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]

3rd November, 2021, Berlin

Advisors: Lisa Maike Steyer, Almond Stöcker

1st Examiner: Prof. Dr. Sonja Greven

2nd Examiner: Prof. Dr. Nadja Klein

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

Contents

1. Introduction

Statistical Shape Analysis (see e.g. Dryden and Mardia 2016) is a branch of statistics concerned with modelling the geometry of objects. Examples might be outlines of bones and organs, handwritten digits, or the folds of a protein. To capture an object's geometrical information, a common approach is the use of *landmarks*, characteristic points on an object, that "match between and within populations"(Dryden and Mardia 2016, p. 3). However, in recent years an alternative approach has gained in popularity, where objects are represented using curves. This has the advantage of a more flexible representation of an object's geometry, as the analysis is not restricted to a fixed set of discrete points. The curves are usually themselves represented by functions $\beta : [0,1] \to \mathbb{R}^k$, which, for example for $k = 2$, might describe the outlines of an object in an image. As each object corresponds to one observation, this opens up a connection to the branch of statistics concerned with observations that are whole functions: Functional Data Analysis (see e.g. Ramsay and Silverman 2005).

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

Definition 1.1 (Shape)**.** "[A]ll the geometrical information that remains when location, scale and rotational effects are removed from an object"(Dryden and Mardia 2016, p. 1).

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

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

The *Full Procrustes Mean* is a different shape mean concept, which is widely used when working with landmark data, and which has particularly nice properties in two dimensions, when identifying $\mathbb{R}^2$ with $\mathbb{C}$ (see Dryden and Mardia 2016, Chap. 8). When working with planar curves, its calculation can be shown to be related to an eigenfunction problem of the complex covariance surface of the observed curves. This offers an advantage when working in the challenging setting of sparsely and irregularly sampled curves, as appropriate smoothing techniques for estimation of covariance surfaces in this setting are already known. Here in particular, Cederbaum, Scheipl, and Greven 2018 offers a method for efficient covariance smoothing in the sparse setting.

The aim of this thesis is to extend existing methods for elastic mean estimation of sparse and irregularly sampled curves, as proposed by Steyer, A. Stöcker, and Greven 2021 and implemented in the `R` package `elasdics` (Steyer 2021), to also include invariance with respect to rotation and scaling. The later will be achieved by generalizing the concept of the *Full Procrustes Mean* from landmark to functional data and by iteratively applying full Procrustes mean estimation, rotation-alignment and parametrisation-alignment, leading to the estimation of *Elastic Full Procrustes Means*. To make use of the nice properties of the Procrustes mean in two dimensions, analysis will be restricted to the case of planar curves. Here, smoothing techniques for sparse estimation of the complex covariance surfaces, as available in the `R` package `sparseFLMM` (Cederbaum, Volkmann, and A. Stöcker 2021), will be used.

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

2. Elastic Full Procrustes Means for Planar Curves

As a starting point, it is important to establish a notational and mathematical framework for the treatment of planar shapes. While the restriction to the 2D case might seem a major one, it still covers all shape data extracted from e.g. imagery and is therefore very applicable in practice. The outline of a 2D object may be naturally represented by a planar curve $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x(t), y(t))^T$, where $x(t)$ and $y(t)$ are the scalar-valued *coordinate functions*. Calculations in two dimensions, and in particular the derivation of the full Procrustes mean, are greatly simplified by using complex notation. Going forward, we will therefore identify $\mathbb{R}^2$ with $\mathbb{C}$ and always use complex notation when representing a planar curve:

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

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

2.1. Equivalence Classes and Shape Invariance

As mentioned in the introduction, shape is usually defined by its invariance under the transformations of scaling, translation, and rotation. When considering the shape of curves, we additionally have to take into account invariance with respect to re-parametrisation. This can be seen, by noting that the curves $\beta(t)$ and $\beta(\gamma(t))$, with some re-parametrisation or *warping function* $\gamma : [0,1] \to [0,1]$ monotonically increasing and differentiable, have the same image and therefore represent the same geometrical object. We can say that the actions of translation, scaling, rotation, and re-parametrisation are *equivalence relations* with respect to shape, as each action leaves the shape of the curve untouched and only changes the way it is represented. The shape of a curve can then be defined as the respective *equivalence class*, i.e. the set of all possible shape preserving transformations of the curve. As two equivalence classses are neccessarily either disjoint or identical, we can consider two curves as having the same shape, if they are elements of the same equivalence class (see Srivastava and Klassen 2016, p. 40).

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

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

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

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

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

[TODO: Wording and sentence structure not super clear in this paragraph.] In a next step, we can consider how these transformations act in concert and whether they *commute*, that is, whether the order of applying the transformations changes outcomes. Consider for example the actions of the rotation and scaling product group $\mathbb{R}^+ \times [0, 2\pi]$ given by $((\lambda, \theta), \beta) \xmapsto{\text{Scl+Rot}} \lambda e^{i\theta}\beta$. These clearly commute, because the order of applying rotation or scaling do not make a difference, as $\lambda(e^{i\theta}\beta) = e^{i\theta}(\lambda\beta)$. However, the joint actions of scaling and translation do not commute, as $\lambda(\beta + \xi) \neq \lambda\beta + \xi$, with the same holding for the joint action of rotation and translation. As the order of translating and rotating or scaling matters, one usually takes the translation to act on the already scaled and rotated curve. The joint action defined using this ordering is usually called an *Euclidean similarity transformation*.

Definition 2.1 (Euclidean similarity transformation (Dryden and Mardia 2016, p. 62))**.** We define an *Euclidean similarity transformation* of a curve $\beta : [0,1] \to \mathbb{C}$ as the joint action of scaling, rotation, and translation by

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

with $\xi \in \mathbb{C}$, $\lambda \in \mathbb{R}^+$, and $\theta \in [0, 2\pi]$.

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

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

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

The *shape space* $\mathcal{S}$ is then given by the corresponding quotient space

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

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

2.2. The Elastic Full Procrustes Distance for Planar Curves

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

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

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

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

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

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

paragraph!] (see Srivastava and Klassen 2016, p. 88).

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

Definition 2.3 (SRV function of a planar curve (Srivastava, Klassen, et al. 2011))**.** The *SRV function* (SRVF) of an absolutely continuous planar curve $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ is given by

$$q(t) = \begin{cases} \frac{\dot{\beta}(t)}{\sqrt{\|\dot{\beta}(t)\|}} & \dot{\beta}(t) \neq 0 \\ 0 & \dot{\beta}(t) = 0 \end{cases},$$

where $q \in \mathbb{L}^2([0,1],\mathbb{C})$ and $\dot{\beta}(t) = \frac{\partial \beta(t)}{\partial t}$.

[TODO: Write more in detail about this being bijective between $\mathcal{AC}$ and $\mathbb{L}^2$]!

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

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

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

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

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

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

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

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

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

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

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

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

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

or when equivalently optimizing over all possible transformations as

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

We can simplify the optimization, by considering that both rotation and warping act by isometries on the elastic metric, which means in practice it is enough to optimize over the rotation and warping of only one of the curves. For warping, we can reformulate the optimization as a problem over the relative parametrisation between the curves, as $\inf_{\gamma_1,\gamma_2\in\Gamma}||(q_1\circ\gamma_1)\sqrt{\dot{\gamma_1}} - (q_2\circ\gamma_2)\sqrt{\dot{\gamma_2}}|| = \inf_{\gamma_1,\gamma_2\in\Gamma}||q_1 - (q_2 \circ (\gamma_2 \circ$

$\gamma_1^{-1}))\sqrt{\dot{(\gamma_2 \circ \gamma_1^{-1})}}\| = \inf_{\gamma \in \Gamma} \|q_1 - (q_2 \circ \gamma)\sqrt{\dot{\gamma}}\|$. And smilarly for rotation, we can optimize over the relative rotation between both curves, as $\inf_{\theta_1, \theta_2 \in [0,2\pi]} \|e^{i\theta_1} q_1 - e^{i\theta_2} q_2\| = \inf_{\theta_1, \theta_2 \in [0,2\pi]} \|q_1 - e^{i(\theta_2 - \theta_1)} q_2\|) = \inf_{\theta \in [0,2\pi]} \|q_1 - e^{i\theta} q_2\|$. Taken together these lead to

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

which, however, still has the problem of not being isometric with respect to scaling. **[TODO: Rewrite! A distance cannot be isometric.]**

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

Definition 2.4 (Elastic full Procrustes distance)**.** The *elastic full Procrustes distance* between the shapes $[\beta_1]$, $[\beta_2]$ of two continuously differentiable planar curves β_1, $\beta_2 \in \mathcal{AC}([0,1], \mathbb{C})$ is given by

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

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

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

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

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

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

$$(\omega^*, \gamma^*) = \underset{\omega \in \mathbb{C}, \gamma \in \Gamma}{\operatorname{argmin}} \|z_1 - \omega(z_2 \circ \gamma)\sqrt{\dot{\gamma}}\| \tag{2.1}$$

The usual strategy is to optimize over the sets of parameters individually and then to iterate through both solutions. **[TODO: Citation, write more here!]** Let us first consider the optimization with respect to rotation and scaling. For fixed $\gamma \in \Gamma$ with $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$ we can write Eq. 2.1 as

$$\omega^* = \underset{\omega \in \mathbb{C}}{\operatorname{argmin}} \|z_1 - \omega \widetilde{z}_2\|, \tag{2.2}$$

which can be solved analytically.

Lemma 2.2. i.) *For a fixed $\gamma \in \Gamma$, the optimal scaling and rotation solving Eq. 2.2 is*

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

ii.) *The optimization problem in Definition 2.4 can then be reduced to*

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

| *Proof.* See A.1.1 in the appendix. □

Remark. For fixed $\gamma \in \Gamma$, we can use the first part of Lemma 2.2 to calculate the optimal rotation and scaling alignment of $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$ onto z_1. The alignment $\widetilde{z}_2^P = \langle \widetilde{z}_2, z_1 \rangle \widetilde{z}_2$ is called the *Procrustes fit* of $\widetilde{z}_2$ onto z_1. The second part of Lemma 2.2 will be useful for mean calculation in the next section.

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

$$\gamma^* = \underset{\gamma \in \Gamma}{\operatorname{argmin}} \; \|z_1 - (z_2^P \circ \gamma)\sqrt{\dot{\gamma}}\|. \tag{2.3}$$

This is a well known problem and usually solved numerically by minimizing a cost function $H[\gamma] = \int_0^1 \|z_1(t) - z_2^P\,(\gamma(t))\,\sqrt{\dot{\gamma}(t)}\|\,dt$ using a dynamic programming algorithm (DPA) or gradient based methods (see Srivastava, Klassen, et al. 2011). In this thesis we will use the methods laid out in Steyer, A. Stöcker, and Greven 2021, for solving Eq. 2.3 in the setting of sparse and irregularly sampled curves.

Anmerkung für Lisa Ich habe mich noch gefragt ob man anstatt Iteration über Eq. 2.2 und Eq. 2.3 die Distanz in Lemma 2.2 ii.) auch direkter optimieren kann.

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

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

[TODO: Explain iterative procedure in abit more detail! Basically as in Srivastava, Klassen, et al. 2011.]

2.3. The Elastic Full Procrustes Mean for Planar Curves

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

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

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

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

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

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

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

Here, $z_i = \frac{q_i}{\|q_i\|}$ are the observed normalized SRV curves, ω_i the rotation and scaling alignment to μ_q and γ_i the re-parametrisation alignment to μ_q. We can further simplify this by solving the optimization over ω_i using Lemma 2.2 ii.

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

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

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2 \colon \|\mu_q\|=1}{\operatorname{argmax}} \sum_{i=1}^{N} \sup_{\gamma_i \in \Gamma} \big\langle \mu_q, (\gamma_i, z_i) \big\rangle \, \big\langle (\gamma_i, z_i), \mu_q \big\rangle$$

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

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

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

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

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

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

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

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

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

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

$$\hat{\mu}_q = \underset{\mu_q \in \mathbb{L}^2,\, \|\mu_q\|=1}{\operatorname{argmax}} \left\langle \mu_q, C\mu_q \right\rangle \tag{2.4}$$

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

$$Cu = \lambda u \quad \Leftrightarrow \quad \int_0^1 C(s,t)u(t)\, dt = \lambda u(s)\,,$$

where $\lambda = \left\langle \mu_q,\, C\mu_q \right\rangle$ is the target function to maximize. **[TODO: Nochmal überprüfen ob das so stimmt. Evtl. complex conjugate irgendwo?]** For normalized eigenfunctions

$u_1, u_2, \dots$ and corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots$ of $C(s,t)$, the solution $\hat{\mu}_q(t)$ is given by the leading normalized eigenfunction $u_1(t)$ of $C(s,t)$ (see Ramsay and Silverman 2005, pp. 153, 397).

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

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

3. Estimation Strategy for Sparse and Irregular Observations

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

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

3.1. Discrete Treatment of SRV Curves

A natural first consideration might be how to calculate SRV curves from sparse observations. We defined the SRV curve of a function $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ as $q = \frac{\dot{\beta}}{\sqrt{\|\dot{\beta}\|}}$ (for $\dot{\beta} \neq 0$). This means that if we want to calculate the SRV curve, we have to calculate the derivative of β. As we never observe the whole function β but only a set discrete points $\beta(t_1), \dots, \beta(t_m)$, this is already not straight forward, as we cannot simply calculate a pointwise derivative. However, following Steyer, A. Stöcker, and Greven 2021, we treat a discretely observed curve β as piecewise linear between its observed corners $\beta(t_1), \dots, \beta(t_m)$, which allows us to calculate a piecewise constant derivative on the intervalls $[t_j,\, t_{j+1}]$, $j = 1, \dots, m-1$. Let us first consider the case of unwarped observations.

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

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

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

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

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

[TODO: Nochmal überlegen mit γ^{-1} und γ. Gerade ist das hier glaube ich falsch.] Let us now consider a warping function γ. The warped discrete derivative is given by $\Delta(\beta \circ \gamma)\big|_{[\gamma(t_j),\gamma(t_{j+1})]} = \frac{\beta(\gamma(t_{j+1}))-\beta(\gamma(t_j))}{\gamma(t_{j+1})-\gamma(t_j)}$. The corresponding warped SRV curve is then given by $(q \circ \gamma)\sqrt{\dot{\gamma}}\big|_{[\gamma(t_j),\gamma(t_{j+1})]} = \frac{\beta(\gamma(t_{j+1}))-\beta(\gamma(t_j))}{\sqrt{\gamma(t_{j+1})-\gamma(t_j)}\cdot\sqrt{\|\beta(\gamma(t_{j+1}))-\beta(\gamma(t_j))\|}}$. **[TODO: Das hier zuende formulieren.]**

3.2. Efficient Estimation using Hermitian Covariance Smoothing

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

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

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

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

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

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

with $\Xi_{\Re}, \Xi_{\Im} \in \mathbb{R}^{K \times K}$ and $\Xi = \Xi_{\Re} + i\Xi_{\Im}$, under the constraints that $\Xi_{\Re}^T = \Xi_{\Re}$ and $\Xi_{\Im}^T = -\Xi_{\Im}$. In this thesis I estimate $\Xi_{\Re}$ and $\Xi_{\Im}$ using the `R` (**Rcore**) package `mgcv` (Wood 2017). For efficient estimation, two `mgcv` smooths from the package `sparseFLMM` (Cederbaum, Volkmann, and A. Stöcker 2021) are used, which implement and generalize the approach proposed by Cederbaum, Scheipl, and Greven 2018 for symmetric and skew-symmetric tensor product p-splines. **[TODO: Ein Absatz zu**

p-splines? Penalty etc.]

3.3. Estimating the Full Procrustes Mean in a Fixed Basis

[TODO: Explain smoothing!] Although observations might be sparse and irregular, we want to model the mean as a whole function. In FDA this is usally achieved by estimating the mean function using a specific set of *basis functions* $b_1, \dots, b_K$ that are chosen so that the mean can be approximated arbitrarily well as a linear combination

$$\mu(t) = \sum_{k=1}^{K} \theta_k b_k(t),$$

as long as K is large enough (see Ramsay and Silverman 2005, Ch. 3). As we are dealing with complex valued functions, we use complex coefficients $\theta_k \in \mathbb{C}$ with real valued basis functions.

[TODO: Motivate b-spline basis using Lisa's paper!]

After choosing a basis representation, we want to estimate complex coefficients $\theta_j \in \mathbb{C}$ so that the Full Procrustes mean of SRV curves is given by $\hat{\mu}(t) = \sum_{j=1}^{k} \hat{\theta}_j b_j(t) = b^T \hat{\theta}$:

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

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

$$S_{kl} = \sum_{i=1}^{n} \int_0^1 \bar{b}_k(t) q_i(t) dt \int_0^1 \bar{q}_i(s) b_l(s) ds$$

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

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

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

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

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

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

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

One may solve this by using Lagrange optimization with the Langrangian

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

3.4. Numerical Integration of the Procrustes Fits

[TODO: Derivation as with Lisa.]

4. Empirical Applications

4.1. Mean Estimation for Simulated Spirals

4.2. Classification of Hand-written Digits

4.3. Mean Differences of Tounge Shapes in a Phonetics Study

5. Summary

Bibliography

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

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

Cederbaum, J., A. Volkmann, and A. Stöcker (2021). *sparseFLMM: Functional Linear Mixed Models for Irregularly or Sparsely Sampled Data*. R package version 0.4.0. url: `https://CRAN.R-project.org/package=sparseFLMM`.

Dryden, I. L. (2019). *`shapes` package*. Contributed package, Version 1.2.5. R Foundation for Statistical Computing. Vienna, Austria. url: `http://www.R-project.org`.

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

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.

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

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

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

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

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

Steyer, L. (2021). *elasdics: Elastic Analysis of Sparse, Dense and Irregular Curves*. R package version 0.1.1. url: `https://CRAN.R-project.org/package=elasdics`.

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

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

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.

A. Appendix

A.1. Proofs and Derivations

A.1.1. Derivation of Lemma 2.2

[TODO: Make dependence on γ explicit!]

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

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

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

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

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

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

[TODO: Make this part clearer!] Lemma 2.2 ii.) follows from $\kappa_\gamma^2 = |\langle z_1, \widetilde{z}_2 \rangle|^2 = \langle z_1, \widetilde{z}_2 \rangle \langle \widetilde{z}_2, z_1 \rangle$ and $\widetilde{z}_2 = (z_2 \circ \gamma)\sqrt{\dot{\gamma}}$. Lemma 2.2 i.) follows from $\lambda e^{i\theta} = \kappa_\gamma e^{-i\phi_\gamma} = \overline{\langle z_1, \widetilde{z}_2 \rangle}$. □

A.2. Discussion of Possible Extensions to Closed Curves

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

B. Supplementary Materials

B.1. Dataset Replication Guide

B.2. Implementation Notes

—Discarded—

Math-Basics Recap

Scalar Products

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

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

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

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

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

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

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

Functional Scalar Products

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

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

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

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

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

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

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

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

which yields the coefficients a_k:

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

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

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

We can then write the scalar product as:

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

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

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

FDA-Basics Recap

As discussed in the last section we can express a function f in its *basis function expansion* using a set of basis functions ϕ_k with $k = 1, \dots, K$ and a set of coefficients $c_1, \dots, c_K$ (both possibly $\mathbb{C}$ valued e.g. in the case of $2D$-curves)

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

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

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

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

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

Smoothing by Regression

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

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

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

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

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

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

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

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

Common Basis Representations

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

Polygonal Basis **[...]**

Bivariate Functional Data

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

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

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

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