Elastic Full Procrustes Means for Sparse and Irregular Planar Curves

Masters Thesis

Manuel Pfeuffer[1]

12th October 2020, Berlin

Advisors: Prof. Sonja Greven, Lisa Maike Steyer, Almond Stöcker

[1] pfeuferm@hu-berlin.de, Matriculation Number: 577668

Contents

1 Math-Basics Recap

1.1 Scalar Products

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

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

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

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

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

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

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

1.2 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:

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

2 FDA-Basics Recap

As discussed in the last section we can express a function f in its *basis function expansion* using a set of basis functions ϕ_k with $k = 1, \ldots, K$ and a set of coefficients $c_1, \ldots, c_K$

(both possibly $\mathbb{C}$ valued e.g. in the case of $2D$-curves)

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

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

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

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

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

2.1 Smoothing by Regression

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

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

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

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

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

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

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

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

2.2 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 **[. . .]**

2.3 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*

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

3 The Full Procrustes Mean for Planar Curves

Let β be a continuous planar curve. It can be represented in a parameterized form in $\mathbb{R}^2$ as

$$\beta : [0,1] \to \mathbb{R}^2, \quad \beta(t) = (x(t), y(t))\,,$$

where x, y are scalar-valued *coordinate functions* of β, parametrized by t. We can equivalently represent a planar curve using complex numbers as

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

with the added benefit that complex notation often simplifies calculations in the 2D case.

For a set of planar curves $\beta_1, \ldots, \beta_n : [0,1] \to \mathbb{C}$, either centered with $\langle \beta_i, \mathbb{1} \rangle$ or with no relative translation to each other, the *full Procrustes mean* $\hat{\mu}$ is then defined as the curve minimizing the sum of squared *full Procrustes distances* from each β_i to an unknown unit size mean configuration μ, that is

$$\begin{aligned}
\hat{\mu} &= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmin}} \sum_{i=1}^{n} d_F^2(\mu, \beta_i) \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmin}} \sum_{i=1}^{n} 1 - \frac{\langle \mu, \beta_i \rangle \langle \beta_i, \mu \rangle}{\langle \mu, \mu \rangle \langle \beta_i, \beta_i \rangle} \quad \text{s.t. } ||\mu|| = 1
\end{aligned}$$

which we can be further simplified by normalizing $\beta_i := \frac{\beta_i}{||\beta_i||}$ and using $\langle \mu, \mu \rangle = 1$

$$\hat{\mu} = \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \langle \mu, \beta_i \rangle \langle \beta_i, \mu \rangle \quad \text{s.t. } ||\mu|| = 1.$$

The expression for $d_F^2(\mu, \beta_i)$ in the case of planar curves is derived in appendix A.1.

3.1 The SRV Framework

Instead of working with the original curve β, for calculation of *elastic* means it is advantageous to work with its corresponding *square root velocity curve* given by

$$q : [0,1] \to \mathbb{C}, \quad q(t) = \frac{\dot{\beta}(t)}{\sqrt{||\dot{\beta}(t)||}} \quad \text{for } \dot{\beta}(t) \neq 0$$

where original curve β can be obtained up to translation by back transformation via $\beta(t) = \beta(0) + \int_0^t q(s)||q(s)||ds$. Moreover, if the original curve is of unit length the SRV curve will be automatically normalized:

$$||q|| = \sqrt{\langle q, q\rangle} = \sqrt{\int_0^1 \overline{q(t)} q(t)\, dt} = \sqrt{\int_0^1 |q(t)|^2\, dt} = \sqrt{\int_0^1 |\dot{\beta}(t)|\, dt} = \sqrt{1} = 1.$$

3.2 The Full Procrustes mean

Consider a set of planar SRV curves $q_1, \ldots, q_n : [0,1] \to \mathbb{C}$ of unit length $||q_i|| = 1$ for all i. The *full Procrustes mean* $\hat{\mu}$ is given by

$$\begin{aligned}
\hat{\mu} &= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \langle \mu, q_i\rangle \langle q_i, \mu\rangle \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \sum_{i=1}^{n} \int_0^1 \overline{\mu(t)} q_i(t)\, dt \int_0^1 \overline{q_i(s)} \mu(s)\, ds \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \int_0^1 \int_0^1 \overline{\mu(t)} \underbrace{\left(\sum_{i=1}^{n} q_i(t) \overline{q_i(s)} \right)}_{:= n\hat{C}(s,t)} \mu(s)\, dtds \quad \text{s.t. } ||\mu|| = 1 \\
&= \underset{\mu:[0,1]\to\mathbb{C}}{\operatorname{argmax}} \int_0^1 \overline{\mu(t)} \int_0^1 \hat{C}(s,t) \mu(s)\, dsdt \quad \text{s.t. } ||\mu|| = 1
\end{aligned}$$

with the solution given by the eigenfunction corresponding to the largest eigenvector of the complex empirical covariance function $\hat{C}(s,t) = n^{-1} \sum_{i=1}^{n} q_i(t) \overline{q_i(s)}$.

3.3 The Full Procrustes Mean in a fixed basis

To avoid having to sample the estimated covariance surface $\hat{C}(s,t)$ on a large grid when calculating its leading eigenfunction, it might be preferable to calculate this eigenfunction from the vector of basis coefficients directly. After choosing a basis representation $b = (b_1, \ldots, b_k)$ with $b_j : \mathbb{R} \to \mathbb{R}$ real-valued basis functions, we want to estimate complex coefficients $\theta_j \in \mathbb{C}$ so that the Full Procrustes mean of SRV curves is given by $\hat{\mu}(t) = \sum_{j=1}^{k} \hat{\theta}_j b_j(t) = b^T \hat{\theta}$:

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

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

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

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

We can then further simplify S_{kl}

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

where g_{ij}, $i, j = 1, \dots, k$ 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 $\hat{\Xi} = (\hat{\xi}_{ij})_{i,j=1,\dots,k}$:

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

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 Estimation of the covariance surface $C(s,t)$

Consider the following model for independent curves

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

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

A Additional Derivations

A.1 Derivation of the Full Procrustes Distance for Functional Data

Consider two curves $\beta_1, \beta_2 : [0,1] \to \mathbb{C}$ with $\langle \beta_1, \mathbb{1} \rangle = \langle \beta_2, \mathbb{1} \rangle = 0$ where $\mathbb{1}$ is the constant function $\mathbb{1}(t) = 1$ for all $t \in [0,1]$. Then β_1 and β_2 can be considered to be centered as

$$\langle \beta_1, \mathbb{1} \rangle = \int_0^1 \bar{\beta_1}(t)\mathbb{1}(t)dt = \int_0^1 \bar{\beta_1}(t)dt = \int_0^1 (y(t) + ix(t))dt = \underbrace{\int_0^1 y(t)dt}_{\overset{!}{=}0} + i \underbrace{\int_0^1 x(t)dt}_{\overset{!}{=}0} = 0$$

Then the full procrustes distance of β_1, β_2 is given by their minimum distance controlling for translation $\gamma \in \mathbb{C}$, and scaling and rotation $\omega = be^{i\theta} \in \mathbb{C}$:

$$\begin{aligned}
d_F^2 &= \min_{\omega,\gamma \in \mathbb{C}} ||\beta_1 - \gamma\mathbb{1} - \omega\beta_2||^2 \\
&= \min_{\omega,\gamma \in \mathbb{C}} \langle \beta_1 - \gamma\mathbb{1} - \omega\beta_2, \beta_1 - \gamma\mathbb{1} - \omega\beta_2 \rangle \\
&= \min_{\omega,\gamma \in \mathbb{C}} \langle \beta_1 - \omega\beta_2, \beta_1 - \omega\beta_2 \rangle - \underbrace{\langle \beta_1, \gamma\mathbb{1} \rangle}_{=0} - \underbrace{\langle \gamma\mathbb{1}, \beta_1 \rangle}_{=0} + \underbrace{\langle \gamma\mathbb{1}, \omega\beta_2 \rangle}_{=0} + \underbrace{\langle \omega\beta_2, \gamma\mathbb{1} \rangle}_{=0} + \underbrace{\langle \gamma\mathbb{1}, \gamma\mathbb{1} \rangle}_{=||\gamma\mathbb{1}||^2} \\
&\overset{\gamma=0}{=} \min_{\omega \in \mathbb{C}} \langle \beta_1, \beta_1 \rangle + \langle \omega\beta_2, \omega\beta_2 \rangle - \langle \beta_1, \omega\beta_2 \rangle - \langle \omega\beta_2, \beta_1 \rangle \\
&= \min_{\omega \in \mathbb{C}} \langle \beta_1, \beta_1 \rangle + |\omega|^2 \langle \beta_2, \beta_2 \rangle - \omega \langle \beta_1, \beta_2 \rangle - \overline{\omega} \langle \beta_2, \beta_1 \rangle
\end{aligned}$$

To find $\omega \in \mathbb{C}$ that minimizes $||\beta_1 - \omega\beta_2||^2$ we first consider the part of the problem dependent on θ. We need to solve

$$\min_{\omega \in \mathbb{C}} -\omega \langle \beta_1, \beta_2 \rangle - \overline{\omega} \langle \beta_2, \beta_1 \rangle = \max_{\omega \in \mathbb{C}} \omega \langle \beta_1, \beta_2 \rangle + \overline{\omega} \langle \beta_2, \beta_1 \rangle$$

by using $\omega = be^{i\theta}$ and $\langle \beta_1, \beta_2 \rangle = ae^{i\phi}$:

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

$$\overset{\theta=-\phi}{=} \max_{b\in\mathbb{R}^+} 2ba$$

and using $\theta = -\phi$ the original mimization problem therefore simplifies to

$$\begin{aligned}
d_F^2 &= \min_{b\in\mathbb{R}^+} \langle\beta_1,\beta_1\rangle + b^2\langle\beta_2,\beta_2\rangle - 2ba \\
\frac{\partial d_F^2}{\partial b} &= 2b\langle\beta_2,\beta_2\rangle - 2a \overset{!}{=} 0 \\
\Rightarrow \quad b &= \frac{a}{\langle\beta_2,\beta_2\rangle}
\end{aligned}$$

And for the *full Procrustes distance* it follows that

$$d_F^2 = \langle\beta_1,\beta_1\rangle - \frac{a^2}{\langle\beta_2,\beta_2\rangle} = \langle\beta_1,\beta_1\rangle - \frac{\langle\beta_1,\beta_2\rangle\langle\beta_2,\beta_1\rangle}{\langle\beta_2,\beta_2\rangle}$$

As this expression is not symmetric in β_1 and β_2 we can take the curves to be of unit length with $\tilde{\beta}_j = \frac{\beta_j}{||\beta_j||}$, $j = 1,2$ with $||\beta_j|| = \sqrt{\langle\beta_j,\beta_j\rangle}$, so that $\langle\tilde{\beta}_1,\tilde{\beta}_1\rangle = \langle\tilde{\beta}_2,\tilde{\beta}_2\rangle = 1$ and obtain a suitable measure of distance:

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

References

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

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

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

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.