MOX–Report No. 07/2010

Functional clustering and alignment methods with applications

Laura M. Sangalli, Piercesare Secchi,
Simone Vantini, Valeria Vitelli

MOX, Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox@mate.polimi.it http://mox.polimi.it

Functional clustering and alignment methods with applications

Laura M. Sangalli, Piercesare Secchi, Simone Vantini, Valeria Vitelli

April 27, 2010

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
laura.sangalli@polimi.it, piercesare.secchi@polimi.it,
simone.vantini@polimi.it, valeria.vitelli@mail.polimi.it

Keywords: Functional data analysis, curve alignment, curve clustering, k-mean algorithm, k-medoid algorithm.

AMS Subject Classification: 62H30, 68T10

Abstract

We consider the issue of classification of functional data and, in particular, we deal with the problem of curve clustering when curves are misaligned. In the proposed setting, we aim at jointly aligning and clustering the curves, via the solution of an optimization problem. We describe an iterative procedure for the solution of the optimization problem, and we detail two alternative specifications of the procedure, a k-mean version and a k-medoid version. We illustrate via applications to real data the robustness of the alignment and clustering procedure under the different specifications.

1 Introduction

Unsupervised classification (or clustering) methods are an important topic in statistics, with many applications in various fields; the aim of such techniques is to classify a sample of data into homogeneous groups, without having any a priori knowledge about the true underlying clustering structure. Here, we shall consider the problem of *clustering of functional data*, and in particular of curves. For an introduction to the statistical analysis of functional data, see the books by Ramsay and Silverman [1] and by Ferraty and Vieu [2].

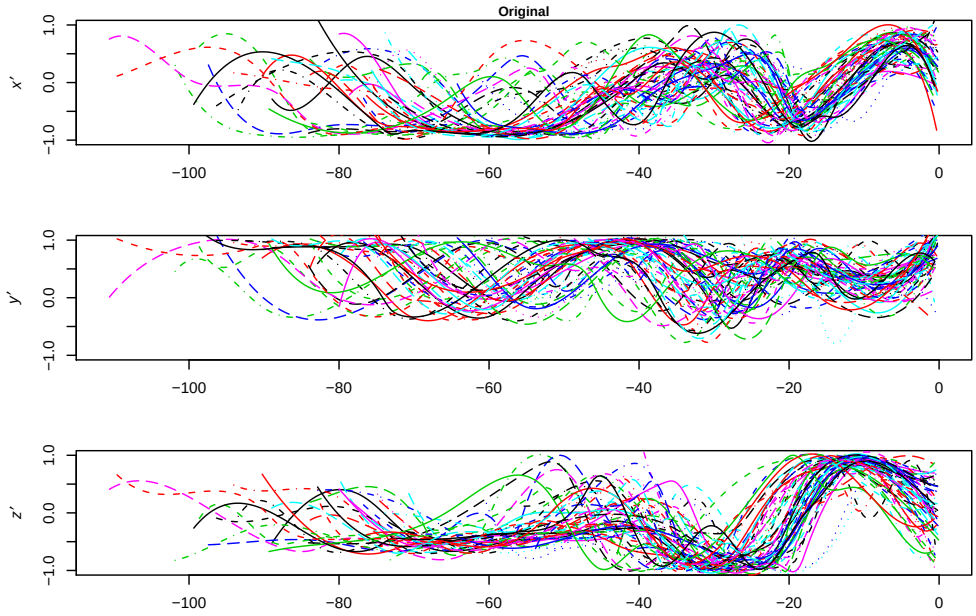

Figure 1: Estimated first derivatives, x', y', z', of the three spatial coordinates of the 65 ICA centerlines.

Many methods for curve clustering have been proposed in the literature on functional data analysis. For example, Shimizu and Mizuta [3], Tarpey and Kinadeter [4] and Tokushige et al. [5] propose a generalization of k-mean clustering algorithms for functional data, as a way to solve the problem of principal points estimation. In Cuesta-Albertos and Fraiman [6], a robust k-mean clustering procedure is developed, based on the idea of "impartial trimming", which proves to be useful for high dimensional data. Another k-mean algorithm for functional data can be found in Chiou and Li [7], where the efficiency of the clustering procedure is improved thanks to the use of a non-parametric random-effect model.

When dealing with clustering of curves, we need to consider a problem which is peculiar to functional data, namely the possible *misalignment* of the data. An instance of this issue is given by the data in Figure 1, which are related to the three spatial coordinates of 65 Internal Carotid Artery (ICA) centerlines (the picture in particular shows the estimated first derivatives of the three-dimensional centerlines). Clustering of these data is of interest for the identification of ICA's with different morphological shapes. On the other hand, the evident misalignment of the data acts as a confounding factor when trying to cluster the data, and the above cited k-mean algorithms fail to give efficient results (see [8]). This highlights the need for a clustering procedure which is able to jointly deal with data alignment, decoupling the variability due to data misalignment (phase variability) and the variability due to the shape (amplitude variability).

The problem of curve alignment (or curve registration) has been considered by a number of authors. Lawton et al. [9] and Altman and Villarreal [10] face

this problem using self-modelling non-linear regression methods, Lindstrom and Bates [11] develop non-linear mixed-effects models, and Ke and Wang [12] merge the above approaches in the unifying framework of semiparametric non-linear mixed-effects models. A different line of research, advocated by J. O. Ramsay, is followed by Ramsay and Li [13], Ramsay and Silverman [1], James [14], Kaziska and Srivastava [15] and Sangalli et al. [16], who define suitable similarity indexes between curves and thus align the curves, maximizing their similarities by means of a Procrustes procedure.

Following the latter line of research, in Sangalli et al. [8] we proposed a procedure which is able to jointly cluster and align a set of functional data. We stated the problem of joint clustering and alignment of functional data as an optimization problem, and we proposed an iterative procedure for its solution. This procedure was thus specified in a k-mean algorithm. Here, we describe an alternative specification of the procedure, in a k-medoid algorithm version. This new version approximates more directly the original optimization problem, and is potentially less sensitive to the presence of anomalous data. Moreover, this alternative version gives us the possibility of testing the robustness of our alignment and clustering procedure under different algorithm specifications. See Boudaoud et al. [17], Liu and Yang [18] and Liu and Muller [19] for other recent approaches to the problem of clustering of misaligned functional data.

The paper is organized as follows. In Section 2 we introduce a proper framework for the problem of clustering and alignment, defining phase and amplitude variability. In Section 3 we state the problem of curve clustering and alignment as an optimization problem, and we propose an iterative procedure for its solution. In Section 4 we describe two approaches to the template identification step in the procedure proposed in Section 3, obtaining a k-mean and a k-medoid specification of the iterative procedure. The two subsequent sections are devoted to the application of the two algorithm versions to real data, with the aim of testing the robustness of the proposed clustering and alignment procedure; in particular, Section 5 describes the clustering and alignment analysis of the data shown in Figure 1, concerning three-dimensional vascular geometries, while in Section 6 the procedure is tested on a benchmark dataset, the Berkeley Growth Study dataset. Some concluding considerations are drawn in Section 7. All analyses of real datasets are performed in R [20].

2 Defining phase and amplitude variabilities

The variability among two or more curves can be though of as having two components: *phase variability* and *amplitude variability*. Heuristically, phase variability is the one that can be eliminated by suitably aligning the curves, and amplitude variability is the one that remains among the curves once they have been aligned. Consider a set $\mathcal{C}$ of curves $\mathbf{c}(s) \colon \mathbb{R} \to \mathbb{R}^d$. Aligning $\mathbf{c}_1 \in \mathcal{C}$ to $\mathbf{c}_2 \in \mathcal{C}$ means finding a warping function $h(s) \colon \mathbb{R} \to \mathbb{R}$, of the abscissa parameter s, such that the

two curves $\mathbf{c}_1 \circ h$ and $\mathbf{c}_2$ are the most similar (with $(\mathbf{c} \circ h)(s) := \mathbf{c}(h(s))$). It is thus necessary to specify a similarity index $\rho(\cdot,\cdot) : \mathcal{C} \times \mathcal{C} \to \mathbb{R}$ that measures the similarity between two curves, and a class W of warping functions h (such that $\mathbf{c} \circ h \in \mathcal{C}$, for all $\mathbf{c} \in \mathcal{C}$ and $h \in W$) indicating the allowed transformations for the abscissa. Aligning $\mathbf{c}_1$ to $\mathbf{c}_2$, according to (ρ, W), means finding $h^* \in W$ that maximizes $\rho(\mathbf{c}_1 \circ h, \mathbf{c}_2)$. This procedure decouples phase and amplitude variability without loss of information: phase variability is captured by the optimal warping function h^*, whilst amplitude variability is the remaining variability between $\mathbf{c}_1 \circ h^*$ and $\mathbf{c}_2$. Note that the choice of the couple (ρ, W) defines what is meant by phase variability and amplitude variability.

Many similarity indexes for measuring similarity between functions have been considered in the literature on functional data analysis; for a proficient mathematical introduction to the issue see the book by Ferraty and Vieu [2]. Sangalli et al. [16, 8] proposed the following bounded similarity index between two curves $\mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$, where $\mathcal{C} = \{\mathbf{c} : \mathbf{c} \in L^2(\mathbb{R}; \mathbb{R}^d), \mathbf{c}' \in L^2(\mathbb{R}; \mathbb{R}^d), \mathbf{c}' \neq 0\}$

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = \frac{1}{d} \sum_{p=1}^{d} \frac{\int_{\mathbb{R}} c'_{1p}(s) c'_{2p}(s) ds}{\sqrt{\int_{\mathbb{R}} c'_{1p}(s)^2 ds} \sqrt{\int_{\mathbb{R}} c'_{2p}(s)^2 ds}}, \tag{1}$$

with c_{ip} indicating the pth component of $\mathbf{c}_i$, $\mathbf{c}_i = \{c_{i1}, \ldots, c_{id}\}$; geometrically, (1) represents the average of the cosines of the angles between the derivatives of homologous components of $\mathbf{c}_1$ and $\mathbf{c}_2$. The two curves are said to be similar when the index assumes its maximal value 1; for the similarity index defined in (1), this happens when the two curves are identical except for shifts and dilations of their components

$$\rho(\mathbf{c}_1, \mathbf{c}_2) = 1 \quad \Leftrightarrow \quad \begin{array}{l} \text{for } p = 1, \ldots, d, \ \exists \theta_{0p} \in \mathbb{R}, \theta_{1p} \in \mathbb{R}^+ : \\ c_{1p}(s) = \theta_{0p} + \theta_{1p} c_{2p}(s). \end{array} \tag{2}$$

The choice of this similarity index comes along with the following choice for the class W of warping functions of the abscissa

$$W = \{h : h(s) = ms + q \ \text{ with } \ m \in \mathbb{R}^+, q \in \mathbb{R}\} \tag{3}$$

i.e., the group of strictly increasing affine transformations.

The couple (ρ, W) defined in (1) and (3) satisfies the following properties (a)-(c) that we deem to be minimal requirements for coherence:

(a) The similarity index ρ is bounded, with maximum value equal to 1. Moreover, ρ is

reflexive: $\rho(\mathbf{c}, \mathbf{c}) = 1, \ \forall \ \mathbf{c} \in \mathcal{C}$;
symmetric: $\rho(\mathbf{c}_1, \mathbf{c}_2) = \rho(\mathbf{c}_2, \mathbf{c}_1), \ \forall \ \mathbf{c}_1, \mathbf{c}_2 \in \mathcal{C}$;
transitive: $\big[\rho(\mathbf{c}_1, \mathbf{c}_2) = 1 \wedge \rho(\mathbf{c}_2, \mathbf{c}_3) = 1\big] \Rightarrow \rho(\mathbf{c}_1, \mathbf{c}_3) = 1$
$\forall \ \mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3 \in \mathcal{C}$.

(b) The class of warping functions W is a convex vector space and has a group structure with respect to function composition $\circ$.

(c) The index ρ and the class W are consistent in the sense that, if two curves $\mathbf{c}_1$ and $\mathbf{c}_2$ are simultaneously warped along the same warping function $h \in W$, their similarity does not change

$$\rho\left(\mathbf{c}_1, \mathbf{c}_2\right) = \rho\left(\mathbf{c}_1 \circ h, \mathbf{c}_2 \circ h\right), \quad \forall\ h \in W. \tag{4}$$

This guarantees that it is not possible to obtain a fictitious increment of the similarity between two curves $\mathbf{c}_1$ and $\mathbf{c}_2$ by simply warping them simultaneously to $\mathbf{c}_1 \circ h$ and $\mathbf{c}_2 \circ h$.

Together, (b) and (c) imply the following property

(d) For all h_1 and $h_2 \in W$,

$$\rho\left(\mathbf{c}_1 \circ h_1, \mathbf{c}_2 \circ h_2\right) = \rho\left(\mathbf{c}_1 \circ h_1 \circ h_2^{-1}, \mathbf{c}_2\right) = \rho\left(\mathbf{c}_1, \mathbf{c}_2 \circ h_2 \circ h_1^{-1}\right).$$

This means that a change in similarity between $\mathbf{c}_1$ and $\mathbf{c}_2$ obtained by warping simultaneously $\mathbf{c}_1$ and $\mathbf{c}_2$, can also be obtained by warping the sole $\mathbf{c}_1$ or the sole $\mathbf{c}_2$.

Moreover, the couple (ρ, W) defined in (1) and (3) satisfies the additional auxiliary property

(e) Let W^d be the set of all transformations $\mathbf{r} : \mathbb{R}^d \longrightarrow \mathbb{R}^d$ such that

$$\mathbf{x} = (x_1, ..., x_d) \in \mathbb{R}^d \quad \longmapsto \quad \mathbf{r}(\mathbf{x}) = (r_1(x_1), \ldots, r_d(x_d)) \in \mathbb{R}^d,$$

with $r_1, \ldots, r_d \in W$. Then, for all $\mathbf{r}_1$ and $\mathbf{r}_2 \in W^d$,

$$\rho(\mathbf{r}_1(\mathbf{c}_1), \mathbf{r}_2(\mathbf{c}_2)) = \rho(\mathbf{c}_1, \mathbf{c}_2)\ .$$

In words, the similarity index between two curves is unaffected by strictly increasing affine transformations of one or more components of the curves.

For $\mathbf{c} = \{c_1, \ldots, c_d\} \in \mathcal{C}$, assume the existence of $\boldsymbol{\varphi} = \{\varphi_1, \ldots, \varphi_d\} \in \mathcal{C}$ and of a parameter vector $\boldsymbol{\theta} = (\theta_{01}, \ldots, \theta_{0d}, \theta_{11}, \ldots, \theta_{1d}, \theta_2, \theta_3)$, with $\theta_{0p} \in \mathbb{R}$ and $\theta_{1p} \in \mathbb{R}^+$ for $p = 1, \ldots, d$, $\theta_2 \in \mathbb{R}$, $\theta_3 \in \mathbb{R}^+$, such that

$$c_p(s) = \theta_{0p} + \theta_{1p}\varphi_p(\theta_2 + \theta_3 s) \qquad \text{for } p = 1, \ldots, d. \tag{5}$$

We shall write $\mathbf{c} \in \mathrm{SIM}(\boldsymbol{\varphi})$, since the condition (5) means that $\mathbf{c}$ falls within a *shape invariant model* (SIM), with *characteristic shape curve* $\boldsymbol{\varphi}$. For $d = 1$, SIM models were introduced by Lawton et al. [9]. For further details, see Kneip and

Gasser [21]. SIM models are strongly connected with the couple (ρ, W) defined in (1) and (3). Indeed,

$$\exists h \in W : \rho(\mathbf{c} \circ h, \boldsymbol{\varphi}) = 1 \quad \Leftrightarrow \quad \mathbf{c} \in \mathrm{SIM}(\boldsymbol{\varphi}); \tag{6}$$

this follows directly from (2) and (3). Note that, thanks to property (d), the roles of $\mathbf{c}$ and $\boldsymbol{\varphi}$ can be swapped. Now, consider a set of N curves $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\} \subset \mathcal{C}$, such that $\mathbf{c}_i \in \mathrm{SIM}(\boldsymbol{\varphi})$ for $i = 1, \ldots, N$; then, the following property follows immediately:

(f) For all $\mathbf{c}_i, \mathbf{c}_j$, with $i, j = 1, \ldots, N$, $\exists h_i \in W,\ h_j \in W$ such that

$$\begin{aligned} \rho\left(\mathbf{c}_i \circ h_i, \mathbf{c}_j \circ h_j\right) &= \rho\left(\mathbf{c}_i \circ h_i, \boldsymbol{\varphi}\right) = \\ &= \rho\left(\mathbf{c}_j \circ h_j, \boldsymbol{\varphi}\right) = 1 \quad \forall\ i, j = 1, \ldots, N. \end{aligned}$$

3 Curve clustering when curves are misaligned

Consider the problem of clustering and aligning a set of N curves $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$ with respect to a set of k template curves $\underline{\boldsymbol{\varphi}} = \{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k\}$ (with $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\} \subset \mathcal{C}$ and $\underline{\boldsymbol{\varphi}} \subset \mathcal{C}$). For each template curve $\boldsymbol{\varphi}_j$ in $\underline{\boldsymbol{\varphi}}$, define the domain of attraction

$$\Delta_j(\underline{\boldsymbol{\varphi}}) = \{\mathbf{c} \in \mathcal{C} : \sup_{h \in W} \rho(\boldsymbol{\varphi}_j, \mathbf{c} \circ h) \geq \sup_{h \in W} \rho(\boldsymbol{\varphi}_r, \mathbf{c} \circ h), \forall\ r \neq j\}, \quad j = 1, \ldots, k. \tag{7}$$

Moreover, define the labeling function

$$\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}) = \min\{r : \mathbf{c} \in \Delta_r(\underline{\boldsymbol{\varphi}})\}. \tag{8}$$

Note that $\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}) = j$ means that the similarity index obtained by aligning $\mathbf{c}$ to $\boldsymbol{\varphi}_j$ is at least as large as the similarity index obtained by aligning $\mathbf{c}$ to any other template $\boldsymbol{\varphi}_r$, with $r \neq j$. Thus $\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}_i)}$ indicates a template the curve $\mathbf{c}$ can be best aligned to and hence $\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c})$ a cluster which $\mathbf{c}$ should be assigned to.

Case of known templates. If the k templates $\underline{\boldsymbol{\varphi}} = \{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k\}$ were known, then clustering and aligning the set of N curves $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$ with respect to $\underline{\boldsymbol{\varphi}}$ would simply mean to assign $\mathbf{c}_i$ to the cluster $\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}_i)$ and align it to the corresponding template $\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}_i)}$, for $i = 1, \ldots, N$.

Here we are interested in the more complex case when the k templates are unknown.

Case of unknown templates. Ideally, if our aim is clustering and aligning the set of N curves $\{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$ with respect to k unknown templates, we should first solve the following optimization problem

(i) find $\underline{\boldsymbol{\varphi}} = \{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k\} \subset \mathcal{C}$ and $\underline{\mathbf{h}} = \{h_1, \ldots, h_N\} \subset W$ such that

$$\frac{1}{N}\sum_{i=1}^{N} \rho(\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}},\mathbf{c}_i)}, \mathbf{c}_i \circ h_i) \geq \frac{1}{N}\sum_{i=1}^{N} \rho(\boldsymbol{\psi}_{\lambda(\underline{\boldsymbol{\psi}},\mathbf{c}_i)}, \mathbf{c}_i \circ g_i), \tag{9}$$

for any other set of k templates $\underline{\boldsymbol{\psi}} = \{\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k\} \subset \mathcal{C}$ and any other set of N warping functions $\underline{\mathbf{g}} = \{g_1, \ldots, g_N\} \subset W$,

and then, for $i = 1, \ldots, N$,

(ii) assign $\mathbf{c}_i$ to the cluster $\lambda(\underline{\boldsymbol{\varphi}}, \mathbf{c}_i)$ and warp $\mathbf{c}_i$ along h_i.

The optimization problem (i) describes a search both for the set of optimal k templates, and for the set of optimal N warping functions. Note that the solution $(\underline{\boldsymbol{\varphi}}, \underline{\mathbf{h}})$ to (i) has mean similarity $\frac{1}{N}\sum_{i=1}^{N} \rho(\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}},\mathbf{c}_i)}, \mathbf{c}_i \circ h_i)$ equal to 1 if and only if it is possible to perfectly align and cluster in k groups the set of N curves, i.e. if and only if there exists $\underline{\mathbf{h}} = \{h_1, \ldots, h_N\} \subset W$ and a partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ of $\{1, \ldots, N\}$ in k elements, such that $\rho(\mathbf{c}_i \circ h_i, \mathbf{c}_j \circ h_j) = 1$ for all i and j belonging to the same element of $\mathcal{P}$. Because of (6), this is equivalent to the existence of k characteristic shape curves, $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k$, such that

$$\forall\ i = 1, \ldots, N, \quad \exists\ l_i \in \{1, \ldots, k\} : \mathbf{c}_i \in \mathrm{SIM}(\boldsymbol{\varphi}_{l_i}). \tag{10}$$

In this case the optimization problem (i) is solved by setting $\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}},\mathbf{c}_i)} \equiv \boldsymbol{\varphi}_{l_i}$.

It should also be noted that, thanks to property (c), if $\{\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_k\}$ and $\{h_1, \ldots, h_N\}$ provide a solution to (i), then also $\{\boldsymbol{\varphi}_1 \circ g_1, \ldots, \boldsymbol{\varphi}_k \circ g_k\}$ and $\{h_1 \circ g_{\lambda(\underline{\boldsymbol{\varphi}},\mathbf{c}_1)}, \ldots, h_N \circ g_{\lambda(\underline{\boldsymbol{\varphi}},\mathbf{c}_N)}\}$ is a solution to (i), for any $\{g_1, \ldots, g_k\} \subset W$. Moreover, this solution identifies the same clusters (i.e., is associated to the same partition $\mathcal{P} = \{P_1, \ldots, P_k\}$ of $\{1, \ldots, N\}$).

The optimization problem (i) is not analytically solvable in its complete generality. For this reason, in [8] we proposed to simultaneously deal with (i) and (ii) via an iterative clustering and alignment algorithm, which alternates *template identification steps* and *assignment and alignment steps*. In the template identification step we estimate the set of k templates associated to the k clusters identified at the previous assignment and alignment step; in the assignment and alignment step, we align the N curves to the set of the k templates obtained in the previous template identification step, and we assign each of the curves to one of the k clusters. As we shall see, the proposed clustering and alignment procedure also considers the problem of non-uniqueness of the solution, by targeting a specific solution via a *normalization step*.

3.1 Clustering and alignment iterative procedure

Let $\underline{\boldsymbol{\varphi}}^{[q-1]} = \{\boldsymbol{\varphi}_1{}^{[q-1]}, \ldots, \boldsymbol{\varphi}_k{}^{[q-1]}\}$ be the set of templates after iteration $q-1$, and $\{\mathbf{c}_1{}^{[q-1]}, \ldots, \mathbf{c}_N{}^{[q-1]}\}$ be the N curves aligned and clustered to $\underline{\boldsymbol{\varphi}}^{[q-1]}$. At the qth iteration the algorithm performs the following steps.

Template identification step. For $j = 1, \ldots, k$, the template of the jth cluster, $\boldsymbol{\varphi}_{j\,[q]}$, is estimated using all curves assigned to cluster j at iteration $q-1$, i.e. all curves $\mathbf{c}_{i\,[q-1]}$ such that $\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]}, \mathbf{c}_{i\,[q-1]}) = j$. Ideally, the template $\boldsymbol{\varphi}_{j\,[q]}$ should be estimated as the curve $\boldsymbol{\varphi} \in \mathcal{C}$ that maximizes the within-cluster total similarity

$$\sum_{i:\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]}, \mathbf{c}_{i\,[q-1]})=j} \rho(\boldsymbol{\varphi}, \mathbf{c}_{i\,[q-1]}), \tag{11}$$

i.e., $\boldsymbol{\varphi}_{j\,[q]}$ should be the functional median, or Fréchet median, associated to the similarity ρ.

Assignment and alignment step. The set of curves $\{\mathbf{c}_{1\,[q-1]}, \ldots, \mathbf{c}_{N\,[q-1]}\}$ is clustered and aligned to the set of templates $\underline{\boldsymbol{\varphi}}_{[q]} = \{\boldsymbol{\varphi}_{1\,[q]}, \ldots, \boldsymbol{\varphi}_{k\,[q]}\}$: for $i = 1, \ldots, N$, the i-th curve $\mathbf{c}_{i\,[q-1]}$ is aligned to $\boldsymbol{\varphi}_{\lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \mathbf{c}_{i\,[q-1]})}$ and the aligned curve $\tilde{\mathbf{c}}_{i\,[q]} = \mathbf{c}_{i\,[q-1]} \circ h_{i\,[q]}$ is assigned to cluster $\lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \mathbf{c}_{i\,[q-1]}) \equiv \lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \tilde{\mathbf{c}}_{i\,[q]})$.

Normalization step. After each assignment and alignment step, we also perform a normalization step. In detail, for $j = 1, \ldots, k$, all the $N_{j\,[q]}$ curves $\tilde{\mathbf{c}}_{i\,[q]}$ assigned to cluster j are warped along the warping function $(\bar{h}_{j\,[q]})^{-1}$, where

$$\bar{h}_{j\,[q]} = \frac{1}{N_{j\,[q]}} \sum_{i:\lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \tilde{\mathbf{c}}_{i\,[q]})=j} h_{i\,[q]} \tag{12}$$

obtaining $\mathbf{c}_{i\,[q]} = \tilde{\mathbf{c}}_{i\,[q]} \circ (\bar{h}_{j\,[q]})^{-1} = \mathbf{c}_{i\,[q-1]} \circ h_{i\,[q]} \circ (\bar{h}_{j\,[q]})^{-1}$. In this way, at each iteration, the average warping undergone by curves assigned to cluster j is the identity transformation $h(s) = s$. Indeed:

$$\frac{1}{N_{j\,[q]}} \sum_{i:\lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \mathbf{c}_{i\,[q]})=j} \left(h_{i\,[q]} \circ (\bar{h}_{j\,[q]})^{-1}\right)(s) = s, \qquad j = 1, \ldots, k. \tag{13}$$

The normalization step is thus used to select, among all candidate solutions to the optimization problem, the one that leaves the average locations of the clusters unchanged, thus avoiding the drifting apart of clusters or the global drifting of the overall set of curves. Note that the normalization step preserves the clustering structure chosen in the maximization step, i.e., $\lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \tilde{\mathbf{c}}_{i\,[q]}) = \lambda(\underline{\boldsymbol{\varphi}}_{[q]}, \mathbf{c}_{i\,[q]})$ for all i.

The algorithm is initialized with a set of initial templates, $\underline{\boldsymbol{\varphi}}_{[0]} = \{\boldsymbol{\varphi}_{1\,[0]}, \ldots, \boldsymbol{\varphi}_{k\,[0]}\} \subset \mathcal{C}$, and with $\{\mathbf{c}_{1\,[0]}, \ldots, \mathbf{c}_{N\,[0]}\} = \{\mathbf{c}_1, \ldots, \mathbf{c}_N\}$, and stopped when, in the assignment and alignment step, the increments of the similarity indexes are all lower than a fixed threshold.

4 Template identification

Whilst the assignment and alignment step and the normalization step are straightforward, the template identification step is more troublesome, since identification of the template $\boldsymbol{\varphi}_{j\,[q]}$, as the curve $\boldsymbol{\varphi} \in \mathcal{C}$ that maximizes the total similarity (11), cannot be easily dealt with. For this reason, in [8] we proposed to estimate the template $\boldsymbol{\varphi}_{j\,[q]}$ as a loess, with Gaussian kernel and appropriate smoothness parameter, of the curves assigned to cluster j at iteration $q-1$ (i.e. all curves $\mathbf{c}_{i\,[q-1]}$ such that $\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]}, \mathbf{c}_{i\,[q-1]}) = j$). See [8] for details on the implementation. The algorithm obtained with this specification for the template identification step was named *k-mean alignment*, in analogy with the k-mean clustering algorithms for multivariate and for functional data.

On the other hand, the fact that the template is estimated by loess of the curves assigned to the cluster, instead of the curve that maximize the total similarity (11), raises doubts about a possible distortion of the algorithm. Moreover, estimating the template by loess of the curves assigned to the cluster might make this step sensitive to the presence of anomalous data.

For this reason, we propose here an alternative specification of the template identification step, that constitutes a direct approximation to the maximization of the total similarity (11). In particular, we restrict the set over which the maximization is carried out, limiting the search to functions of the sample. The template $\boldsymbol{\varphi}_{j\,[q]}$ is thus estimated as the curve $\boldsymbol{\varphi}$, among all curves assigned to cluster j at iteration $q-1$ (i.e. all curves $\mathbf{c}_{i\,[q-1]}$ such that $\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]}, \mathbf{c}_{i\,[q-1]}) = j$), that maximizes the total similarity (11)

$$\boldsymbol{\varphi}_{j\,[q]} = \underset{\mathbf{c}_{i\,[q-1]}:\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]},\mathbf{c}_{i\,[q-1]})=j}{\arg\max} \sum_{k:\lambda(\underline{\boldsymbol{\varphi}}_{[q-1]},\mathbf{c}_{k\,[q-1]})=j} \rho\big(\mathbf{c}_{i\,[q-1]}, \mathbf{c}_{k\,[q-1]}\big).$$

Note that the k curves selected as templates, in the template identification step, shall skip the subsequent assignment and alignment step and normalization step.

The algorithm obtained with this specification for the template identification step will be named *k-medoid alignment*, in analogy with the k-medoid clustering algorithms described for instance by [22] and [23]. The same analogy suggests that this novel specification of the template identification step is robust to the presence of anomalous data. In particular, the comparison of the clustering results obtained with the two alternative algorithm specifications, the k-mean and the k-medoid version, might indicate accidental anomalous data, as we shall see in the applications to real data described in the following sections. Moreover, this alternative version will also give us the possibility of testing the robustness of our alignment and clustering procedure under different algorithm specifications.

5 An application to the analysis of three-dimensional cerebral vascular geometries

In this section, k-mean and k-medoid alignment are used to improve upon the exploratory statistical analyses of the AneuRisk Project[1] dataset (previous analyses are detailed in [16, 24, 8]). The AneuRisk Project is a joint research program that aims at evaluating the role of vascular geometry and hemodynamics in the pathogenesis of cerebral aneurysms. The data considered in the analysis here presented are the three spatial coordinates (in mm) of 65 Internal Carotid Artery (ICA) centerlines, measured on a fine grid of points along a curvilinear abscissa (in mm), decreasing from the terminal bifurcation of the ICA towards the heart. Estimates of these three-dimensional curves are obtained by means of three-dimensional free-knot regression splines, as described in Sangalli et al. [25]. Details about the elicitation of discrete observations from row data can be found in Antiga et al. [26]. Figure 1 displays the first derivatives, x', y' and z', of the estimated centerlines. We are interested in clustering the three-dimensional centerlines, with the aim of identifying ICA's with different morphological shapes. Since the shape of the ICA influences the pathogenesis of cerebral aneurysms through its effects on the hemodynamics, such a classification could in fact be helpful in the determination of the risk level of a given patient.

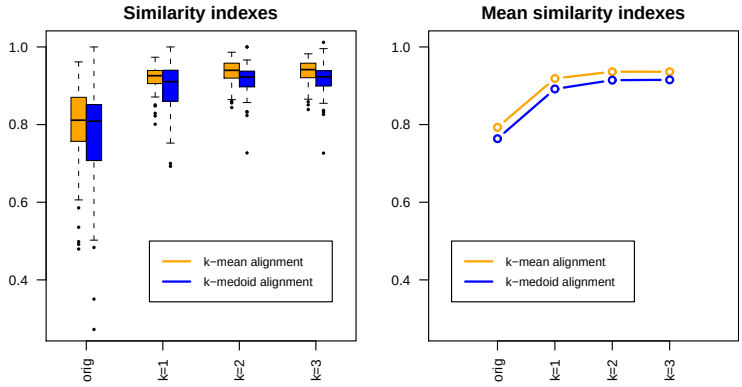

Figure 2: Left: in orange (blue), boxplots of similarity indexes between the original centerlines and their mean (medoid) curve, "orig", and boxplots of similarity indexes between the k-mean aligned (k-medoid aligned) centerlines and their estimated templates, for $k = 1, 2, 3$. Right: corresponding means of similarity indexes.

In the medical literature (see, e.g., [27]) ICA's are classified in Γ-shaped, Ω-

[1]The project involves MOX Laboratory for Modeling and Scientific Computing (Dip. di Matematica, Politecnico di Milano), Laboratory of Biological Structure Mechanics (Dip. di Ingegneria Strutturale, Politecnico di Milano), Istituto Mario Negri (Ranica), Ospedale Niguarda Ca' Granda (Milano), and Ospedale Maggiore Policlinico (Milano), and is supported by Fondazione Politecnico di Milano and Siemens Medical Solutions Italia.

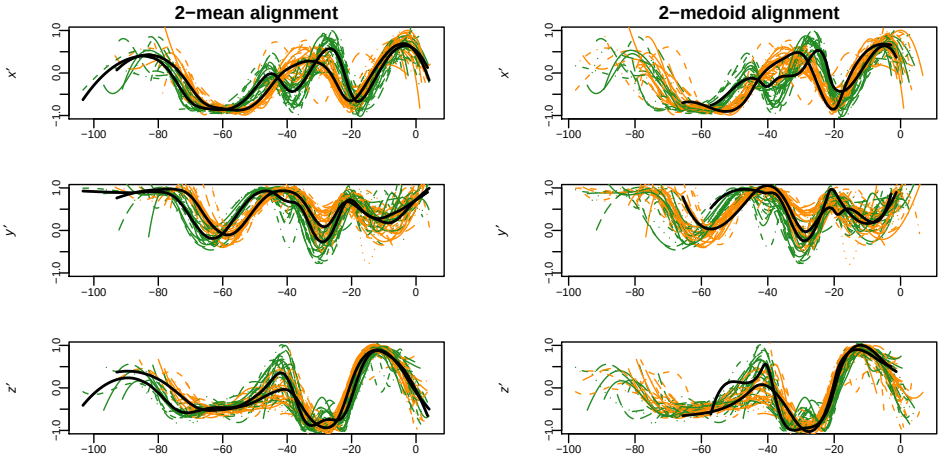

Figure 3: First derivatives of 2-mean (left) and 2-medoid (right) aligned centerlines, with superimposed first derivatives of estimated templates (black lines); the curve color depends on cluster assignment.

shaped, and S-shaped ICA's, according to the form of their distal part, which may resemble the letters Γ, Ω or S, in presence of zero, one, or two siphons, respectively. In [8], using k-mean alignment, we were able to identify a cluster of Ω-shaped ICA's and a cluster S-shaped ICA's, in the AneuRisk dataset. We want here to verify if this clustering result is confirmed by application of k-medoid alignment, and recognize accidental anomalous data.

Figure 2, left, shows in orange the boxplots of the similarity indexes between the original centerlines and their estimated loess (indicated in the figure as "orig"), and the boxplots of the similarity indexes between the k-mean aligned centerlines and the associated estimated templates, for $k = 1, 2, 3$; the same panel also displays in blue the boxplots of the similarity indexes between the original centerlines and their estimated medoid, and the boxplots of the similarity indexes obtained by k-medoid alignment, for $k = 1, 2, 3$. Figure 2, right, displays the corresponding means of the similarity indexes, in orange for k-mean alignment and in blue for k-medoid alignment. Note that both 1-mean alignment and 1-medoid alignment lead to a large increase in the similarities, with respect to the similarities of the original curves, but a further considerable gain can be obtained by setting $k = 2$ in the clustering and aligning procedure, both 2-mean and 2-medoid, whereas an eventual choice of $k = 3$ is not justified by an additional increase in the similarities. Thus, k-medoid alignment, likewise k-mean alignment, suggests the presence of $k = 2$ shape characteristic curves within the analyzed centerlines.

Figure 3 compares the first derivatives of 2-mean and 2-medoid aligned centerlines (left and right, respectively). The two clusters identified by 2-medoid alignment, similarly to the two clusters identified by 2-mean alignment, can be described as the Ω-shaped ICA's cluster (orange) and S-shaped ICA's cluster

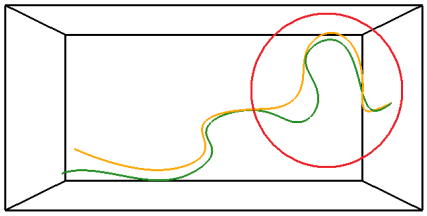

Figure 4: Three-dimensional image of the estimated templates found by 2-mean alignment of ICA centerlines. The template of the orange cluster is a prototype of Ω-shaped ICA (one siphon), whereas the one of the green cluster is a prototype of S-shaped ICA (two siphons).

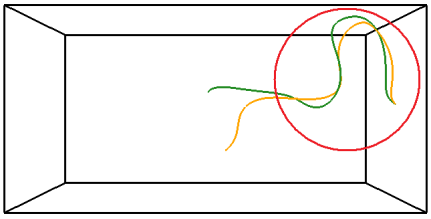

Figure 5: Three-dimensional image of the estimated templates found by 2-medoid alignment of ICA centerlines. The template of the orange cluster is a prototype of Ω-shaped ICA (one siphon in the distal part), whereas the one of the green cluster is a prototype of S-shaped ICA (two siphons in the distal part).

(green). This can be better appreciated in Figures 4 and 5 that give three-dimensional visualizations of the two estimated template curves, obtained by 2-mean and 2-medoid alignment, respectively. In both figures, in fact, the template of the orange cluster is a prototype of Ω-shaped ICA (one siphon in the distal part), whereas the template of the green cluster is a prototype of S-shaped ICA (two siphons in the distal part).

Finally, Table 1 compares the cluster assignments obtained by 2-mean and 2-medoid alignment ("Ω" stands for the cluster of Ω-shaped ICA's, and "S" for the one of S-shaped ICA's). This table shows that only 4 out of the 65 ICA's are differently clustered by the two algorithms. Two of these four data might in fact be interpreted as anomalous data, one because of the shortness of the observed centerline, and the other because of the morphological shape, that resembles a Γ-shaped ICA rather than a Ω or a S-shaped one.

Table 1: Comparison between cluster assignments obtained by 2-mean and 2-medoid alignment algorithms ("Ω" stands for the Ω-shaped cluster and "S" stands for the S-shaped one).

		2-medoid	
		Ω	S
2-mean	Ω	34	1
	S	3	27

6 An application to the analysis of growth data

In this section we apply k-mean and k-medoid alignment for the analysis of a benchmark data set in the functional data analysis literature: the 93 growth curves from Berkeley Growth Study (see Tuddenham and Snyder [28]). These data have been previously considered by a number of authors (see for example Ramsay and Li [13], Ramsay and Silverman [1], James [14], and references therein); in particular, in this paper we will improve upon the analysis illustrated in Sangalli et al. [8].

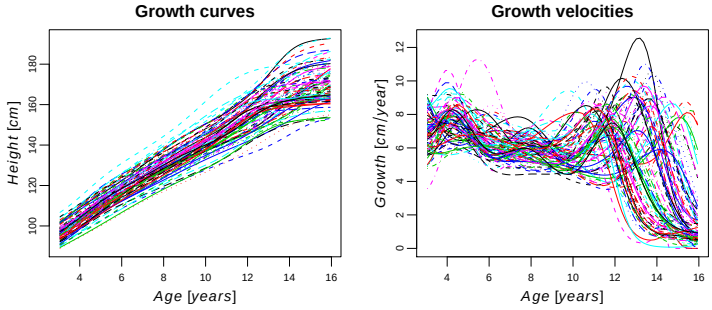

Figure 6: Growth curves of 93 children from Berkeley Growth Study data (left) and corresponding growth velocities (right).

The heights (in cm) of the 93 children in the data set are measured quarterly from 1 to 2 years, annually from 2 to 8 years and biannually from 8 to 18 years. The growth curves are estimated by means of monotonic cubic regression splines (see Ramsay and Silverman [1]), implemented using the R function `smooth.monotone` available in the `fda` package [29]. Figure 6 shows the estimated growth curves and their derivatives, the growth velocities. Looking at the growth velocities, it is apparent that the children follow a similar growth course, but that each child has a personal biological clock.

Figure 7 shows 1-mean and 2-mean aligned growth curves, the corresponding growth velocities and warping functions; Figure 8 shows the corresponding results obtained by k-medoid alignment. Figure 9, left, displays in orange (blue), the boxplots of the similarity indexes between the original growth curves and their mean (medoid) curve, indicated with "orig", and the boxplots of the similarity indexes between the k-mean aligned (k-medoid aligned) growth curves and their estimated templates, for $k = 1, 2, 3$. The right panel of Figure 9 displays the corresponding means of similarity indexes. From inspection of the similarity indexes, both k-mean and k-medoid alignment suggest the presence of just one characteristic curve, since the choice of $k = 2$ is not payed off by a reasonable gain in the similarities.

Since, out of the 93 children, 39 are boys and 54 are girls, we might wonder if the analysis points out some differences among them (notice that here we are not performing any supervised classification of boys and girls). Figure 10

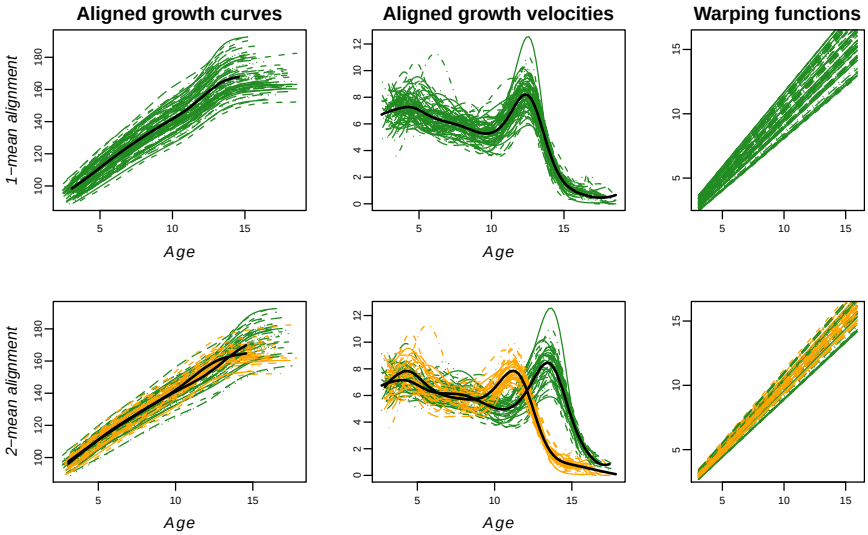

Figure 7: Results of k-mean alignment of growth curves, for k=1; 2: aligned growth curves (with superimposed estimated templates, in black) and corresponding growth velocities (with superimposed first derivatives of estimated templates, in black), together with warping functions. The colors of aligned curves and warping functions depend on cluster assignment.

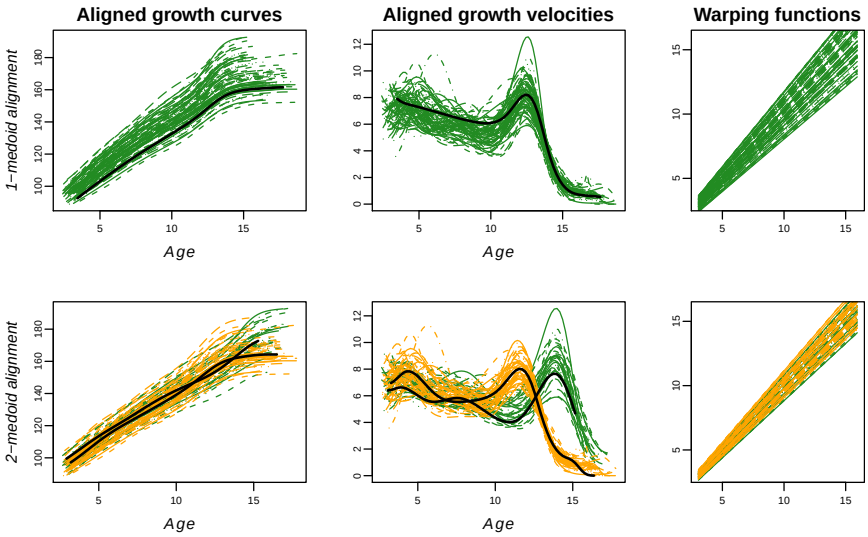

Figure 8: Results of k-medoid alignment of growth curves, for k=1; 2: aligned growth curves (with superimposed estimated templates, in black) and corresponding growth velocities (with superimposed first derivatives of estimated templates, in black), together with warping functions. The colors of aligned curves and warping functions depend on cluster assignment.

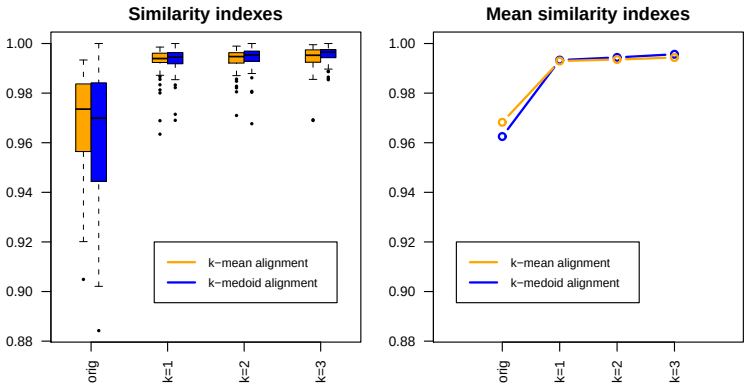

Figure 9: Left: in orange (blue), boxplots of similarity indexes between the original growth curves and their mean (medoid) curve, "orig", and boxplots of similarity indexes between the k-mean aligned (k-medoid aligned) growth curves and their estimated templates, for $k = 1, 2, 3$. Right: corresponding means of similarity indexes.

is obtained from Figure 8 (top panels) displaying in blue the 1-medoid aligned growth curves of boys, and the corresponding growth velocities and warping functions, and in pink the ones of girls. The warping functions show a pretty neat separation of boys and girls in the phase; this highlights that the biological clocks of boys and girls run at different speeds, and in particular that boys start to grow later, having warping functions with smaller intercepts, and grow slowler, having warping functions with smaller slopes. The left panel of Figure 10 also shows that, once the biological clocks of the children have been aligned, the height of boys stochastically dominates the one of girls for any registered biological age. Finally, boys seem also to have a more pronounced growth, especially during puberty, as highlighted by their more prominent growth velocity peak. All these features are in complete agreement with the results obtained by 1-mean alignment, and discussed in detail in [8].

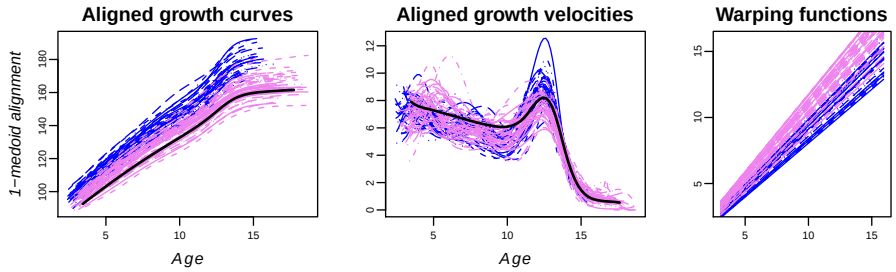

Figure 10: Figure obtained from Figure 8 (top panels) displaying in blue the growth curves, growth velocities and warping functions of boys and in pink the ones of girls.

Table 2: Left: results of 2-mean bivariate clustering of slopes and intercepts of the warping functions obtained by 1-mean alignment. Right: results of 2-medoid bivariate clustering of slopes and intercepts of the warping functions obtained by 1-medoid alignment. Cluster assignment vs gender.

2-mean		clusters	
		1	2
gender	F	44	10
	M	1	38

2-medoid		clusters	
		1	2
gender	F	43	11
	M	1	38

Table 3: A comparison between cluster assignments obtained by 2-mean and 2-medoid bivariate clustering of slopes and intercepts of the warping functions obtained by 1-mean alignment and 1-medoid alignment of growth curves.

		2-medoid	
		1	2
2-mean	1	44	1
	2	0	48

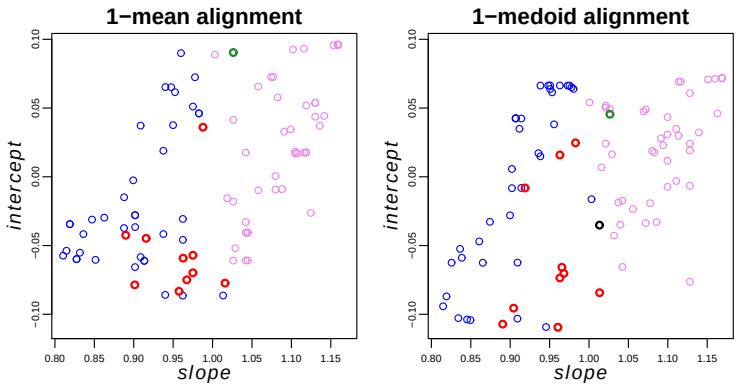

Figure 11: Slopes and intercepts of the warping functions resulting from 1-mean alignment (left) and 1-medoid alignment (right) of growth curves. Blue circles correspond to boys and pink circles to girls. Red (green) circles are girls (boys) which in both cases have been assigned to the male (female) prevalent cluster when the corresponding bivariate clustering procedure is applied. The black circle is the only mismatching case between the two procedures (see Table 3).

The grouping structure of the warping functions obtained by 1-mean and 1-medoid alignment of the growth curves, can be explored by a coherent unsupervised classification of their slopes and intercepts, i.e. by 2-mean and 2-medoid bivariate clustering respectively. The results of these unsupervised classifications are shown in Table 2. Note that both 2-mean clustering of the warping functions obtained by 1-mean alignment of growth curves, and 2-medoid clustering of the warping functions obtained by 1-medoid alignment of growth curves, assign 1 boy to the female prevalent cluster, and respectively 10 and 11 girls to the male prevalent cluster. From inspection of Table 3, which compares the cluster assignments of the two procedures, we conclude that this boy and the 10 girls are exactly the same. Thus, both algorithms agree that these ten girls have biological clocks closer to those of boys, and that the boy has a biological clock closer to those of girls. This fact is evident in Figure 11, which displays the slopes and intercepts of the warping functions (pink for girls and blue for boys), and highlights in red the ones of the ten girls and in green the one of the boy. Figure 11 also displays in black the only mismatch between the two algorithms, which has in fact a biological clock borderline between the two groups.

7 Discussion

We have considered the issue of classification of functional data and, in particular, we have dealt with the problem of curve clustering when curves are misaligned. We have described an alternative specification of the clustering and alignment procedure proposed in [8]. This novel version, named k-medoid alignment algorithm, approximates more directly the clustering and alignment optimization problem, and is potentially less sensitive to the presence of anomalous data. Moreover, this alternative version has given us the possibility of testing the robustness of our alignment and clustering procedure under different algorithm specifications.

References

[1] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*. Springer, 2005.

[2] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis*. Springer, 2006.

[3] N. Shimizu and M. Mizuta, Functional clustering and functional principal points, *LNAI*, no. 4693, pp. 501–508, 2007.

[4] T. Tarpey and K. K. J. Kinateder, Clustering functional data, *Journal of Classification*, no. 20, pp. 93–114, 2003.

[5] S. Tokushige, H. Yadohisa, and K. Inada, Crisp and fuzzy k-means clustering algorithms for multivariate functional data, *Computational Statistics*, no. 22, pp. 1–16, 2007.

[6] J. A. Cuesta-Albertos and R. Fraiman, Impartial trimmed k-means for functional data, *Computational Statistics and Data Analysis*, no. 51, pp. 4864–4877, 2007.

[7] J. M. Chiou and P. L. Li, Functional clustering and identifying substructures of longitudinal data, *J. R. Statist. Soc. Series B*, no. 69, pp. 679–699, 2007.

[8] L. M. Sangalli, P. Secchi, S. Vantini, and V. Vitelli, K-mean alignment for curve clustering, *Computational Statistics and Data Analysis*, 2010, doi:10.1016/j.csda.2009.12.008.

[9] W. H. Lawton, E. A. Sylvestre, and M. S. Maggio, Self modeling nonlinear regression, *Technometrics*, no. 14, pp. 513–532, 1972.

[10] N. S. Altman and J. C. Villarreal, Self-modelling regression for longitudinal data with time-invariant covariates, *Canad. J. Statist.*, no. 32, pp. 251–268, 2004.

[11] M. J. Lindstrom and D. M. Bates, Nonlinear mixed effects models for repeated measures data, *Biometrics*, no. 46, pp. 673–687, 1990.

[12] C. Ke and Y. Wang, Semiparametric nonlinear mixed-effects models and their applications, *J. Amer. Statist. Assoc.*, no. 96, pp. 1272–1298, 2001.

[13] J. O. Ramsay and X. Li, Curve registration, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, no. 60, pp. 351–363, 1998.

[14] G. M. James, Curve alignment by moments, *The Annals of Applied Statistics*, no. 1, pp. 480–501, 2007.

[15] D. Kaziska and A. Srivastava, Gait-based human recognition by classification of cyclostationary processes on nonlinear shape manifolds, *Journal of the American Statistical Association*, no. 102, pp. 1114–1128, 2007.

[16] L. M. Sangalli, P. Secchi, S. Vantini, and A.Veneziani, A case study in exploratory functional data analysis: Geometrical features of the internal carotid artery, *J. Amer. Statist. Assoc.*, vol. 104, pp. 37–48, 2009.

[17] S. Boudaoud, H. Rix, and O. Meste, Core shape modelling of a set of curves, *Computational Statistics and Data Analysis*, no. 54, pp. 308–325, 2010.

[18] X. Liu and M. C. K. Yang, Simultaneous curve registration and clustering for functional data, *Computational Statistics and Data Analysis*, no. 53, pp. 1361–1376, 2009.

[19] X. Liu and H. G. Muller, Modes and clustering for time-warped gene expression profile data, *Bioinformatics*, vol. 19, no. 15, pp. 1937–1944, 2003.

[20] R. D. C. Team, R: A language and environment for statistical computing, *R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0,* `http://www.R-project.org`, 2006.

[21] A. Kneip and T. Gasser, Convergence and consistency results for self-modeling nonlinear regression, *The Annals of Statistics*, vol. 16, no. 1, pp. 82–112, 1988.

[22] L. Kaufman and P. Rousseeuw, *Finding Groups in Data.* Wiley Series in Probability and Mathematical Statistics, 1990.

[23] T. Hastie, R. Tibshirani, and J.Friedman, *The Elements of Statistical Learning.* Springer, 2001.

[24] L. M. Sangalli, P. Secchi, and S. Vantini, Explorative functional data analisys for 3d-geometries of the inner carotid artery, in *Functional and Operatorial Statistics* (S. Dabo-Niang and F. Ferraty, eds.), pp. 289–296, Springer, Contributions to Statistics, 2008.

[25] L. M. Sangalli, P. Secchi, S. Vantini, and A.Veneziani, Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines, *Journal of the Royal Statistical Society Ser. C, Applied Statistics*, no. 58, pp. 285–306, 2009.

[26] L. Antiga, M. Piccinelli, L. Botti, B. Ene-Iordache, A. Remuzzi, and D. Steinman, An image-based modeling framework for patient-specific computational hemodynamics, *Medical and Biological Engineering and Computing*, pp. 1097–1112, 2008.

[27] H. Krayenbuehl, P. Huber, and M. G. Yasargil, Krayenbuhl/yasargil cerebral angiography, *Thieme Medical Publishers, 2nd ed.*, 1982.

[28] R. D. Tuddenham and M. M. Snyder, Physical growth of california boys and girls from birth to age 18, Tech. Rep. 1, University of California Publications in Child Development, 1954.

[29] J. O. Ramsay and H. Wickham, fda: Functional data analysis, *r package version 1.1.8*, 2007.

MOX Technical Reports, last issues

Dipartimento di Matematica "F. Brioschi",
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

07/2010 Laura M. Sangalli, Piercesare Secchi, Simone Vantini, Valeria Vitelli:
Functional clustering and alignment methods with applications

06/2010 Jordi Alastruey, Tiziano Passerini, Luca Formaggia, Joaquim Peiró:
The effect of visco-elasticity and other physical properties on aortic and cerebral pulse waveforms: an analytical and numerical study

05/2010 Matteo Longoni, A.C.I. Malossi, Alfio Quarteroni, Andrea Villa:
A complete model for non-Newtonian sedimentary basins in presence of faults and compaction phenomena

04/2010 Marco Discacciati, Paola Gervasio, Alfio Quarteroni:
Heterogeneous mathematical models in fluid dynamics and associated solution algorithms

03/2010 P.E. Farrell, Stefano Micheletti, Simona Perotto:
A recovery-based error estimator for anisotropic mesh adaptation in CFD

02/2010 Pietro Barbieri, Niccolo' Grieco, Francesca Ieva, Anna Maria Paganoni and Piercesare Secchi:
Exploitation, integration and statistical analysis of Public Health Database and STEMI archive in Lombardia Region

01/2010 G.M. Porta, S. Perotto, F. Ballio:
Anisotropic Mesh Adaptation Driven by a Recovery-Based Error Estimator for Shallow Water Flow Modeling

35/2009 C. D'Angelo, P. Zunino:
Robust numerical approximation of coupled Stokes and Darcy flows applied to vascular hemodynamics and biochemical transport

34/2009 P. F. Antonietti, P. Houston:
A Class of Domain Decomposition Preconditioners for hp-Discontinuous Galerkin Finite Element Methods

33/2009 S. Vantini:
On the Definition of Phase and Amplitude Variability in Functional Data Analysis