MODELLISTICA E CALCOLO SCIENTIFICO

MODELING AND SCIENTIFIC COMPUTING

MOX–Report No. 33/2009

On the Definition of Phase and Amplitude Variability in Functional Data Analysis

S. Vantini

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

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

On the Definition of Phase and Amplitude Variability in Functional Data Analysis

Simone Vantini

MOX– Modellistica e Calcolo Scientifico
Dipartimento di Matematica "F. Brioschi"
Politecnico di Milano
via Bonardi 9, 20133 Milano, Italy
simone.vantini@polimi.it

Keywords: registration, alignment, synchronization, warping.

Abstract

We introduce a mathematical framework in which a functional data registration problem can be soundly and coherently set. In detail, we show that the introduction of a metric/semi-metric and of a group of warping function respect to which the metric/semi-metric is invariant is the key to a clear and not ambiguous definition of phase and amplitude variability. Moreover, an amplitude-to-total variability index is proposed. This index turns to be useful in practical situations to measure to what extent amplitude variability affects the data and to compare the effectiveness of different registration methods.

1 Introduction

The problem of data registration is often encountered in recent statistical literature. All papers devoted to this issue can be parted in two groups: the ones positioned within the *longitudinal data analysis* and the ones positioned within the *functional data analysis*. The distinction between these two types of data is quite subtle and fuzzy: roughly speaking, we refer to longitudinal data for data that have low within-subject signal-to-noise ratio and small within-subject sample size and to functional data for data that have high within-subject signal-to-noise ratio and large within-subject sample size. See the discussion following Ke and Wang (2001) for a detailed account about this issue.

Some examples relevant to the former approach can be found in Lawton et al. (1972), Lindstrom and Bates (1990), Ke and Wang (2001), and Altman and Villarreal (2004). Instead some examples relevant to the latter approach can be found in Ramsay and Li (1998), Kneip et al. (2000), Ramsay and Silverman (2005), James (2007), Kaziska and Srivastava (2007), and Sangalli et al. (2009).

Works belonging to the former line of research stem from the very old and sound tradition of classical regression analysis and they have cleverly been put together in a unique theoretical framework by Ke and Wang (2001). Works belonging to the latter one do not rely on a very established tradition. Indeed they come within a very young branch of statistics: the so-called *Functional Data Analysis* (e.g Ramsay and Silverman 2005; Ferraty and Vieu 2006) or with a more general term *Object Oriented Data Analysis* (e.g. Wang and Marron 2007). Within this approach data are considered as realizations of random variables taking values in a ∞-dimensional functional space. The present work is within the latter line of research.

Data registration is often the first key to a successful functional data analysis. Although many successful methods have already been proposed in literature, a clear theoretical analysis about the soundness and the meaningfulness of the problem of functional data registration is still missing. This work aims at being one of the first attempts to put in a coherent mathematical framework this key problem of functional data analysis. Recently, also in Kneip and Ramsay (2008), some effort has been done in this direction. Even if the latter work and the present one differ by many issues (indeed, the driving idea of Kneip and Ramsay (2008) is the concept of amplitude convex space, while here the driving idea is the concept of phase equivalence classes), the basic assumptions enabling both approaches appear to be non-conflicting, hopefully leaving space for a possible future integration of the two.

The paper is structured as follows: in Section 2 the issue of registering a function with respect to another one is tackled. In detail: in Subsection 2.1 the necessary mathematical framework, that will be used also in the following sections, is introduced; in Subsection 2.2 the problem of registering a function with respect to another is declined in the introduced mathematical framework and an amplitude-to-total variability ratio α^2 is proposed; in Subsection 2.3 a discussion about the number of equivalence classes related to phase variability is undertaken. In Sections 3 and 4 the theory developed in Section 2 is generalized to the problem of registering a set of functions in presence and absence of a reference function, respectively. In Section 5 the theory is further extended to deal with the registration of functions when a semi-metric is used in place of a metric. In Section 6 a couple of real applications presented in literature are presented and discussed in the light of the theory here presented. Finally in

Section 7 the key points of the theory are synthetically pointed out and some hints for future research are proposed.

It must be clear to the reader that by means of this work we do not claim to cover all possible approaches to functional data registration; we rather aim at providing to statisticians a clear mind-set through which setting and comparing different registration methods.

2 Registration of a pair of functions

In this section we deal with the easiest case of functional data registration: this is the problem of registering two functions one with respect to the other one.

2.1 Mathematical framework

In order for a functional data registration problem to be meaningful and mathematically consistent according to our theory, some basic properties, of the set F which the functional data belong to, and of the set W of warping functions, are demanded:

a) $F = \{f : \Omega \subseteq \mathbb{R}^p \longrightarrow \Psi \subseteq \mathbb{R}^q\}$ is a metric space according to a metric $d : F \times F \longrightarrow \mathbb{R}_0^+$, i.e. $\forall\ f_1,\ f_2$, and $f_3 \in F$:

$$f_1 = f_2 \Leftrightarrow d(f_1, f_2) = 0\ , \quad (1a)$$

$$d(f_1, f_2) = d(f_2, f_1)\ , \quad (1b)$$

$$d(f_1, f_3) \leq d(f_1, f_2) + d(f_2, f_3)\ ; \quad (1c)$$

b) W is a subgroup - with respect to ordinary composition $\circ$ - of the group of the continuous automorphisms: $\Omega \subseteq \mathbb{R}^p \longrightarrow \Omega \subseteq \mathbb{R}^p$;

c) $\forall f \in F$ and $\forall h \in W$ we have that $f \circ h \in F$;

d) Given any couple of elements $f_1, f_2 \in F$ and an element $h \in W$, the distance between f_1 and f_2 is invariant under the composition of f_1 and f_2 with h, i.e.:

$$d(f_1, f_2) = d(f_1 \circ h, f_2 \circ h)\ , \quad (2)$$

we will refer to this property as W*-invariance* of d.

We will now present some results derived from properties (a)–(d) which enable the definition of a semi-metric d_W (determined by the metric d and the group W) on the space F and thus a partition of the space F in to a quotient set $\mathcal{F}$, that are easily interpretable in terms of phase and amplitude variability (Subsection 2.2). To make this discussion clearer, all proofs are reported in the Appendix.

The basic assumption upon which it is possible to build the theory is the following:

Assumption 2.1. *Given* f_1 *and* $f_2 \in F$*:* $\exists \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2)$.

Note that in some situations it might be difficult to check for the existence of the minimum. We thus provide a lemma that can be useful to deal with this kind of situations. Indeed this lemma gives a sufficient condition for the existence of a solution of the former minimization problem.

Lemma 2.2. *If* W *is compact and* $\forall f \in F$ *the map* $f\circ : W \to F$ *is continuous, then* $\forall f_1, f_2 \in F$, $\exists \ \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2)$.

Note that the minimizing couple, if existing, is never unique (except for $W = \{\mathbf{1}\}$). Indeed, because of the W-invariance of d, if (h_1,h_2) is a minimizing couple, any other couple of the form $(h_1 \circ h, h_2 \circ h)$ with $h \in W$ is still a minimizing couple. Thus, without loss of generality, h_1 (or h_2) can be fixed equal to a convenient element of W - for instance $\mathbf{1}$ - and h_2 (or h_1) consequently.

We can now define the function d_W as follows:

Definition 2.3. $d_W(f_1, f_2) := \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2)$.

The following lemmas enlighten some important properties of the quantity $d_W(f_1, f_2)$ just defined. In particular, Lemma 2.4 guarantees that, given f_1 and f_2, $d_W(f_1, f_2)$ is bounded, while Lemmas 2.5 and 2.6 characterize the lower and the upper bounds.

Lemma 2.4. $0 \le d_W(f_1, f_2) \le d(f_1, f_2)$.

Lemma 2.5. $d_W(f_1, f_2) = 0 \Leftrightarrow \exists h_1, h_2 \in W$ *such that* $f_1 \circ h_1 = f_2 \circ h_2$.

Lemma 2.6. $d_W(f_1, f_2) = d(f_1, f_2) \Leftrightarrow h_1 = h_2 = \mathbf{1}$ *is a minimizing couple.*

Previous results make it easy to prove that the function d_W is actually a semi-metric, as the following theorem states.

Theorem 2.7. $d_W : F \times F \longrightarrow \mathbb{R}_0^+$ *is a semi-metric, i.e.* $\forall$ f_1, f_2, *and* $f_3 \in F$*:*

$$f_1 = f_2 \Rightarrow d_W(f_1, f_2) = 0 \ , \tag{3a}$$

$$d_W(f_1, f_2) = d_W(f_2, f_1) \ , \tag{3b}$$

$$d_W(f_1, f_3) \le d_W(f_1, f_2) + d_W(f_2, f_3) \ . \tag{3c}$$

The following corollary just states that the equivalence classes of the quotient set induced by the semi-metric d_W on the set F coincide with the orbits of the action of the group W on the set F. This is made possible by the W-invariance of the original metric d.

Corollary 2.8. *The semi-metric d_W induces an equivalence relation $\doteq$ between elements of F defined as*

$$f_1 \doteq f_2 \Leftrightarrow d_W(f_1, f_2) = 0 \ , \tag{4}$$

or equivalently

$$f_1 \doteq f_2 \Leftrightarrow \exists h_1, h_2 \in W \text{ such that } f_1 \circ h_1 = f_2 \circ h_2 \ ; \tag{5}$$

and thus defines a quotient set $\mathcal{F} := F/\doteq$.

We are now moving toward the definition of a distance between the equivalence classes being coherent with the original distance d. In particular, the following lemma states that, given two equivalence classes, the semi-distance d_W between two functions, each one belonging to one of the two classes, is always the same.

Lemma 2.9. $f_1 \doteq \bar{f}_1$ *and* $f_2 \doteq \bar{f}_2 \Rightarrow d_W(f_1, f_2) = d_W(\bar{f}_1, \bar{f}_2)$.

Let $[f]$ indicate the equivalence class $\in \mathcal{F}$ which f belongs to, and let $\bar{f}_1$ and $\bar{f}_2$ be any couple of functions $\in F$ belonging respectively to the equivalence classes $[f_1]$ and $[f_2]$. We are now able to define the function $d_{\mathcal{F}}$ as follows:

Definition 2.10. $d_{\mathcal{F}}([f_1], [f_2]) := d_W(\bar{f}_1, \bar{f}_2)$.

Lemma 2.9 ensures that, given two classes $[f_1]$ and $[f_2]$, $d_{\mathcal{F}}([f_1], [f_2])$ is univocally defined. Indeed, given two equivalence classes $[f_1]$ and $[f_2]$, the semi-distance $d_W(\bar{f}_1, \bar{f}_2)$ between a generic element $\bar{f}_1 \in [f_1]$ and a generic element $\bar{f}_2 \in [f_2]$ is always the same.

The following theorem is probably the most important result of this work; in particular, it states that $d_{\mathcal{F}}$ makes the quotient set $\mathcal{F}$ be a metric set.

Theorem 2.11. $d_{\mathcal{F}} : \mathcal{F} \times \mathcal{F} \longrightarrow \mathbb{R}_0^+$ *is a metric, i.e.* $\forall$ $[f_1]$, $[f_2]$, *and* $[f_3] \in \mathcal{F}$:

$$[f_1] = [f_2] \Leftrightarrow d_{\mathcal{F}}([f_1], [f_2]) = 0 \ , \tag{6a}$$

$$d_{\mathcal{F}}([f_1], [f_2]) = d_{\mathcal{F}}([f_2], [f_1]) \ , \tag{6b}$$

$$d_{\mathcal{F}}([f_1], [f_3]) \leq d_{\mathcal{F}}([f_1], [f_2]) + d_{\mathcal{F}}([f_2], [f_3]) \ . \tag{6c}$$

Note that the introduction of a W-invariant metric d is not crucial in the definition of the equivalence classes, indeed those classes can be simply and directly defined as the orbits induced by the action of the group W on the set F. On the other hand, the introduction of a W-invariant metric d is essential to the identification of the orbits with the equivalence classes induced by the semi-metric d_W, and thus also essential to enable the definition of a distance $d_{\mathcal{F}}$ between orbits that is consistent with the original distance d.

2.2 The problem of registration revisited

We are now ready to formalize in this mathematical framework the problem of a registration of a pair of functions f_1 and f_2.

Definition 2.12. Functions $\tilde{f}_1 \in [f_1]$ and $\tilde{f}_2 \in [f_2]$ are said to be *mutually-registered representatives* of equivalence classes $[f_1]$ and $[f_2]$ (or in more familar terms simply *mutually-registered*) if and only if $d(\tilde{f}_1, \tilde{f}_2) = d_{\mathcal{F}}([f_1], [f_2])$.

In other words two functions are mutually-registered representatives of their equivalence classes if and only if the distance between the two functions coincides with the distance between their respective equivalence classes. By Definitions 2.3, 2.10, and 2.13, we have the equivalent definition of *mutually-registered representatives* of $[f_1]$ and $[f_2]$:

Definition 2.13. Given f_1 and $f_2 \in F$ and a minimizing couple h_1 and $h_2 \in W$ (i.e. h_1 and h_2 such that $d(f_1 \circ h_1, f_2 \circ h_2) = d_{\mathcal{F}}([f_1], [f_2])$), $\tilde{f}_1 = f_1 \circ h_1$ and $\tilde{f}_2 = f_2 \circ h_2$ are said to be *mutually-registered representatives* of $[f_1]$ and $[f_2]$.

Note that even if both f_1 and $\tilde{f}_1 \in [f_1]$, and both f_2 and $\tilde{f}_2 \in [f_2]$, only $d(\tilde{f}_1, \tilde{f}_2) = d_{\mathcal{F}}([f_1], [f_2])$ while $d(f_1, f_2) \geq d_{\mathcal{F}}([f_1], [f_2])$.

Note that since, given a couple of elements f_1 and $f_2 \in F$, there is not a unique minimizing couple h_1 and h_2, there is not a unique couple $\tilde{f}_1$ and $\tilde{f}_2$ of mutually-registered representatives of $[f_1]$ and $[f_2]$. It is worth mentioning two special couples of mutually-registered representatives of $[f_1]$ and $[f_2]$: the one corresponding to $h_1 = \mathbf{1}$ and the one corresponding to $h_2 = \mathbf{1}$. In the former case $\tilde{f}_1 = f_1$, while in the latter case $\tilde{f}_2 = f_2$.

Definition 2.14. Given a couple f_1 and f_2, and a couple of mutually-registered representatives $\tilde{f}_1$ and $\tilde{f}_2$ such that $\tilde{f}_1 = f_1$ and $h_1 = \mathbf{1}$, $\tilde{f}_2$ is said to be a *f_1-registered representative* of $[f_2]$ (or in less formal but more familiar terms $\tilde{f}_2$ is said to be a *registered* version of f_2 with respect to f_1). We will refer to it as $\tilde{f}_{2\to 1}$ and to the corresponding warping function as $h_{2\to 1}$. The definition of $\tilde{f}_{1\to 2}$ and $h_{1\to 2}$ is analogous.

Note that the uniqueness of $\tilde{f}_{2\to1}$ and $\tilde{f}_{1\to2}$ cannot be guaranteed in general. In particular, if $\tilde{f}_{2\to1}$ and $\tilde{f}_{1\to2}$ are unique (like in most practical cases), their definition can be made more explicit. Indeed, under the assumption of uniqueness, since an f_1-registered representative of $[f_2]$ is an element $\in [f_2]$ minimizing the distance with f_1, we have that:

$$\begin{aligned}\tilde{f}_{2\to1} &= \arg\min_{f\in[f_2]} d(f, f_1) \ , \\ \tilde{f}_{1\to2} &= \arg\min_{f\in[f_1]} d(f, f_2) \ ,\end{aligned}$$

with $h_{1\to2} = (h_{2\to1})^{-1}$.

According to this framework registering a function $f_1 \in F$ with respect to a function $f_2 \in F$ - according to a metric d and a class of warping functions W - simply means replacing f_1 with $\tilde{f}_{1\to2}$. Just to keep the notation as simple as possible, in the rest of the paper, we will assume, without loss of generality, $\tilde{f}_{2\to1}$ and $\tilde{f}_{1\to2}$ to be unique. Note that we are not talking about the uniqueness of the minimizing couple $(\tilde{f}_1, \tilde{f}_2)$, that is instead intrinsically non unique.

The introduction of a quotient set $\mathcal{F}$ over F (dependent on the choices for d and W) is the key to a clear and not ambiguous definition of *Phase Variability* and *Amplitude Variability.* We are quite sure to come across the heuristic sense of many authors, by defining the phase variability as the one that can occur between functions belonging to the same equivalence class, i.e. the variability within equivalence classes; note that if f_1 and f_2 belong to the same equivalence class we have that $0 = d_{\mathcal{F}}([f_1], [f_2])$. Coherently, the amplitude variability is the variability between functions not belonging to the same equivalence class and not imputable to phase variability, i.e. the variability between equivalence classes; we can thus say that the difference between f_1 and f_2 is imputable only to amplitude variability in the case $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$. Given the fact that $0 \leq d_{\mathcal{F}}([f_1], [f_2]) \leq d(f_1, f_2)$, we can define an amplitude-to-total variability ratio bounded between 0 and 1, useful in practical situations and measuring to what extent phase and amplitude variability contribute to total variability:

$$\alpha^2 = \frac{d^2_{\mathcal{F}}([f_1], [f_2])}{d^2(f_1, f_2)} \ ;$$

and then we can simply characterize the two extreme situations as follows:

- presence of phase variability only, when $\alpha^2 = 0$, i.e. $d_{\mathcal{F}}([f_1], [f_2]) = 0$;
- presence of amplitude variability only, when $\alpha^2 = 1$, i.e. $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$.

The two extreme situations can be equivalently characterized as follows:

- presence of phase variability only, when $\tilde{f}_{2\to 1} \equiv f_1$ (and thus also $\tilde{f}_{1\to 2} \equiv f_2$);

- presence of amplitude variability only, when $\tilde{f}_{2\to 1} \equiv f_2$ (and thus also $\tilde{f}_{1\to 2} \equiv f_1$).

2.3 How many equivalence classes?

Given a set F and a metric d, the quotient set $\mathcal{F}$ depends only on the group W; to emphasize this dependency, in this subsection we will use the notation $\mathcal{F}_W$ to indicate the quotient set associated to the group W, and the notation $\mathcal{P}(\mathcal{F}_W)$ to indicate its powerset.

It is easy to prove that if W is replaced by a sub-group W', the number of equivalence classes can only increase, i.e.:

$$W' \subset W \implies \mathcal{P}(\mathcal{F}_{W'}) \supseteq \mathcal{P}(\mathcal{F}_W) \ .$$

Equivalently, if W is replaced by a sup-group W' (such that d is also W'-invariant), the number of equivalence classes can only decrease, i.e.:

$$W' \supset W \implies \mathcal{P}(\mathcal{F}_{W'}) \subseteq \mathcal{P}(\mathcal{F}_W) \ .$$

More generally, within a functional data analysis the replacement of the group W with the group $W' \subset W$ [$W' \supset W$] might cause the partitioning [merging] of former equivalence classes (associated to W) into new classes (associated to W'). This kind of variability, that occurs between new [old] classes associated to W' [W] being subsets of the same old [new] class associated to W [W']), is exactly the variability that according to W' [W] is considered as part of the amplitude variability while according to W [W'] is considered as part of phase variability.

In other words, given d, choosing W is the same as defining phase variability. It is worth mentioning the two extreme situations for the choice of W:

- $W = \{\mathbf{1}\}$: in this case each element of F is equivalent only to itself, i.e., $F \equiv \mathcal{F}$. We are thus assuming that no phase variability is present within functional data;

- $W = F$: in this case all elements of F are equivalent, i.e., only one equivalence class exists coinciding with the whole set F. We are thus assuming that no amplitude variability is present within the functional data. Note that this case can occur only if F is a sub-group of the group of the continuous automorphisms: $\Omega \subseteq \mathcal{R}^p \longrightarrow \Omega \subseteq \mathcal{R}^p$.

3 Registration of a set of functions in presence of a target function

We have just shown that, under the introduced framework, registering f_2 with respect to f_1 means replacing f_2 with a function $\tilde{f}_{2\to 1} \in [f_2]$ whose distance to f_1 is minimal. In the same framework, it is straightforward to define the registration of a set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to a target function f_0. Indeed registering the set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to f_0 means replacing the set $\{f_i\}_{i=1,2,\ldots,n}$ with the set $\{\tilde{f}_{i\to 0}\}_{i=1,2,\ldots,n}$ (or simply $\{\tilde{f}_i\}_{i=1,2,\ldots,n}$) whose distances to f_0 are minimal over the relevant equivalence classes:

$$\{f_i\}_{i=1,2,\ldots,n} \longmapsto \{\tilde{f}_i = \arg\min_{f\in[f_i]} d(f_0, f)\}_{i=1,2,\ldots,n} \ .$$

In other words, registering the set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to f_0 consists in finding in $[f_1]$, $[f_2]$, $\ldots$, $[f_n]$, n functions that are the closest to f_0 respectively.

Also in this case, we can define an amplitude-to-total variability ratio:

$$\alpha^2 = \frac{\sum_{i=1}^n d_{\mathcal{F}}^2([f_i],[f_0])}{\sum_{i=1}^n d^2(f_i, f_0)} \ ;$$

we can then simply characterize the two extreme situations as follows:

- presence of phase variability only, when $\alpha^2 = 0$;
- presence of amplitude variability only, when $\alpha^2 = 1$.

The two extreme situations can be equivalently characterized as follows:

- presence of phase variability only, when for $i = 1, 2, \ldots, n : \tilde{f}_i \equiv f_0$;
- presence of amplitude variability only, when for $i = 1, 2, \ldots, n : \tilde{f}_i \equiv f_i$.

In order to help the reader, in Figure 1, a schematic representation of the mathematical framework introduced is reported.

4 Registration of a set of functions in absence of a target function

In most practical problems the focus is on registering a set $\{f_i\}_{i=1,2,\ldots,n}$ with respect to "itself", since a target function f_0 is generally not available. Also in this case, it is still meaningful to talk about registration: roughly speaking, it is straightforward to assert that registering the set $\{f_i\}_{i=1,2,\ldots,n}$ would consist in replacing the set $\{f_i\}_{i=1,2,\ldots,n}$ with a set $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \in [f_1] \times [f_2] \times \cdots \times [f_n]$ of functions that are "closest as possible".

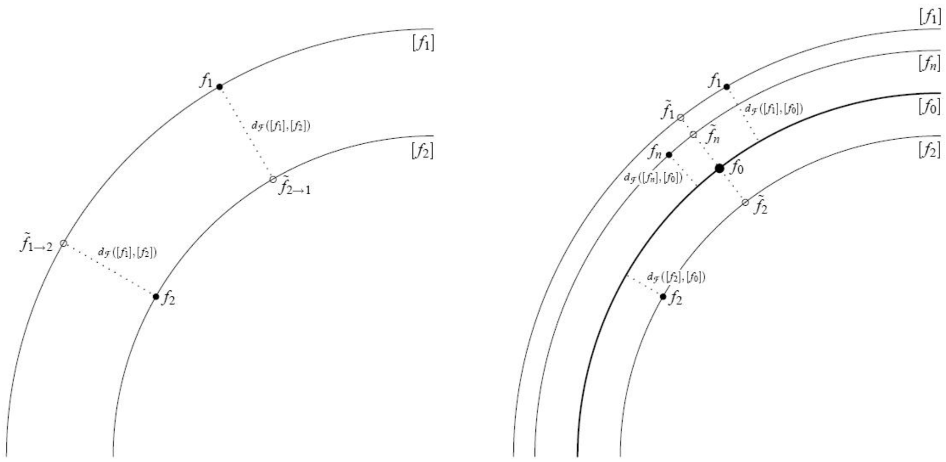

Figure 1: Schematic representation of the mathematical framework introduced: registration of a couple of functions on the left, and registration of a set of functions with respect to a target function f_0 on the right. Black dots refer to non-registered functions, white dots to registered functions, and circumferences to equivalence classes.

A natural approach to formalize the notion of "closest as possible" is to introduce an auxiliary reference function $\hat{f}_0 \in F$ such that $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ is the solution of the following minimization problem:

$$\min_{\tilde{f}_i \in [f_i] \ \wedge \ \hat{f}_0 \in F} \left(\sum_{i=1}^{n} d^2(\tilde{f}_i, \hat{f}_0) \right) . \tag{7}$$

In other words, registering a set $\{f_i\}_{i=1,2,\ldots,n}$ means registering each function of the set with respect to the sample Frechet mean of the registered set. The following lemma guarantees the existence of a solution of the minimization problem (7) under the same assumption of Lemma 2.2.

Lemma 4.1. *If W is compact and $\forall f \in F$ the map $f\circ : W \to F$ is continuous, then a solution of the problem (7) exists.*

Note that, also in this case, the solution is never unique (except for $W = \{\mathbf{1}\}$ where the solution might be unique). Indeed, because of the W-invariance of d, if $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ is a solution of the minimization problem (7) any other set of the form $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$ with $h \in W$ is still a solution.

Also in this case, we can define an amplitude-to-total variability ratio:

$$\alpha^2 = \frac{\sum_{i=1}^{n} d^2_{\mathcal{F}}([f_i], [\hat{f}_0])}{\sum_{i=1}^{n} d^2(f_i, \hat{f}_0)} .$$

Note that in this case (i.e. when a target function f_0 is not present but needs to be estimated) some care is needed to correctly compute the α^2 index:

- Firstly, note that α^2 compares the deviations of the registered functions from the Frechet mean of the same registered functions (numerator) with the deviations of the original functions from the Frechet mean of the registered functions (denominator) and not with the deviations of the original functions from the Frechet mean of the same original functions as one might expect. This mistake has been made - more or less explicitly - in many works dealing with the registration of functional data, providing of course an underestimation of the total variability and consequently an overestimation of the contribution of the amplitude variability to the total variability (e.g. Sangalli et al. 2009) or even bringing to meaningless situations in which the amplitude variability seems to be greater than the total one (e.g. Kneip and Ramsay 2008).

- Secondly, note that the α^2 ratio is not invariant under a joint warping of the solution set $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$ along the same warping function h. Indeed, even if $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$ is still a solution of (7), in the computation of α^2 the numerator does not change while the denominator may change from $\sum_{i=1}^{n} d^2(f_i, \hat{f}_0)$ to $\sum_{i=1}^{n} d^2(f_i, \hat{f}_0 \circ h)$. It is natural to assume that among all possible solution of the minimization problem (7) the one that is "closest as possible" to the original situation is the natural candidate to be the "right one". Formally, it means that given a solution $\{\tilde{f}_i\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0\}$, the solution $\{\tilde{f}_i \circ h\}_{i=1,2,\ldots,n} \cup \{\hat{f}_0 \circ h\}$ to be used to compute the α^2 ratio is the one minimizing the total variability, i.e. given $[\hat{f}_0]$, the following constraint on $\hat{f}_0$ has to be introduced in the minimization problem (7) in order to identify the correct solution:

$$\sum_{i=1}^{n} d^2(f_i, \hat{f}_0) = \min_{f \in [\hat{f}_0]} \left(\sum_{i=1}^{n} d^2(f_i, f) \right) . \tag{8}$$

  Thus, the constrained minimization problem can be restate in a simpler way as follows. Given a set $\{f_i\}_{i=1,2,\ldots,n}$ find the reference class $[\hat{f}_0]$ such that:

$$[\hat{f}_0] = \arg \min_{[f] \in \mathcal{F}} \left(\sum_{i=1}^{n} d^2_{\mathcal{F}}([f_i], [f]) \right)$$

  and take as representatives of the equivalence classes those functions $\{\tilde{f}_i\}_{i=1,2,\ldots,n}$ that are registered with respect to $\hat{f}_0$, that is that function belonging the reference class $[\hat{f}_0]$ such that its average squared distance from the original functions is minimal.

  Neglecting constraint (8) brings of course to an overestimation of the total variability and consequently an underestimation of the contribution of the amplitude variability to the total variability. This constraint essentially avoid the drifting apart of the registered functions from the original ones. Similar constraints have been used in the literature for the same purpose. For instance, in both Sangalli et al. (2009) and Kneip and Ramsay (2008) the constraint $\sum_{i=1}^{n} h_i = \mathbf{1}$ was used. Unfortunately the latter constraint, even if heuristically equivalent to constraint (8), does not appear to be theoretically coherent with the theory here proposed.

If we take care of the points discussed above, also in this case, we can simply characterize the two extreme situations as follows:

- presence of phase variability only, when $\alpha^2 = 0$;
- presence of amplitude variability only, when $\alpha^2 = 1$.

The two extreme situations can be equivalently characterized as follows:

- presence of phase variability only, when for $i = 1, 2, \ldots, n : \tilde{f}_i \equiv f_0$;
- presence of amplitude variability only, when for $i = 1, 2, \ldots, n : \tilde{f}_i \equiv f_i$.

Solving the minimization problem (7) might be of course not trivial. Even if any numerical minimization method can be used to approximate the solution, the proof of Lemma 4.1 suggests all methods belonging to the family of the so-called *Procrustes fitting criteria* to be good candidates to solve this minimization problem. In particular, in the same way of Ramsay and Li (1998), Kneip et al. (2000), and Sangalli et al. (2009), an iterative search of a minimum can be performed alternating minimization and expectation steps:

Minimization: $\{\tilde{f}_i^{[k+1]}\}_{i=1,2,\ldots,n} = \left\{\arg\min_{\tilde{f}_i \in [f_i]} \left(\sum_{i=1}^{n} d^2(\tilde{f}_i, \hat{f}_0^{[k]})\right)\right\}_{i=1,2,\ldots,n}$.
In these steps, each function of the set $\{f_i\}_{i=1,2,\ldots,n}$ is registered with respect to the Frechet mean of the set $\{\tilde{f}_i^{[k]}\}_{i=1,2,\ldots,n}$;

Expectation: $\hat{f}_0^{[k+1]} = \arg\min_{\hat{f}_0 \in F} \left(\sum_{i=1}^{n} d^2(\tilde{f}_i^{[k+1]}, \hat{f}_0)\right)$.
In these steps, the Frechet mean of the set $\{\tilde{f}_i^{[k+1]}\}_{i=1,2,\ldots,n}$ is computed.

The algorithm can be initialized identifying $\hat{f}_0^{[0]}$ with the Frechet mean of the initial set $\{f_i\}_{i=1,2,\ldots,n}$. Moreover, since $\sum_{i=1}^{n} d^2(\tilde{f}_i^{[k]}, \hat{f}_0^{[k]})$ can only decrease as k increases and it is lower bounded by 0, the algorithm can be stopped when

$$\sum_{i=1}^{n} d^2(\tilde{f}_i^{[k]}, \hat{f}_0^{[k]}) - \sum_{i=1}^{n} d^2(\tilde{f}_i^{[k+1]}, \hat{f}_0^{[k+1]}) < \epsilon \; .$$

Note that a small decrement in $\sum_{i=1}^{n} d^2(\tilde{f}_i^{[k+1]}, \hat{f}_0^{[k]})$ can be associated with big changes in $\tilde{f}_i^{[k+1]}$. This is compatible with the W-invariance of d and the non-uniqueness of the solution of the minimization problem (7). Indeed in some cases, the algorithm might approach the set of all possible solutions targeting at each step a different solution. This should not create any concern since after any expectation step the function $\hat{f}_0^{[k+1]}$ has to be replaced by a suitable equivalent function $\hat{f}_0^{[k+1]} \circ h$ satisfying constraint (8).

5 Ancillary variability

In many situations, d is not a metric but a semi-metric, i.e. condition (1a) is relaxed to:

$$f_1 = f_2 \in F \Rightarrow d(f_1, f_2) = 0$$

This means that two functions f_1 and $f_2 \in F$ such that $f_1 \neq f_2$ even if $d(f_1, f_2) = 0$ can exist.

In this frequent case, the presented theory still holds if F is replaced with $\bar{F}$, where $\bar{F}$ is the quotient set $F/\odot$ defined by the following equivalence relation:

$$f_1 \odot f_2 \Leftrightarrow d(f_1, f_2) = 0 \ .$$

It is important to point out that if d is a semi-metric, a further kind of variability is evident: the *Ancillary Variability*. Thus, when d is a semi-metric we can coherently define ancillary, phase and amplitude variability as follows:

- Ancillary variability is the one that can occur between functions belonging to the same equivalence class of $\bar{F}$;
- Phase variability is the one that can occur between equivalence classes of $\bar{F}$ belonging to the same equivalence class of $\mathcal{F}$;
- Amplitude variability is the one that can occur between different equivalence classes of $\mathcal{F}$.

Also in this case we can characterize some extreme situations:

- presence of ancillary variability only, when $d(f_1, f_2) = 0$;
- presence of phase and ancillary variability only, when $d_{\mathcal{F}}([f_1], [f_2]) = 0$;
- presence of amplitude and ancillary variability only, when $d_{\mathcal{F}}([f_1], [f_2]) = d(f_1, f_2)$.

Note that in the definition of the index α^2, the ancillary variability contributes neither to amplitude nor to total variability. Indeed according to d, it is actually a non-variability. For this reason, the easiest approach to functional data registration in these cases should be that of setting the analysis, from the very beginning, in terms of elements of $\bar{F}$ and induced metric $d_{\bar{F}}$ rather than in terms of the original elements of F and the original semi-metric d.

6 Examples presented in literature

The theory hereby developed is able to put in a unique theoretical framework many approaches to functional data registration that have already appeared in the literature. As examples, we illustrate in the light of the present work, two recent papers in which a registration of complex functional data is performed: Sangalli et al. (2009) and Kaziska and Srivastava (2007).

In Sangalli et al. (2009) a registration of 65 Human Internal Carotid Artery centerlines $\subset \mathbb{R}^3$ is performed. In Figure 2, the first derivatives of these centerlines before and after registration are reported. It is easy to identify in this work the set F, the group W, and the semi-metric d:

$$
\begin{aligned}
F &= \{f \in C^1(\mathbb{R};\mathbb{R}^3) : f(s) \not\equiv c \ \text{ with } \ c \in \mathbb{R}^3\} \ , \\
W &= \{h \in C^1(\mathbb{R};\mathbb{R}) : h(s) = ms + q \ \text{ with } \ m \in \mathbb{R}^+, q \in \mathbb{R}\} \ , \\
d(f_1, f_2) &= \sqrt{1 - \frac{1}{3} \sum_{k=x,y,z} \frac{< f'_{1k} \ , \ f'_{2k} >_{L^2(\Omega)}}{||f'_{1k}||_{L^2(\Omega)} ||f'_{2k}||_{L^2(\Omega)}}} \ .
\end{aligned}
$$

The corresponding notions of ancillary, phase, and amplitude variability are thus implicitly assumed:

- Ancillary variability is the one that can occur between functions that are equal up to an increasing affine transformation of the ordinate, i.e.:

$$\exists A_k \in \mathbb{R}^+, B_k \in \mathbb{R} : f_{1k}(s) = A_k f_{2k}(s) + B_k \ .$$

- Phase variability is the one that can occur between functions that are equal up to an increasing affine transformation of the abscissa, i.e.:

$$\exists m \in \mathbb{R}^+, q \in \mathbb{R} : f_1(s) = f_2(ms + q).$$

- Amplitude variability is the one that cannot been removed by the data by means of increasing affine transformations of neither the ordinate nor the abscissa.

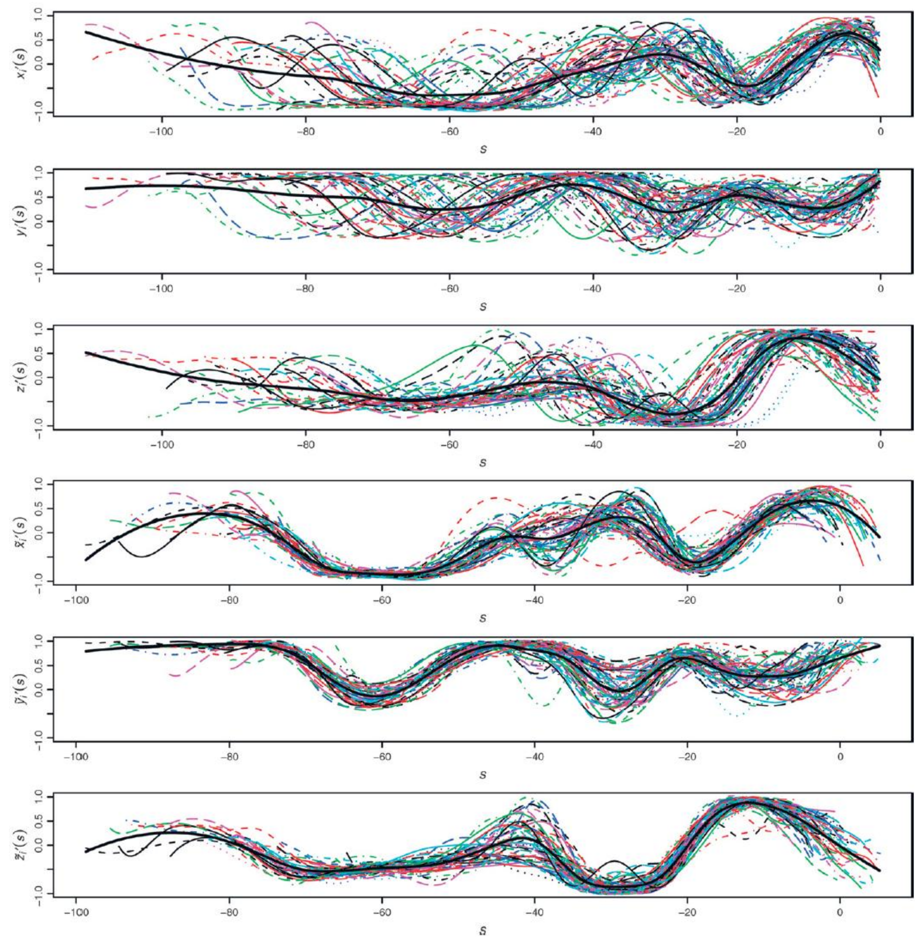

Figure 2: The 65 first derivatives f'_{ix}, f'_{iy} and f'_{iz} before registration (top) and the 65 first derivatives $\tilde{f}'_{ix}$, $\tilde{f}'_{iy}$ and $\tilde{f}'_{iz}$ after registration (bottom). The first derivatives of the estimated Frechet mean $\hat{f}_0$ are reported in black. See Sangalli et al. (2009) for details.

Moreover, we are able to compute $\alpha^2 = 33\%$, i.e. the amplitude variability accounts for nearly just 1/3 of the variability of the 65 Internal Carotid Artery centerlines. This a further evidence of the necessity of a registration procedure within this functional data analysis (Figure 2).

It is worth noticing that the group W used in Sangalli et al. (2009) is not compact and thus in general their choice for W and d cannot guarantee the existence of a solution for the associated minimization problem. On the other hand, the lack of compactness of W should not rise any concern in practical situations. Indeed, in real applications, data are usually only slightly misaligned and therefore a marked local minimum is usually present. So, by simply introducing some "reasonable" constraints on the warping functions (i.e. large enough to gather the local minimum and small enough not to obtain an artificial minimum at the boundary of the constraint) we can re-obtain a meaningful and well posed minimization problem.

Kaziska and Srivastava (2007), in the first part of their work regarding the analysis of human shapes, deal with the registration of simple closed (i.e. periodic) bi-dimensional curves $\subset \mathbb{R}^2$. Also in this work, it is easy to identify the set F, the group W, and the metric d:

$$\begin{aligned} F &= \mathcal{C} \ , \\ W &= \mathbb{S}^1 \times \mathcal{D} \ , \\ d(f_1, f_2) &= d_{\mathcal{C}}(f_1, f_2) \ ; \end{aligned}$$

where $\mathcal{C}$ (the preshape space in the work) is the set of all continuous 2π-periodic functions mapping $[0, 2\pi]$ in a closed curve $\subset \mathbb{R}^2$ of length 2π and average direction π; $\mathbb{S}^1$ is the group of the translation of the abscissa; $\mathcal{D}$ is the group of the automorphisms $[0, 2\pi] \longmapsto [0, 2\pi]$; the distance $d_{\mathcal{C}}$, even if not clearly stated, appears to be the distance induced by the usual inner product $<, >_{L^2([0,2\pi])}$, i.e. $d_{\mathcal{C}}(f_1, f_2) = ||f_1 - f_2||_{L^2([0,2\pi])}$; for completeness, the quotient set $\mathcal{F}$ is indicated in the work as $\mathcal{S}$ (the shape space).

It is important to mention that, as declared by the authors, the original functions do not belong to $\mathcal{C}$; the original curves are indeed preprocessed – i.e. rotated, translated, and scaled in $\mathbb{R}^2$ – such that their average length is 2π and their average direction is π. In this case the ancillary variability (here the one imputable to rotation, translation, and scaling in $\mathbb{R}^2$) is explicitly removed before the analysis.

The corresponding notions of ancillary, phase, and amplitude variability come straightforward:

- As just pointed out, there is no ancillary variability between functions $\in \mathcal{C}$ since this is removed before the analysis;
- Phase variability is the one that can occur between functions representing the same curve $\in \mathbb{R}^2$ by means of a different parameterization, i.e. functions that are equal up to a change in the origin of the abscissa and an automorphism of the abscissa;
- Amplitude variability is the one that can occur between functions representing different curves $\in \mathbb{R}^2$ after the phase variability has been removed, i.e. after the two abscissas have been matched by means of a joint translation and automorphism $h \in \mathbb{S}^1 \times \mathcal{D}$ minimizing $d_{\mathcal{C}}(f_1, f_2 \circ h)$.

Note that the choice $d = d_{\mathcal{C}}$ and $W = \mathcal{D}$ is not fair according to the theory here developed, in the sense that they do not satisfy (d), i.e. $d_{\mathcal{C}}$ is not $(\mathbb{S}^1 \times \mathcal{D})$-invariant (it is actually only $\mathbb{S}^1$-invariant, thanks to the periodicity of functions belonging to $\mathcal{C}$). This means that two curves can be made arbitrarily more/less close simply by jointly changing the parameterization of the two curves. There are two main consequences due to the non $\mathcal{D}$-invariance of $d_{\mathcal{C}}$: firstly the quotient set $\mathcal{F}$ is not formally defined and thus the phase and amplitude variability are not clearly defined; secondly, if the registration procedure is anyway performed, it will be likely to find meaningless and unusable results from a practical point of view. This degeneracy problem occurring with the joint use of the group of automorphisms and of the L^2-norm has already been noticed in Ramsay and Silverman (2005) and more recently in Kneip and Ramsay (2008).

Kaziska and Srivastava (2007), like Ramsay and Silverman (2005), get out of this degeneracy problem by introducing in the minimization process a penalization term. Indeed, even if in the declared theoretical framework the registration problem is introduced as the minimization over $h \in \mathbb{S}^1 \times \mathcal{D}$ of the functional $||f_1 - f_2 \circ h||_{L^2([0,2\pi])}$, in the search for the solution, the functional $\lambda||f_1 - f_2 \circ h||^2_{L^2([0,2\pi])} + (1-\lambda)||h'||^2_{L^2([0,2\pi])}$ is actually minimized.

In the light of the present work, we are convinced that the necessity of introducing a reasonable penalization term or reasonable constraints to make meaningful the results of a registration procedure is in general an evidence of a mismatch between the phase variability as actually defined by W and the phase variability as thought by the statistician.

For this reason, we think that the correct way to get out of degeneracy problems is not to introduce a penalization term or constraints (we are aware that, even if not theoretically sound, these solutions are however practically easy and efficient) but to replace W with a suitable group and to redefine phase and amplitude variability consequently. Of course this second approach is definitely not trivial and poses new challenging questions for future research.

7 Conclusions

We introduced a mathematical framework in which a functional data registration problem can be soundly and coherently set. In detail, we showed that the introduction of a metric/semi-metric and a compact group of warping function respect to which the metric/semi-metric is invariant is the key to a clear and not ambiguous definition of phase and amplitude variability. Moreover, we proposed the amplitude-to-total variability ratio α^2. This index turns to be useful in practical situations to measure to what extent amplitude variability affects the data and to compare the effectiveness of different registration methods.

We think that it might be of interest in future to find out pairs of compact sub-groups of the group of the automorphisms and invariant metrics/semimetrics able to report, in formal terms, the ideas about phase and amplitude variability of the "experts" and/or for which an exact or approximate decomposition of total variability in amplitude and phase variability of the form "$\alpha^2 + \phi^2 = 1$" is possible.

Finally, we want to thank professors Piercesare Secchi and Marco Fuhrman of the Department of Mathematics - Politecnico di Milano for their useful comments.

Appendix

Proof of Lemma 2.2. d is continuous since triangular inequality (1c) implies $|d(f_1, f_2) - d(f_1, f_3)| \leq d(f_2, f_3)$; the maps $f_1\circ$ and $f_2\circ$ are demanded to be continuous; thus $d(f_1 \circ h_1, f_2 \circ h_2)$ is continuous in h_1 and h_2. Moreover $d(f_1 \circ h_1, f_2 \circ h_2)$ is lower bounded (≥ 0). Since W is compact, the extreme value theorem ensures the minimum to exist. □

Proof of Lemma 2.4: lower bound. $d(f_1 \circ h_1, f_2 \circ h_2) \geq 0 \; \forall h_1, h_2 \in W \Rightarrow \min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2) \geq 0$. □

Proof of Lemma 2.4: upper bound. $\min_{h_1,h_2 \in W} d(f_1 \circ h_1, f_2 \circ h_2) \leq d(f_1, f_2)$ since $d(f_1, f_2) = d(f_1 \circ \mathbf{1}, f_2 \circ \mathbf{1})$. □

Proof of Lemma 2.5: ⇒. By Definition 2.3, $d_W(f_1, f_2) = 0$ implies that exists a couple (h_1,h_2) such that $d(f_1 \circ h_1, f_2 \circ h_2) = 0$; (1a) implies that $d(f_1 \circ h_1, f_2 \circ h_2) = 0 \Rightarrow f_1 \circ h_1 = f_2 \circ h_2$. □

Proof of Lemma 2.5: ⇐. (1a) implies that $f_1 \circ h_1 = f_2 \circ h_2 \Rightarrow d(f_1 \circ h_1, f_2 \circ h_2) = 0$; since 0 is also the lower bound, this couple is also a minimizing couple, then $d_W(f_1, f_2) = 0$. □

Proof of Lemma 2.6. Proof is immediate by Definition 2.3. □

Proof of Theorem 2.7: (3a). From (1a), $f_1 = f_2 \Rightarrow d(f_1, f_2) = 0$; from Lemma 2.4, $d(f_1, f_2) = 0 \Rightarrow d_W(f_1, f_2) = 0$. □

Proof of Theorem 2.7: (3b). (3b) descends from (1b), indeed if ($\bar{h}_1$, $\bar{h}_2$) is a minimizing couple for $d(f_1 \circ h_1, f_2 \circ h_2)$, then ($\bar{h}_2$, $\bar{h}_1$) is a minimizing couple for $d(f_2 \circ h_2, f_1 \circ h_1)$ providing the same minimum. □

Proof of Theorem 2.7: (3c). (3c) descends from (1c) and from the W-invariance of d. Let $(\bar{h}_1, \bar{h}_2)$ and $(\bar{h}_2, \bar{h}_3)$ be minimizing couples for $d(f_1 \circ h_1, f_2 \circ h_2)$ and $d(f_2 \circ h_2, f_3 \circ h_3)$ respectively, i.e.:

$$d_W(f_1, f_2) = d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2) \ , \tag{9a}$$

$$d_W(f_2, f_3) = d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3) \ . \tag{9b}$$

As already stressed, because of the W-invariance of d and without loss of generality, $\bar{h}_2$ of the former couple can be fixed equal to $\bar{h}_2$ of the latter couple. The couple $(\bar{h}_1, \bar{h}_3)$ is not in general a minimizing couple for $d(f_1 \circ h_1, f_3 \circ h_3)$, thus:

$$d_W(f_1, f_3) \leq d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3) \ . \tag{10}$$

(1c) applied to $f_1 \circ \bar{h}_1$, $f_2 \circ \bar{h}_2$, and $f_3 \circ \bar{h}_3$, provides that:

$$d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3) \leq d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2) + d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3) \ . \tag{11}$$

Finally - by chaining (10), (11), (9a) and (9b) - (3c) is obtained:

$$\begin{aligned} d_W(f_1, f_3) &\leq d(f_1 \circ \bar{h}_1, f_3 \circ \bar{h}_3) \leq \\ \leq d(f_1 \circ \bar{h}_1, f_2 \circ \bar{h}_2) + d(f_2 \circ \bar{h}_2, f_3 \circ \bar{h}_3) &= d_W(f_1, f_2) + d_W(f_2, f_3) \end{aligned}$$

□

Proof of Corollary 2.8. $\doteq$ is an equivalence relation. Indeed reflexivity, symmetry, and transitivity of the $\doteq$, trivially descend from (3a), (3b), and (3c) respectively. Lemma 2.5 implies the equivalence between enunciations (4) and (5). □

Proof of Lemma 2.9. Firstly, let us prove that $f_2 \doteq \bar{f}_2 \Rightarrow d_W(\bar{f}_1, f_2) = d_W(\bar{f}_1, \bar{f}_2)$. (3c) ensures that $d_W(\bar{f}_1, f_2) \leq d_W(\bar{f}_1, \bar{f}_2) + d_W(f_2, \bar{f}_2)$; $f_2 \doteq \bar{f}_2$ implies $d_W(f_2, \bar{f}_2) = 0$; thus $d_W(\bar{f}_1, f_2) \leq d_W(\bar{f}_1, \bar{f}_2)$. Analogously, permuting f_2 and $\bar{f}_2$, also $d_W(\bar{f}_1, \bar{f}_2) \leq d_W(\bar{f}_1, f_2)$. Thus $d_W(\bar{f}_1, f_2) = d_W(\bar{f}_1, \bar{f}_2)$.

Analogously, we can obtain that $f_1 \doteq \bar{f}_1 \Rightarrow d_W(f_1, f_2) = d_W(\bar{f}_1, f_2)$; thus $(f_2 \doteq \bar{f}_2$ and $f_1 \doteq \bar{f}_1) \Rightarrow d_W(f_1, f_2) = d_W(\bar{f}_1, \bar{f}_2)$. □

Proof of Theorem 2.11: (6a). Sufficient condition: $[f_1] = [f_2]$ implies that $\forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2] \Rightarrow \bar{f}_1 \doteq \bar{f}_2$; moreover $\bar{f}_1 \doteq \bar{f}_2 \Rightarrow d_W(\bar{f}_1, \bar{f}_2) = 0$; but by definition $d_W(\bar{f}_1, \bar{f}_2) = d_{\mathcal{F}}([f_1], [f_2])$ and thus $d_{\mathcal{F}}([f_1], [f_2]) = 0$. Necessary condition: $d_{\mathcal{F}}([f_1], [f_2]) = 0$ implies that $d_W(\bar{f}_1, \bar{f}_2) = 0 \ \forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$; thus $\forall \bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$ we have that $\bar{f}_1 \doteq \bar{f}_2$; thus any $\bar{f}_1 \in [f_1]$ is equivalent to any $\bar{f}_2 \in [f_2]$ and viceversa, that means that $[f_1] = [f_2]$. □

Proof of Theorem 2.11: (6b). Let us take two elements $\bar{f}_1$ and $\bar{f}_2 \in F$ such that $\bar{f}_1 \in [f_1]$ and $\bar{f}_2 \in [f_2]$; by definition $d_{\mathcal{F}}([f_1],[f_2]) = d_W(\bar{f}_1,\bar{f}_2)$ and $d_{\mathcal{F}}([f_2],[f_1]) = d_W(\bar{f}_2,\bar{f}_1)$; moreover (3b) ensures $d_W(\bar{f}_1,\bar{f}_2) = d_W(\bar{f}_2,\bar{f}_1)$ and thus (6b) is proven. □

Proof of Theorem 2.11: (6c). Let us take three elements $\bar{f}_1$, $\bar{f}_2$, and $\bar{f}_3 \in F$ such that $\bar{f}_1 \in [f_1]$, $\bar{f}_2 \in [f_2]$, and $\bar{f}_3 \in [f_3]$; by definition $d_{\mathcal{F}}([f_1],[f_3]) = d_W(\bar{f}_1,\bar{f}_3)$, $d_{\mathcal{F}}([f_1],[f_2]) = d_W(\bar{f}_1,\bar{f}_2)$, and , $d_{\mathcal{F}}([f_2],[f_3]) = d_W(\bar{f}_2,\bar{f}_3)$; moreover (3c) ensures $d_W(\bar{f}_1,\bar{f}_3) \le d_W(\bar{f}_1,\bar{f}_2) + d_W(\bar{f}_1,\bar{f}_3)$ and thus (6c) is proven. □

Proof of Lemma 4.1: $[f_1], [f_2], \ldots, [f_n]$ are compact sets $\subseteq F$ since W is compact and for $i = 1, 2, \ldots, n$, $f_i \circ : h \in W \longmapsto (f_1 \circ h) \in F$ is continuous. The functional $\sum_{i=1}^n d^2(\tilde{f}_i, \hat{f}_0)$ is lower bounded, so $\inf_{\tilde{f}_i \in [f_i] \,\wedge\, \hat{f}_0 \in F} \left(\sum_{i=1}^n d^2(\tilde{f}_i, \hat{f}_0) \right) \ge 0$. The compactness of set $[f_i]$ guarantees that inferior limit occurs in correspondence of functions $\tilde{f}_i$ belonging to the set $[f_i]$. Moreover in correspondence of the inferior limit, we have that $\hat{f}_0$ belongs to the sample Frechet mean set of a set of function belonging to a compact set, i.e. $\bigcup_{i=1,2,\ldots,n}[f_i]$. It is known that the sample Frechet mean set is non-empty and compact. So we can find an element $\hat{f}_0$ belonging to the sample Frechet mean set in correspondence of which the functional takes value equal to its inferior limit. Thus the inferior limit is also a minimum. □

References

Altman, N. S. and Villarreal, J. C. (2004), "Self-modelling regression for longitudinal data with time-invariant covariates," *Canad. J. Statist.*, 32, 251–268.

Ferraty, F. and Vieu, P. (2006), *Nonparametric functional data analysis*, Springer Series in Statistics, New York: Springer, theory and practice.

James, G. M. (2007), "Curve alignment by moments," *The Annals of Applied Statistics*, 1, 480–501.

Kaziska, D. and Srivastava, A. (2007), "Gait-Based Human Recognition by Classification of Cyclostationary Processes on Nonlinear Shape Manifolds," *Journal of the American Statistical Association*, 102, 1114–1124(11).

Ke, C. and Wang, Y. (2001), "Semiparametric nonlinear mixed-effects models and their applications," *J. Amer. Statist. Assoc.*, 96, 1272–1298, with comments and a rejoinder by the authors.

Kneip, A., Li, X., MacGibbon, K. B., and Ramsay, J. O. (2000), "Curve registration by local regression," *Canad. J. Statist.*, 28, 19–29.

Kneip, A. and Ramsay, J. O. (2008), "Combining Registration and Fitting for Functional Models," *Journal of the American Statistical Association*, 103, 1155–1165.

Lawton, W. H., Sylvestre, E. A., and Maggio, M. S. (1972), "Self Modeling Nonlinear Regression," *Technometrics*, 14, 513–532.

Lindstrom, M. J. and Bates, D. M. (1990), "Nonlinear mixed effects models for repeated measures data," *Biometrics*, 46, 673–687.

Ramsay, J. O. and Li, X. (1998), "Curve registration," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60, 351–363.

Ramsay, J. O. and Silverman, B. W. (2005), *Functional Data Analysis*, Springer New York NY, 2nd ed.

Sangalli, L. M., Secchi, P., Vantini, S., and Veneziani, A. (2009), "A Case Study in Exploratory Functional Data Analysis: Geometrical Features of the Internal Carotid Artery," *Journal of the American Statistical Association*, 104, 37–48.

Wang, H. and Marron, J. S. (2007), "Object oriented data analysis: sets of trees," *Ann. Statist.*, 35, 1849–1873.

MOX Technical Reports, last issues

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

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

32/2009 S. Micheletti, S. Perotto:
Anisotropic adaptation via a Zienkiewicz-Zhu error estimator for 2D elliptic problems

31/2009 Paola F. Antonietti, Paolo Biscari, Alaleh Tavakoli, Marco Verani, Maurizio Vianello:
Theoretical study and numerical simulation of textiles

30/2009 Z. C. Xuan, T. Lassila, G. Rozza, A. Quarteroni:
On computing upper and lower bounds on the outputs of linear elasticity problems approximated by the smoothed finite element method

29/2009 P.F. Antonietti, A. Pratelli:
Finite Element Approximation of the Sobolev Constant

28/2009 L. Gaudio, A. Quarteroni:
Spectral Element Discretization of Optimal Control Problems

27/2009 L. Mirabella, F. Nobile, A. Veneziani:
A robust and efficient conservative technique for simulating three-dimensional sedimentary basins dynamics

26/2009 M. Longoni, A.C.I. Malossi, A. Villa:
A robust and efficient conservative technique for simulating three-dimensional sedimentary basins dynamics

25/2009 P.E. Farrell, S. Micheletti, S. Perotto:
An anisotropic Zienkiewicz-Zhu a posteriori error estimator for 3D applications

24/2009 F. Di Maio, P. Secchi, S. Vantini, E. Zio:
Optimized Fuzzy C-Means Clustering and Functional Principal Components for Post-Processing Dynamic Scenarios in the Reliability Analysis of a Nuclear System