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]

Berlin, 14th December, 2021

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

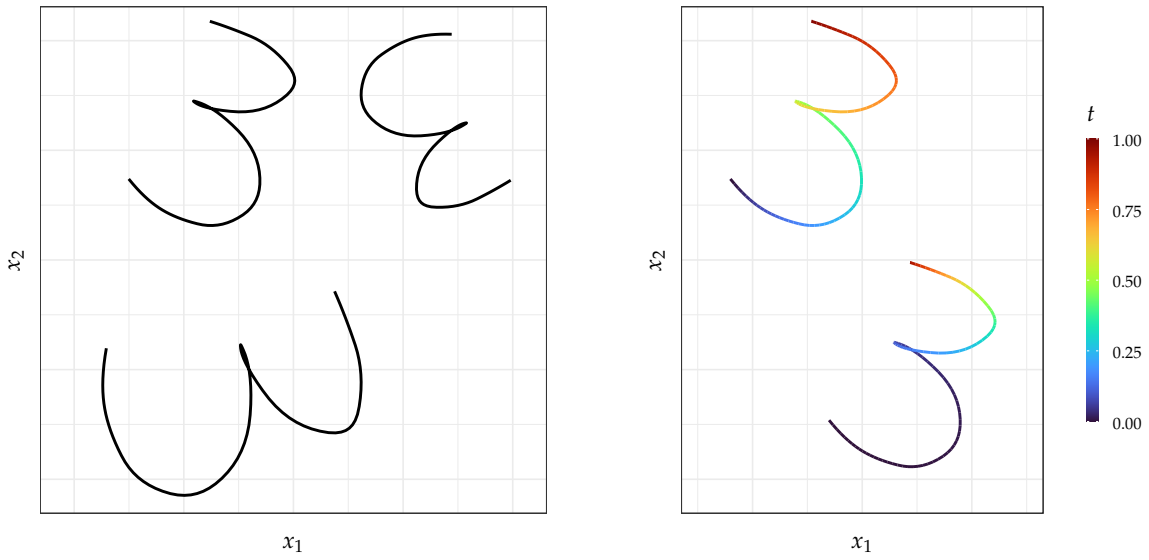
Todo list

1. Introduction

Statistical Shape Analysis (see e.g. Dryden and Mardia 2016) is the branch of statistics concerned with modelling geometric 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 geometric 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 geometric 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 continuous 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; Wang, Chiou, and Müller 2016).

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



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

(b) Original (left) and re-parameterised digit (right; $t \mapsto t^5$). Colour indicates the value of the pasteurisation $t \in [0, 1]$ 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.2. Original dataset collected by Anderson 1997.

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 Fig. 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 parameterisation $t \in [0, 1]$. As illustrated in Fig. 1.1b, curves $\beta(t)$ and $\beta(\gamma(t))$, with some re-parameterisation 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 parameterisation 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*

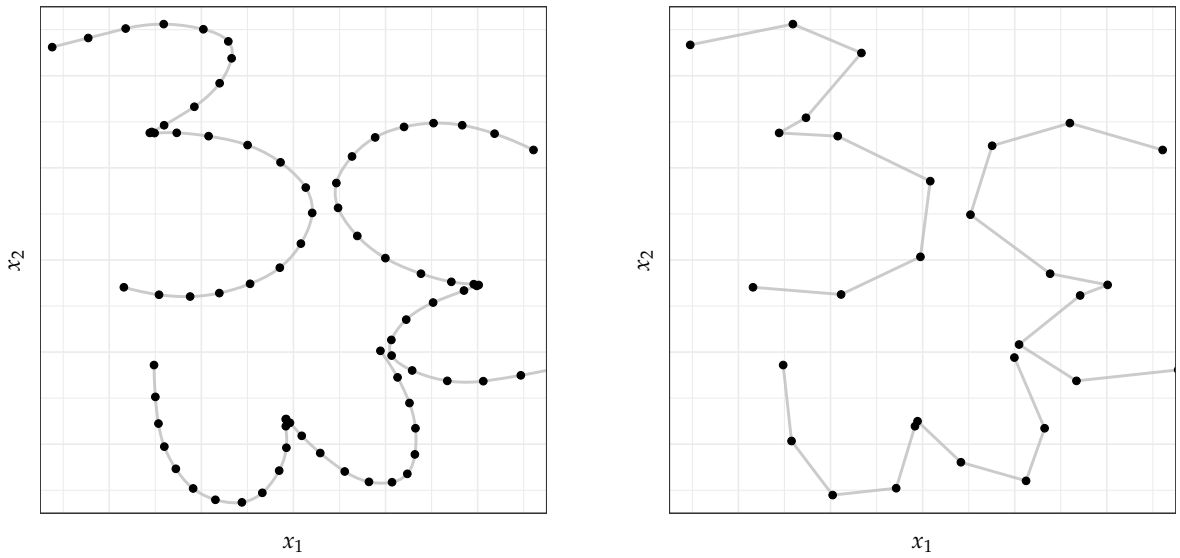


Figure 1.2.: Dense (left) and sparse (right) observations of the same three digits. Data: see Fig. 1.1, with the smooth curves sampled on a dense (left) and sparse (right) grid.

(re-parameterisation 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 continuous curve $\beta : [0,1] \to \mathbb{R}^2$, the curve itself is usually only observed as a discrete set of points $\beta(t_1), \beta(t_2), \ldots, \beta(t_m)$. As shown on the left side of Fig. 1.2 this is no problem when the number m of observed points is high and the whole length of the curve is densely 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 in-variance under scaling, rotation and translation (see Srivastava, Klassen, et al. 2011), these approaches

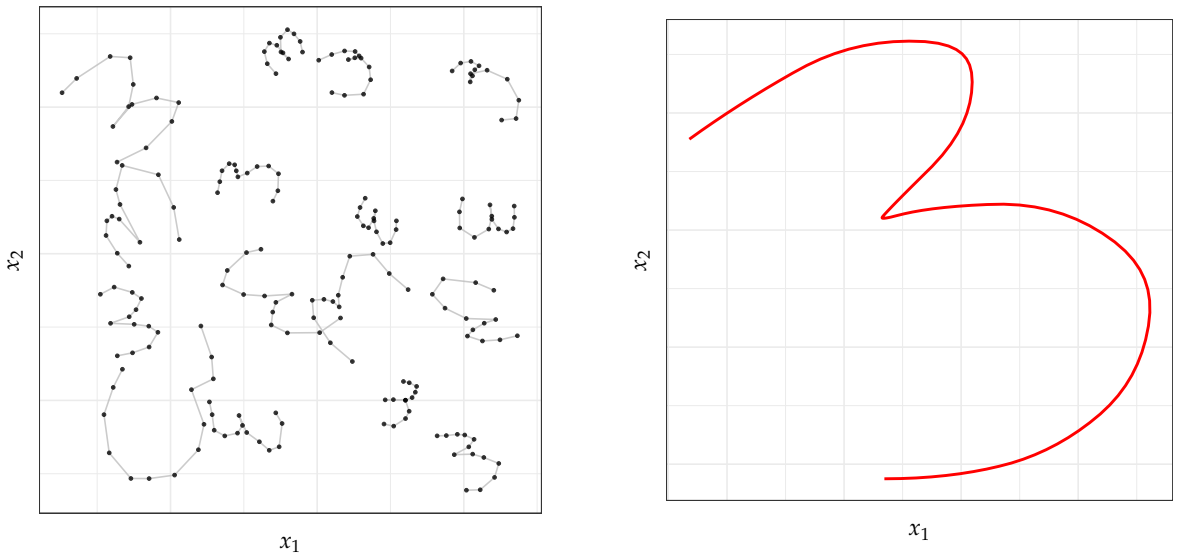


Figure 1.3.: Elastic full Procrustes mean function (right) estimated from sparse and irregular observations (left) using a 13 knot linear B-spline basis, with a 2nd order roughness penalty on SRV level. Data: Original (un-smoothed) `digits3.dat` with additional random rotation, scaling and translation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

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 Fig. 1.3, is to extend existing methods for elastic mean estimation of sparse and irregularly sampled curves, as proposed by Steyer, Stöcker, and Greven 2021 and implemented in the package `elasdics` (Steyer 2021) for the R programming language (R Core Team 2021), to also include in-variance with respect to rotation and scaling. The later will be achieved by generalising the concept of the full Procrustes mean from landmark to functional data and by iteratively applying full Procrustes mean estimation, rotation-alignment and parameterisation-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 Stöcker 2021) will be used. While Steyer, Stöcker, and Greven also propose methods for elastic mean estimation over closed curves, given by $\beta : \mathbb{S}^1 \to \mathbb{C}$, this thesis will only consider means for open curves.

The thesis is organised as follows. In Chapter 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 Chapter 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 Chapter 4 the proposed methods will be verified and applied using simulated and empirical datasets. Finally, all results will be summarised in Chapter 5.

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 curve $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x_1(t), x_2(t))^\top$, 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 Fig. 2.1, and always use complex notation when representing a planar curve:

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

We will assume the curves to be absolutely continuous, denoted as $\beta \in \mathcal{AC}([0,1], \mathbb{C})$, guaranteeing 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.

2.1. Equivalence Classes and Shape

As mentioned in the introduction, shape is usually defined by its in-variance under the transformations of scaling, translation, and rotation. When considering the shape of curves, we additionally have to take into account in-variance 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 Fig. 1.1b). 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

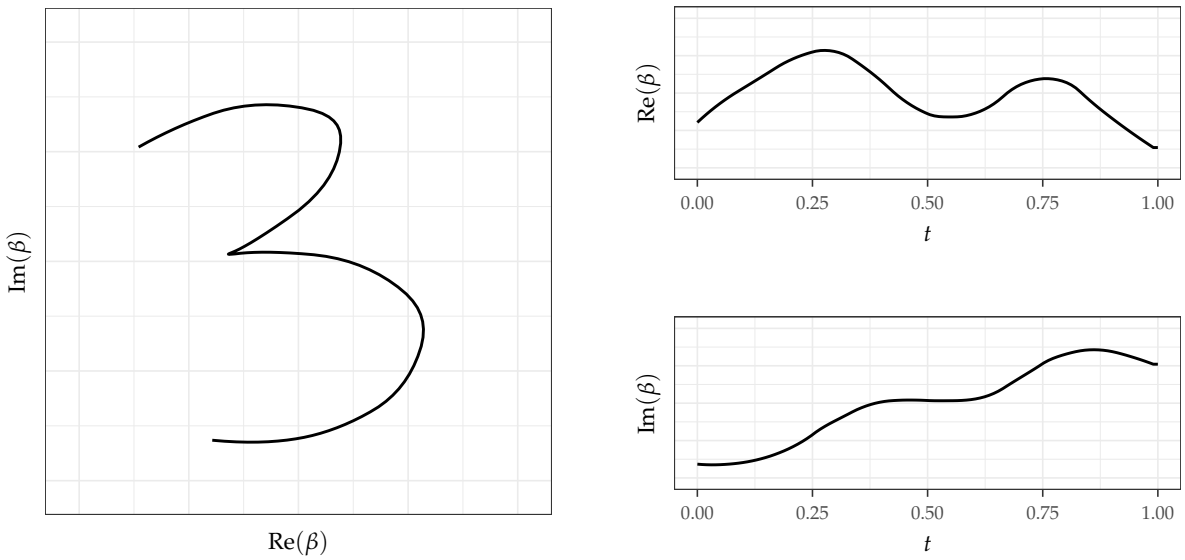


Figure 2.1.: Example of a planar curve (left) with respective coordinate functions (right) using complex notation, where $\mathrm{Re}(\beta) = x_1$ and $\mathrm{Im}(\beta) = x_2$ denote the *real* and *imaginary* parts of β. Data: see Fig. 1.1.

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

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 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 $(\xi, \beta) \overset{\text{Trl}}{\longmapsto} \beta + \xi$, 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 $\widetilde{\xi} \in \mathbb{C}$ such 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 $\widetilde{\lambda} \in \mathbb{R}^+$ such 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 $(\theta, \beta) \overset{\text{Rot}}{\longmapsto} e^{i\theta}\beta$, 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 $(\gamma, \beta) \overset{\text{Wrp}}{\longmapsto} \beta \circ \gamma$, 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 $((\lambda, \theta), \beta) \overset{\text{Scl+Rot}}{\longmapsto} \lambda e^{i\theta}\beta$, 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 $((\xi, \lambda, \theta), \beta) \overset{\text{Eucl}}{\longmapsto} \lambda e^{i\theta}\beta + \xi$ (see Dryden and Mardia 2016, p. 62). Considering the action of *warping* or re-parameterisation, 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 parameterisation. 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 continuous 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-parameterisations

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

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

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 not straight-forward. A common approach is to map each equivalence class $[\beta]$ to a suitable representative $\widetilde{\beta}$, so that the distance calculation in shape space can be identified with a (much simpler) distance calculation over the representatives in an 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 $[\beta_1]_{\text{Trl}}, [\beta_2]_{\text{Trl}}$. They might be uniquely mapped to their centered elements $\widetilde{\beta}_i^{\text{Trl}} = \beta_i - \overline{\beta}_i \in [\beta_i]_{\text{Trl}}$ for $i = 1, 2$. The distance between the centered elements then defines a distance that is invariant under translation $d_{\text{Trl}}([\beta_1]_{\text{Trl}}, [\beta_2]_{\text{Trl}}) = \|\widetilde{\beta}_1^{\text{Trl}} - \widetilde{\beta}_2^{\text{Trl}}\|$, which is equal to the minimal distance when optimizing over the translation group. Similarly, a distance that is invariant under scaling might be defined over the normalised elements $\widetilde{\beta}_i^{\text{Scl}} = \frac{\beta_i}{\|\beta_i\|} \in [\beta_i]_{\text{Scl}}$ for $i = 1, 2$, as $d_{\text{Scl}}([\beta_1]_{\text{Scl}}, [\beta_2]_{\text{Scl}}) = \|\widetilde{\beta}_1^{\text{Scl}} - \widetilde{\beta}_i^{\text{Scl}}\|$. When considering in-variance under rotation, we can first note that no "standardisation" procedure comparable to normalising and centring exists for the case of rotation. Instead of mapping $[\beta]_{\text{Rot}}$ 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 $d_{\text{Rot}}([\beta_1]_{\text{Rot}}, [\beta_2]_{\text{Rot}}) = \min_{\widetilde{\beta}_2^{\text{Rot}} \in [\beta_2]_{\text{Rot}}} \|\beta_1 - \widetilde{\beta}_2^{\text{Rot}}\| = \min_{\theta \in [0, 2\pi)} \|\beta_1 - e^{i\theta}\beta_2\|$, when keeping one curve fixed and rotationally aligning the other curve (see e.g Stöcker and Greven 2021).

The Full Procrustes Distance

The three approaches can be combined to formulate the family of *Procrustes* distances, which are invariant under Euclidean similarity transform. The *partial Procrustes distance* is defined as the minimising distance $d_{PP}([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}}) = \min_{\theta \in [0, 2\pi)} \|\widetilde{\beta}_1 - e^{i\theta}\widetilde{\beta}_2\|$, when rotationally aligning the centred and normalised 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 Definition 2.2) includes an additional alignment over scaling, leading to a slightly different geometrical interpretation (see

Dryden and Mardia 2016, pp. 77–78). Finally, the *Procrustes distance* given by $d_P([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}}) = \arccos|\langle\widetilde{\beta}_1, \widetilde{\beta}_2\rangle|$ defines a geodesic distance in the space of curves modulo Euclidean similarity transforms. As earlier approaches to elastic shape mean estimation, such as Srivastava, Klassen, et al. 2011, have been focused on calculating *intrinsic* means, they mostly use the 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—as will become clear later on—calculating a norm $\|\beta\| = \int_0^1 \|\beta(t)\|\, dt$ may already present a challenge. In this thesis, we will therefore only consider the full Procrustes distance and calculate *extrinsic* shape means. Note that in Definition 2.2 the optimisation over scaling $\lambda \in \mathbb{R}^+$ and rotation $\theta \in [0, 2\pi)$ was combined into a single optimisation over *rotation and scaling* $\omega = \lambda e^{i\theta} \in \mathbb{C}$.

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

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

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

By using a similar proof for complex-valued landmark data in Dryden and Mardia 2016, Chap 8 as a blueprint, we can show that Eq. (2.1) has the following analytical solution.

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 $\widetilde{\beta}_1$.*

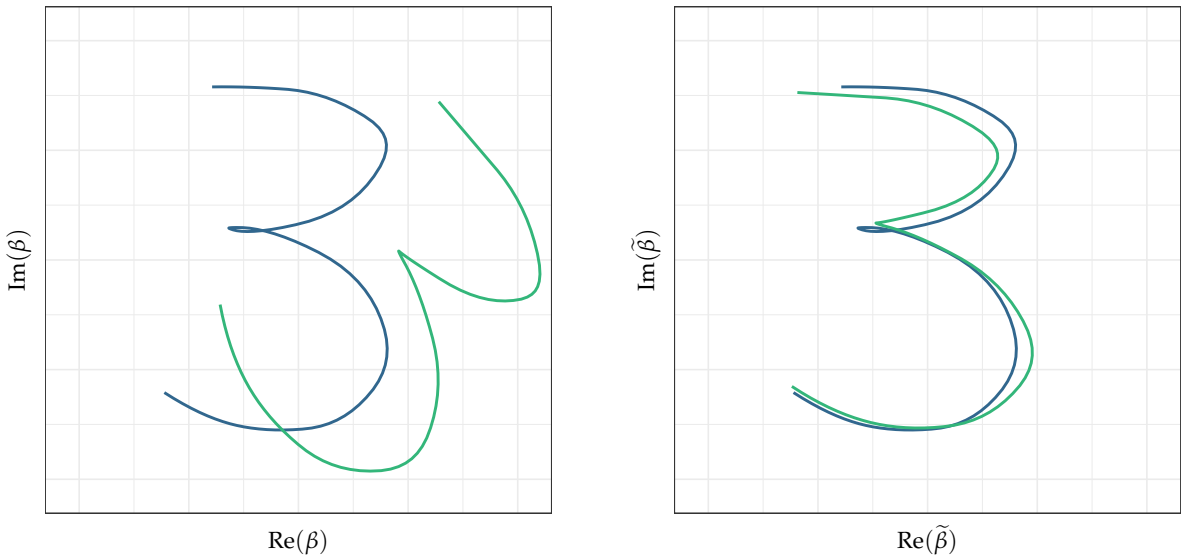


Figure 2.2.: Procrustes fit (right; normalised and centred) of two example curves (left). The Procrustes fit of β_2 (green) onto β_1 (blue) is given by $\widetilde{\beta}_2^p = \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle \widetilde{\beta}_2$. Data: See Fig. 1.1.

Proof. See Appendix A.1.1. □

Fig. 2.2 shows an example of two curves that where aligned by minimising their full Procrustes distance using Lemma 2.1.

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 optimising their distance over the space of warping functions Γ. A usual choice would be to optimise over the $\mathbb{L}^2$-distance by $\inf_{\gamma \in \Gamma} \|\beta_1 - \beta_2 \circ \gamma\|_{\mathbb{L}^2}$, where the $\mathbb{L}^2$-norm and scalar-product are given by $\|f\|_{\mathbb{L}^2} = \sqrt{\int_0^1 \overline{f(t)} f(t) \, \mathrm{d}t}$ and $\langle f, g \rangle_{\mathbb{L}^2} = \int_0^1 \overline{f(t)} g(t) \, \mathrm{d}t$ with $\overline{z} = \mathrm{Re}(z) - \mathrm{i}\mathrm{Im}(z)$ denoting the complex conjugate of $z \in \mathbb{C}$. However, as optimising over re-pasteurisation using the $\mathbb{L}^2$-distance has problems relating to the so called *pinching effect* and *inverse-inconsistency*, this does not define a proper distance. Here, the later means that aligning the pasteurisation of one curve to another by $\inf_{\gamma \in \Gamma} \|\beta_1 - \beta_2 \circ \gamma\|_{\mathbb{L}^2}$ may yield different results than $\inf_{\gamma \in \Gamma} \|\beta_2 - \beta_1 \circ \gamma\|_{\mathbb{L}^2}$ (see Srivastava and Klassen 2016, pp. 88–90).

A solution proposed in Srivastava, Klassen, et al. 2011 is to replace the $\mathbb{L}^2$-distance with an *elastic distance*, on which warping acts by isometry. 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 fulfil 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, e.g.,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 Fig. 2.3 for an example SRV curve of a digit '3'.

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

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

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 $\dot{\beta}_i(t) = \frac{\mathrm{d}\beta_i(t)}{\mathrm{d}t}$ for $i = 1, 2$.

Unlike the optimisation over rotation in the definition of the full Procrustes distance, no analytical solution exists for the optimisation over warping in Eq. (2.3). Instead, it is usually solved numerically, by minimising a cost function $H[\gamma] = \int_0^1 \|q_1(t) - q_2(\gamma(t)) \sqrt{\dot{\gamma}(t)}\| \, \mathrm{d}t$ using a dynamic programming algorithm (see e.g. Srivastava and Klassen 2016, p. 152) or gradient based methods (see e.g. Steyer, Stöcker, and Greven 2021).

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

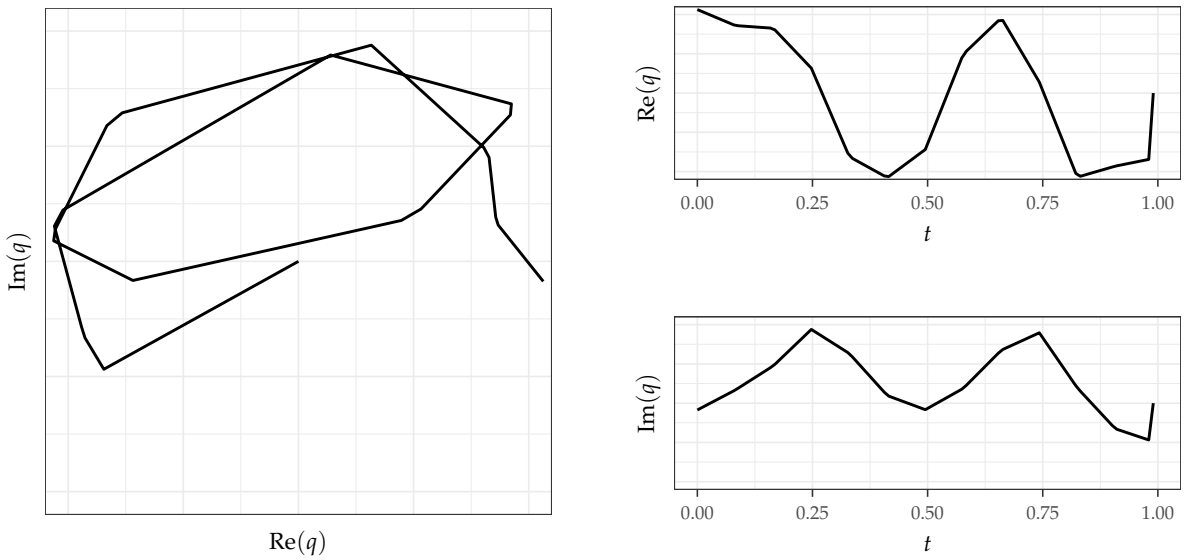


Figure 2.3.: SRV function (left) of the planar curve in Fig. 2.1 with respective SRV coordinate functions (right). Note that the polygon-like look of the SRV curve is an artefact of the (linear) smoothing applied to the original data on SRV level (see Appendix A.2). Data: see Fig. 1.1.

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 Lemma 2.2.

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

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

(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, that scaling translates to SRV level by $\sqrt{\cdot}$. It is in particular 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 Definition 2.2) was defined over the normalised and centred curves, the SRV curves are already translation invariant so additional centring is not necessary (see Lemma 2.2 i.). We will therefore define the *elastic full Procrustes distance* as the minimal distance, when aligning the scaling, rotation, and warping of the normalised SRV curves $\widetilde{q} = \frac{q}{\|q\|}$. Note that when the original curve β is of unit length $L[\beta] = \int_0^1 |\dot{\beta}(t)| \, \mathrm{d}t = 1$ the SRV curve $q = \frac{\dot{\beta}}{\|\dot{\beta}\|}$ will be normalised, as

$$\|q\| = \sqrt{\int_0^1 |q(t)|^2 \, \mathrm{d}t} = \sqrt{\int_0^1 |\dot{\beta}(t)| \, \mathrm{d}t} = \sqrt{L[\beta]}. \tag{2.5}$$

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

$$d([\beta_1], [\beta_2]) = \inf_{\omega \in \mathbb{C}, \, \gamma \in \Gamma} \|\widetilde{q}_1 - \omega(\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\|_{\mathbb{L}^2}, \tag{2.6}$$

with normalised 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 elastic full Procrustes distance, we need to solve the joint optimisation problem over $\mathbb{C} \times \Gamma$

$$\left(\omega^{\mathrm{opt}}, \gamma^{\mathrm{opt}}\right) = \operatorname*{argmin}_{\omega \in \mathbb{C}, \, \gamma \in \Gamma} \|\widetilde{q}_1 - \omega(\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\|_{\mathbb{L}^2}, \tag{2.7}$$

so that the elastic full Procrustes distance is given as the $\mathbb{L}^2$-distance of the optimally aligned normalised SRV curves

$$d([\beta_1], [\beta_2]) = \|\widetilde{q}_1 - \omega^{\mathrm{opt}}(\widetilde{q}_2 \circ \gamma^{\mathrm{opt}})\sqrt{\dot{\gamma}^{\mathrm{opt}}}\|_{\mathbb{L}^2}. \tag{2.8}$$

Following Srivastava, Klassen, et al. 2011 we adapt an iterative procedure, where in each step k we optimise over the sets of parameters individually, iterating through both solutions until some form of convergence is reached. Let us first consider the optimisation over $\omega \in \mathbb{C}$ for a fixed $\gamma^{(k)} \in \Gamma$.

Γ nicht voll-ständig

$$\omega^{(k)} = \underset{\omega \in \mathbb{C}}{\operatorname{argmin}} \, \| \widetilde{q}_1 - \omega (\widetilde{q}_2 \circ \gamma^{(k)}) \sqrt{\dot{\gamma}^{(k)}} \|_{\mathbb{L}^2} \,, \tag{2.9}$$

Eq. (2.9) is equivalent to the optimisation problem of the full Procrustes distance defined in Definition 2.2. Following Lemma 2.1 ii.), the solution is given by the Procrustes fit of $(\widetilde{q}_2 \circ \gamma^{(k)}) \sqrt{\dot{\gamma}^{(k)}}$ onto $\widetilde{q}_1$ with

$$\omega^{(k)} = \langle (\widetilde{q}_2 \circ \gamma^{(k)}) \sqrt{\dot{\gamma}^{(k)}}, \widetilde{q}_1 \rangle \tag{2.10}$$

For fixed rotation and scaling $\omega^{(k)} \in \mathbb{C}$ the optimisation problem over $\gamma \in \Gamma$ is given by

$$\gamma^{(k+1)} = \underset{\gamma \in \Gamma}{\operatorname{arginf}} \, \| \widetilde{q}_1 - (\omega^{(k)} \widetilde{q}_2 \circ \gamma) \sqrt{\dot{\gamma}} \|_{\mathbb{L}^2} \,. \tag{2.11}$$

Again,Eq. (2.11) is equivalent to the optimisation problem of the elastic distance defined in Definition 2.3, when aligning the parameterisation of the normalised SRV curve $\widetilde{q}_1$ and the rotation and scaling aligned, normalised SRV curve $\omega^{(k)} q_2$. A solution $\gamma^{(k+1)}$ can be found by applying known optimisation techniques such as a dynamical programming algorithm or a gradient based method. In this thesis we will use the methods laid out in Steyer, Stöcker, and Greven 2021 and implemented in the `R` package `elasdics` (Steyer 2021) for solving Eq. (2.11) in the setting of sparse and irregularly sampled curves.

Algorithm 2.1 (Elastic full Procrustes distance). *β_1, β_2 absolutely continuous planar curves with SRV curves $q_1, q_2 \in \mathbb{L}^2([0,1], \mathcal{C})$ and normalised SRV curves $\widetilde{q}_i = \frac{q_i}{\|q_i\|}$. Set $\gamma^{(0)}(t) = t$ as the initial parameterisation alignment. Set $k = 0$.*

1. *Set $\widetilde{q}^{(k)} = (\widetilde{q} \circ \gamma^{(k)}) \sqrt{\dot{\gamma}^{(k)}}$*
2. *Calculate $\omega^{(k)} = \langle \widetilde{q}_2^{(k)}, \widetilde{q}_1 \rangle$. **Stop** if $k > 1$ and $\| \omega^{(k)} \widetilde{q}^{(k)} - \omega^{(k-1)} \widetilde{q}^{(k-1)} \| < \epsilon$*
3. *Solve $\gamma^{(k+1)} = \operatorname{argmin}_{\gamma \in \Gamma} \| \widetilde{q}_1 - (\omega^{(k)} \cdot \widetilde{q}_2 \circ \gamma) \sqrt{\dot{\gamma}} \|_{\mathbb{L}^2}$.*
4. *Set $k = k + 1$ and return to Step 1.*

The elastic full Procrustes distance is given as $d([\beta_1],[\beta_2]) = \|\widetilde{q}_1 - \omega^{(k)}(\widetilde{q}_2 \circ \gamma^{(k)})\sqrt{\dot{\gamma}^{(k)}}\|_{\mathbb{L}^2}$.

2.3. The Elastic Full Procrustes Mean for Planar Curves

We now want to use the elastic Full Procrustes distance to calculate shape means for sets of planar curves. Again, we assume all curves to be absolutely continuous $\beta_i \in \mathcal{AC}([0,1], \mathbb{C})$ with corresponding SRV curves $q_i \in \mathbb{L}^2([0,1], \mathbb{C})$, $i = 1, \ldots, N$. We can take into account in-variance with respect to shape-preserving transformations by defining the mean as a minimizer over the sum of squared elastic full Procrustes distances between the shape of each curve and a mean shape. 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).

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

$$[\hat{\mu}] = \operatorname*{arginf}_{[\mu]\in\mathcal{S}} \sum_{i=1}^{N} d_{EF}([\mu],[\beta_i])^2\,, \tag{2.12}$$

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

In practice, we will always solve Eq. (2.12) directly on SRV level, by using the definition of the elastic full Procrustes distance (see Definition 2.4) and writing it as an optimisation problem over a normalised SRV mean function.

$$\hat{\mu}_q = \operatorname*{argmin}_{\mu_q\in\mathbb{L}^2,\,\|\mu_q\|=1} \sum_{i=1}^{N} \left(\inf_{\omega_i\in\mathbb{C},\gamma_i\in\Gamma} \|\mu_q - \omega_i(\widetilde{q}_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}\| \right)^2 . \tag{2.13}$$

The estimated normalised SRV mean $\hat{\mu}_q$ then defines a representative unit-length mean $\hat{\mu} \in [\hat{\mu}]$ by integration $\hat{\mu}(t) = \hat{\mu}(0) + \int_0^t \hat{\mu}_q(s)\|\hat{\mu}_q(s)\|\,ds$ (compare Eq. (2.5)), which is unique up to translation $\hat{\mu}(0)$. When re-constructing $\hat{\mu}$ from $\hat{\mu}_q$, one may decide to set $\hat{\mu}(0)$ to a certain value depending on the application. In particular, setting $\hat{\mu}(0) = 0$, so that the mean curve starts at the origin, makes sense when the object represented by the mean curve has a 'natural' starting point shared across all objects of this type. An

example, explored in Section 4.4 are tongue shapes, which all connect to the back of the mouth on one end. Another possibility would be to choose a $\hat{\mu}(0)$ that centres the mean curve, by setting $\hat{\mu}(0) = \int_0^1 \int_0^t \hat{\mu}_q(s)\|\hat{\mu}_q(s)\|\, ds\, dt$. 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 in visualising multiple curves or in comparing multiple class mean shapes, as those do not typically share the same centre or starting point.

Turning back to the calculation of $\hat{\mu}_1$, we can simplify Eq. (2.13) by applying the following Lemma, which uses the analytical solution for the optimisation over rotation in the full Procrustes distance (see Lemma 2.1 i.).

Lemma 2.3. *Let β_1, β_2 be two absolutely continuous planar curves with corresponding shape $[\beta_1], [\beta_2]$. Let $\widetilde{q}_1, \widetilde{q}_2$ be the respective normalised SRV curves. The elastic full Procrustes distance is given by*

$$d([\beta_1], [\beta_2]) = \inf_{\gamma \in \Gamma} \quad \sqrt{1 - \langle \widetilde{q}_1, (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}} \rangle \langle (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}, \widetilde{q}_1 \rangle} \tag{2.14}$$

Proof. This follows from applying Lemma 2.1 i.) to Eq. (2.6), keeping γ fixed. □

Then Eq. (2.13) can be rewritten as

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2, \|\mu_q\|=1}{\operatorname{argmin}} \sum_{i=1}^{N} \inf_{\gamma_i \in \Gamma} \left(1 - \langle \mu_q, (\widetilde{q}_i \circ \gamma_i)\sqrt{\dot{\gamma}_i} \rangle \langle (\widetilde{q}_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}, \mu_q \rangle\right) \tag{2.15}$$

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2, \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \sup_{\gamma_i \in \Gamma} \langle \mu_q, (\widetilde{q}_i \circ \gamma_i)\sqrt{\dot{\gamma}_i} \rangle \langle (\widetilde{q}_i \circ \gamma_i)\sqrt{\dot{\gamma}_i}, \mu_q \rangle \tag{2.16}$$

and we end up with a two step optimisation problem consisting of an outer optimisation over μ_q and an inner optimisation over the set $\{\gamma_i\}_{i=1,\ldots,N}$. Similarly to the approaches discussed in Srivastava and Klassen 2016 and to Steyer, Stöcker, and Greven 2021, we solve this iteratively by iterative *template based alignment* (see e.g. Srivastava and Klassen 2016, p. 271): In each step, the mean $\hat{\mu}_q$ is estimated while keeping the parameterisations γ_i fixed, after which the γ_i are updated by calculating the warping alignment of the full Procrustes fit of each $\widetilde{q}_i$ onto $\hat{\mu}_q$.[1] Mean estimation

[1] Note that this is *not* the same as performing an elastic full Procrustes fit of $\widetilde{q}_i$ onto $\hat{\mu}_q$ in each step k,

and warping alignment are iterated until the mean shape has converged.

Let us consider the outer optimisation problem in step k for a fixed set of warping functions $\{\gamma_i^{(k)}\}_{i=1,\dots,N}$ with corresponding warping aligned normalised SRV curves $\widetilde{q}_i^{(k)} = (\widetilde{q}_i \circ \gamma_i^{(k)})\sqrt{\dot{\gamma}_i^{(k)}}$. Note that if no warping alignment has happened yet, we can always set $\gamma_i^{(0)}(t) = t$ for all $i = 1, \dots, N$ as a starting value. The problem we have to solve is

$$\hat{\mu}_q^{(k)} = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \left\langle \mu_q, \widetilde{q}_i^{(k)} \right\rangle \left\langle \widetilde{q}_i^{(k)}, \mu_q \right\rangle . \tag{2.17}$$

We can reformulate this by writing out the complex functional scalar products $\langle f, g \rangle = \int_0^1 \overline{f(t)} g(t)\, \mathrm{d}t$ for functions $f, g \in \mathbb{L}^2([0,1], \mathbb{C})$, where $\overline{f(t)}$ denotes the complex conjugate of $f(t)$.

$$\hat{\mu}_q^{(k)} = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \int_0^1 \int_0^1 \overline{\mu_q(s)} \widetilde{q}_i^{(k)}(s) \overline{\widetilde{q}_i^{(k)}(t)} \mu_q(t)\, \mathrm{d}s\, \mathrm{d}t \tag{2.18}$$

$$\hat{\mu}_q^{(k)} = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu_q(s)} \left(\sum_{i=1}^{N} \widetilde{q}_i^{(k)}(s) \overline{\widetilde{q}_i^{(k)}(t)} \right) \mu_q(t)\, \mathrm{d}s\, \mathrm{d}t \tag{2.19}$$

We can identify the inner term as proportional to a sample estimator $\check{C}^{(k)}(s,t) = \frac{1}{N} \sum_{i=1}^{N} \widetilde{q}_i^{(k)}(s) \overline{\widetilde{q}_i^{(k)}(t)}$ of the population covariance surface of the normalised SRV curves $C^{(k)}(s,t) = \mathbb{E}[\widetilde{q}^{(k)}(s) \overline{\widetilde{q}^{(k)}(t)}]$, when noting that $\mathbb{E}[\widetilde{q}^{(k)}(t)] = 0$ for all $t \in [0,1]$ due to rotational symmetry.

$$\hat{\mu}_q^{(k)} = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} N \cdot \int_0^1 \int_0^1 \overline{\mu_q(s)} \check{C}^{(k)}(s,t) \mu_q(t)\, \mathrm{d}s\, \mathrm{d}t \tag{2.20}$$

By replacing $\check{C}^{(k)}(s,t)$ by its expectation $C^{(k)}(s,t)$, we can analogously formulate an estimator on the population level.

$$\mathbb{E}[\mu_q^{(k)}] = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu_q(s)} C^{(k)}(s,t) \mu_q(t)\, \mathrm{d}s\, \mathrm{d}t \tag{2.21}$$

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

because the warping and Procrustes alignments are themselves not iterated over.

p. 153).

$$\mathbb{E}[\mu_q^{(k)}] = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \left\langle \mu_q, C^{(k)} \mu_q \right\rangle \tag{2.22}$$

This is a well known problem in the context of functional principal component analysis (FPCA), but typically only for real-valued covariance operators. From $\overline{C(s,t)} = \overline{\mathbb{E}[\widetilde{q}(s)\overline{\widetilde{q}(t)}]} = \mathbb{E}[\widetilde{q}(t)\overline{\widetilde{q}(s)}] = C(t,s)$ 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 optimisation problem then reduces to an eigenfunction problem

$$C^{(k)} u^{(k)} = \lambda^{(k)} u^{(k)} \quad \Leftrightarrow \quad \int_0^1 C^{(k)}(s,t) u^{(k)}(t)\, dt = \lambda^{(k)} u^{(k)}(s)\,, \tag{2.23}$$

where $\lambda^{(k)} = \left\langle \mu_q,\, C^{(k)} \mu_q \right\rangle$ is the target function to maximise. For normalised eigenfunctions $u_1^{(k)}, u_2^{(k)}, \ldots$ and corresponding eigenvalues $\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \ldots$ of $C^{(k)}(s,t)$, the expectation $\mathbb{E}[\hat{\mu}_q^{(k)}(t)]$ is given by the leading normalised eigenfunction $u_1^{(k)}(t)$ of $C^{(k)}(s,t)$ (see Ramsay and Silverman 2005, pp. 153, 397).

We will estimate $C^{(k)}(s,t)$ using the methods for covariance estimation from sparse and irregular observations laid out in Chapter 3. Given such an estimate $\hat{C}^{(k)}(s,t)$, we can calculate the elastic full Procrustes mean by the following algorithm.

Algorithm 2.2 (Elastic full Procrustes mean)**.** *Let $\{\beta_i\}_{i=1,\ldots,N}$ be a set of planar curves with corresponding SRV curves $\{q_i\}_{i=1,\ldots,N}$. Let $\widetilde{q}_i = \frac{q_i}{\|q_i\|}$. Set $\gamma_i^0(t) = t$ for all $i = 1, \ldots, N$ as the initial parametrisation alignment. Set $k = 0$.*

1. *For $i = 1, \ldots, N$: Set $\widetilde{q}_i^{(k)} = \left(\widetilde{q}_i \circ \gamma_i^{(k)}\right) \cdot \sqrt{\dot{\gamma}_i^{(k)}}$.*
2. *Estimate $\hat{C}^{(k)}(s,t)$ from $\left\{\widetilde{q}_i^{(k)}\right\}_{i=1,\ldots,N}$.*
3. *Estimate $\hat{u}_1^{(k)}$ by eigendecomposition of $\hat{C}^{(k)}(s,t)$.*
4. *Set $\hat{\mu}_q^{(k)}$ as $\hat{u}_1^{(k)}$.* **Stop** *if $k > 1$ and $\|\hat{\mu}_q^{(k)} - \hat{\mu}_q^{(k-1)}\| < \epsilon$.*
5. *For $i = 1, \ldots, N$: Calculate $\omega_i^{(k)} = \left\langle \widetilde{q}_i^{(k)},\, \hat{\mu}_q^{(k)} \right\rangle$.*
6. *For $i = 1, \ldots, N$: Solve $\gamma_i^{(k+1)} = \operatorname{argmin}_{\gamma \in \Gamma} \|\hat{\mu}_q^{(k)} - \omega_i^{(k)} (\widetilde{q}_i \circ \gamma) \sqrt{\dot{\gamma}}\|$.*
7. *Set $k = k + 1$ and return to Step 1.*

3. Mean Estimation for Sparse and Irregular Observations

So far, we have considered estimation of the elastic full Procrustes mean in a setting, where each curve β_i is assumed fully observed. This is usually not the case in practice, as each observation β_i may itself only be observed at a finite number of discrete points $\beta_i(t_{i1}), \ldots, \beta_i(t_{in_i})$. Additionally, the number of observed points per curve n_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 Algorithm 2.2, this chapter 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 $C^{(k)}(s,t)$ from sparse observations. In Section 3.3, calculation of the leading eigenfunction $\hat{u}_1^{(k)}$ of $C^{(k)}(s,t)$ in a fixed basis will be shown. Section 3.4 deals with the estimation of the optimal rotation and scaling alignment $\omega_i^{(k)} = \langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle$, where $\widetilde{q}_i^{(k)}$ is a sparsely observed normalised SRV curve, while $\hat{\mu}_q^{(k)}$ is a smooth SRV mean function. Note that the final warping alignment step in Algorithm 2.2 is solved by using methods for warping alignment of sparse and irregular curves provided in Steyer, Stöcker, and Greven 2021.

3.1. Discrete Treatment of SRV Curves

As a first step, we need to calculate the *normalised SRV curves* $\widetilde{q}_i = \frac{q}{\|q\|}$ from sparse observations. As the SRV curve of $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ is defined as $q = \dot{\beta}/\sqrt{\|\dot{\beta}\|}$ (for $\dot{\beta} \neq 0$), we have to be able to calculate a derivative of β. However, as we never observe the whole function β but only a discrete set of points $\beta(t_1), \ldots, \beta(t_n)$, as seen in Fig. 3.1, we cannot simply calculate a point-wise derivative. Following Steyer, Stöcker, and

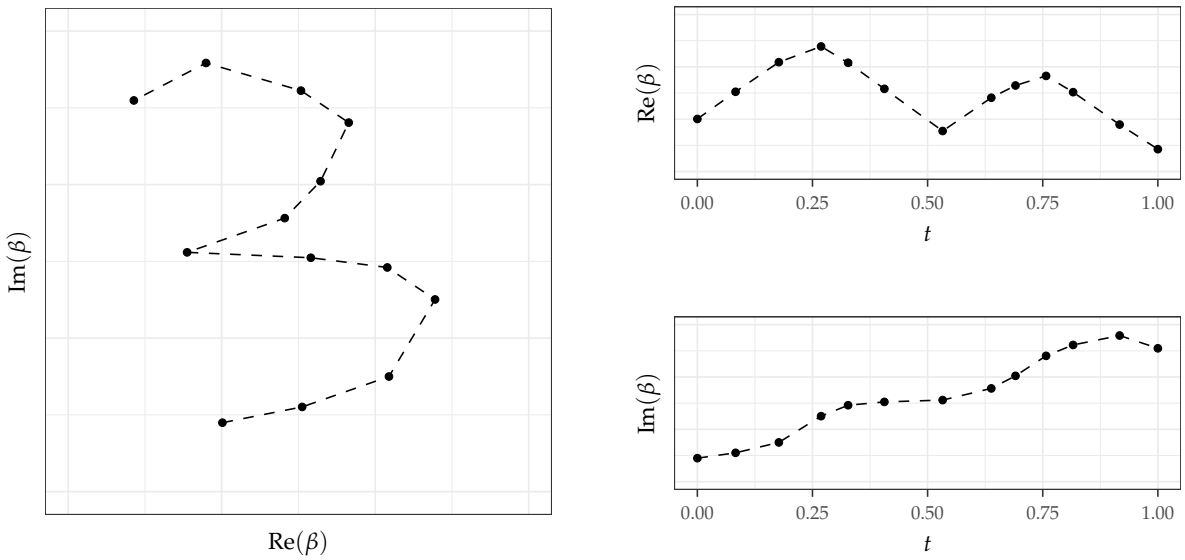


Figure 3.1.: Example of a sparse and irregularly observed digit '3'. Data: `digits3.dat`.

Greven 2021, we may treat a discretely observed curve β as piece-wise linear between its observed corners $\beta(t_1), \ldots, \beta(t_n)$, which allows us to calculate a piece-wise constant derivative on the intervals $[t_j, t_{j+1}]$ for $j = 1, \ldots, n-1$. As usually only the image $\beta(t_1), \ldots, \beta(t_n)$ but not the parametrisation $t_1, \ldots, t_n$ is observed, it is first necessary to construct an initial parameterisation. A common choice is an *arc-length-parametrisation*, where we set $t_j = l_j/l$ with $l_j = \sum_{k=1}^{j-1} |\beta(t_{k+1}) - \beta(t_k)|$ the polygon-length up to point j for $j \leq 2$ with $l_1 = 0$ and $l = l_n$.

Consider the piece-wise-constant derivative $\Delta\beta\big|_{[t_j,t_{j+1}]} = \frac{\beta(t_{j+1})-\beta(t_j)}{t_{j+1}-t_j}$, which assumes that β is linear between its observed corners. The corresponding SRV curve q can then similarly be treated as piece-wise constant $q\big|_{[t_j,t_{j+1}]} = q_j$ with

$$q_j = \Delta\beta\big|_{[t_j,t_{j+1}]} \Big/ \sqrt{\|\Delta\beta\big|_{[t_j,t_{j+1}]}\|} = \frac{\beta(t_{j+1}) - \beta(t_j)}{\sqrt{t_{j+1} - t_j} \cdot \sqrt{\|\beta(t_{j+1}) - \beta(t_j)\|}} \tag{3.1}$$

the constant *square-root-velocity* of β between its corners $\beta(t_j)$ and $\beta(t_{j+1})$. As shown in Steyer, Stöcker, and Greven 2021, Fig. 3, treating the SRV curves as piece-wise-constant functions can lead to over-fitting, 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 piecwise-constant derivative $\Delta\beta\big|_{[t_j,t_{j+1}]}$ at the centre $s_j = \frac{t_{j+1}-t_j}{2}$ of the interval $[t_j, t_{j+1}]$. Here, this will be used for approximating

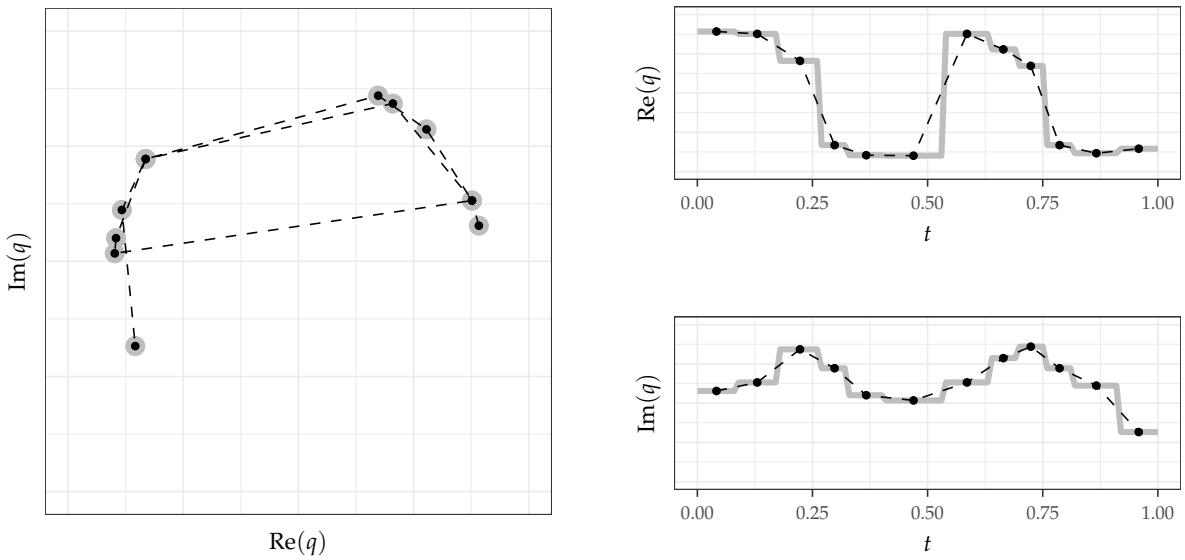


Figure 3.2.: Discretely approximated (black, dashed) and piece-wise constant (gray) SRV curve of the digit '3' in Fig. 3.1. Data: `digits3.dat`.

observations $q(s_j) \approx q_j$ of the SRV curve q in the covariance estimation step. See Fig. 3.2 for a visualisation of both approaches. Finally, we can approximate the normalised SRV curve $\widetilde{q} = q/\|q\|$ using the polygon-length l of β by $\widetilde{q}_j = q_j/\sqrt{l}$ (see Eq. (2.5)). When considering the warped normalised SRV curve $(q \circ \gamma)\sqrt{\dot{\gamma}}$, the warped discrete derivative is given by $\Delta(\beta \circ \gamma)\big|_{[\gamma^{-1}(t_j),\gamma^{-1}(t_{j+1})]} = \frac{\beta(t_{j+1})-\beta(t_j)}{\gamma^{-1}(t_{j+1})-\gamma^{-1}(t_j)}$. The corresponding warped SRV curve is then given by $(q \circ \gamma)\sqrt{\dot{\gamma}}\big|_{[\gamma^{-1}(t_j),\gamma^{-1}(t_{j+1})]} = \frac{1}{\sqrt{\gamma^{-1}(t_{j+1})-\gamma^{-1}(t_j)}} \cdot \frac{\beta(t_{j+1})-\beta(t_j)}{\sqrt{\|\beta(t_{j+1})-\beta(t_j)\|}}$ (see Steyer, Stöcker, and Greven 2021). Note that this does not change the normalisation as re-parameterisation is norm-preserving on SRV level.

3.2. Efficient Estimation using Hermitian Covariance Smoothing

Given approximate observations of the warped normalised SRV curves $\widetilde{q}_i^{(k)}(s_{ij})$ for $j = 1, \ldots, n_i - 1$ and $i = 1, \ldots, N$, where n_i denotes the number of observed points per curve, we want to estimate the warping aligned complex covariance surface $C^{(k)}(s,t) = \mathbb{E}[\widetilde{q}^{(k)}(s)\overline{\widetilde{q}^{(k)}(t)}]$. We can treat this estimation as a smoothing problem, by constructing responses $y_{ilm}^{(k)} = \widetilde{q}_i^{(k)}(s_{il})\overline{\widetilde{q}_i^{(k)}(s_{im})}$ and treating the pairs s_{il}, s_{im} as covariates s and t (see Yao, Müller, and Wang 2005). Smoothing the responses $y_{ilm}^{(k)}$ gives an estimate $\hat{C}^{(k)}(s,t)$ of $C^{(k)}(s,t)$, as each response has expectation $\mathbb{E}[y_{ilm}^{(k)}|s_{il}, s_{im}] = C^{(k)}(s_{il}, s_{im})$.

We carry out the smoothing in a flexible *penalised tensor product spline* basis

$$C^{(k)}(s,t) = b(s)^\top \Xi^{(k)} b(t) \tag{3.2}$$

where $b(s) = (b_1(s), \ldots, b_K(s))$ denotes the vector of a spline basis and $\Xi^{(k)}$ is a $K \times K$ coefficient matrix to be estimated under a roughness penalty to prevent over-fitting. As $C^{(k)}(s,t)$ is complex, we choose the spline basis to be real-valued with $b_j : [0,1] \to \mathbb{R}$ for $j = 1, \ldots, K$ and the coefficient matrix to be complex-valued with $\Xi^{(k)} \in \mathbb{C}^{K \times K}$ without loss of generality. The exact choice of basis and penalty will be discussed in Section 3.3.

Taking into account the symmetry properties of the covariance surface by considering every unique pair (s_{il}, s_{im}) only once allows for more efficient estimation, as shown in Cederbaum, Scheipl, and Greven 2018. In the complex case the covariance surface is Hermitian with $C^{(k)}(s,t) = \overline{C^{(k)}(t,s)}$, which means we can decompose the estimation into two separate regression problems over the symmetric real and skew-symmetric imaginary parts of $C^{(k)}(s,t)$. We estimate the two models

$$\mathbb{E}[\mathrm{Re}(y^{(k)})|s,t] = b(s)^\top \Xi^{(k)}_{\mathrm{Re}} b(t) \tag{3.3}$$

$$\mathbb{E}[\mathrm{Im}(y^{(k)})|s,t] = b(s)^\top \Xi^{(k)}_{\mathrm{Im}} b(t)\,, \tag{3.4}$$

with $\Xi^{(k)}_{\mathrm{Re}}, \Xi^{(k)}_{\mathrm{Im}} \in \mathbb{R}^{K \times K}$ and $\Xi^{(k)} = \Xi^{(k)}_{\mathrm{Re}} + \mathrm{i}\Xi^{(k)}_{\mathrm{Im}}$, under the constraints that $(\Xi^{(k)}_{\mathrm{Re}})^\top = \Xi^{(k)}_{\mathrm{Re}}$ and $(\Xi^{(k)}_{\mathrm{Im}})^\top = -\Xi^{(k)}_{\mathrm{Im}}$. In this thesis $\Xi^{(k)}_{\mathrm{Re}}$ and $\Xi^{(k)}_{\mathrm{Im}}$ are estimated using the `gam` function from the `R` package `mgcv` (Wood 2017), where the smoothing parameters are selected via restricted maximum likelihood (REML) estimation. Two `mgcv` smooths provided in the package `sparseFLMM` (Cederbaum, Volkmann, and Stöcker 2021) are used for efficient Hermitian smoothing, which generalise the approach proposed by Cederbaum, Scheipl, and Greven 2018 for symmetric tensor product P-splines to the skew-symmetric case. Note that `mgcv` automatically adds a sum-to-zero constraint to a specified basis (see Wood 2017, p. 175), which makes it is necessary to transform the coefficient matrices $\widetilde{\Xi}^{(k)}_{\mathrm{Re}}$, $\widetilde{\Xi}^{(k)}_{\mathrm{Im}}$ recovered from `gam` with an appropriate transformation matrix D. The coefficient matrices in the specified basis are given by $\Xi^{(k)}_{\mathrm{Re/Im}} = D \cdot \widetilde{\Xi}^{(k)}_{\mathrm{Re/Im}}$, where D may be calculated from the constrained and unconstrained design matrices via $X = D \cdot \widetilde{X}$.

3.3. Estimation of the Elastic Full Procrustes Mean in a Fixed Basis

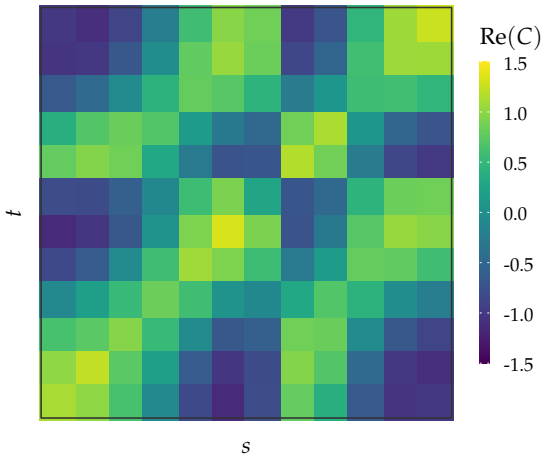
Our goal is to estimate a smooth mean function, which might be constructed with the same uni-variate basis $b(s)$ used in the tensor product basis of the covariance surface, so that the mean is given by $\mu_q(s) = b(s)^\top \theta$. We can choose an appropriate basis $b(s)$ by considering which, e.g., smoothness properties we want the estimated mean to have. In this thesis we will use penalised B-spline basis functions (P-splines), which are piece-wise polynomials of degree l, fused at m knots $\{\kappa_j\}_{j=1,\dots,m}$, where the p-th order differences between coefficients of neighbouring splines are penalised in the covariance estimation (see Fahrmeier et al. 2013, Chap. 8.1). It should be noted that by using a penalty, the number and location of knots do not have an influence on the estimated function when their number is high enough and they are evenly distributed. Furthermore, because of results relating to the identifiable of spline SRV curves modulo warping and translation, only piece-wise linear ($l = 1$) and piece-wise constant ($l = 0$) B-splines will be considered in the estimation (see Steyer, Stöcker, and Greven 2021). See Fig. 3.3 for examples of a covariance surface estimated using piece-wise constant and piece-wise linear P-splines. This still provides the flexibility of estimating either polygonal or smooth means, as the mean function on original curve level is calculated by integration (see Fig. 3.4).

To calculate the elastic full Procrustes Mean, a functional eigenvalue problem on the estimated covariance surface $\hat{C}(s,t) = b(s)^\top \hat{\Xi} b(t)$ had to be solved (omitting '$\cdot^{(k)}$' in this section). Remember that the elastic full Procrustes mean (for fixed warping) is estimated by the solution to the optimisation problem

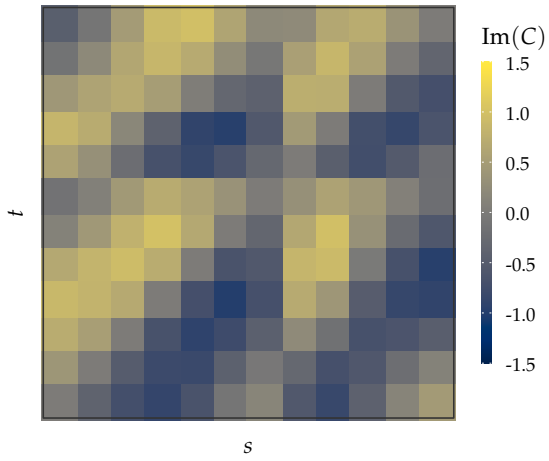
$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu_q(s)} \hat{C}(s,t) \mu_q(t) \,\mathrm{d}s \,\mathrm{d}t \ . \tag{3.5}$$

Then, estimating the mean $\hat{\mu}_q(s) = b(s)^\top \theta$ reduces to estimating the vector of coefficients $\theta = (\theta_1, \dots, \theta_K) \in \mathbb{C}^K$ with

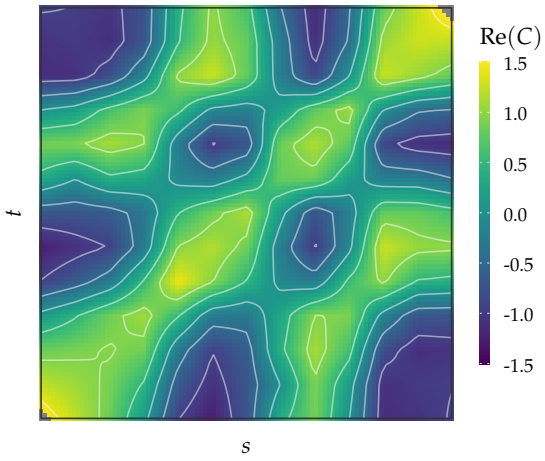
$$\hat{\theta} = \underset{\theta \in \mathbb{C}^K,\, \|b^\top \theta\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \theta^H b(s) b(s)^\top \hat{\Xi} b(t) b(t)^\top \theta \,\mathrm{d}s \,\mathrm{d}t \tag{3.6}$$



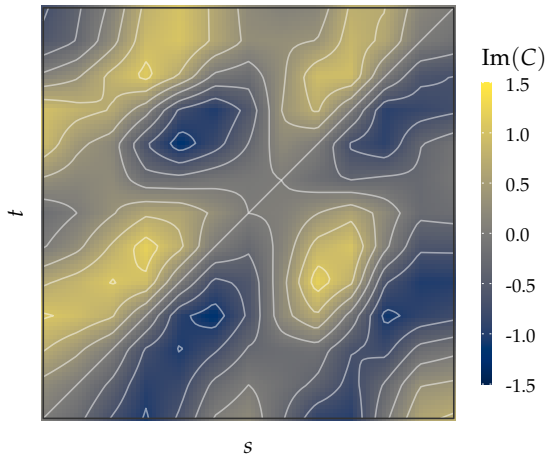
(a) Symmetric real part ($l = 0$)



(b) Skew-symmetric imaginary part ($l = 0$)



(c) Symmetric real part ($l = 1$)



(d) Skew-symmetric imaginary part ($l = 1$)

Figure 3.3.: Complex covariance surface on SRV curve level. Estimated using 13 equidistant knots, a 2nd order penalty and piece-wise constant (top) or piece-wise linear (bottom) B-splines. Data: `digits3.dat`

$$= \underset{\theta \in \mathbb{C}^K, \|b^\top \theta\| = 1}{\operatorname{argmax}} \theta^H \left(\int_0^1 b(s) b(s)^\top \, \mathrm{d}s \right) \hat{\Xi} \left(\int_0^1 b(t) b(t)^\top \, \mathrm{d}t \right) \theta \tag{3.7}$$

$$= \underset{\theta \in \mathbb{C}^K, \theta^H G \theta = 1}{\operatorname{argmax}} \theta^H G \hat{\Xi} G \theta \tag{3.8}$$

where $(\cdot)^H = \overline{(\cdot)}^\top$ denotes the conjugate transpose and G is the $K \times K$ Gram matrix with entries given by the basis products $g_{ij} = \langle b_i, b_j \rangle$. For an orthonormal basis the Gram matrix is an identity matrix as $\langle b_i, b_j \rangle = \delta_{ij}$, however, this is not the case for many basis representations such as, e.g.,the B-spline basis. In thesis the `R` package `orthogonalsplinebasis` (Redd 2015) is used to calculate Gram matrices for B-splines analytically, using the methods laid out in Redd 2012.

Having reduced the functional eigenvalue problem to a multivariate eigenvalue problem over the covariance coefficient matrix, we may solve it using Lagrange optimisation with the following Lagrangian:

$$\mathcal{L}(\theta, \lambda) = \theta^H G \hat{\Xi} G \theta - \lambda(\theta^H G \theta - 1) \tag{3.9}$$

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

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

By multiplying everything out and using $\hat{\Xi}_{\mathrm{Re}}^\top = \hat{\Xi}_{\mathrm{Re}}$ and $\hat{\Xi}_{\mathrm{Im}}^\top = -\hat{\Xi}_{\mathrm{Im}}$ we get

$$\begin{aligned}\mathcal{L}(\theta_{\mathrm{Re}}, \theta_{\mathrm{Im}}, \lambda) = \theta_{\mathrm{Re}}^\top G \hat{\Xi}_{\mathrm{Re}} G \theta_{\mathrm{Re}} + \mathrm{i}\theta_{\mathrm{Re}}^\top G \hat{\Xi}_{\mathrm{Im}} G \theta_{\mathrm{Re}} + \theta_{\mathrm{Im}}^\top G \hat{\Xi}_{\mathrm{Im}} G \theta_{\mathrm{Re}} - \theta_{\mathrm{Re}}^\top G \hat{\Xi}_{\mathrm{Im}} G \theta_{\mathrm{Im}} \\ + \theta_{\mathrm{Im}}^\top G \hat{\Xi}_{\mathrm{Re}} G \theta_{\mathrm{Im}} + \mathrm{i}\theta_{\mathrm{Im}}^\top G \hat{\Xi}_{\mathrm{Im}} G \theta_{\mathrm{Im}} + \lambda \theta_{\mathrm{Re}}^\top G \theta_{\mathrm{Re}} + \lambda \theta_{\mathrm{Im}}^\top G \theta_{\mathrm{Im}} - \lambda .\end{aligned}$$

Differentiation with respect to θ_{Re} and θ_{Im} yield

$$\frac{\partial \mathcal{L}}{\partial \theta_{\mathrm{Re}}} = 2G\hat{\Xi}_{\mathrm{Re}} G\theta_{\mathrm{Re}} - 2G\hat{\Xi}_{\mathrm{Im}} G\theta_{\mathrm{Im}} - 2\lambda G\theta_{\mathrm{Re}} \stackrel{!}{=} 0 \tag{3.10}$$

$$\frac{\partial \mathcal{L}}{\partial \theta_{\mathrm{Im}}} = 2G\hat{\Xi}_{\mathrm{Re}} G\theta_{\mathrm{Im}} + 2G\hat{\Xi}_{\mathrm{Im}} G\theta_{\mathrm{Re}} - 2\lambda G\theta_{\mathrm{Im}} \stackrel{!}{=} 0 \tag{3.11}$$

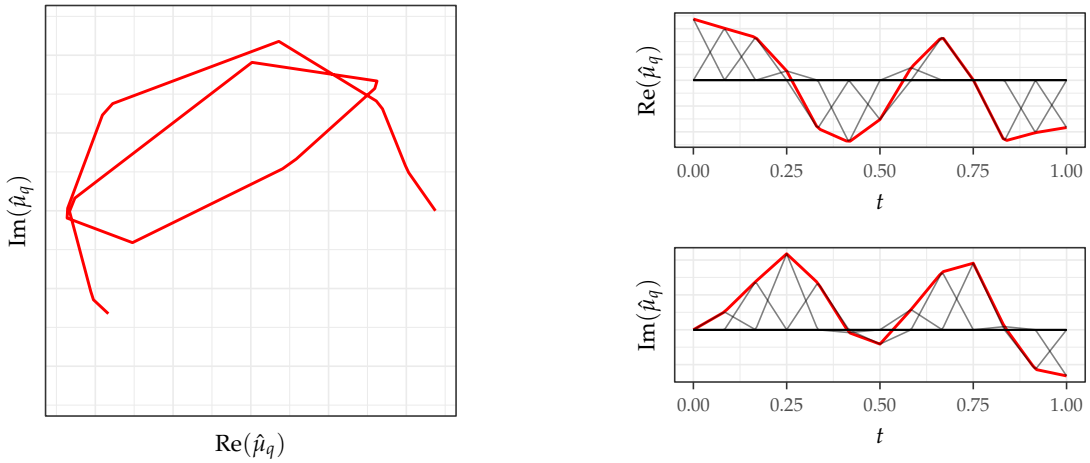
with the additional constraint $\theta_{\mathrm{Re}}^\top G\theta_{\mathrm{Re}} + \theta_{\mathrm{Im}}^\top G\theta_{\mathrm{Im}} = 1$. We can simplify this further and multiply Eq. (3.11) by i, leading to

$$\hat{\Xi}_{\mathrm{Re}} G\theta_{\mathrm{Re}} - \hat{\Xi}_{\mathrm{Im}} G\theta_{\mathrm{Im}} = \lambda\theta_{\mathrm{Re}} \tag{3.12}$$

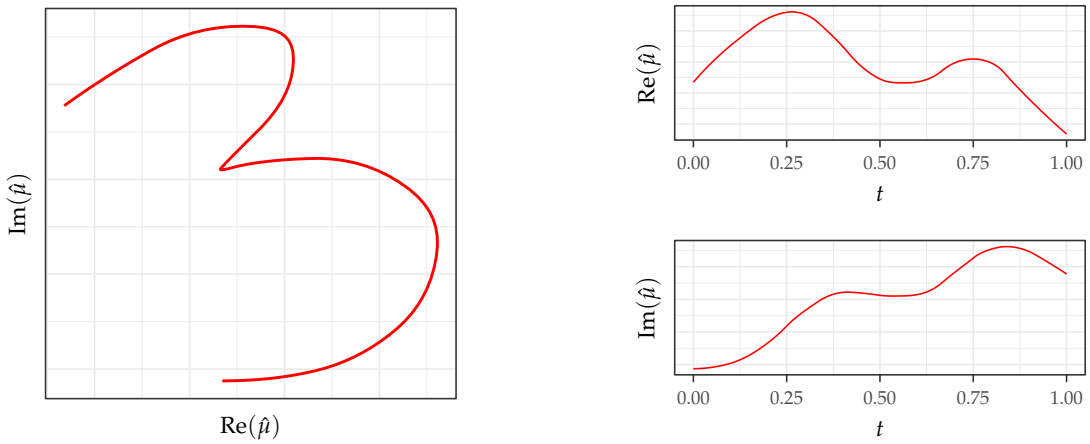
$$\mathrm{i}\hat{\Xi}_{\mathrm{Re}} G\theta_{\mathrm{Im}} + \mathrm{i}\hat{\Xi}_{\mathrm{Im}} G\theta_{\mathrm{Re}} = \mathrm{i}\lambda\theta_{\mathrm{Im}} . \tag{3.13}$$

Adding both equations finally leads to

$$(\hat{\Xi}_{\mathrm{Re}} + \mathrm{i}\hat{\Xi}_{\mathrm{Im}}) G(\theta_{\mathrm{Re}} + \mathrm{i}\theta_{\mathrm{Im}}) = \lambda(\theta_{\mathrm{Re}} + \mathrm{i}\theta_{\mathrm{Im}}) \tag{3.14}$$



(a) Mean function $\hat{\mu}_q(t) = b(t)^\top \hat{\theta}$ on SRV level with scaled linear B-spline basis functions (right).



(b) Mean function $\hat{\mu}(t) = \int_0^t \hat{\mu}_q(s)\|\hat{\mu}_q(s)\| \, \mathrm{d}s$ on (unit-length) original curve level.

Figure 3.4.: Elastic full Procrustes mean of handwritten digits '3'. Estimated using 13 equidistant knots, a 2nd order penalty and piece-wise linear B-splines. Data: `digits3.dat`

$$\hat{\Xi} G \theta = \lambda \theta \tag{3.15}$$

which is an eigenvalue problem on the product of the complex coefficient matrix and the Gram matrix. Note that this mirrors the result by Reiss and Xu 2020 for FPCA on real-valued tensor product spline coefficient matrices. Multiplying by $\theta^H G$ from the left and using $\theta^H G \theta = 1$ yields $\lambda = \theta^H G \hat{\Xi} G \theta$, i.e. the eigenvalues λ correspond to the target function to maximise (see Eq. (3.8)). It follows that the estimate for the coefficient vector of the elastic full Procrustes mean $\hat{\theta}$ is given by the eigenvector of the leading eigenvalue of $\hat{\Xi} G$ or likewise of $\hat{\Xi}^{(k)} G$, when taking into account the warping alignment in step k of Algorithm 2.2.

3.4. Numerical Integration of the Warping-Aligned Procrustes Fits

Given an estimated mean function $\hat{\mu}_q^{(k)} = b(s)^\top \hat{\theta}^{(k)}$, we need to calculate the Procrustes alignment $\omega_i^{(k)} = \langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle$, $i = 1, \ldots, N$, of each curve onto the current mean, before we can update the warping alignment (see Algorithm 2.2). In the sparse and irregular setting, calculating a scalar product such as $\langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle$ can be a challenge, as we have to evaluate $\widetilde{q}_i^{(k)}(t)$ at every $t \in [0, 1]$.

$$\langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle = \int_0^1 \langle \widetilde{q}_i^{(k)}(t), \hat{\mu}_q^{(k)}(t) \rangle \, \mathrm{d}t \tag{3.16}$$

Following Steyer, Stöcker, and Greven 2021 and as discussed in Section 3.1, this may be approximated by treating $\widetilde{q}_i^{(k)}$ as piece-wise constant between the warping-aligned corners of the original curve β_i, with the normalised and warping aligned values given by

$$\widetilde{q}_i^{(k)}\Big|_{\left[\left(\gamma_i^{(k)}\right)^{-1}(t_{ij}), \left(\gamma_i^{(k)}\right)^{-1}(t_{ij+1})\right]} = \widetilde{q}_{ij}^{(k)} \tag{3.17}$$

with $j = 1, \ldots, m_i - 1$ for m_i the number of observed points of the original curve β_i and where

$$\widetilde{q}_{ij}^{(k)} = \frac{1}{\sqrt{L[\beta_i]}} \cdot \frac{1}{\sqrt{\left(\gamma_i^{(k)}\right)^{-1}(t_{ij+1}) - \left(\gamma_i^{(k)}\right)^{-1}(t_{ij})}} \cdot \frac{\beta_i(t_{ij+1}) - \beta_i(t_{ij})}{\sqrt{\|\beta_i(t_{ij+1}) - \beta_i(t_{ij})\|}}. \tag{3.18}$$

Given a fully observed smooth mean function $\hat{\mu}_q^{(k)}$ we can then calculate the scalar product as

$$\langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle = \int_0^1 \langle \widetilde{q}_i^{(k)}(t), \hat{\mu}_q^{(k)}(t) \rangle \, \mathrm{d}t = \sum_{j=0}^{m_i - 1} \int_{\left(\gamma_i^{(k)}\right)^{-1}(t_{ij})}^{\left(\gamma_i^{(k)}\right)^{-1}(t_{ij+1})} \langle \widetilde{q}_{ij}^{(k)}, \hat{\mu}_q^{(k)}(t) \rangle \, \mathrm{d}t \, . \tag{3.19}$$

An alternative approach is discussed in Appendix A.2. There, instead of treating the SRV curves as piece-wise-constant, they are smoothed in the mean basis $\widetilde{q}^{(k)} \approx b(s)^\top \hat{\theta}_i^{(k)}$, where the coefficient vector $\hat{\theta}_i^{(k)}$ is estimated in way that penalises deviations from the mean covariance-structure. The scalar product is then given by $\langle \widetilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle \approx (\hat{\theta}_i^{(k)})^H G \hat{\theta}^{(k)}$. While this method succeeds in estimating very natural looking smooth SRV curves (see **??**), their integral is not guaranteed to go through the observed points

on data curve level. Because of these and other theoretical concerns this approach did not end up getting used in this thesis.

4. Verification and Application using Simulated and Empirical Datasets

In this chapter the proposed methods will be applied and verified. Section 4.1 offers a comparison between the elastic full Procrustes mean, the elastic mean and the full Procrustes mean over two datasets, with the goal of illustrating some properties of the elastic full Procrustes mean. Section 4.2 briefly discusses the effect of the penalty parameter on the estimation. Section 4.3 mentions pitfalls when comparing the mean to the elastic full Procrustes fits and discusses outliers. Finally, Section 4.4 applies the proposed methods to an empirical dataset of tongue contours, which was kindly provided by Prof. Dr. Marianne Pouplier.

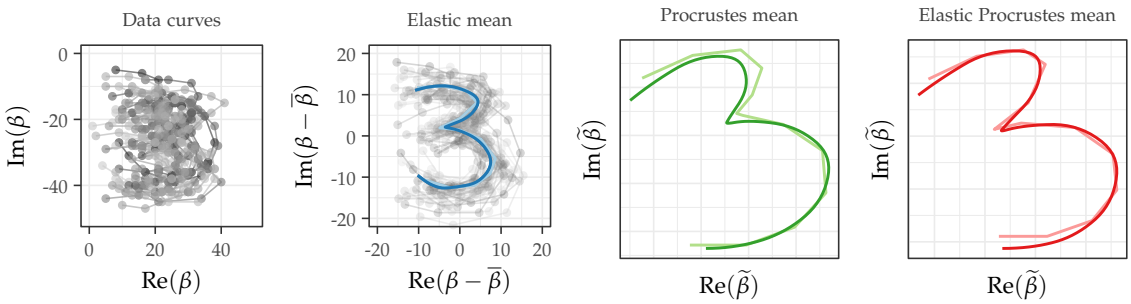
4.1. Comparison to the Elastic and the Full Procrustes Mean

In this section we will estimate and compare the elastic mean, the full Procrustes mean and the elastic full Procrustes mean for two datasets. The first dataset is the `digits3.dat` dataset provided in the `shapes` package (Dryden 2019) and originally collected by Anderson 1997, which has already been used throughout this thesis for illustrative purposes. It consist of 30 handwritten digits '3', each of which was sampled in a regular fashion at 13 points along the digit, leading to sparse but somewhat regular observations. The second dataset consists of ten simulated spirals, each of which is a sample of the curve $\beta(t) = t\cos(13t) + \mathrm{i} \cdot t\sin(13t)$, evaluated $m_i \in [10, 15]$ times over a noisy grid, with additional noise applied to the output, leading to sparse and irregular observations. The code simulating these spirals was adapted from the `compute_elastic_mean` function's documentation in the `elasdics` package (Steyer 2021). Because the spirals "speed up" w.r.t. t towards the end ($t = 1$), they start out quite densely sampled but become increasingly sparser, making the mean estimation, especially close to $t = 1$, a challenge. The datasets are considered in two settings: In the first, each dataset is considered as is, which means that all curves are centred

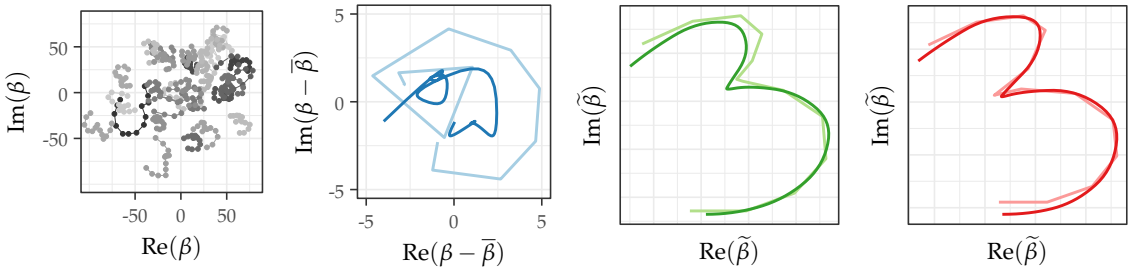
and similarly aligned. In the second, for each curve β_i a random Euclidean similarity transform with translation $\xi_i \sim \mathcal{U}([\xi_{\min}, \xi_{\max}])$, rotation $\theta_i \sim \mathcal{U}([0, 2\pi))$ and scaling $\lambda_i \sim \mathcal{U}([0.5, 1.5])$ is drawn, where $\xi_{\min}, \xi_{\max}$ are set respectively to ± 60 and ± 2 for the digits and spirals. The curves are then transformed by $\lambda_i e^{\mathrm{i}\theta_i}\beta_i + \xi_i$.

The four sets of data curves and means, are show in Fig. 4.1. Here, the elastic mean (blue) is estimated using `compute_elastic_mean` from the `elasdics` package. The full Procrustes (green) and the elastic full Procrustes mean (red) are estimated using the methods proposed in Chapters 2 and 3, where for the full Procrustes mean the estimation is stopped before the warping alignment step during the first iteration. Note that the full Procrustes mean calculated in this way is not exactly a minimizer of the sum of squared full Procrustes distance defined in Definition 2.2, but instead is a minimizer of the sum of squared full Procrustes distances on SRV level. When comparing the different mean types in Fig. 4.1 we can see that, unlike the elastic mean, both Procrustes means are invariant with respect to all Euclidean similarity transforms, as the estimated mean is the same for the transformed and original datasets. However for this very same reason, both Procrustes means hold no information about the scale or rotation of the original curves, as they are of unit-length and have a rotation dependent on the eigendecomposition of the covariance surface. The elastic mean is invariant only with respect to re-parameterisation and translation, so that its scale and rotation match the original data curves. It can therefore be meaningfully plotted together with the original curves, when they are centred.

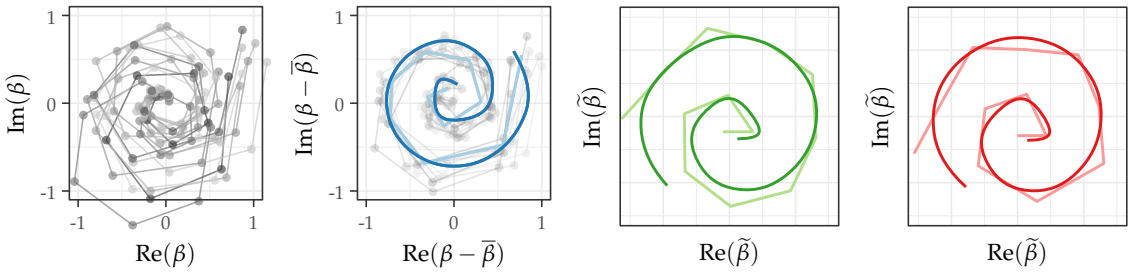
Apart from illustrating these properties, Fig. 4.1 provides two important validation checks for the estimation procedure proposed in Chapters 2 and 3. Firstly, as stated, the estimated elastic full Procrustes mean is invariant to all Euclidean similarity transforms, as the mean shapes do not change with transformations of the input curves. Secondly, when considering untransformed curves, the estimated elastic full Procrustes mean shapes are very comparable to elastic mean shapes estimated with the method proposed in Steyer, Stöcker, and Greven 2021. This is especially notable when comparing the means in Fig. 4.1 (a) where the prominent "notch" in the centre of the mean shape is similarly pronounced for the elastic and the elastic full Procrustes means, but not for the (non-elastic) full Procrustes mean. Taken together this shows that the proposed mean estimation method provides elastic mean estimates in the



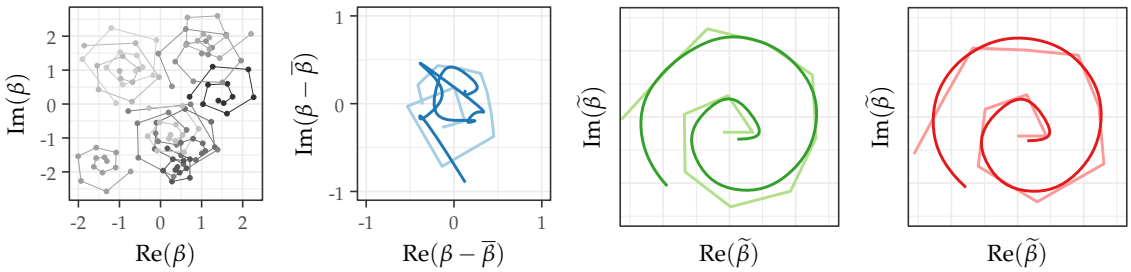
(a) Means for `digits3.dat`.



(b) Means for `digits3.dat` with random Euclidean similarity transform applied.



(c) Means for simulated sparse spirals.



(d) Means for simulated sparse spirals with random Euclidean similarity transform applied.

Figure 4.1.: Comparison of three mean types: Elastic mean (blue), full Procrustes mean (green) and elastic full Procrustes mean (red), estimated over four sets of data curves (grey). Each mean is estimated as polygonal (light, 16 knots) and smooth (dark, 13 knots), where in the estimation of the two Procrustes means a 2nd order penalty was applied. Data: See Section 4.1

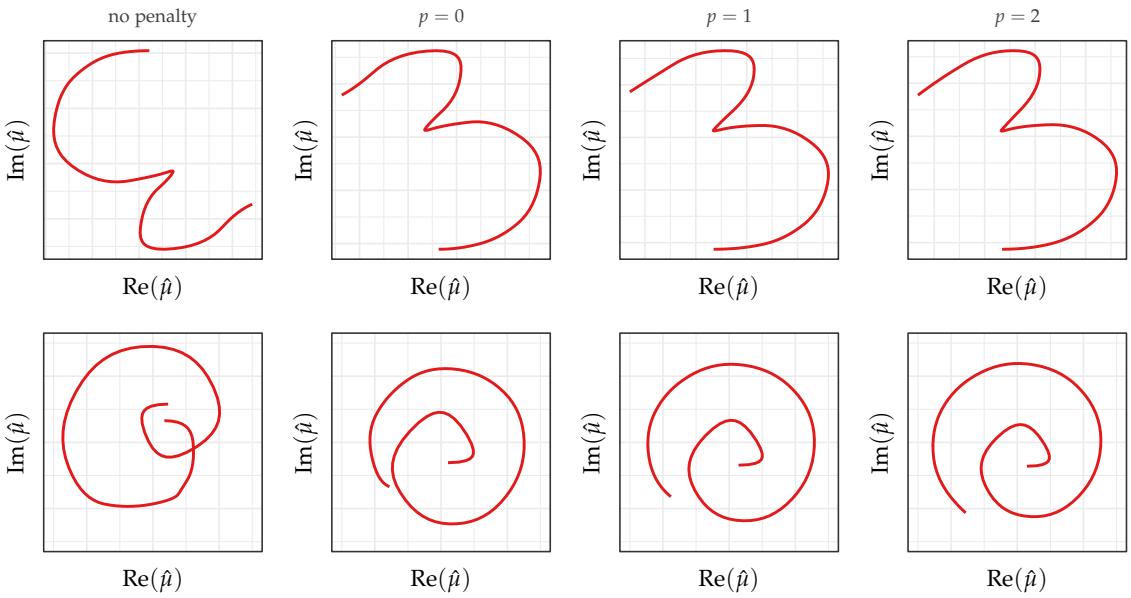


Figure 4.2.: Elastic full Procrustes mean under different penalties. Estimated using no penalty (left) and order 0/1/2 penalties (centre-left/centre-right/right) respectively, as well as 13 equidistant knots and linear B-splines on SRV level. Data: See Figs. 4.1b and 4.1d

setting of sparse and irregular curves, which are invariant to all shape-preserving transformations.

4.2. Effect of the Penalty Parameter on Estimated Means

The elastic full Procrustes mean is given by the leading eigenfunction of the complex covariance surface, which was estimated using tensor product P-splines. As a consequence, the order of the roughness penalty applied in the estimation of the covariance surface directly influences the shape of the estimated mean function. This can be seen in Fig. 4.2, where a smooth mean was estimated for different penalties. Here, in particular, the spiral mean shapes help to illustrate the penalty effect. As mentioned in Section 4.1, the spirals are evaluated in a way that makes them very sparse towards the end, leading to unstable mean estimates in that region. By penalising p-th order differences between neighbouring coefficients, the penalty helps to stabilise the covariance estimation (and thereby the mean) in regions where observations are sparse, where the shape of the function is consequently dominated by the penalty.

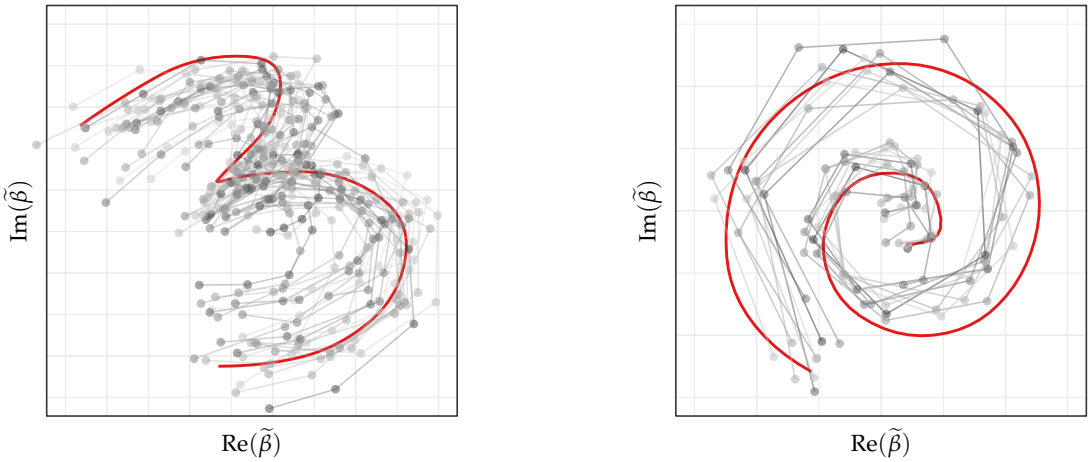
The mean function on the far left in Fig. 4.2 was estimated using no penalty. We can see that the spiral shape is estimated well in the beginning (central part), where

observations are dense, but becomes increasingly unstable and wriggly towards the end (outer part). The mean function on the centre-left was estimated using a zero order penalty. A zero order penalty on a B-spline basis can be interpreted as a ridge-penalty on the basis coefficients, i.e. the basis coefficients (and therefore the estimated function) get shrunk towards zero in areas where observations are sparse. It is important to note that the penalised covariance estimation is performed on SRV level and not on data curve level, which means the ridge penalty shrinks the estimated SRV mean towards zero. However, this does not imply that the estimated mean on data curve level also gets shrunk towards zero. In fact, a ridge penalty on SRV level only shrinks the *absolute velocity* of the mean on data curve level towards zero, but should not directly influence the *direction* of the mean curve. This can be seen, when comparing the zero order penalty spiral mean (centre-left) to the higher order penalty spiral means (right) and noting that, as a consequence of the decreased absolute velocity, it is relatively "shorter" towards the end when compared to the beginning of the spiral. The higher order penalties may be interpreted as smoothing the SRV mean function towards a polynomial of degree $p-1$ in areas where the penalty dominates, which means a constant function for the first order penalty and a linear function for the second order penalty (see e.g. Fahrmeier et al. 2013, p. 435). Looking at the means for $p = 1, 2$ in Fig. 4.2, these differences are already hard to spot. The order two mean (right) is slightly longer towards the end, indicating a speed up, which is caused by the penalisation towards a global linearly (increasing) velocity, compared to a more conservative penalisation towards a global constant velocity.

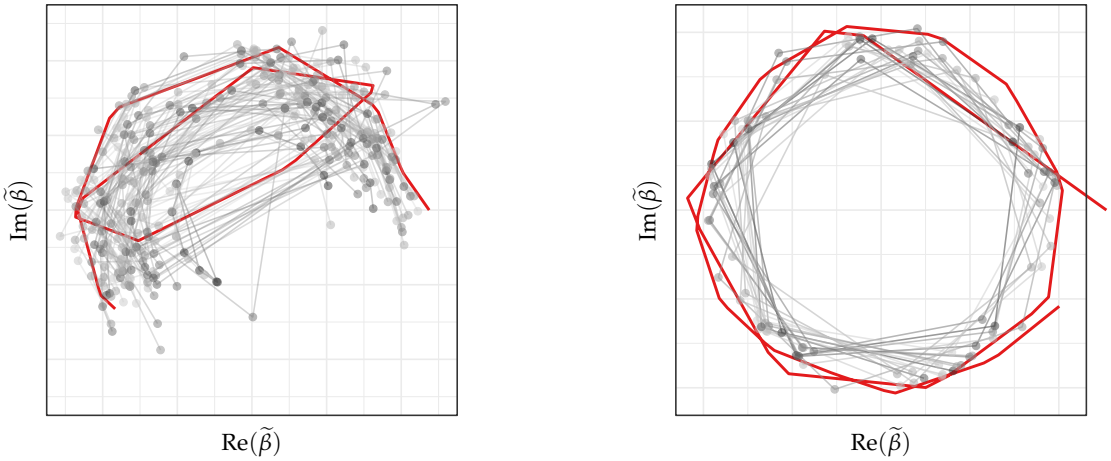
Kann man das so einfach übertrage?

4.3. Elastic Full Procrustes Fits and Outliers

Although the elastic full Procrustes mean does not share the rotation, scale and translation of the input curves, it is still possible to visually compare it to them, by plotting the mean together with the elastic full Procrustes fits. In this thesis, the elastic full Procrustes fits are calculated on SRV level as $\widetilde{q}^{\mathrm{EP}} = (\omega^{\mathrm{opt}} \cdot \widetilde{q} \circ \gamma^{\mathrm{opt}}) \sqrt{\dot{\gamma}^{\mathrm{opt}}}$, where the optimal rotation and scaling alignment $\omega^{\mathrm{opt}} = \lambda^{\mathrm{opt}} e^{\mathrm{i}\theta^{\mathrm{opt}}}$ and optimal warping alignment γ^{opt} of each curve to the mean are taken from the last iteration of the mean estimation step. In Fig. 4.3a the elastic full Procrustes mean and fits are plotted for the digits '3' and simulated spiral datasets discussed in Section 4.1, with random Euclidean



(a) Centred elastic full Procrustes fits (grey) and smooth means (red).

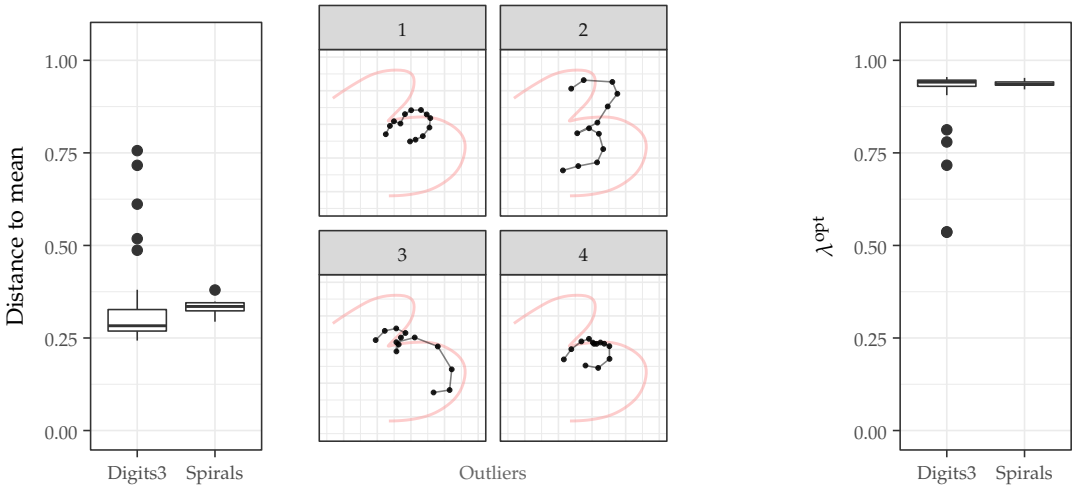


(b) Elastic full Procrustes fits (grey) and piece-wise-linear means (red) on SRV level.

Figure 4.3.: Elastic full Procrustes means and fits for digits (left) and spirals (right). Parameters: Second order penalty and linear B-splines, with 13 (left) and 19 (right) equidistant knots. Data: See Figs. 4.1b and 4.1d

similarity transforms applied to each curve. This alignment works very well for the simulated spirals and is more fuzzy for the digits, due to their greater variability in shape.

In general, one has to be careful when interpreting these plots. For example, it can be seen that the estimated mean does not necessarily follow the "centre of mass" of the aligned curves. One reason for this is that the mean is calculated on SRV level, meaning we may set an arbitrary translation for the mean on data curve level. Here, the Procrustes fits and the mean function are centred, but a natural alternative might be to have all curves start at the origin, with both choices leading to very different, but equally valid, visual representations. Another reason is the scaling alignment, especially with respect to outliers. This is more noticeable in the `digits3.dat`, because



(a) Distribution of elastic full Procrustes distances to the estimated mean (red) with four largest outliers (grey).

(b) Distribution of the optimal scaling alignment parameter of each elastic Procrustes fit.

Figure 4.4.: Elastic full Procrustes distances, outliers and estimated optimal scaling. Parameters: See Fig. 4.3. Data: See Figs. 4.1b and 4.1d

of the greater variation in '3' shapes, where some curves cannot be properly aligned. Four outlier shapes are shown in Fig. 4.4a, which were classified using their distance to the mean (left). One prominent feature of the full Procrustes mean is its relative robustness to outliers, when compared to the normal and partial Procrustes means. This comes from the additional scaling alignment, which allows for shrinkage of any outlier curve towards zero, as can be seen for curves 1 and 4 in Fig. 4.4a. This shrinkage, indicated by a value of $\lambda^{\text{opt}} < 1$ in Fig. 4.4b, does occur for all curves, not only occur for outlier curves, which should not be surprising, as the alignment minimizes the distance to the mean, which always makes it advantageous to shrink the aligned curves slightly rather than scaling them up. Because the warping alignment is performed between the elastic full Procrustes fit and the mean curve, it might be important to better understand how it is affected by bad rotation and scaling alignment. While a bad scaling alignment does not seem to be immediately problematic, as the warping alignment only takes into account the *relative* distances over t between the aligned and the mean curve, a bad rotation alignment will most likely always lead to a bad warping alignment. However, because of the discussed shrinkage, these outliers are only problematic when considering the alignment of single curves, as their influence on the estimated means shape is always limited by a shrinkage to zero.

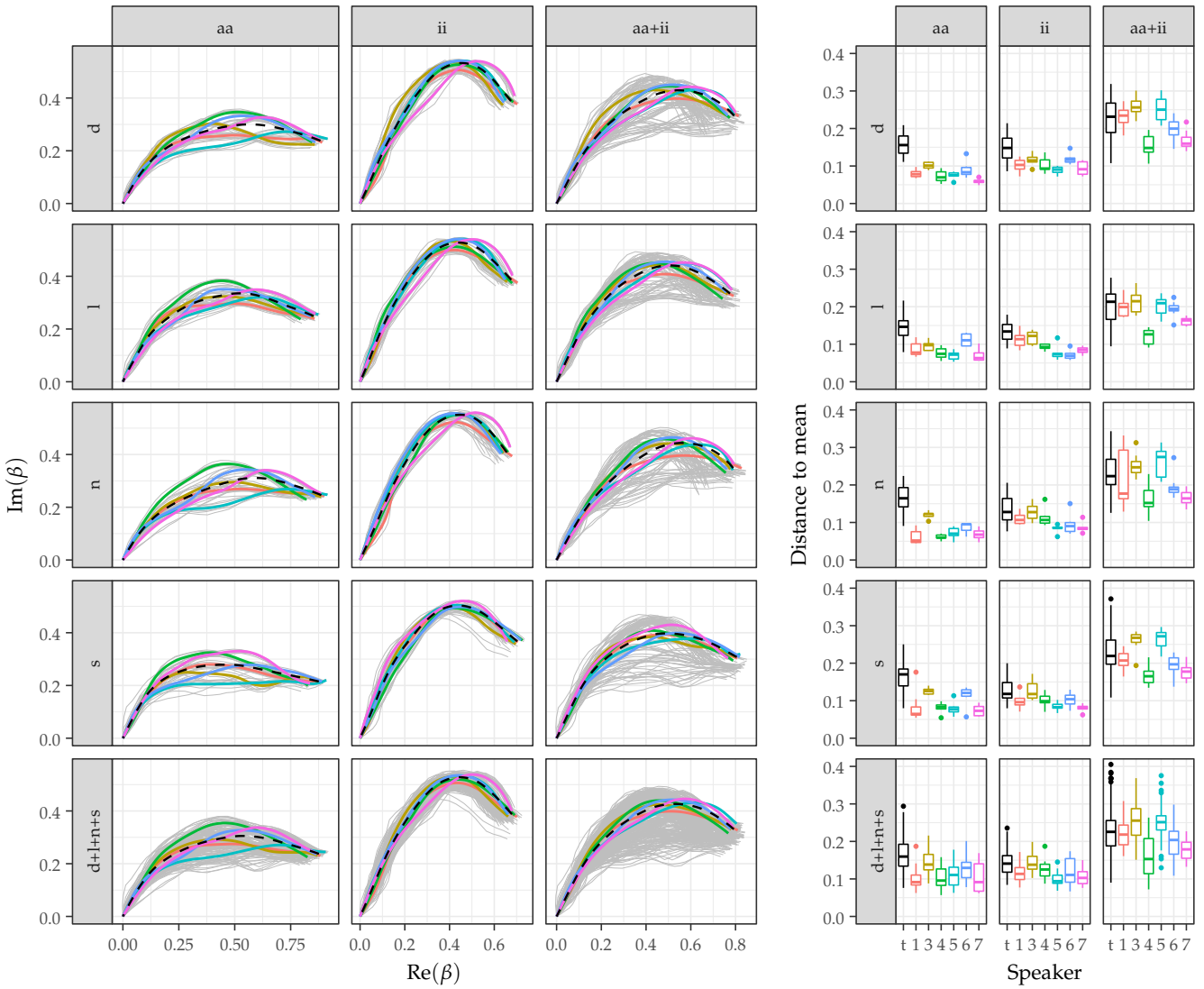
4.4. Analysis of Variations in Tongue Shapes

Mean and distance estimation for planar shapes extracted from imagery is a typical use-case for the proposed methods. In this section a dataset consisting of tongue contours obtained from ultrasound recordings and their phonetic context is analysed. The data was kindly provided by Prof. Dr. Marianne Pouplier and was gathered in an experimental setting from 6 native German speakers, each of whom was recorded speaking the same set of ficticious words over multiple repetitions. Each word is a combination of two flanking vowels such as 'aa' or 'ii' and one consonant such as 'd' or 'l', leading to words such as 'pada', 'pidi', 'pala' or 'pili'. The tongue contours follow the centre of the tongue and are extracted from the ultrasound recordings at times corresponding to the centre of consonant articulation, which is estimated from the acoustic signal. Furthermore, care was taken to observe the tongues over their full length, from their tips to the Hoyid bone (German: Zungenbein), which means we can naturally treat them as planar curves $\beta : [0,1] \to \mathbb{C}$ (see Le and Szydlo 2015). While there is earlier work on functional data approaches to tongue shape analysis (Cederbaum, Pouplier, et al. 2016; Davidson 2006; Pouplier et al. 2014), these approaches do not account for shape invariance in their estimation. Procrustes analysis can account for anatomical differences between speakers, such as their differing sizes, or measurement inconsistencies between iterations, such as differences in orientation of the ultrasound device. At the same time, the tongue is a flexible muscle capable of stretching, bending and compressing itself, providing some indication that an elastic analysis is appropriate. Therefore, the tongue shapes will be analysed using the elastic full Procrustes mean and distance.

The dataset we analyse consists of tongue shapes during articulation of the consonants 'd', 'l', 'n' and 's', for each of which there are multiple observations over two vowel contexts ('aa' and 'ii') and six speakers. The shapes will vary between consonants, vowel context and speaker and also, per speaker, over multiple repetitions. However, it is not immediately clear which factors are relatively more important for tongue shape than others. We are therefore interested in questions such as: Does tongue shape vary more strongly between vowel context than between consonants? Do individual speakers have a consistent tongue shape per word between their repetitions? Does each speakers individual way of speaking influence tongue shape more

strongly than the actual consonant spoken? To answer these questions, the dataset is grouped hierarchically, by consonants and vowel contexts and means are estimated for each grouping structure, one total mean over all observations belonging to the specific combination of consonants and vowels, and 6 speaker-wise means. Then, to get a measure for the variability of tongue shapes inside a specific group, elastic full Procrustes distances to the mean are calculated and their distribution compared.

The results can be seen in Fig. 4.5. The box plots in Fig. 4.5b show the distributions of the elastic full Procrustes fits' distances to their mean in Fig. 4.5a. We can note that although the tongue shapes were measured during consonant articulation, the vowel context has a greater influence on their shape, as grouping by vowel (comparing the right to the other columns) consistently decreases variability more so than grouping by consonant (comparing the bottom to the other rows). When comparing the distance distributions for total means (black) this last part is especially noticeable, as the mean distance hardly changes when grouping by consonant. It seems that without accounting for the individual speaker, the consonant only explains very little in terms of variability of the tongue shapes, as when performing the same comparison over the per-speaker means, we can see that grouping by consonants reduces variability in the dataset slightly. This indicates that there is no strong effect of the spoken consonant on tongue shape over all speakers, but rather that this effect depends on each speaker individually. Finally, the smallest variability is measured over the per-speaker means for repetitions of the same word (central eight plots), indicating that tongue shape is very consistent on a per-speaker basis.



(a) Total (black, dashed) and per-speaker means (coloured) by vowels and consonants, with per-speaker elastic full Procrustes fits (grey). Per-speaker means and -fits are rotationally aligned to the respective total means.

(b) Distribution of elastic full Procrustes distances to respective mean (see Fig. 4.5a) total (black) and by speaker (coloured), where 't' indicates distances to the total mean.

Figure 4.5.: Right column and bottom row correspond to means and distances over all vowels or all consonants respectively, with the bottom right means corresponding to global and global per-speaker means and distances. Parameters: Estimated using 13 equidistant knots, linear B-splines on SRV level and a 2nd order penalty. Data: Prof. Dr. M. Pouplier.

5. Summary and Outlook

In this thesis a method for shape mean estimation of open planar curves was proposed. The estimation method is especially suitable for sparse and irregular observations as it makes use of covariance smoothing methods.

The method and its implementation were validated over toy datasets, where it was shown that the estimated shape means are elastic and invariant with respect to shape preserving transformations of the input curves. The penalty used in the estimation of the Hermitian covariance surface provides stable estimates in regions where observations are sparse, where good results were achieved for first and second order penalties. The estimated mean may be plotted together with the elastic full Procrustes fits, however, care has to be taken when interpreting these plots, as the estimated mean curve is not a simple functional mean of the aligned curves, due to their shrinkage. However, for this same reason the elastic full Procrustes mean is quite robust to outliers, as they tend to get shrunk towards zero, restricting their influence on the mean estimation. Finally, the method was used to analyse variability in an empirical dataset of tongue contours during consonant articualtion. Here it was found that the vowel context has the largest influence on the tongues shape, while the effect of the spoken consonant seems to vary between speakers.

The proposed methods might be improved by including some form of smoothing into the estimation of the elastic full Procrustes fits, such as a variation of the curve smoothing proposed in Appendix A.2, or by directly solving the warping optimisation over the analytical solution to the rotation and scaling alignment. Two important extension that seem quite challanging would be elastic full Procrustes mean estimation for closed curves and for higher-dimensional curves.

Bibliography

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

Cederbaum, J., M. Pouplier, et al. (2016). "Functional linear mixed models for irregularly or sparsely sampled data". In: *Statistical Modelling* 16.1, pp. 67–88.

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`.

Davidson, L. (2006). "Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance". In: *The Journal of the Acoustical Society of America* 120.1, pp. 407–415.

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.

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

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.

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

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

Le, L. and M. Szydlo (2015). *Analyse von Zungenkonturen zur Untersuchung der Konsonantenanpassung an den Vokalkontext*. Projektbericht. Ludwig-Maximilians-Universität München.

Picinbono, B. (1996). "Second-Order Complex Random Vectors and Normal Distributions". In: *IEEE Transactions on Signal Processing* 44, pp. 2637–2640.

Pouplier, M. et al. (2014). "Perceptual and articulatory factors in German fricative assimilation". In: *Proceedings of the 10th International Seminar on Speech Production (ISSP)*. Ed. by S. Fuchs et al., pp. 332–335.

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/.

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.

Redd, A. (2012). "A comment on the orthogonalization of B-spline basis functions and their derivatives". In: *Stat Comput* 22, pp. 251–257.

– (2015). *orthogonalsplinebasis: Orthogonal B-Spline Basis Functions*. R package version 0.1.6. URL: https://CRAN.R-project.org/package=orthogonalsplinebasis.

Reiss, P. T. and Meng Xu (2020). "Tensor product splines and functional principal components". In: *Journal of Statistical Planning and Inference* 208, pp. 1–12.

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].

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.

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

Yao, F., H.-G. Müller, and J.-L. Wang (2005). "Functional Data Analysis for Sparse Longitudinal Data". In: *Journal of the American Statistical Association* 100.470, pp. 577–590.

A. Appendix

A.1. Additional Proofs and Derivations

A.1.1. Derivation of Lemma 2.1

$$\begin{aligned} d_{FP}([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}})^2 &= \min_{\omega \in \mathbb{C}} \|\widetilde{\beta}_1 - \omega \widetilde{\beta}_2\|^2 \\ &= \min_{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi)} \|\widetilde{\beta}_1 - \lambda e^{\mathrm{i}\theta} \widetilde{\beta}_2\|^2 \\ &= \min_{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi)} \langle \widetilde{\beta}_1 - \lambda e^{\mathrm{i}\theta} \widetilde{\beta}_2, \widetilde{\beta}_1 - \lambda e^{\mathrm{i}\theta} \widetilde{\beta}_2 \rangle \\ &= \min_{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi)} \|\widetilde{\beta}_1\|^2 + \lambda^2 \|\widetilde{\beta}_2\|^2 - \lambda (e^{\mathrm{i}\theta} \langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle + e^{-\mathrm{i}\theta} \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle \end{aligned}$$

Define $\langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle = \kappa e^{\mathrm{i}\phi} \in \mathbb{C}$ with $\kappa \in \mathbb{R}^+$, $\phi \in [0, 2\pi)$ and use $\|\widetilde{\beta}_1\| = \|\widetilde{\beta}_2\| = 1$.

$$\begin{aligned} d_{FP}([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}})^2 &= \min_{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi)} 1 + \lambda^2 - \lambda (e^{\mathrm{i}\theta} \kappa e^{\mathrm{i}\phi} + e^{-\mathrm{i}\theta} \kappa e^{-\mathrm{i}\phi}) \\ &= \min_{\lambda \in \mathbb{R}^+, \theta \in [0,2\pi)} 1 + \lambda^2 - \lambda \kappa \left(e^{\mathrm{i}(\theta+\phi)} + e^{-\mathrm{i}(\theta+\phi)} \right) \\ &= \min_{\lambda \in \mathbb{R}^+} 1 + \lambda^2 - \max_{\theta \in [0,2\pi)} 2\lambda\kappa \cos(\theta + \phi) \\ &\overset{\theta^{\text{opt}} = -\phi}{=} \min_{\lambda \in \mathbb{R}^+} 1 + \lambda^2 - 2\lambda\kappa \end{aligned}$$

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

$$d_{FP}([\beta_1]_{\text{Eucl}}, [\beta_2]_{\text{Eucl}})^2 = \left(1 + \kappa^2 - 2\kappa^2\right) = \left(1 - \kappa^2\right)$$

Lemma 2.1 i.) follows by considering $\kappa^2 = |\langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle|^2 = \langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle$. Then

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

Lemma 2.1 ii.) follows by $\omega^{\text{opt}} = \lambda^{\text{opt}} e^{\mathrm{i}\theta^{\text{opt}}} = \kappa e^{-\mathrm{i}\phi} = \overline{\langle \widetilde{\beta}_1, \widetilde{\beta}_2 \rangle} = \langle \widetilde{\beta}_2, \widetilde{\beta}_1 \rangle$.

A.2. Shape-Smoothing Using the Estimated Covariance-Surface

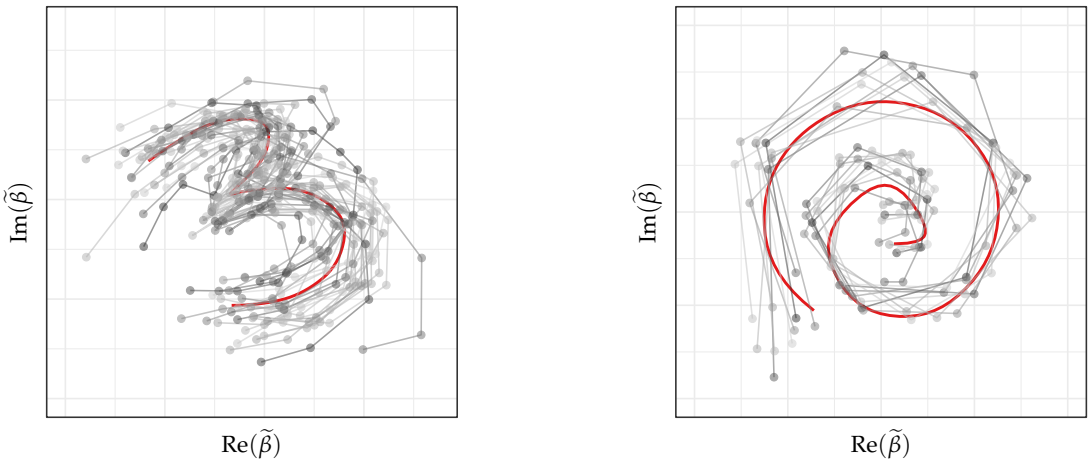
Instead of treating $\widetilde{q}_i$ as piecewise constant, we might want to smooth each curve in the mean basis $b(t)$, so that $\widetilde{q}_i \approx b(s)^\top \hat{\theta}_i$. The optimal rotation and scaling alignment is then simply given by the scalar product $\omega_i = \langle \widetilde{q}_i, \hat{\mu}_q \rangle \approx (\hat{\theta}_i)^H G \hat{\theta}$. To make use of information from the other observations, we can estimate the coefficient vector θ_i in way that penalises deviations from the estimated mean covariance-structure $\hat{\Xi}$. This may be achieved by using a ridge penalty, where we assume $\widetilde{q}_i \sim \mathcal{N}_{\mathbb{C}^k}(B\theta_i, \sigma^2 I_{m_i})$, with B the $m_i \times k$ design matrix and q_i now the stacked vector of observations, and place a complex normal prior $\theta_q \sim \mathcal{N}_{\mathbb{C}^k}(0, \tau^2 \hat{\Xi})$ on the coeficient vector. See e.g. (Picinbono 1996) for information on the complex normal distribution. Here, $\lambda = \frac{\sigma^2}{\tau^2}$ is a hyperparameter controlling the strength of the penalisation.

$$\underset{\theta_i \in \mathbb{C}^k}{\operatorname{argmax}} \ (\widetilde{q}_i - B\theta_i)^H (\widetilde{q}_i - B\theta_i) - \lambda {\theta_i}^H \hat{\Xi}^H \theta_i \tag{A.1}$$

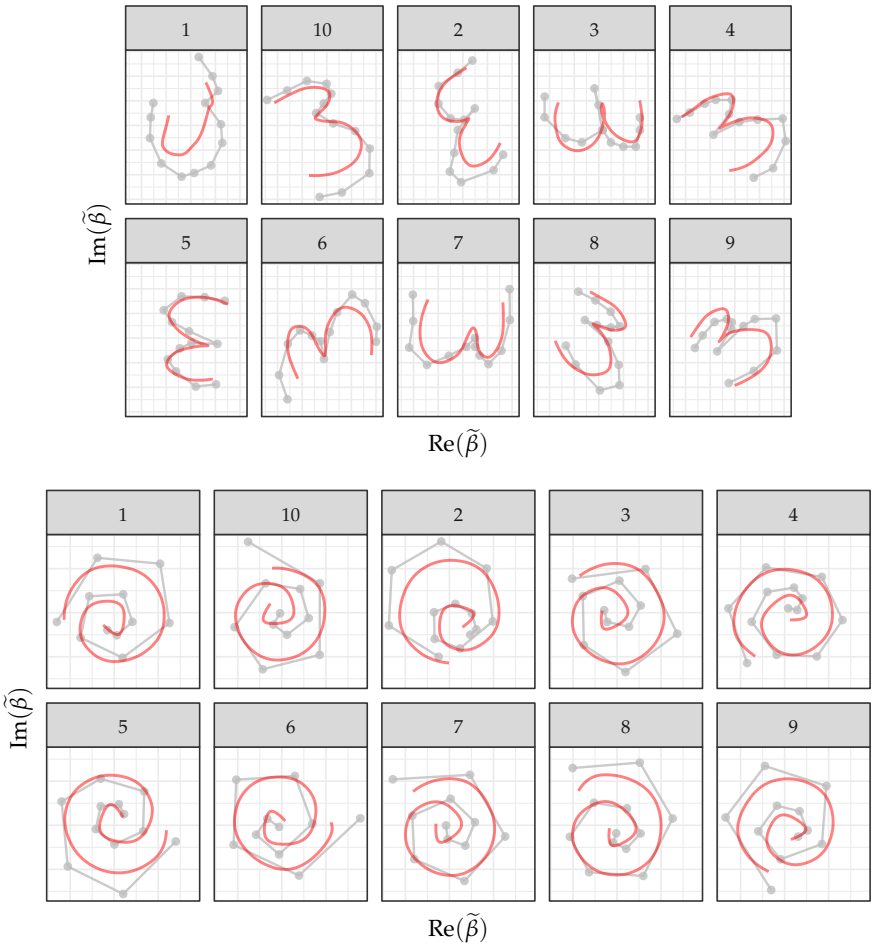
The penalised estimate is then given by the solution to the above optimisation problem, and therefore after some derivation

$$\hat{\theta}_i = (B^T B + \lambda \hat{\Xi}^{-1})^{-1} B^T \widetilde{q}_i. \tag{A.2}$$

The results of this smoothing procedure can be seen in Fig. A.1 for $\lambda = 0.4$. The method succeeds in estimating very natural looking smooth SRV curves, however, the smoothed curves tend to get shrinked towards 0, because of the ridge penalty. While this shrinking may be prevented by normalizing the smoothed SRV curve (leading to unit-length curves on original curve leve), a normal prior is probably simply not the appropriate prior in this setting. For the individual coefficients θ_{ik} it is true that $\mathbb{E}[\theta_{ik}] = 0$ (due to rotational symmetry), however, in general $\mathbb{E}[|\theta_{ik}|] \neq 0$. A distribution more like a Normal distribution lying on a ring around 0 with radius $r = |\theta|$, or a proper shape distribution such as the complex Bingham distribution (see e.g. Dryden and Mardia 2016, Chap. 10) might be more appropriate. Using those, the estimates would not be shrinked towards zero and instead would be smoothed toworads a curve of size equal to the estimated mean. The smoothing could be further improved by providing true interpolation, so that the interpolated curve goes through the



(a) Centred elastic full Procrustes fits (grey) and smooth means (red).



(b) Smooth (red) and original (grey) observations.

Figure A.1.: Elastic full Procrustes means and fits for digits with smoothed observations ($\lambda = 0.4$). Parameters: Second order penalty and linear B-splines, with 13 equidistant knots. Data: See Figs. 4.1b and 4.1d

observed points on original curve level. As the smoothed original curve is calculated by integrating the smoothed SRV curves, this means one has to implement a nonlinear constraint ensuring that the curve integrates to the difference between the observed points on the interval between them.

A.3. Implicit Rotation and Scaling Alignment

Instead of solving for the optimal warping and rotation and scaling alignment iteratively, we might optimse the warping alignment directly over the analytical solution to the rotation and scaling alignment (see Lemma 2.1 i.). Then we have an alternative optimisation problem

$$\begin{aligned}
\gamma^{\text{opt}} &= \underset{\gamma\in\Gamma}{\text{argmin}} \sqrt{1 - \langle \widetilde{q}_1, (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\rangle \langle (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}, \widetilde{q}_1\rangle} \\
&\quad \underset{\gamma\in\Gamma}{\text{argmin}} \sqrt{1 - \langle (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}, \widetilde{q}_1\rangle \langle \widetilde{q}_1, (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\rangle} \\
&= \underset{\gamma\in\Gamma}{\text{argmax}} \langle (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}, \widetilde{q}_1\rangle \langle \widetilde{q}_1, (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\rangle \\
&= \underset{\gamma\in\Gamma}{\text{argmax}} \langle C_1\,(\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}, (\widetilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\rangle
\end{aligned}$$

with C_1 the covariance operator belonging to the covariance function $C_1(s,t) = \widetilde{q}_1(s)\overline{\widetilde{q}_1(t)}$ of $\widetilde{q}_1$ (compare also the derivation of the elastic full Procrustes mean). In the mean estimation step, C_1 can be replaced with the estimated covariance surface, making the rotation and scaling alignment unneccessary and possibly providing a computational advantage. However, this approach does not integrate with existing approaches for warping alignment such as Steyer, Stöcker, and Greven 2021. Further work might focus on solving the above optimisation problem for sparse and irregular planar curves.

Statutory Declaration

I declare that I have not previously submitted the present work for other examinations and that I wrote this work independently. All sources that I have reproduced in either an unaltered or modified form, have been acknowledged as such. I understand that violations of these principles will result in proceedings regarding deception or attempted deception.

Manuel Pfeuffer
Berlin, 14th December, 2021