Würde ich noch einführen - könntest Bsp... die caption dazu benutzen.

2. Elastic Full Procrustes Means for Planar Curves

sec:2

As a starting point, it is important to establish a notational and mathematical framework for the treatment of planar shapes. While the restriction to the 2D case might seem a major one, it still covers all shape data extracted from e.g. imagery and is therefore very applicable in practice. The outline of a 2D object may be naturally represented by a planar curve $\beta : [0,1] \to \mathbb{R}^2$ with $\beta(t) = (x_1(t), x_2(t))^T$, where $x_1(t)$ and $x_2(t)$ are the scalar-valued *coordinate functions*. Calculations in two dimensions, and in particular the derivation of the full Procrustes mean, are greatly simplified by using complex notation. We will therefore identify $\mathbb{R}^2$ with $\mathbb{C}$, as shown in Figure 2.1, and always use complex notation when representing a planar curve:

notational overlap →

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

vielleicht z.B. zumindest eines der „i" nicht kursiv?

We will assume the curves to be absolutely continuous, denoted as $\beta \in \mathcal{AC}([0,1],\mathbb{C})$, guaranteeing ~~us~~ that $\beta(t)$ has an integrable derivative. This is important when working in the square-root-velocity (SRV) framework, as will be discussed in Section 2.2. All considerations will be restricted to the case of open curves, with possible extensions to closed curves $\beta \in \mathcal{AC}(\mathbb{S}^1,\mathbb{C})$ discussed in Section A.2 of the appendix.

2.1. Equivalence Classes and Shape

ec:2-shape

As mentioned in the introduction, shape is usually defined by its invariance under the transformations of scaling, translation, and rotation. When considering the shape of curves, we additionally have to take into account invariance with respect to re-parametrisation. This can be seen, by noting that the curves $\beta(t)$ and $\beta(\gamma(t))$, with some re-parametrisation or *warping function* $\gamma : [0,1] \to [0,1]$ monotonically increasing and differentiable, have the same image and therefore represent the same geometrical object (see Figure 1.1b). We can say that the actions of translation, scaling, rotation, and

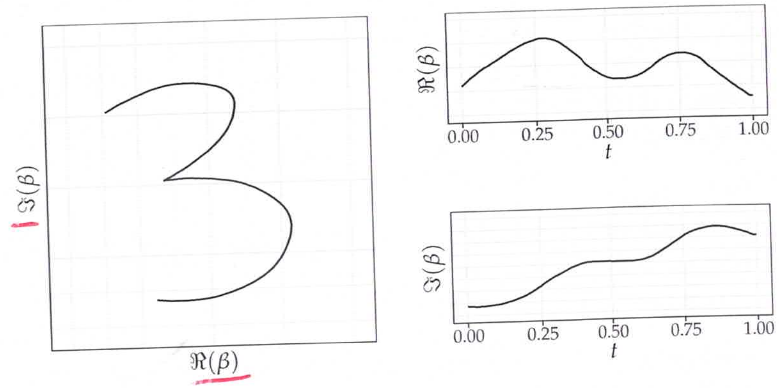


Figure 2.1.: Example of a planar curve (left) with respective coordinate functions (right) using complex notation. Data: see Figure 1.1.

ig:2-curve

re-parametrisation are *equivalence relations* with respect to shape, as each action leaves the shape of the curve untouched and only changes the way it is represented. The shape of a curve can then be defined as the respective *equivalence class*, i.e. the set of all possible shape preserving transformations of the curve. As two equivalence classses are neccessarily either disjoint or identical, we can consider two curves as having the same shape, if they are elements of the same equivalence class (see SRIVASTAVA and KLASSEN 2016, p. 40).

When defining an equivalence class, one has to first consider how each individual transformation acts on a planar curve $\beta : [0,1] \to \mathbb{C}$. This is usually done using the notion of *group actions* and *product groups*, with the later 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 $\beta_1 \overset{\text{Trl}}{\sim} \beta_2$, if there exists a complex scalar $\tilde{\xi} \in \mathbb{C}$ so that $\beta_1 = \beta_2 + \tilde{\xi}$. Then, for some function β, the related equivalence class with respect to translation is given by $[\beta]_{\text{Trl}} = \{\beta + \xi \,|\, \xi \in \mathbb{C}\}$.
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 $\beta_1 \overset{\text{Scl}}{\sim} \beta_2$, if there exists a scalar $\tilde{\lambda} \in \mathbb{R}^+$ so that $\beta_1 = \tilde{\lambda}\beta_2$. An equivalence class is such that ?

0 oder 2π besser ausschließen oder?

$$[\beta]_{\text{Scl}} = \{\lambda\beta \mid \lambda \in \mathbb{R}^+\}.$$

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 $[\beta]_{\text{Rot}} = \{e^{i\theta}\beta \mid \theta \in [0, 2\pi]\}$.

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 $[\beta]_{\text{Wrp}} = \{\beta \circ \gamma \mid \gamma \in \Gamma\}$.

In a next step, we can consider how these transformations act in concert and whether they *commute*, i.e. whether the order of applying the transformations changes outcomes. Consider for example the actions of the *rotation and scaling* product group $\mathbb{R}^+ \times [0, 2\pi]$ given by $((\lambda, \theta), \beta) \xmapsto{\text{Scl+Rot}} \lambda e^{i\theta}\beta$, which clearly commutes as $\lambda(e^{i\theta}\beta) = e^{i\theta}(\lambda\beta)$. On the other hand, the joint actions of *scaling and translation* do not commute, as $\lambda(\beta + \xi) \neq \lambda\beta + \xi$, with the same holding for the joint actions of *rotation and translation*. As the order of translating and rotating or scaling matters, one usually takes the translation to act on the already scaled and rotated curve. The joint action defined using this ordering is called an *Euclidean similarity transformation* with $((\xi, \lambda, \theta), \beta) \xmapsto{\text{Eucl}} \lambda e^{i\theta}\beta + \xi$ (see Dryden and Mardia 2016, p. 62). Considering the action of *warping* or re-parametrization, we can note that it necessarily commutes with all Euclidean similarity transformations as those only act on the image of β, while the former only acts on the parametrization. Putting everything together we can give a formal definition of the shape of a planar curve as the following equivalence class:

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

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

The *shape space* $\mathcal{S}$ is then given by $\mathcal{S} = \{[\beta] \mid \beta \in \mathcal{AC}([0,1], \mathbb{C})\}$.

sec:2-dist

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 already not straight-forward. A common approach is to map each equivalence class $[\beta]$ to a suitable representative, so that the distance calculation in shape space can be identified with a (much simpler) distance calculation over the representatives in the underlying functional space.

To illustrate this, let us first discuss each type of shape-preserving transformation individually, starting with the Euclidean similarity transformations. Consider two equivalence classes with respect to translation $[\beta_1]_{\text{Trl}}, [\beta_2]_{\text{Trl}}$. 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 $i = 1, 2$. We can then define an appropriate distance that is invariant under translation as $d_{\text{Trl}}([\beta_1]_{\text{Trl}}, [\beta_2]_{\text{Trl}}) = \|\widetilde{\beta}_1^{\text{Trl}} - \widetilde{\beta}_2^{\text{Trl}}\|$. Similarly, a distance that is invariant under scaling might be defined over the normalized 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 invariance under rotation, we can first note that no "standardization" procedure compareable to normalizing and centering exists for the case of rotation. Instead of mapping $[\beta]_{\text{Rot}}$ 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 (compare e.g Stöcker and Greven 2021).

Hier fänd ich den Hinweis gut, dass das der minimale Abstand

Hier entspr... wundert mich ein bisschen, dass Du die Distanz nimmst, die eigtl. der partial Procrustes Distance entspricht.

The Full Procrustes Distance

gibt ja mehr als die beiden

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

An irgendeiner Stelle würde ich noch auf die "procrustes distance" ~~als~~ als dritte Alternative verweisen, weil die bislang der 'State of the Art' von Srivastava & Co. ist. ~~[illegible]~~ Vielleicht hast Du das aber auch anderswo?

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

Hier vielleicht sowas einfügen wie "as will later be ~~discussed~~ discussed in more detail," ?

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

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

eq:2-fpdist-(

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

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

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

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

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

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

Proof. See Appendix A.1.1. □

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

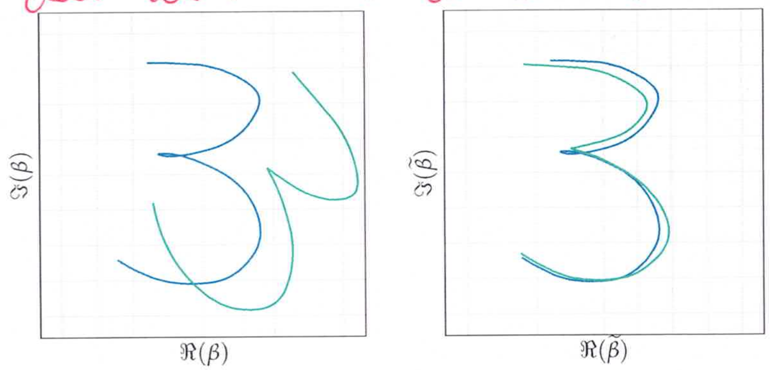


fig:2-pfit

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

The Elastic Distance

When considering warping, we would like to do something similar to rotation in trying to find an optimal warping alignment between two curves β_1, β_2 by optimizing over their distance $\inf_{\gamma \in \Gamma} \|\beta_1 - (\beta_2 \circ \gamma)\|$, where Γ is the space of warping functions. A usual choice would be to optimize this over the $\mathbb{L}^2$-distance between the curves, but in this case the result would not define a proper distance. Optimizing over re-parametrization using the $\mathbb{L}^2$-distance has problems relating to the so called *pinching effect* and *inverse-inconsistency*, where the later means that aligning the parametrisation of one curve to another by $\inf_{\gamma \in \Gamma} \|\beta_1 - \beta_2 \circ \gamma\|$ may yield different results than $\inf_{\gamma \in \Gamma} \|\beta_2 - \beta_1 \circ \gamma\|$ (see Srivastava and Klassen 2016, pp. 88–90).

umgangssprachlich

A solution proposed in Srivastava, Klassen, et al. 2011 is to ditch the $\mathbb{L}^2$-metric in favor of an *elastic metric*, which is isometric with respect to warping. Calculation of this metric, the Fisher-Rao Riemannian metric (Rao 1945), can be greatly simplified by using the *square-root-velocity* (SRV) framework, as the Fisher-Rao metric of two curves can be equivalently calculated as the $\mathbb{L}^2$-distance of their respective SRV curves. As this SRV representation makes use of derivatives, any curve β that has a SRV curve must fulfill some kind of differentiability constraint. Here it is enough to consider only curves that are absolutely continuous $\beta \in \mathcal{AC}([0,1], \mathbb{C})$, which in particular means

Die Formel ~~[illegible]~~ würde ich runterziehen, weil hypothetisch — in oder ~~[illegible]~~ $d(\beta_1, (\beta_2 \circ \gamma))$ umformulieren damit es eine gültige Formulierung bleibt.

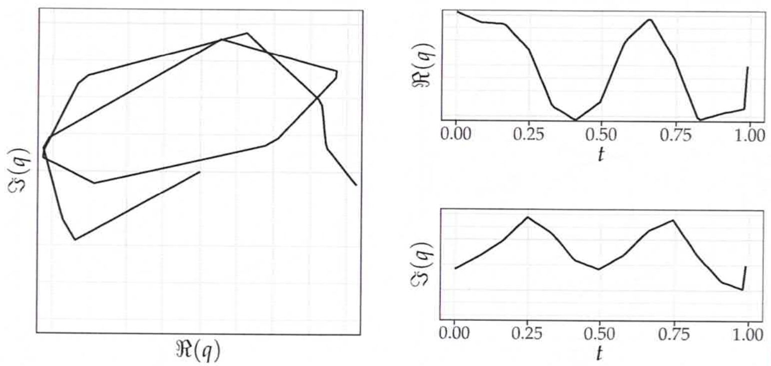


Figure 2.3.: SRV function (left) of the planar curve in Figure 2.1 with respective SRV coordinate functions (right). Note that the polygon-like look of the SRV curve is an artifact of the (linear) smoothing applied to the original data on SRV level (see Appendix A.3). Data: see Figure 1.1.

fig:2-srv

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

def:2-eldist

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

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

eq:2-eldist-c

Hier schreibst Du ~~...~~ $\|\cdot\|_{\mathbb{L}^2}$ aber meist $\|\cdot\|$, Ich glaube, Du solltest ~~das~~ $\mathbb{L}^2$-Skalarprod. und -norm vorher klar einführen (vielleicht gleich zu Beginn von Kapitel 2)

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

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

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

Unlike the optimization over rotation in the definition of the full Procrustes distance (see Eq. 2.1), no analytical solution exists for the optimization over warping (see Eq. 2.3)

würde stattdessen vorschlagen "warping in Eq. \eqref{...}" oder auch ohne "Eq."

In meiner Erfahrung wird die Standard-Ableitung nach einer Variable üblicherweise eher mit $\frac{d\beta_i(t)}{dt}$ bezeichnet als mit \partial.

in the definition of the elastic distance. Instead, Eq. 2.3 is usually solved numerically, by minimizing a cost function $H[\gamma] = \int_0^1 \|q_1(t) - q_2(\gamma(t))\sqrt{\dot\gamma(t)}\|\,dt$ using a dynamic programming algorithm (see e.g. SRIVASTAVA and KLASSEN 2016, p. 152) or gradient based methods (see e.g. STEYER, A. 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 curve β up to translation from its respective SRV curve q by integration $\beta(t) = \beta(0) + \int_0^t q(s)\|q(s)\|ds$. Because the translation of the original curve is usually not of interest from the point of shape analysis, the SRV curve holds all relevant information about the shape of β. This means, in particular, that instead of analysing the shape of β, we can equivalently analyse the shape of q. The shape preserving transformations on original curve level translate to SRV curve level by actions laid out in Lem. 2.2.

Write abit more here. Cita- tions!

n:2-transf

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

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

(see e.g. SRIVASTAVA *and* KLASSEN *2016, p. 142).*

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

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

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 Def. 2.2)

↳ Gleichungen werden üblicherweise in Runden Klammern referenziert.

was defined over the normalized and centered curves, the SRV curves are already translation invariant so additional centering is not necessary. We will therefore define the *elastic full Procrustes distance* as the minimal distance, when aligning the scaling, rotation, and warping of the normalized SRV curves $\tilde{q} = \frac{q}{\|q\|}$. Note that when the original curve β is of unit length $L[\beta] = \int_0^1 |\dot{\beta}(t)| dt = 1$ the SRV curve $q = \frac{\dot{\beta}}{\|\dot{\beta}\|}$ will be normalized, as

$$\|q\| = \sqrt{\int_0^1 |q(t)|^2 \, dt} = \sqrt{\int_0^1 |\dot{\beta}(t)| \, dt} = \sqrt{L[\beta]}. \tag{2.5}$$ eq:2-norm

ef:2-elfpdist **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

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

with normalized SRV representation $\tilde{q}_i = \frac{q_i}{\|q_i\|} \in \mathbb{L}^2([0,1], \mathbb{C})$, 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 optimization problem over $\mathbb{C} \times \Gamma$

$$(\omega^{\text{opt}}, \gamma^{\text{opt}}) = \operatorname*{argmin}_{\omega \in \mathbb{C}, \gamma \in \Gamma} \|\tilde{q}_1 - \omega(\tilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\|_{\mathbb{L}^2}, \tag{2.7}$$ eq:2-elfpd

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

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

$\gamma^{(k)}$ nicht notwendigerw. in Γ

Satzbau →

Following Srivastava, Klassen, et al. 2011 we optimize over the sets of parameters individually and then to iterate through both solutions until some form of convergence is reached. Let us first consider the optimization over $\omega \in \mathbb{C}$ for a fixed $\gamma^{(k)} \in \Gamma$.

→ ~~könnte~~ weil nicht vollständig definiere vielleicht einfach $\bar{\Gamma}$

$$\omega^{(k)} = \operatorname*{argmin}_{\omega \in \mathbb{C}} \|\tilde{q}_1 - \omega(\tilde{q}_2 \circ \gamma^{(k)})\sqrt{\gamma^{(k)}}\|_{\mathbb{L}^2}, \tag{2.9}$$ eq:2-elfpd

$\dot{\gamma}^{(k)}$?

Hier wäre es beim Lesen angenehmer, würdest Du „k" gleich zu Beginn einführen.

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

$$\omega^{(k)} = \langle (\tilde{q}_2 \circ \gamma^{(k)})\sqrt{\dot{\gamma}^{(k)}}, \tilde{q}_1 \rangle \tag{2.10}$$ eq:2-elfpd

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

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

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

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

links & rechts passen nicht zusammen: $\gamma^{(0)}(t) = t$, $\gamma^{(0)} = id$, $\gamma^{(0)}: t \mapsto t$

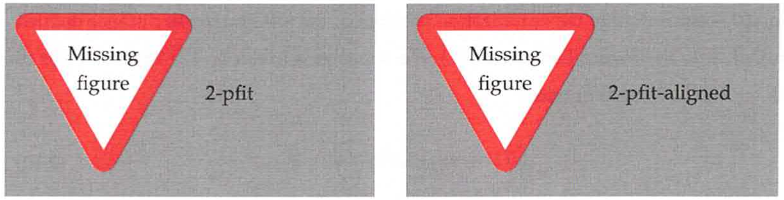
1. *Calculate $\omega^{(k)} = \langle (\tilde{q}_2 \circ \gamma^{(k)})\sqrt{\dot{\gamma}^{(k)}}, \tilde{q}_1 \rangle$. **Stop** if ...*
2. *Solve $\gamma^{(k+1)} = \operatorname{argmin}_{\gamma \in \Gamma} \|\tilde{q}_1 - (\omega^{(k)} \cdot \tilde{q}_2 \circ \gamma)\sqrt{\dot{\gamma}}\|_{\mathbb{L}^2}$.*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

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

Stopping criterion nochmal überlegen.

See Figure 2.4 for an example of two curves that have been aligned by minimizing their elastic full Procrustes distance.

→ Oder hier bei dir absichtlich schon, weil endlich viele Iterationen?



astic-pfit

Figure 2.4.: Elastic Procrustes Fits. Data: See Figure 1.1.

sec:2-mean

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 continous $\beta_i \in \mathcal{AC}([0,1],\mathbb{C})$ with corresponding SRV curves $q_i \in \mathbb{L}^2([0,1],\mathbb{C})$, $i = 1,\dots,N$. We can take into account invariance 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).

def:2-mean **Definition 2.5** (Elastic full Procrustes mean). For a set of curves $\beta_i \in \mathcal{AC}([0,1],\mathbb{C})$, $i = 1,\dots,N$, their *elastic full Procrustes mean* is given by the minimizing shape $[\hat{\mu}]$ with

a ?

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

eq:2-mean-def

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 Def. 2.4) and writing it as an optimization problem over a normalized SRV mean function.

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

eq:2-mean-

The estimated normalized SRV mean $\hat{\mu}_q$ then defines a representative unit-length (see Eq. 2.5) mean $\hat{\mu} \in [\hat{\mu}]$ by integration $\hat{\mu}(t) = \hat{\mu}(0) + \int_0^t \hat{\mu}_q(s)\|\hat{\mu}_q(s)\|\,ds$, 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 4.3 are tounge shapes, which all connect to the back of the mouth on one end. Another possibility would be to choose a $\hat{\mu}(0)$ that centers 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 visualizing multiple curves or in 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 the empirical tounge shapes data in Section 4.3.

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 optimization over rotation in the full Procrustes distance (see Lemma 2.1 i.).

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

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

Proof. The result follows from applying Lemma 2.1 i.) to Eq. 2.6 keeping γ fixed. □

Check Proof: Does it really?

i'd say so :)

Then Eq. 2.13 can be rewritten as

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

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

Würde wieder damit starten erstmal explizit zu sagen, dass über k iteriert werden soll.

and we end up with a two step optimization problem consisting of an outer optimization over μ_q and an inner optimization over the set $\{\gamma_i\}_{i=1,\dots,N}$. Similarly to the approaches discussed in Srivastava and Klassen 2016 and to Steyer, A. Stöcker, and Greven 2021, we solve this by *template based alignment* (see e.g. Srivastava and Klassen 2016, p. 271): In a first step the mean $\hat{\mu}_q$ is estimated while keeping the parametrizations fixed, after which the γ_i are updated by calculating the optimal warping alignment from the elastic full Procrustes fit of each $\tilde{q}_i$ onto $\hat{\mu}_q$. Both steps are iterated until the mean shape has converged.

Check in code!

Let us consider the outer optimization problem for a fixed set of warping functions $\{\gamma_i^{(k)}\}_{i=1,\dots,N}$ with corresponding warping aligned normalized SRV curves $\tilde{q}_i^{(k)} = (\tilde{q}_i \circ \gamma_i^{(k)})\sqrt{\dot{\gamma}_i^{(k)}}$. 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

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

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

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

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

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

Warping? (k)??

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

By replacing $\check{C}^{(k)}(s,t)$ by its expectation $C^{(k)}(s,t)$, we can analogously formulate an

Notation? (k)??

estimator on the population level.

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

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

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

eq:quadr_c

This is a well known problem in the context of functional principal component analysis (FPCA). From $\overline{C(s,t)} = \overline{\mathbb{E}[\tilde{q}(s)\overline{\tilde{q}(t)}]} = \mathbb{E}[\tilde{q}(t)\overline{\tilde{q}(s)}] = C(t,s)$ 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 optimization problem then reduces to an eigenfunction problem

– but typically for real-valued covariance operators.

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

eq:funceig

where $\lambda^{(k)} = \left\langle \mu_q, C^{(k)} \mu_q \right\rangle$ is the target function to maximize. For normalized eigenfunctions $u_1^{(k)}, u_2^{(k)}, \dots$ and corresponding eigenvalues $\lambda_1^{(k)} \geq \lambda_2^{(k)} \geq \dots$ of $C^{(k)}(s,t)$, the expectation $\mathbb{E}[\hat{\mu}_q^{(k)}(t)]$ is given by the leading normalized 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 Section 3. Given such an estimate $\hat{C}^{(k)}(s,t)$, we can calculate the elastic full Procrustes mean by the following algorithm.

algo:mean

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

Siehe weiter vorne

1. *For $i = 1, \dots, N$: Set $\tilde{q}_i^{(k)}(t) = \tilde{q}_i \left(\gamma_i^{(k)}(t)\right) \cdot \sqrt{\dot{\gamma}_i^{(k)}(t)}$.*
2. *Estimate $\hat{C}^{(k)}(s,t)$ from $\left\{\tilde{q}_i^{(k)}\right\}_{i=1,\dots,N}$.*
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 $i = 1, \dots, N$: Calculate $\omega_i^{(k)} = \left\langle \tilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \right\rangle$.*

6. *For $i = 1, \dots, N$: Solve* $\gamma_i^{(k+1)} = \text{argmin}_{\gamma \in \Gamma} \|\hat{\mu}_q^{(k)} - \omega_i^{(k)} (\tilde{q}_i \circ \gamma)\sqrt{\dot{\gamma}}\|$.

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

z.B.: Mean estimation ... ? (sprechenderer Titel)

3. Estimation Strategy for Sparse and Irregular Observations

sec:3

Alg. 2.2 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. 2.2, 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 $C^{(k)}(s,t)$ 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 $\omega_i^{(k)} = \langle \tilde{q}_i^{(k)}, \hat{\mu}_q^{(k)} \rangle$, where $\tilde{q}_i^{(k)}$ is a sparsely observed normalized SRV curve, while $\hat{\mu}_q^{(k)}$ is a smooth SRV mean function. Note that the final warping alignment step in Alg. 2.2 is solved by using methods for warping alignment of sparse and irregular curves provided in Steyer, A. Stöcker, and Greven 2021.

3.1. Discrete Treatment of SRV Curves

3-discrete

A natural first consideration might be how to calculate SRV curves from sparse observations. As the SRV curve of $\beta \in \mathcal{AC}([0,1], \mathbb{C})$ is defined as $q = \frac{\dot{\beta}}{\sqrt{\|\dot{\beta}\|}}$ (for $\dot{\beta} \neq 0$), we need to calculate a derivate of β. However, as we never observe the whole function β, but only a set discrete points $\beta(t_1), \dots, \beta(t_m)$, we cannot simply calculate a pointwise derivative. Following Steyer, A. Stöcker, and Greven 2021, we may treat a discretely observed curve β as piecewise linear between its observed corners

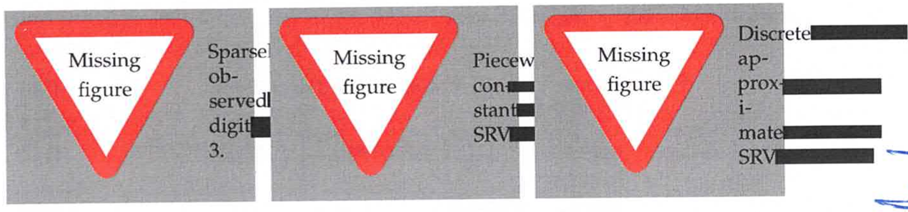


Figure 3.1.: A sparsely observed planar curve (left) with piecewise constant (center) and approximate discrete SRV curve (right). Data: see Figure 1.1.

fig:3-disc

$\beta(t_1), \ldots, \beta(t_m)$, which allows us to calculate a piecewise constant derivative on the intervals $[t_j, t_{j+1}]$ for $j = 1, \ldots, m-1$. As usually only the image $\beta(t_1), \ldots, \beta(t_m)$ but not the parametrisation $t_1, \ldots, t_m$ is observed, it is 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 $l = l_m$.

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

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

the constant *square-root-velocity* of β between its corners $\beta(t_j)$ and $\beta(t_{j+1})$. As shown in Steyer, 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 $\Delta\beta|_{[t_j,t_{j+1}]}$ 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. See Fig. 3.1 for a visualization of both approaches. Finally, we can approximate the normalized SRV curve $\tilde{q} = q/\|q\|$ using the polygon-length l of β by $\tilde{q}_j = q_j/\sqrt{l}$ (see Eq. 2.5).

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

Nochmal überlegen mit γ^{-1}

Ich würde hier kurz sagen, dass Du auch Penalties verwendest und dass die Glättungsparameter über Restricted Maximum Likelihood Estimation geschätzt werden in einem Quasi-Likelihood Ansatz.

3.2. Efficient Estimation using Hermitian Covariance Smoothing

sec:3-cov

In step k of Alg. 2.2, we want to estimate the warping aligned complex covariance surface $C^{(k)}(s,t) = \mathbb{E}[\tilde{q}^{(k)}(s)\overline{\tilde{q}^{(k)}(t)}]$ given approximate observations of the warped, normalized SRV curves $\tilde{q}_i^{(k)}(s_{ij})$ for $j = 1, \ldots, m_i - 1$ and $i = 1, \ldots, N$, where m_i denotes the number of observed points per curve. Following ..., we can treat this estimation as a smoothing problem, by constructing responses $y_{ijk}^{(k)} = \tilde{q}_i^{(k)}(s_{ij})\overline{\tilde{q}_i^{(k)}(s_{ik})}$ and treating the pairs s_{ij}, s_{ik} as covariates s and t. Smoothing the responses $y_{ijk}^{(k)}$ gives an estimate $\hat{C}^{(k)}(\cdot,\cdot)$ of $C^{(k)}(s,t)$, as each response has expectation $\mathbb{E}[y_{ijk}^{(k)}|s_{ij}, s_{ik}] = C^{(k)}(s_{ij}, s_{ik})$. We carry out the smoothing in a very flexible *tensor product spline* basis

Cite!

Index k für step ändern

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

\top statt „T"

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

co. 1.1

Erklärung Penalty? Motivate b-spline/p-spline basis using Lisa's paper? REML?

Taking into account the symmetry properties of the covariance surface by considering every unique pair s_{ij}, s_{ik} 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 seperate regression problems over the symmetric real and skew-symmetric imaginary parts of $C^{(k)}(s,t)$.

(folgt nicht hinreichend daraus sondern bestimmt die Form)

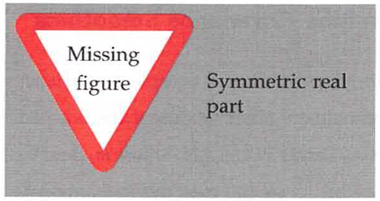
$$\mathbb{E}[\Re(y^{(k)})] = b(s)^T \Xi_{\Re}^{(k)} b(t) \tag{3.3}$$

$$\mathbb{E}[\Im(y^{(k)})] = b(s)^T \Xi_{\Im}^{(k)} b(t) \tag{3.4}$$

with $\Xi_{\Re}^{(k)}, \Xi_{\Im}^{(k)} \in \mathbb{R}^{K \times K}$ and $\Xi^{(k)} = \Xi_{\Re}^{(k)} + i\Xi_{\Im}^{(k)}$, under the constraints that $(\Xi_{\Re}^{(k)})^T = \Xi_{\Re}^{(k)}$ and $(\Xi_{\Im}^{(k)})^T = -\Xi_{\Im}^{(k)}$.

In this thesis $\Xi_{\Re}^{(k)}$ and $\Xi_{\Im}^{(k)}$ are estimated using the gam function from the R package mgcv (Wood 2017). Two mgcv smooths from the package sparseFLMM (Cederbaum, Volkmann, and A. Stöcker 2021) are used for efficient, hermition smoothing, which

Hermitian (auch immer groß oder?)



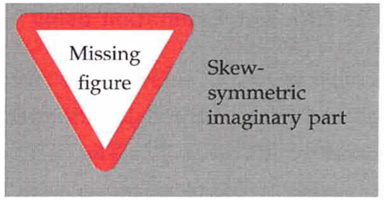


fig:3-cov

Figure 3.2.: Complex covariance surface on SRV curve level. Estimated using **PARAMETERS**. Data: `digits3.dat`

implement and generalize the approach proposed by Cederbaum, Scheipl, and Greven 2018 for symmetric and ~~skew-symmetric~~ tensor product p-splines.

Würde ich hier weglassen, weil die rein symmetrisches Covariance smoothing machen.

sec:3-mean

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

To estimate the elastic full Procrustes Mean, we have to solve a functional eigenvalue problem on the estimated covariance surface $\hat{C}^{(k)}(s,t) = b(s)^t \hat{\Xi}^{(k)} b(t)$. This may be achieved by evaluating $\hat{C}^{(k)}(s,t)$ on a dense grid and performing an eigendecomposition on the matrix of evaluations Alternatively, we can estimate the mean directly Cite in some basis $b(s) = (b_1(s), \dots, b_K(s))$, where a natural choice might be to evaluate mean and covariance surface in the same basis, i.e. $\mu_q^{(k)}(s) = b(s)^T \theta^{(k)}$. For notational clarity, let us formulate everything on the level of the unwarped estimated covariance surface $\hat{C}(s,t)$ for now. This should not matter, as everything translates one to one to the warped estimated covariance surface $C^{(k)}(s,t)$, with the only difference lying in the observations ($\tilde{q}_i^{(k)}$ vs. $\tilde{q}_i$) used for estimating $C(s,t)$.

Ich würde hier einfach nur sagen, wie Du es machst.

Ich glaube, diese Unterscheidung macht hier weniger klar, als sie verkompliziert.

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

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

Given an estimate of the covariance surface $\hat{C}(s,t) = b(s)^T \hat{\Xi} b(t)$, estimating the mean $\hat{\mu}_q$ in a basis $b(s)$, with $\hat{\mu}_q(s) = b(s)^T \theta$, then reduces to estimating the vector of

coefficients $\theta = (\theta_1, \dots, \theta_K) \in \mathbb{C}^K$ with

$$\begin{aligned}
\hat{\theta} &= \underset{\theta \in \mathbb{C}^K, \|b^T\theta\|=1}{\operatorname{argmax}} \int_0^1 \int_0^1 \theta^H b(s) b(s)^T \hat{\Xi} b(t) b(t)^T \theta \, ds \, dt && (3.6) \\
&= \underset{\theta \in \mathbb{C}^K, \|b^T\theta\|=1}{\operatorname{argmax}} \theta^H \left(\int_0^1 b(s) b(s)^T ds \right) \hat{\Xi} \left(\int_0^1 b(t) b(t)^T dt \right) \theta && (3.7) \\
&= \underset{\theta \in \mathbb{C}^K, \theta^H G \theta = 1}{\operatorname{argmax}} \theta^H G \hat{\Xi} G \theta && (3.8)
\end{aligned}$$

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

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

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

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

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

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

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

eq:lagrRe

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

eq:lagrIm

with the additional constraint $\theta_\Re^T G \theta_\Re + \theta_\Im^T G \theta_\Im = 1$.T We can simplify this further and

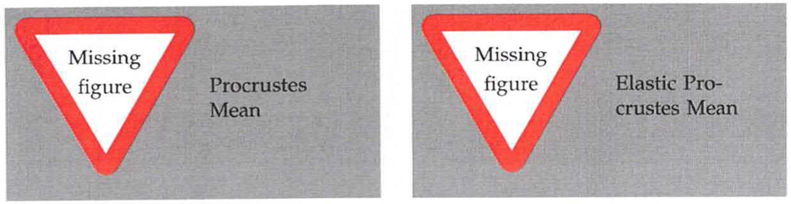


Figure 3.3.: Full Procrustes mean (left) and elastic full Procrustes mean (right) of handwritten digits '3'. Estimated using **PARAMETERS** (blue) and **PARAMETERS** (red). Data: `digits3.dat`

fig:3-mean

multiply Eq. 3.11 by i, leading to

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

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

Adding both equations finally leads to

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

or likewise, using θ and $\hat{\Xi}$

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

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

Cite Reiss.

3.4. Numerical Integration of the Procrustes Fits

ec:3-pfits

Subsection muss noch geschrieben werden. Mean value theorem in the integral, etc.

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

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