Similarly to the approaches discussed in Srivastava and Klassen 2016 and to Steyer, A. Stöcker, and Greven 2021, we solve this by *template based alignment* (see e.g. Srivastava and Klassen 2016, p. 271). In a first step the mean $\hat{\mu}_q$ is estimated while keeping the parametrisations fixed, after which the γ_i are updated by pairwise warping-alignment between z_i and $\hat{\mu}_q$, which is usually achieved by DPA or a gradient based approach. Both steps are iterated until the mean shape has converged. **[TODO: Noch genauer auf warping alignment step eingehen -> muss ja erstmal auch procrustes fit berechnen um warping alignen zu können.]**



(a) Covariance surface on original curve level. **(b)** Covariance surface on SRV curve level.

Figure 2.5.: Complex covariance surface on original and SRV curve level. Data: see Figure 1.1.

fig:2-cov

Let us now consider the outer optimization problem for a fixed set of warping function $\{\gamma_i^*\}_{i=1,\dots,N}$ where we denote the warping aligned normalized SRV curves (γ_i^*, z_i) as $\widetilde{z}_i$. Note that if no warping alignment has happened yet, we can always set $\gamma_i^*(t) = t$ for all $i = 1, \dots, N$ as a starting value. The problem we have to solve is:

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \langle \mu_q, \widetilde{z}_i \rangle \langle \widetilde{z}_i, \mu_q \rangle$$

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \int_0^1 \int_0^1 \overline{\mu_q(s)} \widetilde{z}_i(s) \overline{\widetilde{z}_i(t)} \mu_q(t) \, ds \, dt$$

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu_q(s)} \left(\sum_{i=1}^{N} \widetilde{z}_i(s) \overline{\widetilde{z}_i(t)} \right) \mu_q(t) \, ds \, dt$$

We can identify the inner term as proportional to a sample estimator $\check{C}(s,t) = \frac{1}{N} \sum_{i=1}^{N} z_i(s) \overline{z_i(t)}$ of the population covariance surface $C(s,t) = \mathbb{E}[z(s)\overline{z(t)}]$, when noting that $\mathbb{E}[z(t)] = 0$ for all $t \in [0,1]$ due to rotational symmetry. **[TODO: Hier wird die Notation etwas unsauber. Man hat hier ja einen Cov. Estimator aus den "gewarpten" Kurven. Vielleicht müsste man sich hier nicht nochmal kurz Gedanken machen wie $C(s,t)$ eigentlich aussieht, wenn man berücksichtigt, dass die $z(t)$ auch random gewarpt sind. Vielleicht ist's auch egal.]**

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmax}} \quad N \cdot \int_0^1 \int_0^1 \overline{\mu_q(s)} \check{C}(s,t) \mu_q(t) \, ds \, dt$$

By replacing $\check{C}(s,t)$ by its expectation $C(s,t)$, we can analogously formulate an estimator on the population level. **[TODO: Notation?]**

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 :\, \|\mu_q\|=1}{\operatorname{argmax}} \quad \int_0^1 \int_0^1 \overline{\mu_q(s)} C(s,t) \mu_q(t) \, ds \, dt$$

We can rewrite this again as a functional scalar product by considering the *covariance operator* C with $(C\mu_q)(s) = \int_0^1 C(s,t)\mu_q(t)dt$ (see Ramsay and Silverman 2005, p. 153).

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \quad \langle \mu_q, C\mu_q \rangle \tag{2.8}$$ eq:quadr_c

This is a well known problem in the context of functional principal component analysis (FPCA). From $\overline{C(s,t)} = \overline{\mathbb{E}[z(s)\overline{z(t)}]} = \mathbb{E}[z(t)\overline{z(s)}] = C(t,s)$ it follows that $\langle \mu_q, C\mu_q \rangle = \langle C\mu_q, \mu_q \rangle$ and therefore that C is a *self-adjoint* operator. The optimization problem then reduces to an eigenfunction problem

$$Cu = \lambda u \quad \Leftrightarrow \quad \int_0^1 C(s,t)u(t)\, dt = \lambda u(s)\,, \tag{2.9}$$ eq:funceig

where $\lambda = \langle \mu_q,\, C\mu_q \rangle$ is the target function to maximize. For normalized eigenfunctions $u_1, u_2, \ldots$ and corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$ of $C(s,t)$, the solution $\hat{\mu}_q(t)$ is given by the leading normalized eigenfunction $u_1(t)$ of $C(s,t)$ (see Ramsay and Silverman 2005, pp. 153, 397).

algo:mean **Algorithm 2.1** (Sample elastic full Procrustes mean)**.** *Let $\{\beta_i\}_{i=1,\ldots,N}$ be a sample of planar curves with corresponding SRV curves $\{q_i\}_{i=1,\ldots,N}$. Let $z_i = \frac{q_i}{\|q_i\|}$. Set $\gamma_i^0 = t$ for all $i = 1, \ldots, N$ as the initial parametrisation alignment. Set $k = 0$.*

1. *For $i = 1, \ldots, N$: Set $z_i^k(t) = z_i\left(\gamma_i^k(t)\right) \cdot \sqrt{\dot{\gamma}_i^k(t)}$.*
2. *Estimate $C(s,t) = \mathbb{E}[z(s)\overline{z(t)}]$ from $\{z_i^k\}_{i=1,\ldots,N}$. Call this estimate $\hat{C}^k(s,t)$.*
3. *Set μ_q^k as the leading normalized eigenfunction of $\hat{C}^k(s,t)$.* **Stop** *if μ_q^k is close to μ_q^{k-1}.*
4. *For $i = 1, \ldots, N$: Calculate the optimal rotation and scaling alignment $\omega_i^k = \left\langle z_i^k,\, \mu_q^k \right\rangle$.*

5. *For* $i = 1, \ldots, N$*: Solve* $\gamma_i^{k+1} = \mathrm{argmin}_{\gamma \in \Gamma} \| \mu_q^k - (\omega_i^k \cdot z_i \circ \gamma)\sqrt{\dot{\gamma}} \|$.

6. *Set* $k = k + 1$ *and return to Step 1.*

3. Estimation Strategy for Sparse and Irregular Observations

sec:3

Create figures for everything.

Alg. 2.1 shows an idealized version of the elastic full Procrustes mean estimation, where it is assumed that each curve β_i is fully observed. This is not the case in practice, as each observation β_i is usually itself only observed at a finite number of discrete points $\beta_i(t_{i1}), \ldots, \beta_i(t_{im_i})$. Additionally, the number of observed points per curve m_i might be quite small and the points do not need to follow a common sampling scheme across all curves, a setting which is respectively known as *sparse* and *irregular*.

Following the steps laid out in Alg. 2.1, this section proposes a mean estimation strategy for dealing with sparse and irregular observations. In a first step, the construction of SRV and warped SRV curves from discrete (and possibly sparse) observations will be shown in Section 3.1. Section 3.2 discusses efficient estimation of the complex covariance surface $C(s,t)$ from sparse observations. In Section 3.3 the calculation of the leading eigenfunction u_1 of $C(s,t)$ in a fixed basis will be shown. Section 3.4 deals with the estimation of the scalar product $\omega = \langle z, \mu_q \rangle$, which gives the optimal rotation and scaling alignment. Note that the final warping alignment step in Alg. 2.1 can be solved by using methods for warping alignment of sparse and irregular curves provided in Steyer, A. Stöcker, and Greven 2021.

3.1. Discrete Treatment of SRV Curves

3-discrete

A natural first consideration might be how to calculate SRV curves from sparse observations. We defined the SRV curve of a function $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ as $q = \frac{\dot{\beta}}{\sqrt{\|\dot{\beta}\|}}$ (for $\dot{\beta} \neq 0$). This means that if we want to calculate the SRV curve, we have to calculate the derivative of β. As we never observe the whole function β but only a set discrete points $\beta(t_1), \ldots, \beta(t_m)$, this is already not straight forward, as we cannot simply calculate a pointwise derivative. However, following Steyer, A. Stöcker, and

Greven 2021, we treat a discretely observed curve β as piecewise linear between its observed corners $\beta(t_1), \dots, \beta(t_m)$, which allows us to calculate a piecewise constant derivative on the intervalls $[t_j,\ t_{j+1}]$, $j = 1, \dots, m-1$. Let us first consider the case of unwarped observations.

Initial parameterization Usually only the image $\beta(t_1), \dots, \beta(t_m)$, but not the parametri[redacted] sation $t_1, \dots, t_m$, is observed. Therefore it is first necessary to construct an initial parameterisation. A common choice is the *arc-length-parametrisation*, where we set $t_j = \frac{l_j}{l}$ with $l_j = \sum_{k=1}^{j-1} |\beta(t_{k+1}) - \beta(t_k)|$ the polygon-length up to point j for $j \leq 2$, $l_1 = 0$ and $l_m = l$.

Piecewise-constant SRV curve Consider the discrete derivative $\Delta\beta\big|_{[t_j, t_{j+1}]} = \frac{\beta(t_{j+1}) - \beta(t_j)}{t_{j+1} - t_j}$[redacted] which assumes that β is linear between its observed corners. The corresponding SRV curve q can then be treated as piecewise constant $q\big|_{[t_j, t_{j+1}]} = q_j$ with $q_j = \Delta\beta\big|_{[t_j, t_{j+1}]} \Big/ \sqrt{\|\Delta\beta\big|_{[t_j, t_{j+1}]}\|} = \frac{\beta(t_{j+1}) - \beta(t_j)}{\sqrt{t_{j+1} - t_j} \cdot \sqrt{\|\beta(t_{j+1}) - \beta(t_j)\|}}$ the discrete *square-root-velocity* of β between the corners $\beta(t_j)$ and $\beta(t_{j+1})$.

Approximate discrete SRV curve As shown in Steyer, A. Stöcker, and Greven 2021 (cf. Fig. 3), treating the SRV curves as piecewise-constant functions can lead to overfitting, where the mean shape is estimated too polygon-like. As an alternative they propose to approximate the derivative, by assuming that it attains the value of the discrete derivative $\Delta\beta\big|_{[t_j, t_{j+1}]}$ at the center $s_j = \frac{t_{j+1} - t_j}{2}$ of the interval $[t_j, t_{j+1}]$. Using this, we can construct "approximate observations" $q(s_j) \approx q_j$ of the SRV curve q.

Normalization We can approximate the normalized SRV curve z using the polygon-length l of β by $z_j = q_j / \sqrt{l}$.

[TODO: Eventuell den ganzen Part mit den piecewise-constant curves streichen. Am ende wird sowieso nur die Mittelwert-Approximation benutzt.]

[TODO: Nochmal überlegen mit γ^{-1} und γ. Gerade ist das hier glaube ich falsch.]

Let us now consider a warping function γ. The warped discrete derivative is given by $\Delta(\beta \circ \gamma)\big|_{[\gamma(t_j), \gamma(t_{j+1})]} = \frac{\beta(\gamma(t_{j+1})) - \beta(\gamma(t_j))}{\gamma(t_{j+1}) - \gamma(t_j)}$. The corresponding warped SRV curve is then given by $(q \circ \gamma)\sqrt{\dot\gamma}\big|_{[\gamma(t_j), \gamma(t_{j+1})]} = \frac{\beta(\gamma(t_{j+1})) - \beta(\gamma(t_j))}{\sqrt{\gamma(t_{j+1}) - \gamma(t_j)} \cdot \sqrt{\|\beta(\gamma(t_{j+1})) - \beta(\gamma(t_j))\|}}$.

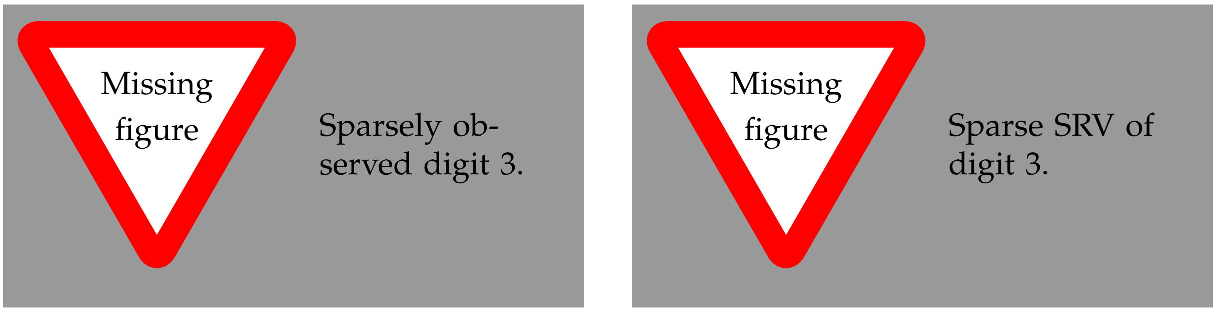


Figure 3.1.: A sparsely observed planar curve (left) with sparse approximate SRV curve (right). Data: see Figure 1.1.

fig:3-disc

[TODO: Das hier zuende formulieren.]

3.2. Efficient Estimation using Hermitian Covariance Smoothing

sec:3-cov

We want to estimate $C(s,t) = \mathbb{E}[z(s)\overline{z(t)}]$ given approximate observations of the normalized SRV curves $z_i(s_{ij})$, with $j = 1, \ldots, m_i - 1$ and $i = 1, \ldots, N$, where m_i denotes the number of observed points per curve. Following **[TODO: Cite]**, we can treat this estimation as a smoothing problem, by constructing responses $y_{ijk} = z_i(s_{ij})\overline{z_i(s_{ik})}$ and treating the pairs s_{ij}, s_{ik} as covariates s and t. Smooothing the responses y_{ijk} gives an estimate $\hat{C}(\cdot, \cdot)$ of $C(s,t)$, as each response has expectation $\mathbb{E}[y_{ijk}] = C(s_{ij}, s_{ik})$. A popular approach **[TODO: Cite]** is to carry out the smoothing in a *tensor product spline* basis

$$C(s,t) = b(s)^T \Xi b(t)$$

where $b(s) = (b_1(s), \ldots, b_K(s))$ denotes the vector of a spline basis and Ξ is a $K \times K$ coefficient matrix to be estimated. As $C(s,t)$ is complex, we restrict the spline basis to be real-valued with $b_k : \mathbb{R} \to \mathbb{R}$ for $k = 1, \ldots, K$ and the coefficient matrix to be complex-valued with $\Xi \in \mathbb{C}^{K \times K}$, without loss of generality.

Considering the symmetry properties of the covariance surface allows for more efficient estimation, as shown in Cederbaum, Scheipl, and Greven 2018 for real valued, symmetric covariance surfaces, by considering every unique pair s_{ij},s_{ik} only once. In the complex case, the covariance surface is hermitian with $C(s,t) = \overline{C(t,s)}$, which means we can decompose the estimation into two seperate regression problems over the symmetric real and skew-symmetric imaginary parts of $C(s,t)$.

$$\mathbb{E}[\Re(y)] = b(s)^T \Xi_{\Re} b(t)$$

$$\mathbb{E}[\Im(y)] = b(s)^T \Xi_{\Im} b(t)$$

with $\Xi_{\Re}, \Xi_{\Im} \in \mathbb{R}^{K\times K}$ and $\Xi = \Xi_{\Re} + i\Xi_{\Im}$, under the constraints that $\Xi_{\Re}^T = \Xi_{\Re}$ and $\Xi_{\Im}^T = -\Xi_{\Im}$. In this thesis I estimate $\Xi_{\Re}$ and $\Xi_{\Im}$ using the `R` (**Rcore**) package `mgcv` (Wood 2017). For efficient estimation, two `mgcv` smooths from the package `sparseFLMM` (Cederbaum, Volkmann, and A. Stöcker 2021) are used, which implement and generalize the approach proposed by Cederbaum, Scheipl, and Greven 2018 for symmetric and skew-symmetric tensor product p-splines. **[TODO: Erklärung p-splines?] [TODO: Motivate b-spline/p-spline basis using Lisa's paper?] [TODO: REML?]**

3.3. Estimating the Elastic Full Procrustes Mean in a Fixed Basis

sec:3-mean

To estimate the elastic full Procrustes Mean, we have to solve a functional eigenvalue problem on the estimated covariance surface $\hat{C}(s,t) = b(s)^t \hat{\Xi} b(t)$. This may be achieved by evaluating $\hat{C}(s,t)$ on a dense grid and performing an eigendecomposition on the matrix of evaluations **[TODO: Cite]**. Alternatively, we can estimate the mean directly in some basis $b(s) = (b_1(s), \dots, b_K(s))$, where a natural choice might be to evaluate mean and covariance surface in the same basis, i.e. $\mu_q(s) = b(s)^T\theta$.

Remember that the elastic full Procrustes mean (for fixed warping) is given by the solution to the optimization problem

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu_q(s)} C(s,t) \mu_q(t) \, ds \, dt \ .$$

Given an estimate of the covariance surface $\hat{C}(s,t) = b(s)^T \hat{\Xi} b(t)$, the mean estimation then reduces to estimating the vector of coefficients $\theta = (\theta_1, \dots, \theta_K) \in \mathbb{C}^K$ with

$$\hat{\theta} = \underset{\theta \in \mathbb{C}^K \colon \|b^T\theta\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \theta^H b(s) b(s)^T \hat{\Xi} b(t) b(t)^T \theta \, ds \, dt$$

$$= \underset{\theta \in \mathbb{C}^K:\, \|b^T\theta\|=1}{\operatorname{argmax}} \quad \theta^H \left(\int_0^1 b(s)b(s)^T\, ds \right) \hat{\Xi} \left(\int_0^1 b(t)b(t)^T\, dt \right) \theta$$

$$= \underset{\theta \in \mathbb{C}^K:\, \theta^H G\theta=1}{\operatorname{argmax}} \quad \theta^H G\hat{\Xi}G\theta$$

where $(\cdot)^H = \overline{(\cdot)}^T$ denotes the conjugate transpose and G is the $K \times K$ Gram matrix with entries given by the basis products $g_{ij} = \langle b_i, b_j \rangle$. In the special case of an orthonormal basis with $\langle b_i, b_j \rangle = \delta_{ij}$ the Gram matrix is an identity matrix, however, this is not the case for many basis representations such as the b-spline basis.

We have reduced the functional eigenvalue problem to a multivariate eigenvalue problem over the covariance coefficient matrix. We might solve this using Lagrange optimization with the following Langrangian:

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

Taking into account that we identified $\mathbb{R}^2$ with $\mathbb{C}$ we can split everything into real and imaginary parts and optimize with respect to $\Re(\theta)$ and $\Im(\theta)$ seperately, to avoid having to take complex derivatives. Using $\theta = \theta_\Re + i\theta_\Im$ and $\hat{\Xi} = \hat{\Xi}_\Re + i\hat{\Xi}_\Im$ we can write

$$\begin{aligned} \mathcal{L}(\theta_\Re, \theta_\Im, \lambda) = &\ (\theta_\Re^T - i\theta_\Im^T)G(\hat{\Xi}_\Re + i\hat{\Xi}_\Im)G(\theta_\Re + i\theta_\Im) - \lambda \left((\theta_\Re^T - i\theta_\Im^T)G(\theta_\Re + i\theta_\Im) - 1 \right) \\ = &\ \theta_\Re^T G\hat{\Xi}_\Re G\theta_\Re + i\theta_\Re^T G\hat{\Xi}_\Im G\theta_\Re + \theta_\Im^T G\hat{\Xi}_\Im G\theta_\Re - \theta_\Re^T G\hat{\Xi}_\Im G\theta_\Im \\ &+ \theta_\Im^T G\hat{\Xi}_\Re G\theta_\Im + i\theta_\Im^T G\hat{\Xi}_\Im G\theta_\Im + \lambda \left(\theta_\Re^T G\theta_\Re + \theta_\Im^T G\theta_\Im - 1 \right) \end{aligned}$$

using $\hat{\Xi}_\Re^T = \hat{\Xi}_\Re$ and $\hat{\Xi}_\Im^T = -\hat{\Xi}_\Im$. Differentiation w.r.t. $\theta_\Re$ and $\theta_\Im$ yields

$$\frac{\partial \mathcal{L}}{\partial \theta_\Re} = 2G\hat{\Xi}_\Re G\theta_\Re - 2G\hat{\Xi}_\Im G\theta_\Im - 2\lambda G\theta_\Re \overset{!}{=} 0 \tag{3.1}$$ eq:lagrRe

$$\frac{\partial \mathcal{L}}{\partial \theta_\Im} = 2G\hat{\Xi}_\Re G\theta_\Im + 2G\hat{\Xi}_\Im G\theta_\Re - 2\lambda G\theta_\Im \overset{!}{=} 0 \tag{3.2}$$ eq:lagrIm

with the additional constraint $\theta_\Re^T G\theta_\Re + \theta_\Im^T G\theta_\Im = 1$.T We can simplify this further and multiply Eq. 3.2 by i, leading to

$$\hat{\Xi}_\Re G\theta_\Re - \hat{\Xi}_\Im G\theta_\Im = \lambda\theta_\Re \tag{3.3}$$

$$i\hat{\Xi}_{\Re}G\theta_{\Im} + i\hat{\Xi}_{\Im}G\theta_{\Re} = i\lambda\theta_{\Im}\,. \tag{3.4}$$

Adding both equations finally leads to

$$(\hat{\Xi}_{\Re} + i\hat{\Xi}_{\Im})G\theta_{\Re} + i(\hat{\Xi}_{\Re} + \hat{\Xi}_{\Im})G\theta_{\Im} = \lambda(\theta_{\Re} + i\theta_{\Im})$$

or likewise, using θ and $\hat{\Xi}$

$$\hat{\Xi}G\theta = \lambda\theta$$

which is an eigenvalue problem on the product of the complex coefficient matrix and the Gram matrix. Multiplying by $\theta^H G$ from the left yields $\lambda = \theta^H G\hat{\Xi}G\theta$, i.e. the eigenvalues correspond to the target function to maximize. It follows that the estimate for the coefficient vector of the elastic full Procrustes mean is given by the eigenvector of the leading eigenvalue of $\hat{\Xi}G$. **[TODO: Cite Reiss.]**

3.4. Numerical Integration of the Procrustes Fits

ec:3-pfits

[TODO: Mean value theorem in the integral, etc.]

$\hat{p}(t) = b(t)^T\hat{\theta}$ estimated Procrustes mean function, q piecewise constant SRV transform with $q|_{[t_j,t_{j+1}]} = q_j \in \mathbb{C}$.

For estimation of the Procrustes fits we need to estimate two scalar products:

$$\hat{q}_p = \frac{\langle q, \hat{p}\rangle}{\langle q, q\rangle} q$$

So far I treat q as piecewise constant (as in Lisa's paper):

$$\langle q, \hat{p}\rangle = \int_0^1 \langle q(t), \hat{p}(t)\rangle\, dt \approx \sum_{j=0}^{m-1} \int_{t_j}^{t_{j+1}} \langle q_j, \hat{p}(t)\rangle\, dt$$

$$\langle q, q\rangle = \int_0^1 \langle q(t), q(t)\rangle\, dt \approx \sum_{j=0}^{m-1} (t_{j+1} - t_j)\, \langle q_j, q_j\rangle$$

4. Empirical Applications

sec:4

4.1. Mean Estimation for Simulated Spirals

:4-spirals

4.2. Classification of Hand-written Digits

c:4-digits

4.3. Mean Differences of Tounge Shapes in a Phonetics Study

:4-tounges

5. Summary

sec:5

Bibliography

`derson1997` Anderson, C. R. (1997). *Object recognition using statistical shape analysis*. PhD thesis. University of Leeds.

`Greven2018` Cederbaum, J., F. Scheipl, and S. Greven (2018). "Fast Symmetric Additive Covariance Smoothing". In: *Computational Statistics & Data Analysis* 120, pp. 25–41.

`sparseFLMM` 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`.

`shapes` 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`.

`Mardia2016` Dryden, I. L. and K. V. Mardia (2016). *Statistical Shape Analysis with Applications in R*. 2nd ed. John Wiley and Sons Ltd.

`erEtAl2013` Fahrmeier, L. et al. (2013). *Regression. Models, Methods and Applications*. Berlin, Heidelberg: Springer.

`rechet1948` Fréchet, M. R. (1948). "Les éléments aléatoires de nature quelconque dans un espace distancié". fr. In: *Annales de l'institut Henri Poincaré* 10.4, pp. 215–310.

`archer1977` Karcher, H. (1977). "Riemannian center of mass and mollifier smoothing". In: *Communications on Pure and Applied Mathematics* 30.5, pp. 509–541.

`endall1977` Kendall, D. G. (1977). "The diffusion of shape". In: *Advances in Applied Probability* 9.3, pp. 428–430.

`R` R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. url: `https://www.R-project.org/`.

`verman2005` Ramsay, J. and B. W. Silverman (2005). *Functional Data Analysis*. Springer series in statistics.

Rao1945 Rao, C. R. (1945). "Information and accuracy attainable in the estimation of statistical parameters". In: *Bulletin of Calcutta Mathematical Society* 37, pp. 81–91.

lassen2016 Srivastava, A. and E. P. Klassen (2016). *Functional and Shape Data Analysis*. New York: Springer.

vaEtAl2011 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.

elasdics 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`.

Steyer2021 Steyer, L., A. Stöcker, and S. Greven (2021). *Elastic analysis of irregularly or sparsely sampled curves*. arXiv: `2104.11039 [stat.ME]`.

oecker2021 Stöcker, A and S. Greven (2021). *Functional additive regression on shape and form manifolds of planar curves*. arXiv: `2109.02624 [stat.ME]`.

ueller2016 Wang, J.-L., J.-M. Chiou, and H.-G. Müller (2016). "Functional Data Analysis". In: *Annual Review of Statistics and Its Application* 3.1, pp. 257–295.

Wood2017 Wood, S.N (2017). *Generalized Additive Models: An Introduction with R*. 2nd ed. Chapman and Hall/CRC.

A. Appendix

app:a

A.1. Proofs and Derivations

pp:a-deriv

A.1.1. Derivation of Lemma 2.1

riv-fpdist

Rewrite proof for on curve level.

Proof. Start with ii.). Let $\widetilde{z}_2 = (\gamma, z_2) = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$, then optimise $d_{EF}([\beta_1],[\beta_2])^2$ over rotation and scaling, keeping γ fixed.

$$\begin{aligned}
d_{EF}([\beta_1],[\beta_2])^2 &= \inf_{\lambda\in\mathbb{R}^+,\theta\in[0,2\pi],\gamma\in\Gamma} \|z_1 - \lambda e^{i\theta}\widetilde{z}_2\|^2 \\
&= \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+,\theta\in[0,2\pi]} \langle z_1 - \lambda e^{i\theta}\widetilde{z}_2, z_1 - \lambda e^{i\theta}\widetilde{z}_2\rangle\right) \\
&= \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+,\theta\in[0,2\pi]} \|z_1\|^2 + \lambda^2\|\widetilde{z}_2\|^2 - \lambda(e^{i\theta}\langle z_1,\widetilde{z}_2\rangle + e^{-i\theta}\langle\widetilde{z}_2, z_1\rangle)\right)
\end{aligned}$$

As $\langle z_1, \widetilde{z}_2\rangle \in \mathbb{C}$, define $\langle z_1,\widetilde{z}_2\rangle = \kappa_\gamma e^{i\phi_\gamma}$, where $(\cdot)_\gamma$ denotes the dependence on γ. Furthermore, use $\|z_{1,2}\| = 1$, which implies $\|\widetilde{z}_2\| = 1$, as re-parametrisation is norm-preserving, when using the elastic metric.

$$\begin{aligned}
\ldots \quad &= \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+,\theta\in[0,2\pi]} 1 + \lambda^2 - \lambda(e^{i\theta}\kappa_\gamma e^{i\phi_\gamma} + e^{-i\theta}\kappa_\gamma e^{-i\phi_\gamma})\right) \\
&= \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+,\theta\in[0,2\pi]} 1 + \lambda^2 - \lambda\kappa_\gamma\left(e^{i(\theta+\phi_\gamma)} + e^{-i(\theta+\phi_\gamma)}\right)\right) \\
&= \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+} 1 + \lambda^2 - \sup_{\theta\in[0,2\pi]} 2\lambda\kappa_\gamma\cos(\theta+\phi_\gamma)\right) \\
&\overset{\theta=-\phi_\gamma}{=} \inf_{\gamma\in\Gamma}\left(\inf_{\lambda\in\mathbb{R}^+} 1 + \lambda^2 - 2\lambda\kappa_\gamma\right)
\end{aligned}$$

From $\frac{\partial}{\partial\lambda}\left(1 + \lambda^2 - 2\lambda\kappa_\gamma\right) = 2\lambda - 2\kappa_\gamma \overset{!}{=} 0$ it follows that $\lambda = \kappa_\gamma$.

$$\ldots \quad = \inf_{\gamma\in\Gamma}\left(1 + \kappa_\gamma^2 - 2\kappa_\gamma^2\right) = \inf_{\gamma\in\Gamma}\left(1 - \kappa_\gamma^2\right)$$

[TODO: Make this part clearer!] Lemma 2.3 ii.) follows from $\kappa_\gamma^2 = |\langle z_1, \widetilde{z}_2 \rangle|^2 = \langle z_1, \widetilde{z}_2 \rangle \langle \widetilde{z}_2, z_1 \rangle$ and $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$. Lemma 2.3 i.) follows from $\lambda e^{i\theta} = \kappa_\gamma e^{-i\phi_\gamma} = \overline{\langle z_1, \widetilde{z}_2 \rangle}$. □

A.2. Discussion of Possible Extensions to Closed Curves

p:a-closed

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

p:a-smooth

Idea: Smooth q in mean basis $b(t)$, so that $\hat{q}(t) = b(t)^T \hat{\theta}_q$. Then the scalar products simplify to

$$\langle q, \hat{p} \rangle \approx \hat{\theta}_q^H G \hat{\theta}$$

$$\langle q, q \rangle \approx \hat{\theta}_q^H G \hat{\theta}_q$$

As q can be sparse, we may want to use the estimated covariance matrix $\hat{\Xi}$ when estimating θ_q in the form of a Normal prior $\theta_q \sim \mathcal{N}_{\mathbb{C}^k}(0, \hat{\Xi})$ leading to

$$\hat{\theta}_q = (B^T B + \Xi^{-1})^{-1} B^T q$$

or equivalently using eigendecompositon $\Xi = V \Lambda V^{-1}$ with $V^H V = \mathcal{I}$ and $\Lambda = diag(\lambda_1, \ldots, \lambda_k)$

$$\hat{\theta}_q = V(V^H B^T B V + \Lambda^{-1})^{-1} V^H B^T q$$

In general how well this works tends to depend on how close the curve is to the estimated mean.

Maybe using a penalty parameter - controling the strength of the regularization - is also thinkable? Something like this:

$$\hat{\theta}_q(\lambda) = V(V^H B^T B V + \lambda \cdot \Lambda^{-1})^{-1} V^H B^T q$$

Thought: A normal prior is probably not appropriate, because while its true that $\mathbb{E}[\theta] = 0$ (due to rotational symmetry) I would think that in general $\mathbb{E}[\|\theta\|] \neq 0$. A distribution more like a Normal distribution lying on a ring around 0 with radius $r = \|\theta\|$ is probably better? Then estimates wouldn't be pressed to zero so strongly, and instead would be pressed to curve of size equal to the mean.

B. Supplementary Materials

app:b

B.1. Dataset Replication Guide

app:b-data

B.2. Implementation Notes

app:b-impl

—Discarded—

Math-Basics Recap

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

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 $n \to \infty$) 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

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

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*

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