Combining Registration and Fitting for Functional Models

A. Kneip and J. O. Ramsay

1

Abstract

A registration method can be defined as a process of aligning features of a sample of curves by monotone transformations of their domain. The aligned curves exhibit only amplitude variation, and the domain transformations, called warping functions, capture the phase variation in the original curves. In this paper, we precisely define a new type of registration process, in which the warping functions optimize the fit of a principal components decomposition to the aligned curves. The principal components are effectively the features that this process aligns. We discuss the relation of registration to closure of a function space under convex operations, and define consistency for registration methods. An explicit decomposition of functional variation into amplitude and phase partitions is defined, and an algorithm for combining registration with principal components analysis is developed, and applied to simulated and real data.

2

1   Introduction to the registration problem

The data that we consider are a sample of n smooth random functions $x_1, \ldots, x_n$ defined over a closed interval on the real line, such as the 21 curves displayed in the top panel of Figure 1. The functions share a common set of shape features, consisting of two peaks separated by a valley. The sizes of the features vary, and we refer to this as *amplitude variation*. The locations of the features also vary from curve to curve, and this is *phase variation*. We define the registration problem as the search for a set of smooth strictly monotonic functions h_i, called its *warping function*, such that the functions of the form

$$y(t) = x[h(t)] = (x \circ h)(t) \tag{1}$$

have their features aligned, where our concept of a feature will include principal components of variation.

Phase variation poses severe problems for the application of functional versions of commonly used multivariate data analyses such as computing pointwise means, variances and correlations; principal components analysis and canonical correlation analyses (Ramsay and Silverman, 2005; Silverman, 1995). For example, the mean function in the top panel of Figure 1 does not seem appropriate as a data summary in the sense that only six of the curves have smaller amplitudes than the mean in either lobe, and the mean has less variation than most of the curves. If we perform a functional principal components analysis of the unregistered curves in Figure 1, the first two eigenvalues account for 98.7% of their variation. While this may seem entirely satisfactory by some standards, an inspection of the curves indicates that the heights of the two lobes seem to vary independently, suggesting that two principal components could yield an even better approximation to these functional data. In fact, the leading two eigenvalues for the curves in the second panel, from which phase variation has been removed by the method
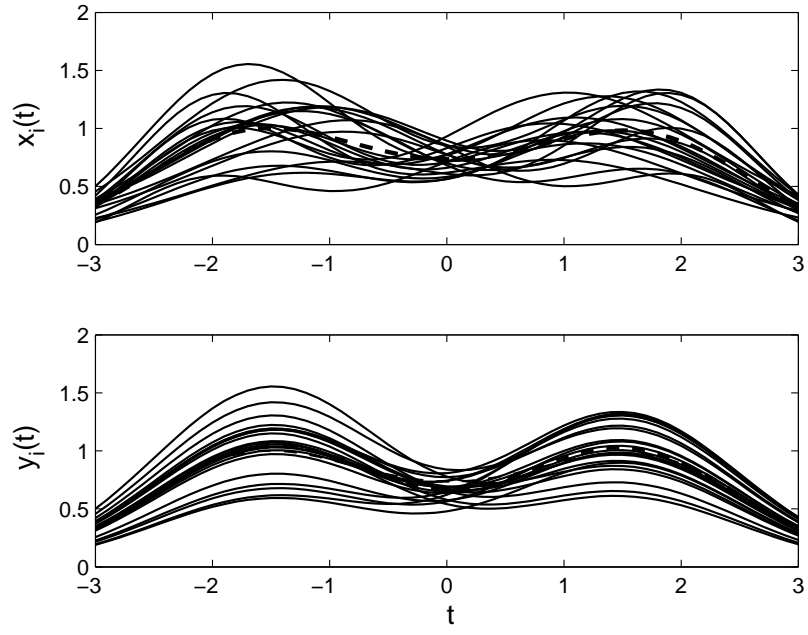
3

Figure 1: The top panel contains 21 random unregistered curves with a two-dimensional structure generated from curves (17) and warping functions (18). The bottom panel contains these curves after registration by the method described in this paper. The heavy dashed lines are the respective pointwise means of these curves.

described in this paper, account for 99.98% of the variation. In Section 2 we will relate these phase variation issues to the failure of unregistered curves to satisfy *convexity*.

Different registration procedures have been proposed in statistical literature. *Landmark registration* removes phase variation by monotonically transforming the domain for each curve so that points specifying the locations of shape features are aligned across curves. *Features* for this purpose may be maxima, minima and crossings of fixed thresholds, possibly defined at the level of one or more derivatives in addition to being visible in each curve. Landmark registration has been studied in depth by Bookstein (1978, 1991), Kneip and Gasser (1992) and Gasser and Kneip (1995).

Other methods not using landmarks have also been developed, partly in response to situations where landmarks are not clearly identifiable in all curves. Algorithms based on dynamic programming have been proposed in the engineering literature, see for example Sakoe and Chiba (1978). Ramsay (1996) and Ramsay and Li (1998) developed the fitting of a general and flexible family of warping functions h_i making use of a regularization technique. The procedure of Kneip, Li, MacGibbon and Ramsay (2000) analogous to local polynomial smoothing for identifying warping functions that register a sample of curves. Wang and Gasser (1997, 1998, 1999), Rønn (2001), and Gervini and Gasser (2004) have also contributed registration methods, and considered some important theoretical issues. Hall, Lee and Park (2007) proposed some summary statistics for amplitude and phase variation.

The success of registration methods has, nevertheless, tended to be assessed in terms of how well they align visible features. But from a conceptual perspective, how features are defined and in what sense they are aligned is rather less central than statistical interpretability. Generally speaking, the technology that we apply for registration must lead to data summaries and models that we consider useful. Based on this insight we formal-

5

ize the registration problem in Section 2 and develop some basic criteria characterizing methodologically consistent and statistically interpretable registration procedures.

Another central theme of the paper is the combination of registration and functional principal components analysis. Scientists often use principal components to identify fundamental modes of variation, as we did with Figure 1, and these components may be considered as a more general type of features,their amplitudes being identified with principal component scores. The hope is that removing phase variation will result in a relatively small number of principal component features doing a better job of accounting for curve to curve variation. In Section 3 we use a variance decomposition to quantify this improvement in fit, and, for example, show that 18% of the variation in the unregistered functions in Figure 1 is due to phase. Silverman (1995), assuming that phase variation only consists of individual time shifts $h(t) = t + \tau$, presented a first approach to this problem. Section 4 describes a computational procedure for registering curves to their principal components of variation. Sections 6 and 7 contain simulated and a real data applications, respectively.

2   Consistency of registration procedures

We consider observations consisting of a sample of smooth random functions $x_1, \ldots, x_n$ defined on a common interval, that we may take as $[0, 1]$ without losing generality. It is assumed that $D^m x_i \in \mathcal{L}^2[0, 1]$ for some $m \geq 1$, so that $x_i \in \mathcal{W}^m[0, 1]$ almost surely, where $\mathcal{W}^m$ is a Sobolev space of order m. Let $\mathcal{X} \subset \mathcal{W}^m[0, 1]$ denote the function space containing these observations, i.e. $P(x_i \in \mathcal{X}) = 1$ and additionally assume that the random functions possess bounded variation, $E[\sup_t x_i(t)^2] < \infty$.

As outlined in the introduction, registration aims to eliminate phase variation by

defining appropriate warping functions h. It then leads to a *synchronized* space $\mathcal{S}$ of registered functions $y = x \circ h$. In the following subsections we formalize this process and identify basic statistical properties to be satisfied by any sensible procedure.

2.1 Warping functions

A *warping function* h is an element of the *convex* space $\mathcal{H} \subset \mathcal{W}^m[0,1]$ of all strictly increasing functions such that $h(0) = 0$, $h(1) = 1$. The *functional inverse* h^{-1} with the property $h^{-1}[h(t)] = (h^{-1} \circ h)(t) = t$ for all t is uniquely defined, and the *identity warping function* $\mathcal{I}$ given by $\mathcal{I}(t) = t$ for all t acts as the unit element $\mathcal{H}$ for functional composition.

A *registration procedure* is formally defined as a measureable mapping $\mathcal{R}$ from $\mathcal{X}$ into $\mathcal{H}$ such that $\mathcal{R}(x) = h$. Either $\mathcal{X}$ or $\mathcal{H}$ may actually be differentiable to a higher degree, so that m is the minimum order of differentiability required of functions of the form $y = x \circ h$. Ramsay (1996) gives a general representation of the class of strictly monotonic differentiable functions.

Mapping $h_i := \mathcal{R}(x_i)$ maps a probability measure on $\mathcal{X}$ to a corresponding measure on the convex space $\mathcal{H}$, and we assume that $E[h_i]$ exists. It is not restrictive to require that a registration procedure is *standardized* in the sense that

$$E[h_i(t)] = E[\mathcal{R}(x_i)(t)] = t \ \text{ for all } \ t \in [0,1]$$

since, if $\mathcal{R}$ is such that $E[\mathcal{R}(x_i)] = \bar{h}$ for some $\bar{h} \in \mathcal{H}$ with $\bar{h} \neq \mathcal{I}$, then a standardized registration procedure $\mathcal{R}^*$ is given by $\mathcal{R}^*(x_i) = \mathcal{R}(x_i) \circ \bar{h}^{-1}$ which transforms x_i into the convex space $\{y \circ \bar{h}^{-1} | \ y \in \mathcal{S}\}$. Of course, when dealing with a finite sample of functions, standardizing a registration procedure has to refer to sample averages instead of expectations.
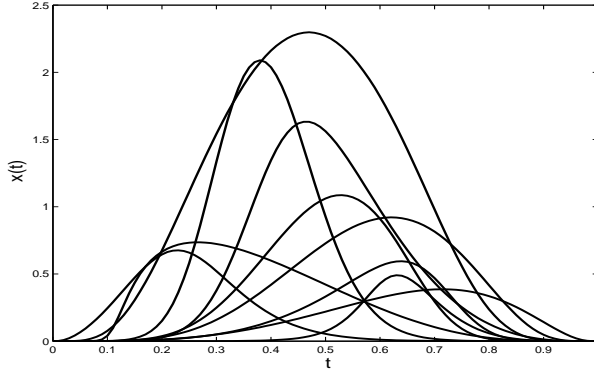
Figure 2: Bell-shaped functions from a non-convex space $\mathcal{X}$.

2.2   Convexity

In practice, the term "registration problem" is usually used in the context of functional data with the property that each curve possesses a well-defined sequence of shape features which are realized, but with substantial individual differences in timing and amplitude. One then generally has to cope with the situation that the corresponding function space $\mathcal{X}$ is not closed under linear operations, and in particular is *not a convex space*. Recall that some function space $\mathcal{S}$ containing functions y_i is said to be convex if, for any $\lambda \in [0, 1]$ and any (y_1, y_2), $\lambda y_1 + (1 - \lambda)y_2 \in \mathcal{S}$. To illustrate the point let us consider a simple example. Figure 2 shows smooth, positive curves x with a single peak located at some point τ_x and satisfying $x(0) = x(1) = 0$ as well as $x(t) > 0$ for $t \in (0, 1)$.

For random function x_i taking values in the corresponding space $\mathcal{X}$ of nonnegative single–peaked functions registration is *necessary* for any sensible statistical analysis. It is easily seen that this is not a convex function space: For two functions $x, x^* \in \mathcal{X}$ with $|\tau_x - \tau_{x^*}| \gg 0$ many linear combinations $\alpha x + \beta x^*$ will lead to curves with more than one extremum, and hence $\alpha x + \beta x^* \notin \mathcal{X}$. As a consequence, we run into the fundamental

8

problem already mentioned in the introduction that standard statistics based on weighted averages of functions x_i will often not provide interpretable results. In particular, the pointwise mean $\mu(t) = E(x_i(t))$ will often not be an element of $\mathcal{X}$.

The above arguments generalize to more complex situations. Generally speaking, registration deals with the situation that the functions x_i take values in some *non-convex* function space $\mathcal{X}$. A basic task of registration procedures then is to transform the functions in such a way that the synchronized space $\mathcal{S}$ of registered functions *is convex*. It will then be possible to analyze registered curves $y_i = x_i \circ h_i$ by calculating a statistically well-interpretable *structural mean* $\nu = E[y_i]$ and by assessing covariance structure via principal components. Note that this also applies to warping functions, since they are monotone and $\mathcal{H}$ is a convex space. The role of convexity has already been emphasized in Liu and Müller (2004), who present an interpretation in terms of bivariate processes, where for each $t \in [0, 1]$ registration assigns a pair $[h(t), y(t)]$ to the pair $[t, x(t)]$.

2.3 Identifiability

Identifiability is a central problem in registration. Unless we have specific knowledge about the data-generating process, there is no unique registration procedure, and when choosing a particular procedure one has to focus on statistical interpretability. For the same space $\mathcal{X}$, different forms of registration may lead to different warping functions and different synchronized spaces $\mathcal{S}$, using suitable restrictions on the structure of $h \in \mathcal{H}$ and/or $y = x \circ h \in \mathcal{S}$.

The identification problem is illustrated in Figure 3 for the simple space $\mathcal{X}$ of nonnegative single–peaked functions. The subset $\mathcal{S}_\tau = \{x \in \mathcal{X} |\; \tau_x = \tau\}$ of all $x \in \mathcal{X}$ possessing their unique extremum at τ is a *convex* subspace of $\mathcal{X}$. Without restriction one may require that $\tau = E(\tau_{x_i})$. For $x \in \mathcal{X}$ *any* $h \in \mathcal{H}$ with $h(\tau) = \tau_x$ will transform x into
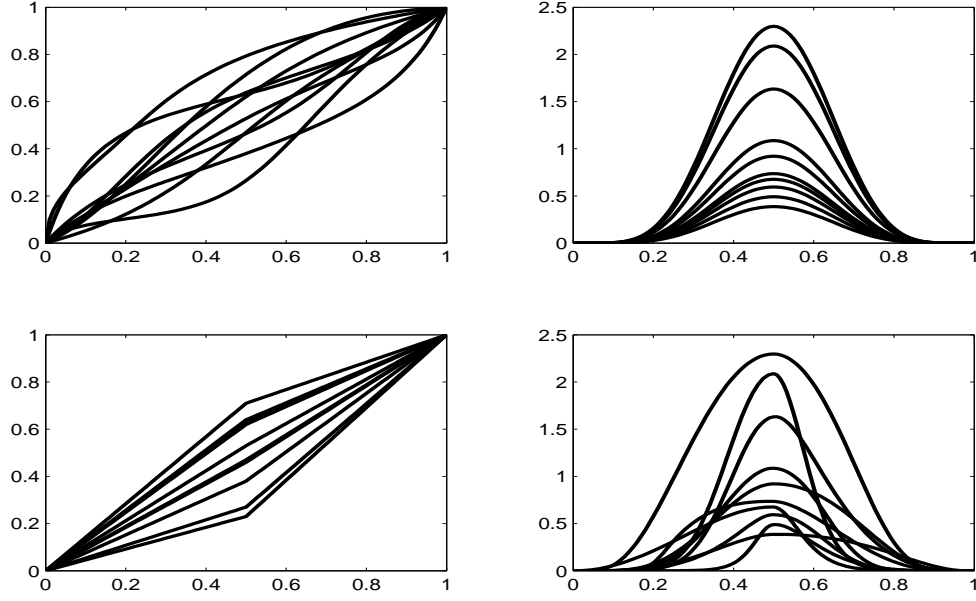
Figure 3: The upper panels contain warping functions on the left and functions in $\mathcal{S}_\xi$ on the right when registering the bell-shaped curves in Figure 2 with respect to a template. The lower panels contain warping functions obtained by piecewise linear interpolation of end and landmark values and functions in $\mathcal{S}_\tau$ when registering by landmark or peak alignment.

an element $y = x \circ h \in \mathcal{S}_\tau$. But the values of h in the intervals $(0, \tau)$ and $(\tau, 1)$ are arbitrary, so that identification may be obtained by a choice of a particularly simple h, as, for example, a piecewise linear interpolant of the points $(0, 0), (\tau, \tau_x), (1, 1)$. This is shown in the lower panels of Figure 3.

Template registration separates amplitude and phase variation by using single mean or template function ξ as a target for registered functions. Considering the space $\mathcal{X}$ of nonnegative single–peaked functions, it is easily checked that for arbitrary $\xi \in \mathcal{X}$ *any* other function $x \in \mathcal{X}$ can be represented in the form $x = \theta \cdot \xi \circ h^{-1}$ for some (unique) $h \in \mathcal{H}$ and $\theta = x(\tau_x)/\xi(\tau_\xi) \in \mathbb{R}^+$. Without restriction one may require that $E[h_i(t)] = t$ and $\xi = E(x_i \circ h_i)$. Then $\tau_\xi = E(\tau_{x_i})$. The synchronized function space $\mathcal{S}_\xi$ then consists of all functions positively proportional to ξ, $\mathcal{S}_\xi = \{\theta \xi | \ \theta \in \mathbb{R}^+\}$, and the sample space $\mathcal{X}$ is equivalent to the space $\mathcal{X}_\xi := \{x | \ x \circ h \in \mathcal{S}_\xi\}$.

The upper panels of Figure 3 show the effects of this registration strategy for a specific choice of ξ. When comparing synchronized spaces $\mathcal{S}_\tau, \mathcal{S}_\xi$ we see that $\mathcal{S}_\xi \subset \mathcal{S}_\tau$ if $\tau_\xi = \tau$. A procedure registering with respect to $\mathcal{S}_\xi$ may lead to very different results compared to registration with respect to $\mathcal{S}_\tau$. Alignment to $\mathcal{S}_\tau$ permits simplicity in the corresponding space $\mathcal{H}_\tau$ of warping functions, while registration to the simpler space $\mathcal{S}_\xi$ is at the expense of complexity in the respective space $\mathcal{H}_\xi$.

The above ideas can be generalized to more complex spaces $\mathcal{X}$. If all $x \in \mathcal{X}$ possess a sequence of k shape features, a natural generalization of $\mathcal{S}_\tau$ is a space $\mathcal{S}_{\tau_1,\ldots,\tau_k}$, where all features are aligned at some locations $\tau_1, \ldots, \tau_k$. Using suitable interpolation schemes to define warping functions, this is basically the strategy adopted by landmark registration.

Generalizing template registration one may tend consider spaces of functions x which can be represented in the form $x(h(t)) = \theta(t)\xi(t)$ for some non-constant amplitude functions $\theta(t)$. But such spaces are too complex to be useful for registration purposes. If for

11

example ξ is a positive function, then for *every* $h \in \mathcal{H}$ and all functions x the relation $x(h(t)) = \theta(t)\xi(t)$ holds with $\theta(t) = x(h(t))/\xi(t)$ and there is no way of identifying a suitable warping function.

A more useful and logical extension of $\mathcal{X}_\xi$ defines the core synchronized space $\mathcal{S}_{\xi_1,\ldots,\xi_p}$ to consist of p basic components of variation, $\xi_j, j = 1, \ldots, p$, along with linear combinations of these. Since this finite dimensional function space is closed under linear operations, it is also closed under convex linear operations. The space of observed functions can then be described by

$$\mathcal{X}_{\xi_1,\ldots,\xi_p} = \{x | (x \circ h) = \sum_j^p \theta_j \xi_j, h \in \mathcal{H}, \ \theta_j \in \mathbb{R}\}, \tag{2}$$

and the random curves presented in the top panel of Figure 1 provide an example with $p = 2$. Corresponding registration methods will be considered in Sections 4 and 5.

2.4 Consistency

We will say that a registration procedure $\mathcal{R}$ is *consistent* if the following conditions are satisfied:

i) $\mathcal{S} := \{x \circ \mathcal{R}(x) | \ x \in \mathcal{X}\}$ is a convex subspace of $\mathcal{X}$.

ii) $\mathcal{R}$ satisfies $\mathcal{R}[x \circ \mathcal{R}(x)] = \mathcal{I}$ for all $x \in \mathcal{X}$ and, hence, $\mathcal{R}(y) = \mathcal{I}$ for all $y \in \mathcal{S}$.

A standardized registration $\mathcal{R}^*$ is consistent if and only if $\mathcal{R}$ is consistent. The goals of registration require that a registration procedure be consistent in order to be statistically interpretable. The convexity condition i) for $\mathcal{S}$ is crucial for any further analysis, but condition ii) has not yet been discussed in the literature and is motivated by the identification problem. It ensures that there exists a well-defined space of functions y which are to be considered free of any phase variation, $\mathcal{S} = \{y \in \mathcal{X} | \ \mathcal{R}(y) = \mathcal{I}\}$.

We want to emphasize that there can be several consistent registration procedures leading to different synchronized spaces, and the two methods presented in Figure 3 provide examples. In a real data analysis the statistician will have to choose the approach which is best suitable in terms of subsequent interpretation of the respective results.

On the other hand, consistency is not a trivial requirement and will only be satisfied if registration procedure and structure of the sample space $\mathcal{X}$ fit together. Potential problems are illustrated by the failure of a well-known method which has frequently been used in practice: Template registration by least squares based on minimizing

$$\int_0^1 \left[x[h(t)] - \xi(t)\right]^2 dt \tag{3}$$

over all possible $h \in \mathcal{H}$. Least squares registration is consistent when no amplitude variation is present so that $\mathcal{X} = \{x | x \circ h = \xi\}$, but the presence of amplitude variation usually leads to warping functions that tend to pinch in the regions over which curves are non-zero (see for example Ramsay and Li, 1998). Indeed, if $\sup_t |x(t)| \neq \sup_t |\xi(t)|$ it can be shown that in general the infimum of (3) is attained by a function $\bar{h}$ possessing jumps and/or constant segments which is an element of the *closure* $\bar{\mathcal{H}}$ of $\mathcal{H}$.

A remedy proposed in the literature is to minimize (3) subject to all h in an appropriate (parametric or nonparametric) subspace $H \subset \mathcal{H}$ which only contains warping functions bounded by constraints, as for example $\int D^2 h(t)^2 dt \leq c$. Unfortunately, this approach necessarily leads to inconsistency of the warping procedure. The algorithm will tend to go as far as possible in direction of $\bar{h} \in \bar{\mathcal{H}}$ and choose a solution $\tilde{h}$ with $\int [D^2 \tilde{h}(t)]^2 dt \approx c$. When the original function x is replaced by the registered curve $y = x \circ \tilde{h}$, the algorithm will provide a new warping function $\tilde{h}^* \neq \mathcal{I}$ such that $\tilde{h} \circ \tilde{h}^*$ is still closer to the true infimum $\bar{h}$.

3 Quantifying amplitude and phase variation

We now assume that a standardized, consistent registration procedure has been defined, so that to each x there corresponds an h and a registered function y. Let $\mu = \mathrm{E}[x]$ denote the *cross-sectional* mean, and let the *structural* mean $\nu = \mathrm{E}[y]$. The *total variation* in $\mathcal{X}$ is

$$MS_{total} = \mathrm{E} \int [x(t) - \mu(t)]^2 \, dt.$$

We first define the constant C as a measure of the independence of warping functions h and the registered functions y:

$$C = \frac{\mathrm{E} \int x^2(t) \, dt}{\mathrm{E} \int y^2(t) \, dt} = \frac{\mathrm{E} \int Dh(t) y^2(t) \, dt}{\mathrm{E} \int y^2(t) \, dt} = 1 + \frac{\int \mathrm{Cov}[Dh(t), y^2(t)] \, dt}{\mathrm{E} \int y^2(t) \, dt}, \qquad (4)$$

using the change of variable of integration from t to $h(t)$, and noting that $\mathrm{E}[Dh] = 1$ for standardized registration as well as $\mathrm{E} \int Dh(t) y^2(t) \, dt = \mathrm{E} \int y^2(t) \, dt + \int \mathrm{Cov}[Dh(t), y^2(t)] dt$. We can conclude that $C = 1$ whenever $Dh(t)$ and $y^2(t)$ are uncorrelated for all t. This is of course necessarily true if Dh and y^2 are independent.

The total variance MS_{total} can be decomposed into a part representing the total variance of the registered curves and a second part explained by phase differences as follows:

$$
\begin{aligned}
MS_{total} &= \mathrm{E} \int [x(t) - \mu(t)]^2 \, dt \\
&= C\mathrm{E} \int y^2(t) \, dt - \int \mu^2(t) \, dt \\
&= C\mathrm{E} \int [y(t) - \nu(t)]^2 \, dt + C \int \nu^2(t) \, dt - \int \mu^2(t) \, dt. \qquad (5)
\end{aligned}
$$

The two terms the last line may be interpreted as

$$MS_{amp} = C\mathrm{E} \int [y(t) - \nu(t)]^2 \, dt \quad \text{and} \quad MS_{phase} = C \int \nu^2(t) \, dt - \int \mu^2(t) \, dt,$$

14

or the *total amplitude variation* and the *total phase variation* of x, respectively.

The amount of variance explained by either component will depend on the specific registration procedure. If no registration is applied, that is $h(t) = t$ for all i, then $C = 1$, $\mu = \nu$ and $MS_{phase} = 0$. The other extreme arises if $\mathcal{X} = \{x | x \circ h = \xi\}$ for some template function ξ. Then $y = \xi = \nu$ and $MS_{amp} = 0$. Furthermore, the functions $y = \nu$ and h then are trivially independent, and $C = 1$. Consequently, in this case

$$\mathrm{E} \int [x(t) - \mu(t)]^2 \, dt = \int \nu^2(t) \, dt - \int \mu^2(t) \, dt.$$

However, faulty registration may lead to $MS_{phase} \leq 0$. In order to clarify this point let us consider a simple example. Suppose that all sample curves are equal, $P(x = \mu) = 1$, and assume that warping functions h are assigned by randomly selecting an element of some arbitrary, finite subset $\{h^{(1)}, \ldots, h^{(m)}\} \subset \mathcal{H}$, $m > 1$. Of course, this absurd approach does not define a consistent registration procedure, but the functions h can be standardized and formally (5) holds for some $C > 0$. Instead of explaining variability, this method obviously *creates* some artificial variation of $y = x \circ h = \mu \circ h$. We obtain $MS_{amp} > MS_{total} = 0$ and $MS_{phase} < 0$.

On the other hand, in view of the discussion in Section 2 it does not make much sense to look for procedures for which sole aim is to maximize the explanatory power of MS_{phase}. For example, the least squares criterion (3) will try to maximize MS_{phase}, but it cannot be considered as a satisfactory registration procedure in general.

4   Registration to principal components

Functional principal component analysis is a powerful tool for describing the amplitude variation of random functions, and is often the next step after applying a registration procedure. Let $\lambda_1 \geq \lambda_2 \geq \ldots$ and $\zeta_1, \zeta_2, \ldots$ denote eigenvalues and corresponding orthonor-

mal eigenfunctions or *principal components* of the covariance operator of y_i, respectively.

Let principal component score $c_{ij} = \int [y_i(t) - \nu(t)]\zeta_j(t)\, dt$, and the approximation of y_i on the basis of the first p principal components be indicated by

$$\nu_i^p(t) = \nu(t) + \sum_j^p c_{ij}\zeta_j(t).$$

For any p the error in the approximation satisfies

$$\mathrm{E}\int [y_i(t) - \nu_i^p(t)]^2\, dt = \sum_{j>p} \lambda_j \leq \mathrm{E}\min_{\alpha_{i1},\dots,\alpha_{ip}} \int [y_i(t) - \nu(t) - \sum_{r=1}^p \alpha_{ir}\psi_r(t)]^2\, dt \qquad (6)$$

for any alternative choice of basis functions $\psi_1,\dots,\psi_p \in \mathcal{L}^2[0,1]$. Moreover, $\mathrm{E}[\nu_i^p(t)] = \nu(t)$ and $\mathrm{E}\int [\nu_i^p(t) - \nu(t)]^2\, dt = \sum_j^p \lambda_j$.

These relations imply the following modification of (5)

$$\mathrm{E}\int [x_i(t) - \mu(t)]^2\, dt$$

$$= C\mathrm{E}\int [\nu_i^p(t) - \nu(t)]^2\, dt + C\int \nu(t)^2\, dt - \int \mu(t)^2\, dt$$

$$+ C\mathrm{E}\int [y_i(t) - \nu_i^p(t)]^2\, dt, \qquad (7)$$

and one can conclude that the total variance MS_{total} of x can be decomposed into

1) $MS_{amp} = C\mathrm{E}\int [\nu_i^p(t) - \nu(t)]^2\, dt$, the portion of the total variance of x_i explained by the first p principal components of y_i,

2) $MS_{phase} = C\int \nu(t)^2\, dt - \int \mu(t)^2\, dt$, the portion of the total variance of x explained by phase variation, and

3) $MS_{res} = C\mathrm{E}\int [y_i(t) - \nu_i^p(t)]^2\, dt$, the unexplained residual variance.

The proportion of variation explained by phase can be expressed by either of

$$R_1^2 = \frac{MS_{phase}}{MS_{total}} \quad \text{or} \quad R_2^2 = \frac{MS_{phase}}{MS_{amp} + MS_{phase}}. \qquad (8)$$

Decompositions (5) as well as (7) may be seen as useful tools for interpreting the results of *any* consistent registration procedure. However, the following proposition shows that, in many applications, one may find a registration procedure such that the unexplained residual variance in (7) is negligible when using a *small* number of principal components of the registered functions y. The notation $\#\text{zeros}(Dx)$ will denote the number of points $\tau \in [0,1]$ with $Dx(\tau) = 0$, while $\#\text{feat}(x)$ will denote the number of strict local minima/maxima corresponding to isolated zero crossings of Dx. Note that by requiring $\#\text{zeros}(Dx_i) < \infty$ we assume that sample curves are piecewise strictly monotone without flat parts.

Proposition 1: Assume that the random functions x_i are such that with probability 1, $\#\text{feat}(x_i) \leq k$ as well as $\#\text{zeros}(Dx_i) \leq k$ for some $0 < k < \infty$. We then obtain:

a) For some $p \leq k + 2$ there exist some smooth functions $\xi_1, \ldots, \xi_p \in \mathcal{W}^m[0,1]$ and a corresponding function space

$$\mathcal{X}_{\xi_1,\ldots,\xi_p} = \{\sum_j^p \theta_j \xi_j \circ g| \; g \in \mathcal{H}, \theta_1, \ldots, \theta_p \in \mathbb{R}\} \tag{9}$$

such that $x_i \in \mathcal{X}_{\xi_1,\ldots,\xi_p}$ with probability 1.

b) For the space $\mathcal{X}_{\xi_1,\ldots,\xi_p}$ given by (9) define a registration procedure $\mathcal{R}$ in the following way: For $x \in \mathcal{X}_{\xi_1,\ldots,\xi_p}$ the warping function $\mathcal{R}(x)$ is the element $h \in H_x$ with the smallest value of $\int_0^1 [h(t) - t]^2 dt$, where H_x is the set of all $h \in \mathcal{H}$ satisfying

$$\int_0^1 \min_{c_1,\ldots,c_p} \left(x[h(t)] - \sum_j^p c_j \xi_j \right)^2 dt = 0. \tag{10}$$

Then $\mathcal{R}$ is a consistent registration procedure, and the synchronized space is given by $\mathcal{S}_{\xi_1,\ldots,\xi_p} = \{\sum_j^p \theta_j \xi_j| \; \theta_1, \ldots, \theta_p \in \mathbb{R}\}$.

A proof is given in Appendix A.

We can conclude that under the conditions of the proposition there exists a consistent registration procedure such that for some $p \leq k + 2$ the registered functions y_i can be written as

$$x_i[h_i(t)] = y_i(t) = \sum_j^p \theta_{ij}\xi_j(t), \quad t \in [0, 1], \tag{11}$$

for some random coefficients $\theta_{ij} \in \mathbb{R}$. But $\mathcal{S}_{\xi_1, \ldots, \xi_p}$ is a p-dimensional linear space and by (6) it is also spanned by the first p principal components $\zeta_1, \ldots, \zeta_p$ of y. The registered curves can therefore be alternatively represented by

$$x_i[h_i(t)] = y_i(t) = \nu(t) + \sum_j^p c_{ij}\zeta_j(t), \quad t \in [0, 1], \tag{12}$$

with $\nu(t) = \mathrm{E}[y_i(t)] = \sum_j^p E(\theta_{ij})\xi_j(t)$ and $MS_{res} = 0$. This provides a decomposition of the variance of curve samples into amplitude variation, modeled by principal components, and phase variation, modeled by different warping functions. Note that generally $\zeta_j \neq \xi_j$. Theoretically appropriate functions ξ_j may be determined from the eigenfunctions of the moment operator with kernel $m_2(t, s) = \mathrm{E}[y_i(t)y_i(s)]$, while, as explained above, the functions ζ_j correspond to the eigenfunctions of the covariance operator with kernel $\sigma(t, s) = E\left[(y_i(t) - \nu(t))(y_i(s) - \nu(s))\right]$. However, if $\nu \in span\{\zeta_1, \ldots, \zeta_p\}$ both representations are equivalent.

The essential condition of Proposition 1 is a bounded number $k < \infty$ of first order features. This is not a trivial assumption. Many standard random functions considered in theoretical literature do not possess this property. For example, the number of possible maxima/minima of Brownian motion paths is unbounded.

In practice, a statistician may look for approximate solutions of (12). These may be found for different values of p. Indeed, as already demonstrated in Figure 3 there will exist different consistent procedures, and a statistician may have the choice between

a smaller value of p at the expense of more complicated warping functions, or a larger value of p together with simpler warping functions. We want to emphasize, however, that an (approximate) solution (12) with respect to a *small* number p of components has the attractive feature of providing a qualitative model explaining the common structure of all $x = y \circ h^{-1}$. The sequence of first order features in all functions x reflects the structure of the corresponding registered functions y, since the monotone transformations h^{-1} preserve the number and amplitudes of the minima/maxima of y.

5 An algorithm for registering to principal components

Our algorithm registers data to decomposition (12). Calculating the components of (12) from data $x_1, \ldots, x_n$, requires the simultaneous calculation of warping functions h_i, mean function ν, principal component scores c_{ij} and functional principal components ζ_j. Our strategy is to iteratively construct the target functions y_i corresponding to a sample of unregistered functions x_i through successive fittings of models of the form (12).

Within iteration number ℓ, $\nu^{(\ell)}$ is first estimated by $n^{-1} \sum_i y_i^{(\ell-1)}$, and then a principal components analysis of the centered functions $y_i^{(\ell-1)} - \nu^{(\ell)}$ is followed by a registration procedure. An incremental warping function $h_{inc,i}^{(\ell)}$ is determined by registering $y_i^{(\ell)} = y_i^{(\ell-1)} \circ h_{inc,i}^{(\ell)}$ to the model value $\hat{y}_i^{(\ell)} = \nu^{(\ell)} + \sum_j^p c_{ij}^{(\ell)} \zeta_j^{(\ell)}$, and $h_i^{(\ell)} = h_i^{(\ell-1)} \circ h_{inc,i}^{(\ell)}$. The amount by which the *deformation function d* value $d_i^{(\ell)}(t) = h_{inc,i}^{(\ell)}(t) - t$ departs from 0 is deliberately kept small, in order to ensure that both the principal component structure and the warping functions evolve smoothly over course of the iterations.

However, the total registration for each curve is accumulated over these local registrations, so that the final total registration can be as large as is required. Once the

estimation of the mean, principal component analysis and its associated registration have been computed, and the decomposition (5) has been evaluated, iteration ℓ is completed, and we then repeat this entire process on iteration $\ell + 1$ until we note that the deformations that are being computed are all of negligible size and that the proportion of variance accounted for $R^2 = MS_{phase}/MS_{total}$ is no longer increasing.

The following pseudo-code specifies this iterative process. It involves calls to three procedures, $DECOMP, PCAAPPROX$, and $REGIST$, and these are described in detail following the pseudo-code.

Iteration 0 :

$$\mathbf{y}^{(0)} = \mathbf{x}$$

$$\mathbf{h}^{(0)} = \mathcal{I}$$

Iteration $\ell > 0$:

$$\nu^{(\ell)} = n^{-1} \sum_i y_i^{(\ell-1)}$$

$$\hat{\mathbf{y}}^{(\ell)} = \nu^{(\ell)} + PCAAPPROX(\mathbf{y}^{(\ell-1)} - \nu^{(\ell)})$$

$$\mathbf{y}^{(\ell)}, \mathbf{h}^{(\ell)} = REGIST(\mathbf{y}^{(\ell-1)}, \hat{\mathbf{y}}^{(\ell)}, \mathbf{h}^{(\ell-1)})$$

$$MS_{amp}^{(\ell)}, MS_{pha}^{(\ell)}, MS_{res}^{(\ell)} = DECOMP(\mathbf{y}^{(\ell)}, \hat{\mathbf{y}}^{(\ell)}, \mathbf{x})$$

Test for convergence and exit if achieved.

Alternatively, we may start the process at iteration 0 by the using $\mathbf{y}$ and $\mathbf{h}$ from a previous registration step, and, for example, the simplicity of the landmark registration process when a few key landmarks can be identified in all curves can be well worth taking advantage of, as we shall do below in Section 7.

5.1 The registration step $REGIST$

The heart of the registration process is the technique used update a function $y_i^{(\ell-1)}$ by reducing its phase variation to produce a registered version $y_i^{(\ell)} \approx \hat{y}_i^{(\ell)}$. We focus on incremental warping functions generating small deformations $|d_i^{(\ell)}(t)|$, where $d_i^{(\ell)}(t) = h_{inc,i}^{(\ell)}(t) - t$. We then have the first order approximation at t

$$
\begin{aligned}
y_i^{(\ell)}(t) &= y_i^{(\ell-1)}[h_{inc,i}^{(\ell)}(t)] \approx y_i^{(\ell-1)}(t) + Dy_i^{(\ell-1)}(t)[h_{inc,i}^{(\ell)}(t) - t] \\
&= y_i^{(\ell-1)}(t) + Dy_i^{(\ell-1)}(t)d_i^{(\ell)}(t),
\end{aligned}
$$

or, alternatively that

$$
\hat{y}_i^{(\ell)}(t) - y_i^{(\ell-1)}(t) \approx Dy_i^{(\ell-1)}(t)d_i^{(\ell)}(t), \tag{13}
$$

We represent deformation $d_i^{(\ell)}$ by the basis function expansion

$$
d_i^{(\ell)}(t) = \sum_{s=1}^{L} d_{is}^{(\ell)}\psi_s(t) = (\mathbf{d}_i^{(\ell)})'\boldsymbol{\psi}(t) \tag{14}
$$

where, because of the restrictions $h_{inc,i}^{(\ell)}(0) = 0$ and $h_{inc,i}^{(\ell)}(1) = 1$, it is required that for each basis function $\psi_\ell(0) = \psi_\ell(1) = 0$. This can be achieved in practice by using a B-spline basis where the first and last basis functions are either discarded or where their coefficients are kept at zero. We must also ensure that $h_{inc,i}^{(\ell)}(t) = d_i^{(\ell)}(t) + t$ is strictly increasing and only takes values within $(0, 1)$, which in turn is ensured if $d_i^{(\ell)}(t)$ is sufficiently small for all t.

The approximate decomposition (13) can be viewed as a varying coefficient functional linear model where the residual function $\hat{y}_i^{(\ell)}(t) - y_i^{(\ell-1)}(t)$ is being predicted by a regression on covariate $Dy_i^{(\ell-1)}$, and where deformation $d_i^{(\ell)}$ is the varying regression coefficient. The varying regression coefficient function $d_i^{(\ell)}$ is estimated by minimizing the integrated squared error

$$
\text{SSE}(\mathbf{d}_i^{(\ell)}) = \int_0^1 [\hat{y}_i^{(\ell)}(t) - y_i^{(\ell-1)}(t) - Dy_i^{(\ell-1)}(t)\boldsymbol{\psi}'(t)\mathbf{d}_i^{(\ell)}]^2 \, dt \tag{15}
$$

21

with respect to deformation coefficient vector $\mathbf{d}_i^{(\ell)}$. We ensure that $d_i^{(\ell)}$ is sufficiently small by appending to this criterion a penalty term that penalizes its departure from zero as follows:

$$\texttt{PENSSE}(\mathbf{d}_i^{(\ell)}|\lambda) = \texttt{SSE}(\mathbf{d}_i^{(\ell)}) + \lambda \int_0^1 [D^2 \boldsymbol{\psi}'(t)\mathbf{d}_i^{(\ell)}]^2 \, dt. \tag{16}$$

The minimizing value of $\mathbf{d}_i^{(\ell)}$ can be expressed analytically, and therefore computed without iteration. Software for this in the languages R, S-PLUS and Matlab can be downloaded through the internet site `www.functionaldata.org`.

The value of the smoothing parameter λ plays a critical role in this algorithm since it determines how slowly the algorithm approaches an optimal solution. If λ is too small, then the $R^2 = MS_{phase}/MS_{total}$ measure (8) will increase for only a few iterations, and then begin to decrease and may even go negative. Some experimentation is required to identify the value of λ that achieves a smooth increase to a maximizing value in an acceptable number of iterations.

The partition of variance into amplitude and phase and residual variation in procedure $DECOMP$ that concludes each major iteration ℓ is described in detail in Section 3. Since the result is a set of summary statistics, high accuracy in computing the integrals involved is not required, and simple methods such as the trapezoidal rule with a fine equally spaced mesh of values has been found to be entirely satisfactory.

6 A simulated data illustration

We generated 21 curves over the interval $[-3, 3]$ of the form

$$y_i^*(t) = z_{i1} \exp[(t - 1.5)^2/2] + z_{i2} \exp[(t + 1.5)^2/2] \tag{17}$$

where the coefficients z_{i1} and z_{i2} were randomly generated from the distribution $N(1, 0.25^2)$. The associated warping functions h_i were

$$h_i(t) = 6\frac{\exp[a_i(t+3)/6] - 1}{\exp[a_i] - 1} - 3, \quad a_i \neq 0, \quad \text{and} \quad t \quad \text{otherwise.} \tag{18}$$

The coefficients a_i were equally spaced between -1 and 1. The curves $x_i(t) = y_i[h_i(t)]$ are shown in the top panel of Figure 1.

The basis functions ϕ_k used to represent the unregistered curves, the registered curves and the principal component curves ζ_j consisted of 23 cubic B-spline basis functions defined by equally spaced knots. The basis functions ψ_ℓ used in (14) were the central five B-spline basis functions from seven cubic B-spline functions with equally spaced knots. All seven basis functions were used to represent the warping functions h_i. The smoothing parameter λ in (16) was set to 10.

Figure 4 shows a variety of measures of performance over 13 iterations. The upper right panel shows the root-mean-squared measure

$$RMSE = \sum_i \int [y_i(t) - x_i(t)]^2 \, dt \tag{19}$$

assessing how closely the true registered curves have been reconstructed by this joint principal components and registration process. The amount of variation accounted for by amplitude, shown in the upper right panel, increases slightly, while that accounted for by phase in the middle right penal increases across all iterations. The $R^2 = MS_{phase}/(MS_{total})$ measure (8) reaches a maximum of 0.18 in the lower left panel. The registered curves are shown in the bottom panel of Figure 1.

We also registered the curves using landmark registration, using as landmarks the locations of the first maximum, the minimum, and the second maximum. The basis for the warping functions consisted of five order 2 or polygonal B-splines, which was capable of interpolating the landmark locations for each curve.
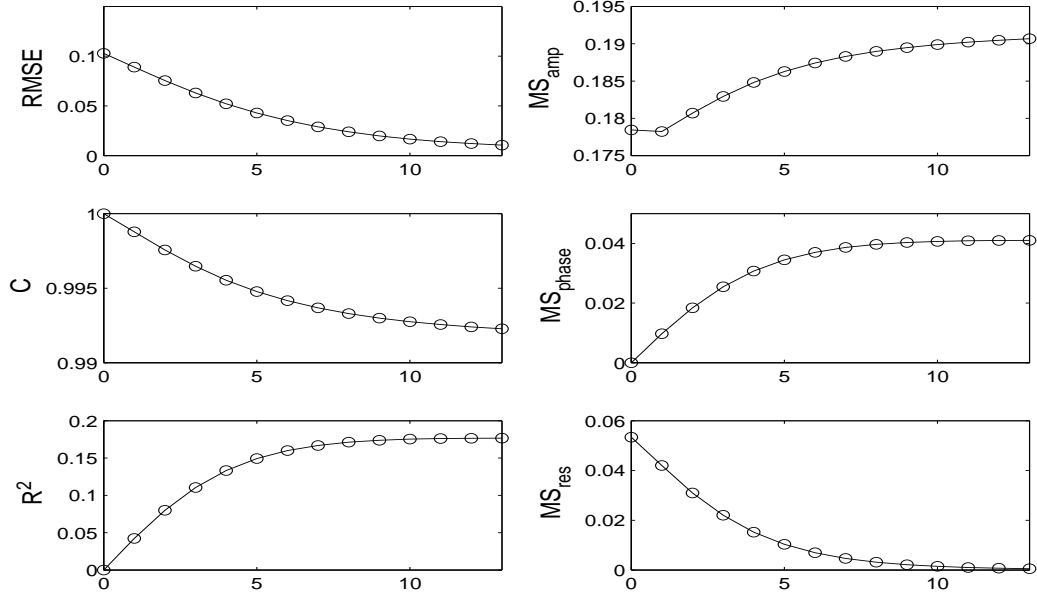
23

Figure 4: Six measures of the performance of the algorithm over 13 iterations. The upper left panel shows the root-mean-squared-error measure (19) of fit to the true registered curve. The center left panel shows the change in the constant C defined in (4). The lower left panel shows the squared correlation measure (8) of the proportion of variation due to phase. The right panels show the three mean square measures for amplitude, phase, and residual, respectively, that make up decomposition (7).

We then assessed the recovery of the curves y_i^* by the squared correlation measure

$$R^2 = \frac{1}{21} \sum_i 1 - \frac{\int [y_i(t) - y_i^*(t)]^2 dt}{\int [y_i^*(t)]^2 dt}.$$

These measures were 0.9997, 0.9982 and 0.9835 for the registration to principal components, landmark registration and the unregistered curves, respectively. The landmark registered curves displayed distortions over regions between landmark locations that were easily visible in plots. The corresponding measures of the recovery of the principal components were 1.0000, 0.9999 and 0.9986 for the first component, and 1.0000, 0.9683 and 0.9654 for the second component, respectively. That is, a principal components analysis of either the landmark registered curves or even the unregistered data identified the first principal component rather well, but the second principal component was only well captured by the registration to principal components algorithm.

Differences between landmark registration and the principal component method are easily explained when looking at the bottom panel of Figure 1, where registered curves still show some variation in the locations of the local minimum around 0. Indeed, location of this minimum depends on the specific linear combination of the two generating components. Differences in locations are thus partially due to *amplitude* differences and not to phase. By definition, landmark registration *exactly* aligns all extrema and thus in this example slightly exaggerates phase differences which results in comparably larger deformations $|h_i(t) - t|$.

7   Principal Components of Registered Handwriting

Ramsay (2000) reported on the analysis of 50 samples of the production of the Chinese script for "statistical science," and we illustrate our combined principal components and registration procedure for these data, which consist of the position of an infra-red detector

attached to the tip of the pen recorded 400 times a second by an OPTOTRAK motion capture system. The objective was information on how the brain's motor control center controls complex movement unfolding rapidly in time, with a special focus on pen tip acceleration since in biological systems, as in physical systems, the expenditure of energy is reflected in the second derivative.

The single sample shown in Figure 5 indicates three specified points at which the pen is lifted off the writing surface as a peak in the vertical Z coordinate, and also shows that these peaks occur early relative to the average time of their occurrences. The phase variation in these and other landmarks are a notable feature in the data, which also show substantial shape variation in the $X - Y$ coordinates over samples.

Each record was smoothed using 605 order 6 B-splines with a penalty on the $\mathcal{L}^2$ norm of the fourth derivative defined by a smoothing parameter value of 10^{-12}. The second derivatives $D^2 X(t)$ and $D^2 Y(t)$ were estimated by the derivatives of the spline approximations, and the tangential acceleration computed as $A(t) = \sqrt{[D^2 X(t)]^2 + [D^2 Y(t)]^2}$. The bottom panel of Figure 5 displays the variation in tangential acceleration along the pen trajectory, with some scripts displaying accelerations of more than 30 meters per second per second, or three times the force of gravity. If sustained, this would put a satellite in orbit in about six minutes. But we also see that these instants of high acceleration are often separated by near zero values where the pen is effectively in free or ballistic motion in both coordinates, and the peak times also appear to be nearly equally spaced, suggesting a strong stationary harmonic component.

Since the data offer several landmarks that are easily visible in each record, it is reasonable to begin with the simpler landmark registration procedure, and we registered the three peaks in the Z-coordinate identified in Figure 5. The acceleration curves were then registered using the warping functions $h_i(t)$ estimated by the landmark reg-
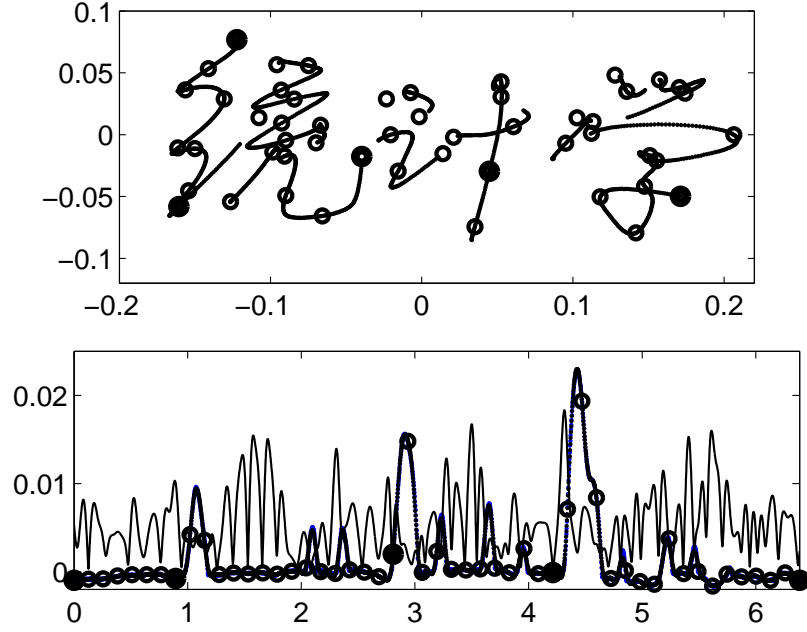
Figure 5: The top panel contains the $X - Y$ coordinates for a sample of Chinese script for "statistical science" in meters. The omitted portions correspond to points with the vertical coordinate Z above a fixed threshold. There are 50 strokes in the script, and circles are placed at 51 equally spaced time points. The heavy solid circles correspond to average times at which three specified peaks in the Z-coordinate occur. The lower panel contains the Z coordinates along with the circles indicating the same events shown in the top panel. The lighter solid line is the horizontal acceleration of the pen along the trajectory in the top panel, here divided by 1000 so as to fit on the plot.

istration. The top two panels in Figure 6 show all 50 acceleration curves before and after landmark registration. We note that the curves become relatively well registered in the vicinity of the average landmark locations 0.8, 2.6 and 4.0 seconds, but that there remains substantial phase variation elsewhere. The variance decomposition yields $C = 1.009, MS_{amp} = 18.4, MS_{phase} = 37.4$ and $R^2 = 0.67$. Although it may seem surprising that about 2/3 of the variation in the top panel of Figure 6 is due to phase, in fact, the middle panel indicates that registered portions of the acceleration records show rather small amplitude variation.

The functions $d_i(t)$ in (14) were expanded in terms of the central 21 of 23 order 4 B-splines with equally spaced knots, so that $d(0) = d(6) = 0$. The $\mathcal{L}^2$ norm of the estimated deformation functions was penalized with a smoothing parameter value of 1000. The variance decomposition yielded $C = 0.986, MS_{amp} = 33.3, MS_{phase} = 14.4$ and $R^2 = 0.30$.

The bottom panel of Figure 6 shows the accelerations registered in this way to three principal components, and we see that the curves are now registered over the entire scripts, and that the amount of amplitude variation for well-defined peaks is remarkably small, so that an estimate of 30% of the original variation being in phase seems reasonable.

Figure 7 plots the common logs of the first 10 eigenvalues for the unregistered, landmark registered and completely registered accelerations. It is common for higher order log-eigenvalues in functional PCA to decay linearly, and for dominant modes of variation to show up as initial points well above this linear trend. We see that the unregistered curves have one dominant component, and inspection of the mean curve perturbed by a multiple of this first eigenfunction shows that this variation captures only how late or early the prominent acceleration peaks are collectively; that is, it captures a simple global phase shift extending over most of a curve. However, the log–eigenvalue distri-
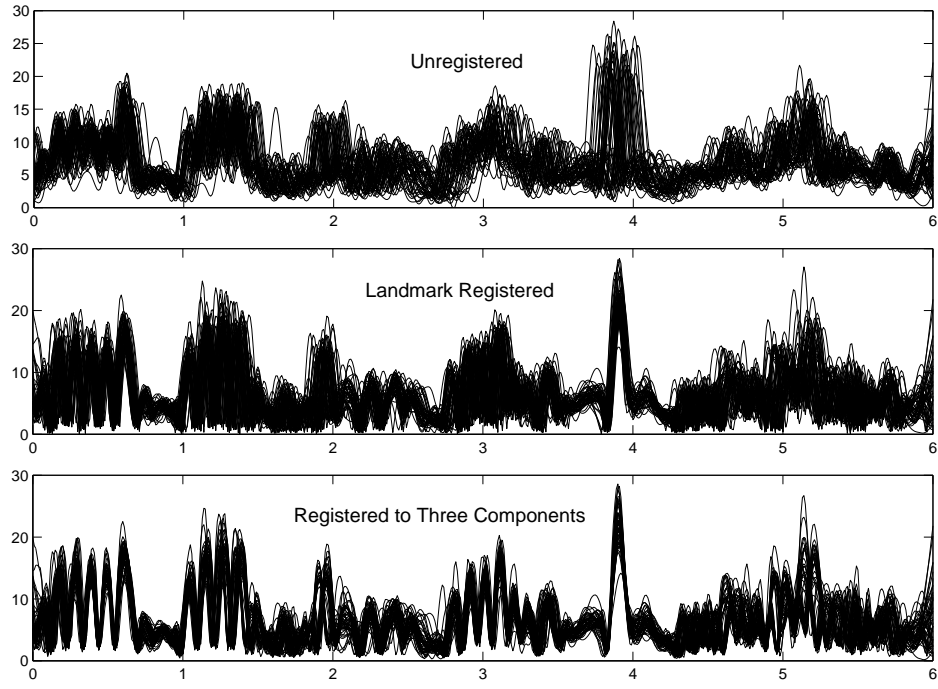
28

Figure 6: The top panel displays the 50 tangential curves prior to registration, the middle panel shows the curves registered using three landmarks in the vertical Z coordinate, and the bottom panel shows the curves after a further registration using the joint registration–principal components analysis described in this paper.
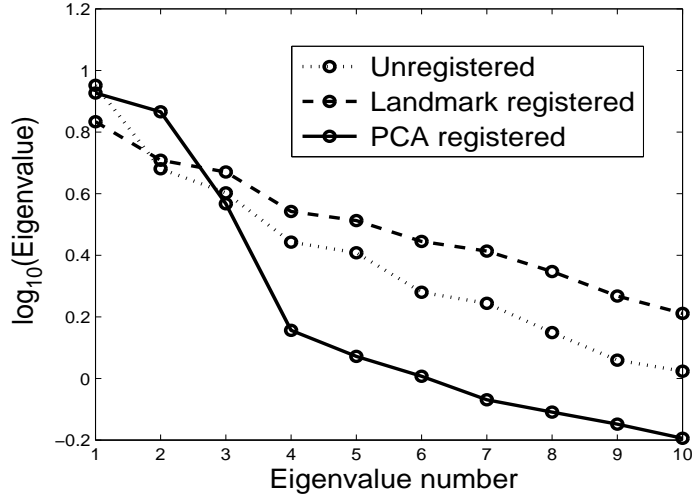
Figure 7: The distribution of the common log of eigenvalues for the three stages of registration.

bution at the landmark-registered phase does not suggest any clear break that would imply a small number of dominant eigenvalues. In other words, landmark registration has merely removed the first mode of variation in the unregistered curves. But three dominant eigenvalues emerge when our algorithm is used, and proportion of variation accounted for by these is 0.57, as opposed to 0.42 for the unregistered curves and 0.32 for landmark registration.

8 Discussion and Conclusions

We have attempted to shed some light on the identifiability of the amplitude/phase partition of variation, and find that registration is necessary if sample functions must be assumed to take values in a non-convex function space. By selecting a specific registration procedure the statistician decides which parts of the variation between curves are

attributed to phase and amplitude. In this context, consistency of a registration procedure seems to be a *necessary* condition for statistical interpretability. However, there will usually exist several consistent procedures leading to different registration results. An important issue of future research is the development of additional criteria to allow us to say that one method is preferable to another.

Given a registration procedure, we have developed a variance decomposition in Section 3 that we have found useful for quantifying the amount of variation in functional data due to phase variation, and may be used as a test for the presence of phase variation. Further work will be necessary to clarify effects of potential correlations between warping and registered functions determining the value of the constant C.

The results of Section 4 show that registering to principal components can be useful registration procedure. Indeed, (12) provides a qualitative model for the data which defines what we call amplitude variation, and explains a typical sequence of peaks and valleys to be found in all sample curves. In this approach principal components themselves may be interpreted as common features whose timings or location can vary from replication to replication.

The heart of the algorithm in Section 5 is the locally linear decomposition (13). The algorithm, like other registration algorithms such as Kneip, et al. (2000), relies on an alternation between fitting the model and registering the data to the fitted model. For example, the Procrustes procedure often used for registering functional data to a common mean function also proceeds by a preliminary estimation of a cross-sectional mean (this being the model), a registration of the observed functions to this mean, a re-estimation of the cross-sectional mean of the registered data, and repetition of the registration process, and so forth.

Our registration algorithm may also be employed with any model that captures a

substantial part of the amplitude variation in the observed curves. That is, we may define as features treatment effects in an experimental design, contributions from functional or multivariate covariates in a regression model and so on. Moreover the empirical basis functions that principal components analysis defines may be replaced by functions spanning the null space of a differential operator in principal *differential* analysis (Ramsay and Silverman, 2005). The algorithm, in effect, provides a way of folding the registration process into virtually any functional data analysis.

Although the registration algorithm appears to work well in the problems that we considered, it seems likely that better approaches can be developed. It should be possible to estimate the principal components and the registration warpings jointly and achieve a smoother and more rapid progress to an optimal solution that is not so sensitive to the smoothing parameter value in (16). We are also uneasy about the use of least squares as a fitting criterion in this registration context, which we show can be inconsistent in Section 2.

We hope that our work will further extend the concept of registration as well as the methodology for achieving it, and thus lead to models allowing for both the amplitude and phase variation that we see in so many of the functional data that we have considered.

9   Appendix A

Proof of Proposition 1: Consider smooth functions x in a Sobolev space $\mathcal{W}^m[0, 1]$ with the property that there exist exactly $k_x \leq k$ points $0 < \tau_1(x) \leq \tau_2(x) \leq \tau_{k_x}(x) < 1$ which are locations of first order features of x. Let $a_0(x) = x(0)$, $a_j(x) = x[\tau_j(x)]$, $j = 1, \ldots, k_x$, and $a_{k_x+1}(x) = x(1)$.

For $p = k + 2$ let $\xi_1, \xi_2, \ldots, \xi_p \in \mathcal{L}^2([0, 1])$ denote a sequence of smooth, $(m + 1)$-times

continuously differentiable basis functions with the property that, for all x with the above property and any possible sequence $a_0(x), a_2(x), \ldots, a_{k_x+1}(x)$ there exists a combination $\alpha_1(x), \ldots, \alpha_p(x)$ of coefficients such that the function

$$z_x(t) = \sum_{r=1}^{p} \alpha_r(x)\xi_r(t)$$

possesses exactly k_x first order features located at some points $0 < \tau_{z,1} \leq \ldots, \leq \tau_{z,k_x} < 1$, and such that $z_x(0) = a_0(x)$, $z_x(\tau_{z,1}) = a_j(x)$, $j = 1, \ldots, k_x$, as well as $z_x(1) = a_{k_x+1}(x)$. An example is a suitable basis of all polynomials of order p on $[0,1]$.

With $\tau_0(x) = \tau_{z,0} = 0$ and $\tau_{k_x+1}(x) = \tau_{z,k_x+1} = 1$, the functions x and z are strictly monotonic within each interval $[\tau_{j-1}(x), \tau_j(x)]$ and $[\tau_{z,j-1}, \tau_{z,j}]$, $j = 1, \ldots, k_x + 1$, respectively. Hence, for each $j = 1, \ldots, k_x + 1$ there exists a strictly monotonic function $z_{x;j}^{-1} : [a_{j-1}(x), a_j(x)] \to [\tau_{z,j-1}, \tau_{z,j}]$ such that $z_{x;j}^{-1}[z_x(t)] = t$ for all $t \in [\tau_{z,j-1}, \tau_{z,j}]$. Defining then $g_j : [\tau_{j-1}(x), \tau_j(x)] \to [\tau_{z,j-1}, \tau_{z,j}]$ by $g_j(t) = z_{x;j}^{-1}[x(t)]$, the above construction implies that with $g(t) = \sum_{j}^{k_x+1} g_j(t)I(t \in [\tau_{z,j-1}, \tau_{z,j}])$ we obtain

$$x(t) = z_x[g(t)] = \sum_{r=1}^{p} \alpha_r(x)\xi_r[g(t)] \text{ for all } t \in [0,1]$$

g is a strictly monotonic function, and $x \in \mathcal{W}^m[0,1]$ together with $(m+1)$-times continuous differentiability of z_x translates into $g \in \mathcal{W}^m[0,1]$. Note that $Dg(t) = \frac{Dx(t)}{Dz_x[g(t)]}$ for $t \notin \{\tau_1(x), \ldots, \tau_{k_x}(x)\}$, and $Dg(\tau_j) = \lim_{t \to \tau_j} \frac{Dx(t)}{Dz_x[g(t)]} = \left(\frac{D^2 x(t)}{D^2 z_x(\tau_{z,j})}\right)^{1/2}$. Higher derivatives can be computed accordingly. Hence, $g \in \mathcal{H}$ which proves assertion a).

The definitions of $\mathcal{X}_{\xi_1,\ldots,\xi_p}$, $\mathcal{R}$ and $\mathcal{S}_{\xi_1,\ldots,\xi_p}$ imply $\{x \circ \mathcal{R}(x) | \ x \in \mathcal{X}_{\xi_1,\ldots,\xi_p}\} = \mathcal{S}_{\xi_1,\ldots,\xi_p}$. Obviously, $\mathcal{S}_{\xi_1,\ldots,\xi_p}$ is a convex space. Therefore, in order to prove consistency of $\mathcal{R}$ it only remains to show that $\mathcal{R}(y) = \mathcal{I}$ for any $y \in \mathcal{S}_{\xi_1,\ldots,\xi_p}$. Thus let $y = \theta_1\xi_1 + \cdots + \theta_p\xi_p$ denote an arbitrary element of $\mathcal{S}_{\xi_1,\ldots,\xi_p}$. Then (10) as well as $\int_0^1 (h(t) - t)^2 dt = 0$ hold with $h = \mathcal{I}$. Consequently, by definition of $\mathcal{R}$, we have $\mathcal{R}(y) = \mathcal{I}$. This completes the proof of the proposition.

References

Bookstein, F. L. (1978) The Measurement of Biological Shape and Shape Change. *Systematic Zoology,* **29,** 102–104.

Bookstein, F. L. (1991) *Morphometric Tools for Landmark Data: Geometry and Biology.* Cambridge: Cambridge University Press.

Gasser, T. and Kneip, A. (1995) Searching for structure in curve samples. *Journal of the American Statistical Association,* **90**, 1179–1188.

Gervini, D. and Gasser, T. (2004) Self–modeling warping functions. *Journal of the Royal Statistical Society, Series B,* **66,** 959–971.

Hall, P., Lee, Y. K., and Park, B. U. (2007) A method for projecting functional data onto a low-dimensional space. *Journal of Computational and Graphical Statistics,* **16,** 799–812.

Kneip, A. and Gasser, T. (1992) Statistical tools to analyze data representing a sample of curves. *Annals of Statistics,* **20**, 1266–1305.

Kneip, A., Li, X., MacGibbon, K. B. and Ramsay, J. O. (2000) Curve registration by local regression. *The Canadian Journal of Statistics,* **28,** 19–29.

Liu, X. and Müller, H-G. (2004) Functional convex averaging and synchronization for time-warped random curves. *Journal of the American Statistical Association,* **99,** 687-699.

Ramsay, J. O. (1996) Estimating smooth monotone functions. *Journal of the Royal Statistical Society, Series B.,* **60**, 365–375.

Ramsay, J. O. (2000) Functional components of variation in handwriting. *Journal of the American Statistical Association*, **95**, 9–15.

Ramsay, J. O. and Li, X. (1998) Curve registration. *Journal of the Royal Statistical Society, Series B.*, **60,** 351–363.

Ramsay, J. O. and Silverman, B. W. (2002) *Applied Functional Data Analysis.* New York: Springer.

Ramsay, J. O. and Silverman, B. W. (2005) *Functional Data Analysis.* New York: Springer.

Rønn, B. B. (2001) Nonparametric maximum likelihood estimation of shifted curves. *Journal of the Royal Statistical Society, Series B.*, **63**, 243–259.

Sakoe, H. and Chiba, S. (1978) Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, **26**, 43–49.

Silverman, B. W. (1995) Incorporating parametric effects into functional principal components analysis. *Journal of the Royal Statistical Society, Series B.*, **57**, 673–689.

Wang, K. and Gasser, T. (1997) Alignment of curves by dynamic time warping. *The Annals of Statistics*, **25,** 1251–1276.

Wang, K. and Gasser, T. (1998) Asymptotic and bootstrap confidence bounds for the structural average of curves. *The Annals of Statistics*, **26,** 872–991.

Wang, K. and Gasser, T. (1999) Synchronizing sample curves nonparametrically. *The Annals of Statistics*, **27,** 439–460.