Functional Convex Averaging and Synchronization for Time-Warped Random Curves

Xueli Liu and Hans-Georg Müller

December 31, 2003

Short title. Functional Convex Averaging.

Xueli Liu is at the Departments of Human Genetics and Biomathematics, David Geffen School of Medicine, University of California, Los Angeles, CA 90095, and Hans-Georg Müller is at the Department of Statistics, University of California, Davis, CA 95616. The authors wish to thank three referees and especially an associate editor for excellent and constructive comments that led to many improvements. This research was partially supported by National Science Foundation Grants DMS99-71602 and DMS02-04869.

Abstract

Data that can be best described as a sample of curves are now fairly common in science and engineering. When the dynamics of development, growth or response over time are at issue, subjects or experimental units may experience events at a different temporal pace. For functional data where trajectories may be individually time-transformed, it is usually inadequate to use commonly employed sample statistics such as the cross-sectional mean or median, or the cross-sectional sample variance. If one observes time-warped curve data, i.e., random curves or random trajectories that exhibit random transformations of the time scale, the usual L^2 norm and metric typically are inadequate. One may then consider subjecting each observed curve to a time transformation in an attempt to reverse the warping of the time scale, prior to further statistical analysis. Dynamic time warping, alignment, curve registration and landmark-based methods have been put forward with the goal of finding adequate empirical time transformations.

Previous analyses of warping have typically not been based on a model where individual observed curves are viewed as realizations of a stochastic process. We propose a functional convex synchronization model, under the premise that each observed curve is the realization of a stochastic process. Monotonicity constraints on time evolution provide the motivation for a functional convex calculus with the goal of obtaining sample statistics such as a functional mean. Observed random functions in warped time space are represented by a bivariate random function in synchronized time space, consisting of a stochastic monotone time transformation function and an unrestricted random amplitude function. Our theory assumes a monotone time warping transformation which maps synchronized time to warped (i.e., observed) time. This leads to the definition of a functional convex average or "longitudinal average", in contrast to the conventional "cross-sectional" average. We discuss various implementations of functional convex averaging and derive a functional limit theorem and asymptotic confidence intervals for functional convex means. The results are illustrated with a novel time warping transformation and extend to commonly-used warping and registration methods such as landmark registration. The methods are applied to simulated data and the Berkeley growth data.

Key words: Alignment, confidence bands, convex calculus, curve registration, functional data analysis, growth curves, nonparametric function estimation, sample statistics, smoothing, stochastic process, warping, weak convergence.

1. Introduction

Repeated measurements taken from the same unit at different times are frequently obtained in the sciences. If sampling is sufficiently dense, the resulting data may be viewed as a sample of curves. These curves may vary in shape, both in amplitude and time progression. For human growth curves, different individuals may experience certain events such as the pubertal growth spurt at different, individually determined times (Gasser et al., 1984a). Hence stochastic models that incorporate individual time scales are of significant interest.

Given sample data, a fundamental problem is to define sample statistics, such as the sample mean, targeting a suitably defined population mean. If curve data contain random time transformations, the definition of an appropriate population mean and *a fortiori* of a sample mean is not evident. For example, the usual pointwise or cross-sectional mean is well-known to be an inadequate estimate of the mean curve in such a situation (Gasser et al. 1984b; Kneip and Gasser 1992; Gasser and Kneip 1995). Time warping, also called curve registration in engineering (Sakoe and Chiba 1978; Ramsay and Silverman 1997), or curve alignment in biology, usually aims at heuristically aligning an individual curve or signal to a given template by warping the time axis of an individual so that the warped curve maximally coincides with the template in a suitable metric.

Landmark registration is a well-known warping method, aligning curves by identifying the timing of salient features, such as peaks, troughs, or inflection points. Curves are then aligned by transforming individual time so that landmark events become synchronized. An implementation is the "Structural Average" of Kneip and Gasser (1992). Individual curves are aligned towards the average location of common features that are observed in the sample of curves. In this well-studied approach, all curves must exhibit the same common features, as these features are critical to determine the alignment.

Silverman (1995) proposed a simple and generally applicable time shift model by assuming that an individually observed curve may be time-shifted according to $X(t+\tau)$, where τ is a random time-shift. Given a current mean $\hat{\mu}$, updating of the shift parameters is done via $\hat{\tau}_i = \text{argmin}_\tau \sum_i \int [X_i(s+\tau) - \hat{\mu}(s)]^2 \, ds$; this can be viewed as a precursor of the "Continuous Monotone Registration Method" (Ramsay and Li 1998).

For the case of non-random functions, Wang and Gasser (1997) proposed an approach based on penalty functions that measure the misalignment between two sample curves. This idea was extended in Wang and Gasser (1998, 1999) to align m sample curves, $m > 2$. Amplitudes

and derivatives of the curves are used to measure the alignment status, which is optimized in an iterative procedure. In other recent developments, nonparametric maximum likelihood (MLE) was introduced by Rønn (2001) and local nonlinear adjustments were proposed by Kneip, Li, MacGibbon and Ramsay (2000) to obtain the warping functions.

A central assumption for the model-based warping approach of Wang and Gasser (1997, 1998, 1999) is that the observed curves are fixed unknown smooth functions. Finite sample constraints on the functions are invoked for consistency results. A more realistic model where the observed curves are viewed as an unrestricted sample of random functions does not yet exist, and this motivates our proposal in this paper, which provides a model-based approach to warping that allows the derivation of asymptotic properties for the resulting estimates.

It is natural to assume that each observed function corresponds to the realization of a random process. This assumption is the point of departure for our investigation. The treatment of sample curves as realizations of a smooth stochastic process was pioneered by Rao (1958) and further developed in a landmark paper by Rice and Silverman (1991), who studied functional principal components, obtained via the Karhunen-Loève representation for stochastic processes (Ash and Gardner 1975). Capra and Müller (1997) and Shiau and Lin (1999) discussed functional warping models where multiplicative random time acceleration factors were coupled with random process expansions, and were related to a covariate.

Our proposed warping models for functional data include random time-synchronizing maps as an integral part. The observed curves are assumed to be generated by a latent bivariate stochastic process, where one component corresponds to the random time warping function, and the other component to a random amplitude function. The time transformation is constrained to be monotone increasing. The fact that monotone functions do not form a vector space, but rather a convex space, motivates the concept of functional convex averaging. Invoking convexity enables us to define convex paths connecting two curves, convex sums and convex averages. Our method potentially provides a framework for virtually all curve registration methods in the sense that given the warping functions for individual processes, the theory developed in this article can be invoked to obtain not only finite-sample convex paths and averages, but also weak convergence and asymptotic inference.

The paper is organized as follows. The underlying model and convex calculus are introduced in section 2. Illustrations of functional convex averaging and selection of time-synchronizing maps are discussed in section 3. Sample statistics based on functional convex averaging, obtained

by applying time-synchronizing transformations, and corresponding asymptotic distribution results are described in section 4. These results are applied to construct pointwise asymptotic confidence regions for the underlying mean curves. Simulation results can be found in section 5, where it is shown that the proposed functional convex averages based on a specific and simple choice of time-synchronization transformations generally improve upon conventional averaging when time-warping is present. A comparative illustration of functional convex averaging with different warping methods for the Berkeley growth data is presented in section 6. Concluding remarks are in section 7, and theoretical developments and proofs are provided in an Appendix.

2. Time Synchronization Model and Functional Convex Sums for Random Curves

Assume $Y(t)$, $t \in I$, where I is an interval, is a real-valued stochastic process. If $E|Y(t)|^2 < \infty$ for $t \in I$ and $E[\int Y^2(t)\,dt] < \infty$, we refer to Y as square integrable on the domain I. For square integrable processes X, Y, the usual inner product on $L^2(I)$ is given by

$$\langle X, Y\rangle = \int_I X(s)Y(s)\,ds,$$

with associated norm $||X|| = \langle X, X\rangle^{1/2}$.

2.1 Model for Functional Convex Synchronization. Consider a class of bivariate stochastic processes

$$\mathcal{S} = \{(X(t), Y(t)), t \in [0,1]\} \subset L^2([0,1]) \times L^2([0,1]),$$

where X, Y are square integrable, $X(\cdot) \in [0,T]$ for a given $T > 0$, $X(0) = 0$, $X(1) = T$, and X is monotone increasing and invertible. We refer to $t \in [0,1]$ as synchronized time and $\mathcal{S}$ as synchronized time space; $X : [0,1] \mapsto [0,T]$ is interpreted as a time-warping transformation that maps synchronized time to warped time, and $Y : [0,1] \mapsto R$ is interpreted as a random amplitude function. We assume that the observed processes $\tilde{Y}(x)$ are obtained from time-synchronized processes by warping the time to the individual's warped time with domain $x \in [0,T]$. We note that warped time in our model corresponds to the actually observed time, while many other authors apply the warping mapping to the observed data. The warped time space of the observed process is

$$\mathcal{W} = \{(x, \tilde{Y}(x)),\ x \in [0,T]\} \subset L^2([0,T]) \times L^2([0,T]).$$

We make the following assumptions:

(A1) An observed process $\{(x, \tilde{Y}(x)),\ x \in [0,T]\} \in \mathcal{W}$, is generated from a latent bivariate process $\{(X(t), Y(t)),\ t \in [0,1]\} \in \mathcal{S}$ through the warping mapping $\psi : \mathcal{S} \mapsto \mathcal{W}$,

$$\psi :\ \{(X(t), Y(t)),\ t \in [0,1]\} \mapsto \{(x, \tilde{Y}(x)),\ x \in [0,T]\}, \tag{1}$$

defined by $\tilde{Y}(x) = Y(X^{-1}(x))$, where $X^{-1}(\cdot)$ denotes the inverse of $X(\cdot)$.

In practice, a sample of observed processes with warped time $\{\tilde{Y}(x), x \in [0,T]\}$, is given. The inverse of ψ, the mapping $\psi^{-1} : \mathcal{W} \mapsto \mathcal{S}$, then produces a corresponding representation in synchronized time,

$$\psi^{-1} :\ \{(x, \tilde{Y}(x)),\ x \in [0,T]\} \mapsto \{(X(t), Y(t)),\ t \in [0,1]\}.$$

The latent bivariate process with synchronized time $\{(X(t), Y(t)),\ t \in [0,1]\}$, that is assumed to be associated with each observed process $\tilde{Y}$, is not unique. While the forward mapping ψ is uniquely determined, its inverse is not: Given $\tilde{Y}(x)$, there are many choices of $(X(t), Y(t))$ such that $\psi : \{(X(t), Y(t)),\ t \in [0,1]\} \mapsto \{(x, Y(X^{-1}(x))),\ x \in [0,T]\}$. For example, for all strictly increasing and invertible functions $f : [0,1] \mapsto [0,1]$, all $\{X(t), Y(t)\} = \{(f(t), 1 - f(t)),\ t \in [0,1]\}$ have the same image under ψ , namely $\psi((X,Y)) = \{(x, 1-x), x \in [0,1]\}$.

This inherent non-uniqueness of the warping mapping ψ reflects a basic but not widely acknowledged non-identifiability problem, that affects all warping procedures, and that motivates our second assumption:

(A2) A family of time-synchronizing mappings

$$\varphi_{\tilde{Y}} :\ [0,T] \ \mapsto [0,1],\ \ x \ \mapsto t(x),$$

indexed by observed processes $\tilde{Y}$, is given, such that each $\varphi_{\tilde{Y}}$ is monotone and invertible, with $\varphi_{\tilde{Y}}(0) = 0,\ \varphi_{\tilde{Y}}(T) = 1$, and $X(t) = \varphi_{\tilde{Y}}^{-1}(t)$.

Note that $\varphi_{\tilde{Y}}$ can be interpreted as mapping the warped time scale x of observed processes $\tilde{Y}$ to synchronized time t. The inverses $\varphi_{\tilde{Y}}^{-1}$ are time-warping functions. Under assumptions (A1) and (A2), we are able to define an inverse ψ^{-1} of ψ from the warped time space $\mathcal{W} = \{\tilde{Y}(x), x \in [0,T]\}$ of observed processes to the synchronized time space $\mathcal{S}$ by

$$\psi^{-1} :\ \{(x, \tilde{Y}(x)),\ x \in [0,T]\} \mapsto \{(X(t), Y(t)),\ t \in [0,1]\} = \{(\varphi_{\tilde{Y}}^{-1}(t), \tilde{Y}(\varphi_{\tilde{Y}}^{-1}(t))),\ \ t \in [0,1]\}.$$

The mapping ψ^{-1} thus uniquely assigns a bivariate process $\{(X(t), Y(t)), t \in [0,1]\} \in \mathcal{S}$ to an observed process $\tilde{Y}(x) \in \mathcal{W},\ x \in [0,T]$.

The functional convex mean curve, the primary target of our analysis, is given by

$$\{(\mu_X(t), \mu_Y(t)), t \in [0,1]\} \in \mathcal{S}, \text{where}\, \mu_X(t) = EX(t),\ \mu_Y(t) = EY(t).$$

Here $\mu_X(\cdot)$ corresponds to mean time transformation and $\mu_Y(\cdot)$ to mean amplitude function. Assuming that μ_X^{-1} exists, we use the notation

$$E_\oplus \tilde{Y} \equiv E_\oplus \tilde{Y}(x) = \mu_Y \circ \mu_X^{-1}(x) \equiv \mu_Y(\mu_X^{-1}(x)),\ x \in [0,T], \tag{2}$$

to represent the functional convex mean curve in $\mathcal{W}$, where $\circ$ denotes the composition of two functions. Note that $E_\oplus \tilde{Y}$ corresponds to the image under ψ of the graph in $\mathcal{S}$ that is generated by mean synchronized time and mean amplitude functions.

2.2 Functional Convex Sums and Averages of Time-Warped Random Curves. Under our assumptions, the time transformation $X(t) = \varphi_{\tilde{Y}}^{-1}(t)$ must always be monotone increasing as it maps absolute time t to warped time $x = X(t)$. Accordingly, the space $\mathcal{S}$ is not a linear space, but is convex: For any $\lambda \in [0,1]$, and $(X_1, Y_1),\ (X_2, Y_2) \in \mathcal{S},\ \{\lambda(X_1(t), Y_1(t)) + (1-\lambda)(X_2(t), Y_2(t)),\ t \in [0,1]\} \in \mathcal{S}$.

A key observation is that means and averages can be defined for random elements in a convex space solely based on convex combinations. Given two observed processes $\tilde{Y}_1, \tilde{Y}_2 \in \mathcal{W}$, and a fixed $0 \le \pi \le 1$, we use the mappings $\psi,\ \psi^{-1}$ to define a convex sum of $\tilde{Y}_1,\ \tilde{Y}_2$ by:

$$\pi\tilde{Y}_1 \oplus (1-\pi)\tilde{Y}_2 = \psi\{\pi\psi^{-1}(\tilde{Y}_1) + (1-\pi)\psi^{-1}(\tilde{Y}_2)\}. \tag{3}$$

The interpretation of this convex sum is as follows: Starting with

$$\psi^{-1}(\tilde{Y}_1) = \{(X_1(t), Y_1(t)),\ t \in [0,1]\},\ \psi^{-1}(\tilde{Y}_2) = \{(X_2(t), Y_2(t)), t \in [0,1]\},$$

one forms the conventional convex combination in the synchronized time space $\mathcal{S}$. This leads to

$$\{(\pi X_1(t) + (1-\pi)X_2(t), \pi Y_1(t) + (1-\pi)Y_2(t)),\ t \in [0,1]\}.$$

Finally, one applies the warping mapping ψ to obtain the graph of $\pi\tilde{Y}_1 \oplus (1-\pi)\tilde{Y}_2$ in $\mathcal{W}$, the space of the observed processes with warped time.

The $\oplus$ operation easily extends to the case of more than two functions. For example, for K observed sample random curves $\tilde{Y}_1, \ldots, \tilde{Y}_K \in \mathcal{W}$, given fixed constants $0 \leq \pi_j \leq 1,\ 1 \leq j \leq K$, such that $\sum_{j=1}^{K} \pi_j = 1$, let $(X_j, Y_j) = \psi^{-1}(\tilde{Y}_j)$. Then define

$$\bigoplus_{j=1}^{K} \pi_j \tilde{Y}_j = \psi\{(\sum_{j=1}^{K} \pi_j X_j(t), \sum_{j=1}^{K} \pi_j Y_j(t)),\ t \in [0,1]\}. \tag{4}$$

The functional convex sample mean of n observed random processes $\tilde{Y}_1, \ldots, \tilde{Y}_n$ is accordingly defined as

$$\bar{\tilde{Y}}_{\oplus} = \bigoplus_{j=1}^{n} \frac{1}{n} \tilde{Y}_j, \tag{5}$$

and is referred to as the *functional convex average*. The functional convex average is obtained by first time-synchronizing the curves, performing a conventional averaging operation on the time-synchronized versions, and then transforming back to warped time. Note that the operation $\oplus$ induces a convex geometry in $\mathcal{W}$: For $\tilde{Y}_1,\ \tilde{Y}_2 \in \mathcal{W}$, the *convex path* connecting $\tilde{Y}_1$ and $\tilde{Y}_2$ is defined by the set of processes $\{\pi \tilde{Y}_1 \oplus (1-\pi)\tilde{Y}_2,\ \ \pi \in [0,1]\}$.

2.3 Functional Convex Norm and Inner Product for Time-Warped Random Curves. Setting $(X, Y) = \psi^{-1}(\tilde{Y})$, define the functional convex norm of $\tilde{Y}$ by

$$\begin{aligned} ||\tilde{Y}||_{\oplus} &= [\int X(t)^2\, dt + \int Y(t)^2\, dt]^{1/2} \\ &= [\int \varphi_{\tilde{Y}}^{-1}(t)^2\, dt + \int \tilde{Y}(\varphi_{\tilde{Y}}^{-1}(t))^2\, dt]^{1/2}. \end{aligned} \tag{6}$$

For two processes $\tilde{Y}_1, \tilde{Y}_2 \in \mathcal{W}$, for which $(X_1, Y_1) = \psi^{-1}(\tilde{Y}_1),\ (X_2, Y_2) = \psi^{-1}(\tilde{Y}_2)$, a corresponding functional convex distance is given by:

$$\begin{aligned} d_{\oplus}(\tilde{Y}_1, \tilde{Y}_2) &= ||\tilde{Y}_1 - \tilde{Y}_2||_{\oplus} \\ &= [\int (X_1(t) - X_2(t))^2\, dt + \int (Y_1(t) - Y_2(t))^2\, dt]^{1/2}. \end{aligned} \tag{7}$$

Norms and metrics on function spaces are understood to be defined on equivalence classes of functions that are almost everywhere equal. These definitions may be generalized by integrating with respect to measures other than ordinary Lebesgue measure, and the measures used may differ between the X and Y components. This allows for the use of various weight functions in defining the functional convex norm and distance.

Proposition 2.1 *The function $d_{\oplus}(\tilde{Y}_1, \tilde{Y}_2) : \mathcal{W} \times \mathcal{W} \mapsto R^+$ in (7) defines a metric in $\mathcal{W}$ and $||\tilde{Y}||_{\oplus}$ in (6) defines a norm in $\mathcal{W}$.*

The proof is in the Appendix. It is of interest to define a scalar product $\langle \cdot, \cdot \rangle_{\oplus}$ that is compatible with the norm in the sense that it satisfies the parallelogram identity,

$$\langle x, y \rangle_{\oplus} = \frac{1}{2}(||x \oplus y||_{\oplus}^2 - ||x||_{\oplus}^2 - ||y||_{\oplus}^2). \tag{8}$$

A scalar product in $\mathcal{W}$ which satisfies (8) is given by

$$\langle \tilde{Y}_1, \tilde{Y}_2 \rangle_{\oplus} = \int X_1(t) X_2(t)\, dt + \int Y_1(t) Y_2(t)\, dt. \tag{9}$$

Proposition 2.2 *The functional $\langle \cdot, \cdot \rangle_{\oplus} : \mathcal{W} \times \mathcal{W} \mapsto R$ in (9) has the properties:*

(a) *It is a scalar product in $\mathcal{W}$, and in particular*
$\langle \tilde{Y}_1, \pi \tilde{Y}_2 \oplus (1-\pi)\tilde{Y}_3 \rangle = \pi \langle \tilde{Y}_1, \tilde{Y}_2 \rangle + (1-\pi)\langle \tilde{Y}_1, \tilde{Y}_3 \rangle$.

(b) $||\tilde{Y}||_{\oplus} = \langle \tilde{Y}, \tilde{Y} \rangle_{\oplus}^{1/2}$.

(c) $\langle \tilde{Y}_1, \tilde{Y}_2 \rangle_{\oplus} = \frac{1}{2}(||\tilde{Y}_1 \oplus \tilde{Y}_2||_{\oplus}^2 - ||\tilde{Y}_1||_{\oplus}^2 - ||\tilde{Y}_2||_{\oplus}^2)$.

The proof is in the Appendix. Note that the metric $d_{\oplus}$ for $\mathcal{W}$ defined as above allows us to define a measure of variation for time-warped processes $\tilde{Y}$,

$$\mathrm{var}_{\oplus}(\tilde{Y}) = E\{d_{\oplus}^2(\tilde{Y}, E_{\oplus}(\tilde{Y}))\}, \tag{10}$$

with the corresponding sample version

$$\hat{\mathrm{var}}_{\oplus}(\tilde{Y}) = \frac{1}{n}\sum_{i=1}^{n} d_{\oplus}^2(\tilde{Y}_i, \bigoplus_{j=1}^{n} \frac{1}{n}\tilde{Y}_j)). \tag{11}$$

These quantities replace the conventional cross-sectional versions that are based on the L^2 distance $d(\tilde{Y}_1, \tilde{Y}_2) = [\int (\tilde{Y}_1(x) - \tilde{Y}_2(x))^2\, dx]^{1/2}$, and are given by

$$\mathrm{var}(\tilde{Y}) = E\{d^2(\tilde{Y}, E\tilde{Y})\}, \tag{12}$$

$$\hat{\mathrm{var}}(\tilde{Y}) = \frac{1}{n}\sum_{i=1}^{n} d^2(\tilde{Y}_i, \bar{\tilde{Y}}). \tag{13}$$

where $\bar{\tilde{Y}} = \frac{1}{n}\sum_{i=1}^{n} \tilde{Y}_i$.

3. Examples of Functional Convex Sums and Time-Synchronization Mappings

In this section we motivate the proposed functional convex calculus, in which functional convex sums and functional convex averaging replace the conventional linear operations and averaging.

We refer to functional convex sums and averages also as time-synchronized sums and averages. Given two observed random curves $\tilde{Y}_1,\ \tilde{Y}_2 \in \mathcal{W}$, and a $\kappa \in [0,1]$, the conventional convex combination of $\tilde{Y}_1,\ \tilde{Y}_2$ is

$$\kappa\tilde{Y}_1 + (1-\kappa)\tilde{Y}_2. \tag{14}$$

In contrast, the proposed functional convex combination is

$$\kappa\tilde{Y}_1 \oplus (1-\kappa)\tilde{Y}_2, \tag{15}$$

where the operator $\oplus$ is defined in section 2.3.

3.1 Synchronizing Maps. Important ingredients for the proposed functional convex calculus are the synchronizing maps $\varphi_{\tilde{Y}}$ [given in (A2) in section 2]. We emphasize that our general results on convex operations, convex averages and asymptotics work for any given synchronizing map. One of the simplest possible choices is area-under-the-curve synchronization, for which the synchronizing map is:

$$\varphi_{\tilde{Y},p}(x) = [\frac{\int_0^x |\tilde{Y}(s)|^p\, ds}{\int_0^T |\tilde{Y}(s)|^p\, ds}]^{1/p}, \tag{16}$$

where $p > 0$ is a given parameter. With increasing p, this map places increasing emphasis on aligning peaks/maxima. Note that for $p = 1$, (16) leads to

$$\frac{\partial t}{\partial x} = \frac{\partial}{\partial x}\varphi_{\tilde{Y}}(x) = |\tilde{Y}(x)| / \int_0^T |\tilde{Y}(s)|\, ds. \tag{17}$$

Therefore the derivatives of the synchronizing map tend to be large near peaks, corresponding to more rapidly advancing synchronized time near such points. More generally, as observed by both Wang and Gasser (1996) and Ramsay and Li (1998), derivatives $\tilde{Y}^{(l)},\ 0 \le l \le q$, are often useful for the selection of a synchronizing time transformation. A corresponding general class of time transformations is as follows: Given positive parameters $\kappa_0,\ \ldots, \kappa_q,\ p_0,\ \ldots,\ p_q$ with $\sum_{l=0}^q \kappa_l = 1$, choose

$$\varphi_{\kappa_l,p_l}(x) = \sum_{l=0}^{q} \kappa_l [\frac{\int_0^x |\tilde{Y}^{(l)}(s)|^{p_l}\, dx}{\int_0^T |\tilde{Y}^{(l)}(s)|^{p_l}\, ds}]^{1/p_l}, \tag{18}$$

of which (16) is a special case.

Another promising version of (16), which is invariant in regard to additive constants and was suggested by a reviewer, is to apply (16) to centered processes

$$\tilde{\tilde{Y}}_i(x) = \tilde{Y}_i(x) - \frac{1}{T}\int_0^T \tilde{Y}_i(s)\, ds. \tag{19}$$

3.2 Examples of Functional Convex Sums. Comparing conventional convex combinations of functions with the proposed functional convex combinations, we choose $\varphi_{\tilde{Y}}$ according to (16), with $p = 1$. The comparisons are illustrated with $\kappa = 0, .25, .5, .75, 1.0$ in (14) and (15).

As a first example, consider two Gaussian peaks with different peak sizes, mean locations, and variances,

$$f_j(x) = \beta_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-\frac{1}{2\sigma_j^2}(x-\mu_j)^2\},\ j = 1, 2,\ x \in [0, T], \tag{20}$$

where $\mu_1 = .5,\ \mu_2 = 1.5,\ \sigma_1 = .15,\ \sigma_2 = .25,\ \beta_1 = 2,\ \beta_2 = 6,\ T = 3.$

From Figure 1, the shapes of the $f_j, j = 1, 2$ are seen to remain virtually invariant under the proposed functional convex combinations (lower panel in Figure 1), each curve having exactly one peak, while conventional convex combinations (upper panel in Figure 1) display multimodal shapes that do not resemble the original curves.

A second example is provided by the convex path connecting a unimodal to a bimodal curve in function space. Let f_1 be as in (20) with $\beta_1 = 2$, while f_2 is a mixture of two Gaussian peaks,

$$f_2(x) = \beta_2 \frac{1}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2\} + \beta_3 \frac{1}{\sqrt{2\pi}\sigma_3} \exp\{-\frac{1}{2\sigma_3^2}(x-\mu_3)^2\},\ x \in [0, T], \tag{21}$$

where $\beta_1 = 2,\ \beta_2 = \beta_3 = 3; \mu_1 = .5,\ \mu_2 = 1,\ \mu_3 = 1.5,\ \sigma_1 = \sigma_2 = \sigma_3 = .15,\ T = 3.$

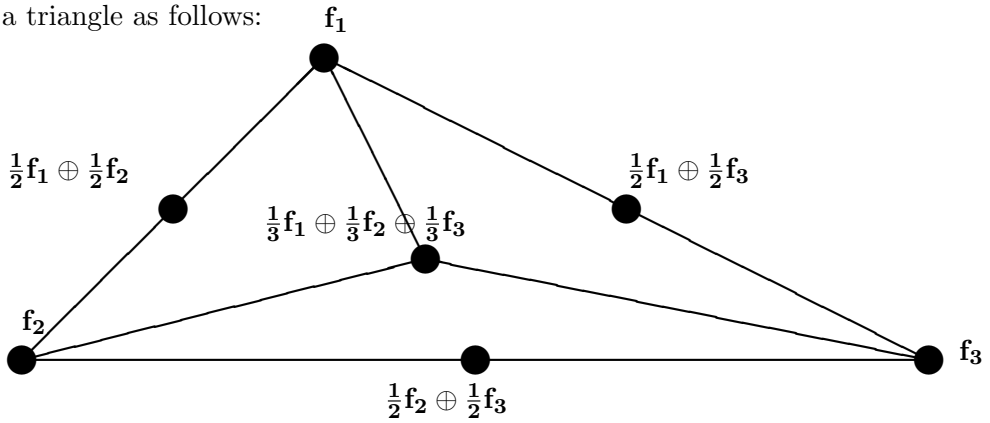
For the functional convex combinations (lower panel in Figure 2), the curves defining the path first have one peak, starting at f_1, then gradually developing a second peak. There are never more than two peaks, whereas for the conventional convex combinations (upper panel in Figure 2), all curves in-between f_1 and f_2 have three peaks, and their shape resembles neither f_1 nor f_2.

As another example, consider three bimodal curves. These are obtained as mixtures of Gaussian curves, defined as follows:

$$\begin{aligned}
f_1(x) &= \beta_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-\frac{1}{2\sigma_1^2}(x-\mu_{11})^2) + \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-\frac{1}{2\sigma_2^2}(x-\mu_{12})^2), \\
f_2(x) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-\frac{1}{2\sigma_2^2}(x-\mu_{21})^2) + \beta_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp(-\frac{1}{2\sigma_1^2}(x-\mu_{22})^2), \\
f_3(x) &= \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-\frac{1}{2\sigma_2^2}(x-\mu_{31})^2) + \frac{1}{\sqrt{2\pi}\sigma_2} \exp(-\frac{1}{2\sigma_2^2}(x-\mu_{32})^2),\ x \in [0, T],
\end{aligned}$$

where $\beta_1 = .5,\ \mu_{11} = 1.5,\ \mu_{12} = 2.5,\ \mu_{21} = 3.5,\ \mu_{22} = 4.5,\ \mu_{31} = 6.0,\ \mu_{32} = 7.5,\ \sigma_1 = .2,\ \sigma_2 = .5,\ T = 9.0.$

These three functions are depicted in Figure 3 (upper panel) from left to right, starting with f_1 on the left. Their functional convex average as well as conventional cross-sectional average is shown in the lower panel. It is obvious that the two averages differ dramatically and only the functional convex average provides a summary that reasonably reflects the shape of the three functions. The functional convex paths leading from each of the original curves $f_j, j = 1,\ 2,\ 3$ to the functional convex averages for pairs of curves and for all three curves can be visualized in a triangle as follows:



We conclude that conventional averages are prone to give misleading impressions about the shape of the underlying curves, and that even very simple choices of synchronizing mappings and associated functional convex averages can lead to substantial improvements.

4. Random Elements in Convex Space and Asymptotics

4.1 Variance-minimizing Property of Functional Convex Means. In analogy to functional convex sample means as defined in (5), one might consider other sample statistics based on functional convex operations to describe the population of random curves. For example, the sample version (11) targets the variation measure (10). We focus here on an investigation of the mean as the most basic statistical functional.

The conventional mean of a sample of functions is the minimizer of the sums of the L^2 distances between the elements in the sample space to a fixed function,

$$\bar{u} = \text{argmin}_{u \in L^2} \sum_{i=1}^{n} d^2(u, u_i) = \frac{1}{n} \sum_{i=1}^{n} u_i. \tag{22}$$

The following results provide analogous characterizations for functional convex means. The proofs are in the Appendix.

Theorem 4.1 *For $\tilde{Y} \in \mathcal{W}$,* $\text{argmin}_{\mu \in \mathcal{W}} E[d_{\oplus}(\mu, \tilde{Y})] = E_{\oplus}\tilde{Y}$, *where $d_{\oplus}$ is as defined in (7).*

The corresponding sample version, analogous to (22), is

Theorem 4.2 *For given arbitrary $\tilde{Y}_1, \ldots, \tilde{Y}_n \in \mathcal{W}$, it holds that*

$$\text{argmin}_{\mu \in \mathcal{W}} \{\sum_{j=1}^{n} d_{\oplus}^2(\mu, \tilde{Y}_i)\} = \bigoplus_{i=1}^{n} \frac{1}{n} \tilde{Y}_i = \bar{\tilde{Y}}_{\oplus}. \tag{23}$$

These results demonstrate the coherence of the proposed convex calculus with the usual properties of population and sample statistics, and justify the use of the convex variance measures, in both population (10) and empirical (11) versions.

4.2 Asymptotic Properties of Functional Convex Averaging. We denote pointwise convergence in distribution by $\xrightarrow{\mathcal{D}}$, and weak convergence in function space by $\Rightarrow$. Suppose we are presented with a sample of i.i.d. observed random processes $\tilde{Y}_1, \ldots, \tilde{Y}_n \in \mathcal{W}$, and the corresponding latent bivariate processes $(X_i, Y_i) = \psi^{-1}(\tilde{Y}_i) \in \mathcal{S},\ i = 1, \ldots, n$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$. The following additional assumptions on the latent bivariate processes $\{(X(t),\ Y(t))\,,\ t \in [0,1]\}$ are needed:

(H1) There exists a constant $C_0 > 0$, such that for $Z = X$ and $Z = Y$,

$$E|Z(t) - EZ(t) - (Z(s) - EZ(s))|^2 \leq C_0\ |t - s|^2,\ t, s \in [0,1]. \tag{24}$$

(H2) The two functions $\mu_X(t) = EX(t)$ and $\mu_Y(t) = EY(t)$ are continuously differentiable in t and there exists a $\delta > 0$, such that $\inf_{t \in [0,1]} |\mu_X^{'}(t)| \geq \delta$.

(H3) For X and Y, $\sup_t \text{var}(Y(t)) < \infty$ and $\sup_t \text{var}(X(t)) < \infty$.

Given $(X, Y) \in \mathcal{S}$, define

$$\gamma_{XX}(s,t) = \text{cov}(X(s), X(t)),\ \gamma_{XY}(s,t) = \text{cov}(X(s), Y(t)),\ \gamma_{YY}(s,t) = \text{cov}(Y(s), Y(t)), \tag{25}$$

and let $\{(G_X(t),\ G_Y(t))\,, t \in [0,1]\}$ be a bivariate Gaussian process with means $EG_X(t) = 0,\ EG_Y(t) = 0$ for $t \in [0,1]$ and covariances

$$\text{cov}(G_X(s), G_X(t)) = \gamma_{XX}(s,t),\ \text{cov}(G_X(s), G_Y(t)) = \gamma_{XY}(s,t),\ \text{cov}(G_Y(s), G_Y(t)) = \gamma_{YY}(s,t).$$

Lemma 4.1 *Under (H1)–(H3),*

$$\sqrt{n}(\bar{X}_n - \mu_X) \Rightarrow G_X,\ \ \sqrt{n}(\bar{Y}_n - \mu_Y) \Rightarrow G_Y,\ \textit{as}\ \ n \to \infty. \tag{26}$$

A consequence is

Theorem 4.3 *Under (H1)–(H3),*

$$\sqrt{n}(\bar{\tilde{Y}}_{\oplus} - E_{\oplus}\tilde{Y}) \Rightarrow \Xi, \ on\ [0,T],$$

where Ξ *is a Gaussian process,*

$$\Xi(x) = G_Y \circ \mu_X^{-1}(x) + \frac{\mu_Y^{'} \circ \mu_X^{-1}(x)}{\mu_X^{'} \circ \mu_X^{-1}(x)} G_X \circ \mu_X^{-1}(x), \ x \in [0,T],$$

and $\mu_X^{'}, \mu_Y^{'}$ *are the first order derivatives of* μ_X *and* μ_Y *respectively.*

We note that the limiting process Ξ has the moments

$$E\Xi(x) = 0, \ x \ \in [0,T], \tag{27}$$

and

$$\begin{aligned} \mathrm{cov}(\Xi(x_1), \Xi(x_2)) &= \gamma_{YY}(\mu_X^{-1}(x_1), \mu_X^{-1}(x_2)) + \frac{\mu_Y^{'} \circ \mu_X^{-1}(x_2)}{\mu_X^{'} \circ \mu_X^{-1}(x_2)} \gamma_{XY}(\mu_X^{-1}(x_2), \mu_X^{-1}(x_1)) \\ &+ \frac{\mu_Y^{'} \circ \mu_X^{-1}(x_1)}{\mu_X^{'} \circ \mu_X^{-1}(x_1)} \frac{\mu_Y^{'} \circ \mu_X^{-1}(x_2)}{\mu_X^{'} \circ \mu_X^{-1}(x_2)} \gamma_{XX}(\mu_X^{-1}(x_1), \mu_X^{-1}(x_2)) \\ &+ \frac{\mu_Y^{'} \circ \mu_X^{-1}(x_1)}{\mu_X^{'} \circ \mu_X^{-1}(x_1)} \gamma_{XY}(\mu_X^{-1}(x_1), \mu_X^{-1}(x_2)), \ x_1, x_2 \in [0,T]. \end{aligned} \tag{28}$$

From tightness, it follows that the limiting process Ξ is continuous and therefore satisfies $\sup_{0 \le x \le T} |\Xi(x)| = O_p(1)$ (compare also Adler 1990). It follows from the Continuous Mapping Theorem that the functional convex sample mean is uniformly consistent for the functional convex mean,

$$\sup_{0 \le x \le T} |\bar{\tilde{Y}}_{\oplus}(x) - E_{\oplus}\tilde{Y}(x)| = O_p(n^{-1/2}). \tag{29}$$

Another consequence of Theorem 4.3 is pointwise normality: For each fixed $x \in [0,T]$, we have

$$\sqrt{n}(\bar{\tilde{Y}}_{\oplus}(x) - E_{\oplus}\tilde{Y}(x)) \xrightarrow{\mathcal{D}} \Xi(x). \tag{30}$$

In order to construct pointwise asymptotic confidence bands based on (30), we need to find an estimate of $\mathrm{var}(\Xi(x))$. Estimates for $\gamma_{XX}(t,t), \gamma_{YY}(t,t)$ and $\gamma_{XY}(t,t)$ can be based on (25), substituting empirical variances and covariances:

$$\hat{v}_{ZZ}(t) := \hat{\gamma}_{ZZ}(t,t) = \frac{1}{n-1} \sum_{i=1}^{n} (Z_i(t) - \bar{Z}_i(t)^2), \tag{31}$$

$$\hat{v}_{XY}(t) := \hat{\gamma}_{XY}(t,t) = \frac{1}{n-1} \sum_{i=1}^{n} (X_i(t) - \bar{X}_i(t))(Y_i(t) - \bar{Y}_i(t)), \tag{32}$$

choosing $Z = X$ and $Z = Y$ in (31). These estimators are based on averaging i.i.d. data and therefore are consistent.

Because the $X_i(t),\ t \in [0,1]$, are monotone increasing and invertible, so is $\bar{X}_n(t),\ t \in [0,1]$, and $\bar{X}_n^{-1}(x),\ x \in [0,T]$, is a reasonable estimator of μ_X^{-1}. By the Law of Large Numbers,

$$|\bar{X}_n(t) - \mu_X(t)| = O_p(n^{-1/2}), \tag{33}$$

and by Lemma $A.1$ and (43) in the Appendix,

$$\sup_{0 \le x \le T} |\bar{X}_n^{-1}(x) - \mu_X^{-1}(x)| = O_p(n^{-1/2}). \tag{34}$$

The derivatives $\mu_X^{'},\ \mu_Y^{'}$ can be estimated by applying local fitting of quadratic polynomials to the data, using the coefficients for the linear terms of the local quadratic fits as derivative estimates. Consistency of these estimates under usual regularity conditions follows from Theorem 3.1 in Fan and Gijbels (1996).

Lemma 4.2 *The variance estimator*

$$\hat{\text{var}}(\Xi(x)) = \hat{v}_{YY}(\hat{\mu}_X^{-1}(x)) + 2\frac{\hat{\mu}_Y^{'} \circ \hat{\mu}_X^{-1}(x)}{\hat{\mu}_X^{'} \circ \hat{\mu}_X^{-1}(x)}\hat{v}_{XY}(\hat{\mu}_X^{-1}(x)) + (\frac{\hat{\mu}_Y^{'} \circ \hat{\mu}_X^{-1}(x)}{\hat{\mu}_X^{'} \circ \hat{\mu}_X^{-1}(x)})^2\hat{v}_{XX}(\hat{\mu}_X^{-1}(x)), \tag{35}$$

based on $\hat{v}_{YY}, \hat{v}_{XY}, \hat{v}_{XX}$ *in (31), (32), is consistent for* $\text{var}(\Xi(x))$.

Combining Lemma 4.2 with the pointwise asymptotic normality result (30), pointwise asymptotic confidence bands for the functional convex mean are obtained as follows:

Corollary 4.1 *A pointwise* $100(1-\alpha)\%$ *asymptotic confidence band is given by*

$$\left[\bar{\hat{Y}}_{\oplus}(x) - \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\text{var}}(\Xi(x))/n},\ \bar{\hat{Y}}_{\oplus}(x) + \Phi^{-1}(1-\alpha/2)\sqrt{\hat{\text{var}}(\Xi(x))/n}\right], \tag{36}$$

with $\hat{\text{var}}(\Xi(x))$ *as in (35), where* Φ *is the c.d.f. of the standard normal distribution.*

Illustrations of these asymptotic confidence bands are provided in the next two sections. We note that the distribution of $\sup_{0 \le x \le T} |\Xi(x)|$ determines the width of the asymptotic simultaneous confidence bands. Since this distribution is unknown, we might instead work with a bootstrap approximation resampling from the original sample of random curves, in situations where simultaneous bands are of interest.

5. Simulation Study

This section contains an outline of the warping algorithm, exploring the specific example of time-synchronization by the area-under-the-curve method, and some simulation-based numerical illustrations. Robustness of the algorithm is explored for situations where the warping function is misspecified.

5.1 Time-Synchronization by Area Under the Curve and Generation of Sample Curves. For a given sample of observed random curves $\tilde{Y}_1, \ldots, \tilde{Y}_n$ in the warped time space $\mathcal{W}$, the first step is to time-synchronize the curves by applying the maps $\varphi_{\tilde{Y}}$ (16) for $p = 1$ if area-under-the-curve synchronization is chosen.

1. The synchronizing functions $\varphi_{\tilde{Y}_i}(x) = \int_0^x |\tilde{Y}_i(s)|\, ds / \int_0^T |\tilde{Y}_i(s)|\, ds$ are inverted numerically on a fine grid, and a time-synchronized sample $(X_i, Y_i) \in \mathcal{S}$ is obtained by $X_i(t) = \varphi_{\tilde{Y}_i}^{-1}(t),\ Y_i(t) = \tilde{Y}_i(\varphi_{\tilde{Y}_i}^{-1}(t)),\ i = 1, \ldots, n.$
2. The sample means $\bar{X}_n(t) = \frac{1}{n}\sum_{i=1}^n X_i(t),\ \bar{Y}_n(t) = \frac{1}{n}\sum_{i=1}^n Y_i(t)$ are obtained; $(\bar{X}_n(t), \bar{Y}_n(t))$ are transformed back via the map ψ (1) to the warped time space. This yields the functional convex average $\bar{\tilde{Y}}_{\oplus}(x) = \bar{Y}_n(\bar{X}_n^{-1}(x)),\ x \in [0, T]$, where $\bar{X}_n$ is inverted numerically.

We specify non-random functions $f_1,\ f_2 \in \mathcal{W},\ f_j(x) = \beta_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-\frac{1}{2\sigma_j^2}(x - \mu_j)^2\},\ \ x \in [0, T],\ j = 1, 2$, where $\mu_1 = .5,\ \beta_1 = 2,\ \sigma_1 = .15,\ \mu_2 = 1,\ \beta_2 = 6,\ \sigma_2 = .15,\ T = 2$. We obtain $(X_j^0(t),\ Y_j^0(t)),\ t \in [0, 1]$, by $X_j^0(t) = \varphi_{f_j}^{-1}(t),\ Y_j^0(t) = f_j(\varphi_{f_j}^{-1}(t)),\ \ j = 1, 2$, where $\varphi_{f_j}(x) = \int_0^x |f_j(s)|\, ds / \int_0^T |f_j(s)|\, ds$. The sample curves are then generated in $\mathcal{S}$ by

$$\begin{pmatrix} X_i(t) \\ Y_i(t) \end{pmatrix} = \begin{pmatrix} \epsilon_{i1} X_1^0(t) \\ \epsilon_{i2} Y_1^0(t) \end{pmatrix} + \begin{pmatrix} (1 - \epsilon_{i1}) X_2^0(t) \\ (1 - \epsilon_{i2}) Y_2^0(t) \end{pmatrix},\ 0 \le t \le 1,\ i = 1, \ldots, n,$$

where we choose mutually independent and uniform $[0, 1]$ pseudo-random numbers $\epsilon_{i1},\ \epsilon_{i2},\ i = 1, \ldots, n$, with $n = 50$. Finally, the sample curves are generated by $\tilde{Y}_i(x) = \psi(X_i, Y_i) = Y_i(X_i^{-1}(x)),\ x \in [0, T]$. The expected value in $\mathcal{S}$, corresponding to the target function, is given by

$$\begin{pmatrix} X^*(t) \\ Y^*(t) \end{pmatrix} = \frac{1}{2}\begin{pmatrix} X_1^0(t) \\ Y_1^0(t) \end{pmatrix} + \frac{1}{2}\begin{pmatrix} X_2^0(t) \\ Y_2^0(t) \end{pmatrix},\ 0 \le t \le 1.$$

In $\mathcal{W}$, the corresponding target function is given by $\tau(x) := \psi(X^*, Y^*),\ x \in [0, T]$.

5.2 Simulation Results. For one sample run, 50 sample curves in $\mathcal{W}$ are generated. These observed random curves all are unimodal, with varying peak sizes and peak locations. Figure 4 illustrates the asymptotic 95% pointwise confidence regions for the population mean curve, for one sample run, based on Corollary 4.1 and (35). The functional convex average is seen to be close to the target in terms of both L^2 distance and general shape. The conventional cross-sectional average function does not fall within the 95% confidence band. The estimated coverage rate for 1000 simulation runs was found to be 93.90% for the 95% nominal level.

Another comparison of interest concerns the variance measures. For 1000 Monte Carlo runs, the conventional estimate (13) is $\hat{\text{var}}(\tilde{Y}) = 11.33$ (S.E. 0.0352), and the functional convex measure (11) is $\hat{\text{var}}_{\oplus}(\tilde{Y}) = 5.47$ (S.E. 0.0223). As expected, the former is inflated by the additional horizontal variation in the time axis. The target value in this example is $\text{var}_{\oplus}(\tilde{Y}) = 5.46$, in good agreement with $\hat{\text{var}}_{\oplus}(\tilde{Y})$. We note that a sizable difference between $\hat{\text{var}}(\tilde{Y})$ (13) and $\hat{\text{var}}_{\oplus}(\tilde{Y})$ (11) indicates that a substantial fraction of the variation observed in the sample curves can be explained by time-warping.

Define the integrated square error (ISE)

$$\text{ISE}(\tilde{Y}, \tau) = \int_0^T (\tilde{Y}(x) - \tau(x))^2 \, dx. \tag{37}$$

For 1000 simulation runs, the average ISE of the functional convex average $\bar{\tilde{Y}}_{\oplus}$ was found to be 32.35 (S.E. 1.217), while the average ISE of the conventional average $\bar{\tilde{Y}}$ was found to be 288.22 (S.E. 2.212). As expected, the functional convex average leads to significantly improved recovery of the target function $\tau(\cdot)$.

In a second simulation study, we choose the generating functions (used in the same way as described in section 5.1) as f_1 in (20) with $\beta_1 = 2$, $\mu_1 = .5$, $\sigma_1 = .15$, and as f_2 in (21). Figure 5 (upper panel) provides the sample of 50 observed curves for one simulation run. Most of the curves have two peaks, with high variability in peak size and location. The lower panel demonstrates the comparison of the functional convex average and the conventional cross-sectional average: The two peaks of the sample curves are blurred to one peak in the conventional cross-sectional average, and the peak size is much lower than the target; in contrast, the functional convex average recovers the shape of the target function quite well. For 100 simulation runs, the conventional variance estimate (13) is $\hat{\text{var}}(\tilde{Y}) = 5.9567$ (S.E. 0.0455), while the functional convex variance estimate (11) comes out at $\hat{\text{var}}_{\oplus}(\tilde{Y}) = 0.5416$ (S.E. 0.0058).

The average ISE of the functional convex average for the 100 Monte Carlo runs was found

to be 26.797 (S.E. 3.373), and for the conventional average the numbers were 211.45 (S.E. 3.378). The functional convex average clearly leads to improved recovery of the target function $\tau(\cdot)$ in these simulation examples, and the functional convex mean provides a reasonable representation of shapes and characteristics of the random curves.

5.3 Robustness of Time-Synchronizing Maps to Misspecification. In order to explore robustness of the proposed method in regard to misspecification of the time-synchronizing maps, we consider the case where simple area-under-the-curve time synchronization is assumed but does not correspond to the true underlying warping function.

Specifically, the data-generating mechanism remains the same as described above, but is based on altered time-synchronizing maps,

$$\varphi_{f_j}(x) = \int_0^x |f_j(s) + c|\, ds / \int_0^T |f_j(s) + c|\, ds, \tag{38}$$

suggested by a reviewer. These maps coincide with area-under-the-curve synchronization for $c = 0$. The difference between assumed and actual maps increases with increasing values of c. We ran simulations over a grid of values for c in order to detect the break-down point, defined as the value of c where the ISE (37) of the functional convex mean (5) using area-under-the-curve synchronization (16) or of its centered version (19) becomes larger than the corresponding value for the cross-sectional average (13).

We found that time-synchronizing maps based on area-under-the-curve synchronization with centering (19) were more robust than those constructed from (16) for large values of c, but worked less well for small values of c. Specific values for the ISE can be found in Table 1.

Table 1: Values of Integrated Square Error (37) obtained for functional convex averages FCA(AUC) corresponding to (16), of the centered version FCA(AUCC) (19) and of cross-sectional averages when generating simulated data for selected values of c in (38), based on 100 simulation runs.

Method	c=0	0.4	0.5	2.2	2.3	3.9	4.0
FCA(AUC)	25.58	44.71	47.39	74.51	75.28	83.98	84.50
FCA(AUCC)	182.96	54.87	46.93	36.65	36.70	43.70	44.20
Cross-sectional	291.98	180.38	167.39	76.78	74.03	44.88	43.72

The functional convex mean with centering was found to break down at around $c = 3.95$,

and the non-centered basic version at around $c = 2.25$. However, the non-centered version performed considerably better for values $c < 0.47$. This may be simply an indication that the centered version of the area-under-the-curve synchronization is particularly misspecified in these cases.

We conclude that while gains are still achievable when the assumed warping mechanism is somewhat different from the true underlying mechanism, simple time-synchronization methods break down when these differences exceed a certain threshold.

6. Application to the Berkeley Growth Data

The proposed model and algorithm are compared with other registration methods, using the Berkeley growth study (Chapter 6, Ramsay and Silverman 2002, and available at ftp://ego.psych.mcgill.ca/pub/ramsay/FDAfuns/Matlab/) as an example. In this study, the heights of 54 girls and 39 boys from age one to age 18 were recorded. The trajectories on which we base our analysis are the velocity curves for the 54 girls, obtained by estimating the derivative of height with respect to time using a local quadratic fit; these velocity curves are shown in the upper panel of Figure 6. The timing of the pubertal growth spurt varies from 8 to 17 years for these girls. This substantial time variation of the most prominent functional feature implies that a substantial fraction of the observed variability in these growth curves is due to variability in the time scale, which therefore needs to be reflected in statistical models. It is clearly of interest to explore time-synchronizing maps for these data.

For each girl, 31 measurements are available. Intervals between measurements range from three months (age 1 to 2 years), yearly (age 3 up to age 8), to half-yearly (age 8 to 18). The major feature of the growth rate for individual girls is peak size and location of the pubertal growth spurt, and it is well-known that cross-sectional means and measures do not adequately represent the growth dynamics (Gasser et al. 1984b).

Opting for the least complex implementation of functional convex averaging, we use simple area-under-the-curve time-synchronizing map $\varphi_{\tilde{Y}}$ (16) with $p = 1$. The corresponding time-warping mappings $\varphi_{\tilde{Y}}^{-1}(\cdot)$ are shown in the lower panel of Figure 6. As would be expected, the time variation and individual differences in timing increase towards the right, with the exception of a small neighborhood around the right endpoint, where all time-warping functions are forced to converge to the last age, which is 18.

Of interest is a comparison of this specific implementation of functional convex averaging with the previously proposed landmark method, and in addition with the more recent continuous monotone registration method of Ramsay and Li (1998). Like functional convex averaging, the latter method is designed for unstructured samples of curves which exhibit structural variability (for example varying number of peaks, highly varying shapes) and therefore it is of interest to compare this established method with the proposed least complex implementation of functional convex averaging.

In contrast, the landmark method serves as a natural benchmark for these regularly structured growth velocity trajectories. The landmark method has been developed exactly for such well-structured and regular functional data for which it is known to perform extremely well (Kneip and Gasser 1992, Gasser and Kneip, 1995). We will therefore compare the behavior of continuous monotone registration, also occasionally referred to as the Procrustes method, with the simple implementation of functional convex averaging, while landmark registration serves as the target.

The comparison of the estimated sample mean curves for functional convex averaging, continuous monotone registration, and the landmark method, is illustrated in Figure 7. We find that all methods reflect both pubertal growth spurt and mid-growth spurt (at around 5 years). There are no differences in the timing of the mid-growth spurt, while such differences surface for the important pubertal growth spurt. The timing of the cross-sectional average and of the continuous monotone registration method coincide, while the same is true for functional convex averaging and landmark registration, but the two sets of methods differ by about six months in terms of where the peak is located. In terms of shape of the pubertal peak, landmark registration and functional convex averaging nearly coincide, while continuous monotone registration and cross-sectional averaging both are different. Regarding the shape of the mid-growth spurt, none of the other methods closely matches the landmark registration. Functional convex averaging (FCA-AUC) is closest to the cross-sectional average. Adjusting the area-under-the-curve mapping slightly (FCA-AUCC) might help to improve the definition of the mid-growth spurt.

In conclusion, all three alignment methods improve on the cross-sectional average. Functional convex averaging is closer to the benchmark presented by landmark registration in defining shape and timing of the important pubertal growth spurt feature, whereas continuous monotone registration performs better in defining the shape of the mid-growth spurt. Overall, functional convex averaging performs well and is the simplest among these curve registration methods.

Calculating convex functional and conventional variance measures (10) and (12), we obtain for the estimate (13) of the conventional cross-sectional measure a value of $\hat{\text{var}}(\tilde{Y}) = 30.02$ and for the estimate (11) of the functional convex measure a value of $\hat{\text{var}}_{\oplus}(\tilde{Y}) = 2.86$. The difference between these estimates provides additional evidence that time-warping explains a substantial fraction of the variability in these data.

7. Discussion and Conclusions

We have proposed a stochastic warping model based on a latent bivariate process and have developed a convex functional calculus to reflect random time-warping in a sample of stochastic processes. Families of time-synchronization mappings that emphasize the alignment of peaks and troughs were explored in more detail, without requiring that observed random curves are well-structured. We found that very simple time-synchronizing transformations in this class such as the area-under-the-curve method prove surprisingly effective. A suggestion for future work is to develop methods for choosing the synchronization map in a data-adaptive manner, for example using the minimization of the estimate $\hat{\text{var}}_{\oplus}(\tilde{Y})$ (11) of the convex functional variation measure $\text{var}_{\oplus}(\tilde{Y})$ (10) as a possible criterion, coupled with an appropriately restricted warping family such as (18).

Simulations and examples indicate that the proposed simple registration method, which is just one implementation of functional convex averaging with specifically chosen time-synchronizing functions, works well and provides interpretable results as well as improved functional means. To some extent this also works in instances in which the simple area-under-curve method does not correspond to the true warping function, as long as the true warping is not too different from the assumed one. The method breaks down when the true underlying time-synchronization functions are substantially different from the assumed ones.

One central concept of our approach is the latent bivariate process that is associated with each individual trajectory, with one component of the bivariate process representing the time-synchronizing function and the other component representing the random amplitude function. In the synchronized time space the two components – time and amplitude – are thus assumed to be separated, while they are commingled in the observed trajectories, due to the time warping. This concept allowed us to derive asymptotic consistency and inference, under the assumption that the correct time-synchronization family of maps is given.

The latent bivariate process concept for warping provides a generally applicable modeling

framework for virtually any curve registration method that can be defined in terms of time-synchronization maps. One can simply incorporate the warping functions (or their inverses) from these registration methods as time-synchronizing maps into our proposed model. The observed sample of curves is then assumed to be generated through a corresponding bivariate latent process model, coupled with a specific class of preferred warping functions. As a consequence, the convex functional calculus extends to these methods as well, and thus provides a fairly general framework for the study of warping.

The asymptotic theory and confidence bands that we developed within our modeling framework can also be extended to other registration methods, notably to landmark registration. One caveat is that our theoretical derivations require independence in the estimated time-synchronization maps. This may make it necessary to ignore or overcome (by means of a training sample) possible dependence in estimated warping functions for those registration algorithms that use the entire sample of curves to determine the warping for each individual curve.

APPENDIX

Let $\rho(Z_1, Z_2) = [\int (Z_1(t) - Z_2(t))^2\, dt]^{1/2}$ for $Z = X$ and $Z = Y$ be a distance (pseudo-metric) in $L^2[0,T]$, for $T > 0$.

Proof of Propostion 2.1. That $d_\oplus(\tilde{Y}_1, \tilde{Y}_2) \geq 0$ and $d_\oplus(\cdot,\cdot)$ is symmetric in its arguments is obvious. By (7),

$$\begin{aligned} d^2_\oplus(\tilde{Y}_1, \tilde{Y}_2) &\leq (\rho(X_1, X_3) + \rho(X_3, X_2))^2 + (\rho(Y_1, Y_3) + \rho(Y_3, Y_2))^2 \\ &= d^2_\oplus(\tilde{Y}_1, \tilde{Y}_3) + d^2_\oplus(\tilde{Y}_3, \tilde{Y}_2) + 2(\rho(X_1, X_3)\rho(X_3, X_2) + \rho(Y_1, Y_3)\rho(Y_3, Y_2)), \end{aligned}$$

and the triangle inequality follows from

$$\begin{aligned} d_\oplus(\tilde{Y}_1, \tilde{Y}_3) d_\oplus(\tilde{Y}_3, \tilde{Y}_2) &\geq (\rho^2(X_1, X_3)\rho^2(X_3, X_2) + \rho^2(Y_1, Y_3)\rho^2(Y_3, Y_2) \\ &+ 2\rho(X_1, X_3)\rho(X_3, X_2)\rho(Y_1, Y_3)\rho(Y_3, Y_2))^{1/2} \\ &= \rho(X_1, X_3)\rho(X_3, X_2) + \rho(Y_1, Y_3)\rho(Y_3, Y_2). \end{aligned}$$

Proof of Proposition 2.2. (b) follows from Proposition 2.1, and (a) and (c) are obtained by substituting the definition of the scalar product $\langle \cdot, \cdot \rangle$ given in (9).

Proof of Theorem 4.1. Note that $EX = \text{argmin}_\mu E\rho^2(\mu, X)$, which is seen by Fubini's theo-

rem and

$$\int E(X(t)-\mu(t))^2\,dt \;=\; \int E(X(t)-EX(t))^2\,dt + \int (EX(t)-\mu(t))^2\,dt.$$

Arguing analogously for Y and setting

$$(\mu_X,\mu_Y) = \text{argmin}_{(\mu_1,\mu_2)\in\mathcal{S}} E(\rho^2(\mu_1,X)+\rho^2(\mu_2,Y)),$$

we find that $\mu_X = EX,\ \mu_Y = EY$, and therefore $\text{argmin}_{\mu\in\mathcal{W}}\, Ed^2_{\oplus}(\mu,\tilde{Y})^2 = \psi(\mu_X,\mu_Y) = E_{\oplus}\tilde{Y}$.

Proof of Theorem 4.2. For $(X_i,Y_i) = \psi^{-1}(\tilde{Y}_i), i=1,\ldots,n$, we need to show that if

$$(X^*,Y^*) = \text{argmin}_{(\mu_1,\mu_2)} \left[\int_0^1 \sum_{i=1}^n (X_i(t)-\mu_1(t))^2\,dt + \int_0^1 \sum_{i=1}^n (Y_i(t)-\mu_2(t))^2\,dt\right],$$

then $(X^*,Y^*) = (\bar{X}_n,\bar{Y}_n)$. Now let $(X^*,Y^*) = (\bar{X}_n,\bar{Y}_n) + (\delta_X,\delta_Y)$ with perturbation functions (δ_X,δ_Y). Note that

$$\begin{aligned}\int \sum_{i=1}^n (X_i(t)-(\bar{X}_n(t)+\delta_x(t)))^2\,dt &= \int \sum_{i=1}^n [(X_i(t)-\bar{X}_n(t))^2 + n\delta_X^2(t)]\,dt \\ &> \int \sum_{i=1}^n (X_i(t)-\bar{X}_n(t))^2\,dt,\end{aligned}$$

if $\int \delta_X^2(t)\,dt > 0$, and analogously for Y. Note that $\bar{\tilde{Y}}_{\oplus} = \bar{Y}_n \circ \bar{X}_n^{-1} = \psi(\bar{X}_n,\bar{Y}_n)$, and thus (23) holds.

Proof of Lemma 4.1. From the Multivariate CLT, for any finite k and fixed $t_1,\ldots,t_k\in[0,1]$,

$$\sqrt{n}\begin{pmatrix}\bar{Y}_n(t_1)-EY(t_1)\\ \vdots \\ \bar{Y}_n(t_k)-EY(t_k)\end{pmatrix} \xrightarrow{\mathcal{D}} N\left(\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},\sum\right). \tag{39}$$

in distribution, where the covariance matrix $\sum = \Sigma_{t_1,t_2,\ldots,t_k} = \text{cov}(Y(t_i),Y(t_j)), i,j=1,\ldots,k$. It remains to show that $\{\sqrt{n}(\bar{Y}_n(\cdot)-EY(\cdot))\}$ is tight. We verify the moment condition [Billingsley (1968), Theorem 15.6], using (H1),

$$\begin{aligned}E\{\sqrt{n}[\bar{Y}_n(t)-EY(t) &- (\bar{Y}_n(s)-EY(s))]\}^2 = E\{\frac{1}{n}\sum_{i=1}^n (Y_i(t)-EY(t)-(Y_i(s)-EY(s)))\}^2 \\ &= E|Y(t)-EY(t)-(Y(s)-EY(s))|^2 \le C_0|t-s|^2.\end{aligned}$$

To prove Theorem 4.3, we make use of the following auxiliary result, the proof of which is omitted.

Lemma A.1 *Assume $f_n, n = 1, 2, \ldots$ is a sequence of invertible functions on a compact set A and f is an invertible function on A such that the derivative f' of f exists and is continuous. If there exists a $\delta > 0$ such that $\inf_{x\in A}|f'(x)| \geq \delta$, then $\sup_{x\in A}|f_n(x) - f(x)| = O(n^{-1/2})$ implies $\sup_{x\in A}|f_n^{-1}(x) - f^{-1}(x)| = O(n^{-1/2})$.*

We refer to $C = C[0, 1]$ as the space of continuous functions on $[0, 1]$ with the topology of uniform convergence, and to $C_0 \subset C$ as the space of those elements of C that are nondecreasing and satisfy $0 \leq \alpha(t) \leq 1$, where $\alpha \in C[0, 1]$. The following result from Billingsley (1968, p. 145) is a central tool in the proof of Theorem 4.3.

Lemma A.2 *Let U be a random element of C, V be a random element of C_0, (U, V) be a random element of $C \times C_0$ with the product topology, and for each n, U_n, V_n be elements of C and C_0 respectively. If*

$$(U_n, V_n) \Rightarrow (U, V), \tag{40}$$

then

$$U_n \circ V_n \Rightarrow U \circ V. \tag{41}$$

Proof of Theorem 4.3. Recall $\{(x, \tilde{Y}(x)),\ x \in [0, T]\} \equiv \{(X(t), Y(t)),\ t \in [0, 1]\}$ for all $\omega \in \Omega$, and that the mapping $\varphi_{\tilde{Y}} : x \mapsto t$ is monotone and invertible. Given a sample $\tilde{Y}_1, \cdots, \tilde{Y}_n \in \mathcal{W}$, $\bar{\tilde{Y}}_{\oplus}(x) = \psi(\bar{X}_n, \bar{Y}_n) = \bar{Y}_n(\bar{X}_n^{-1}(x))$, since the time transformation $\bar{X}_n^{-1}(x)$ is monotone and invertible [see (34)]. Now $\{(x, \bar{\tilde{Y}}_{\oplus}(x)),\ x \in [0, T]\} \equiv \{(\bar{X}_n(t), \bar{Y}_n(t)),\ t \in [0, 1]\}$ implies that $\bar{Y}_n(t) = \bar{\tilde{Y}}_{\oplus}(\bar{X}_n(t)),\ t \in [0, 1]$.

Setting $U_n(t) \equiv \sqrt{n}(\bar{Y}_n(t) - \mu_Y(t)),\ U \equiv G_Y,\ V_n(x) \equiv \bar{X}_n^{-1}(x),\ V(x) \equiv \mu_X^{-1}(x)$, we find by Lemma 4.1 that $U_n \Rightarrow U$. Since the $\tilde{Y}_i$, $i = 1, \ldots, n$ are i.i.d., this implies that the X_i, defined by $X_i(t) = \varphi_{\tilde{Y}_i}^{-1}(t)$, $i = 1, \ldots, n$ are also i.i.d.. Lemma 4.1 and an application of the Continuous Mapping Theorem lead to

$$\sup_{t\in[0,1]} |\bar{X}_n(t) - \mu_X(t)| = O_p(n^{-1/2}), \tag{42}$$

implying $V_n(x) \xrightarrow{p} V(x)$ uniformly for $x \in [0, T]$. Therefore, by Theorem 4.4 of Billingsley (p. 27) and Lemma 4.1 for U_n,

$$(U_n, V_n) \Rightarrow (U,\ V).$$

This guarantees that condition (40) of Lemma $A.2$ is satisfied, and by (41),

$$\sqrt{n}(\bar{Y}_n(\bar{X}_n^{-1}(x)) - \mu_Y(\bar{X}_n^{-1}(x))) \Rightarrow G_Y\ \circ \mu_X^{-1}(x). \tag{43}$$

By a Taylor expansion, for each fixed $x \in [0,T]$,

$$\begin{aligned}\sqrt{n}\{\mu_Y(\bar{X}_n^{-1}(x)) - \mu_Y(\mu_X^{-1}(x))\} &= \sqrt{n}\mu_Y'(\mu_X^{-1}(x))\{\bar{X}_n^{-1}(x) - \mu_X^{-1}(x) \\ &\quad + o_p(|\bar{X}_n^{-1}(x) - \mu_X^{-1}(x)|)\} \\ &= \sqrt{n}\mu_Y'(\mu_X^{-1}(x))(\bar{X}_n^{-1}(x) - \mu_X^{-1}(x)) + o_p(1).\end{aligned}$$

The above remainder term $o_p(1)$ is uniform in $x \in [0,T]$ because of (34). Observe that

$$\begin{aligned}\bar{X}_n^{-1}(x) - \mu_X^{-1}(x) &= \mu_X^{-1}(\mu_X(\bar{X}_n^{-1}(x))) - \mu_X^{-1}(x) \\ &= \frac{1}{\mu_X'(\zeta_n)}(\mu_X(\bar{X}_n^{-1}(x)) - x),\end{aligned}$$

where ζ_n is between $\bar{X}_n^{-1}(x)$ and $\mu_X^{-1}(x)$, using (H2).

Using a similar argument as that leading to (43), again applying Lemma $A.2$, and choosing $U_n(t) \equiv \sqrt{n}(\bar{X}_n(t) - \mu_X(t))$, $U \equiv G_X$, $V_n(x) \equiv \bar{X}_n^{-1}(x)$, $V(x) \equiv \mu_X^{-1}(x)$, we arrive at

$$\sqrt{n}\{\bar{X}_n(\bar{X}_n^{-1}(x)) - \mu_X(\bar{X}_n^{-1}(x))\} = \sqrt{n}(x - \mu_X(\bar{X}_n^{-1}(x))) \Rightarrow G_X \circ \mu_X^{-1}(x). \tag{44}$$

Applying Theorem 4.4 of Billingsley (p. 27),

$$\sqrt{n}(\bar{X}_n^{-1}(x) - \mu_X^{-1}(x)) \Rightarrow \frac{G_X \circ \mu_X^{-1}(x)}{\mu_X' \circ \mu_X^{-1}(x)}. \tag{45}$$

Combining these facts, we find

$$\sqrt{n}(\mu_Y(\bar{X}_n^{-1}(x)) - \mu_Y(\mu_X^{-1}(x)) = \sqrt{n}\frac{\mu_Y'(\mu_X^{-1}(x))}{\mu_X'(\zeta_n)}(x - \mu_X(\bar{X}_n^{-1}(x))) \Rightarrow \frac{\mu_Y' \circ \mu_X^{-1}(x)}{\mu_X' \circ \mu_X^{-1}(x)} G_X \circ \mu_X^{-1}(x). \tag{46}$$

Noting that by (2),

$$\sqrt{n}(\bar{\tilde{Y}}_\oplus(x) - E_\oplus \tilde{Y}(x)) = \sqrt{n}[\bar{\tilde{Y}}_\oplus(x) - \mu_Y(\bar{X}_n^{-1}(x)) + \mu_Y(\bar{X}_n^{-1}(x)) - \mu_Y(\mu_X^{-1}(x))], \tag{47}$$

Theorem 4.3 follows from (43) and (46).

REFERENCES

Adler, R.J. (1990), *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*, Institute of Mathematical Statistics: Lecture Notes–Monograph Series, Vol. 12.

Ash, R.B. and Gardner, M.F. (1975), *Topics in Stochastic Processes*, New York: Academic Press.

Billingsley, P. (1968), *Convergence of Probability Measures*, John Wiley & Sons, Inc.

Capra, W.B. and Müller, H.G. (1997), "An accelerated-time model for response curves," *Journal of the American Statistical Association*, 92, 72–83.

Fan, J. and Gijbels, I. (1996), *Local Polynomial Modelling and Its Applications*, New York: Chapman & Hall.

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

Gasser, Th., Köhler, W., Müller, H.G., Kneip, A., Largo, R., Molinari, L. and Prader, A. (1984a), "Velocity and acceleration of height growth using kernel estimation," *Annals of Human Biology*, 11, 397–411.

Gasser, Th., Müller, H.G., Köhler, W., Molinari, L. and Prader, A. (1984b), "Nonparametric regression analysis of growth curves," *The Annals of Statistics*, 12, 210-229.

Kneip, A. and Gasser, Th. (1992), "Statistical tools to analyze data representing a sample of curves," *The Annals of Statistics*, 16, 82–112.

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

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

Ramsay, J.O. and Silverman, B.W. (1997), *Functional Data Analysis*, New York: Springer.

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

Rao, C.R. (1958), "Some statistical models for comparison of growth curves," *Biometrics*, 14, 1–17.

Rice, J. and Silverman, B.W. (1991), "Estimating the mean and covariance structure nonparametrically when the data are curves," *Journal of the Royal Statistical Society*, Ser. B, 53, 233–243.

Rønn, B.B. (2001), "Nonparametric maximum likelihood estimation for shifted curves," *Journal of the Royal Statistical Society*, Ser. 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.

Shiau, J.J.H. and Lin, H.H. (1999), "Analyzing accelerated degradation data by nonparametric regression," *IEEE Transactions on Reliability*, 48, 149-158.

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

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

Wang, K. and Gasser, Th. (1998), "Asymptotic and bootstrap confidence bounds for the structural average of curves," *The Annals of Statistics*, 26, 972–991.

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

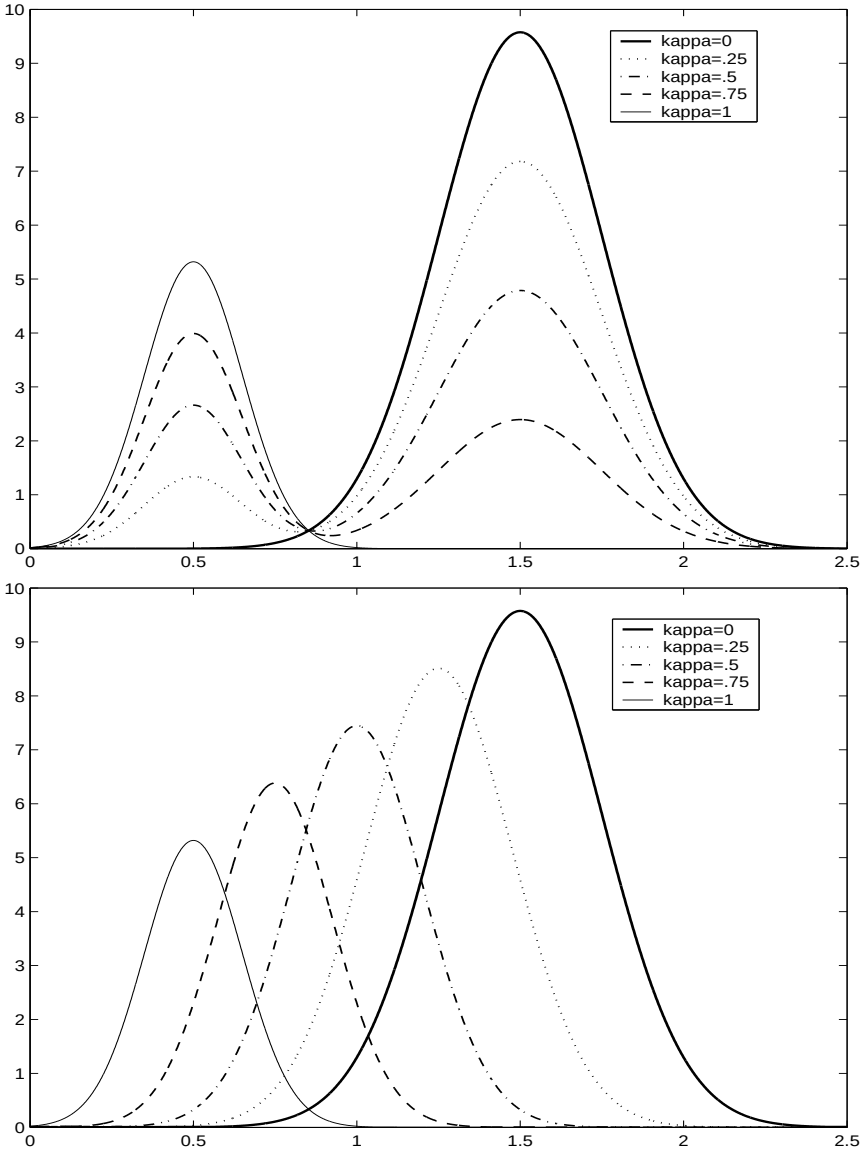


Figure 1: *Conventional* $\kappa f_1 + (1-\kappa) f_2$ *and functional* $\kappa f_1 \oplus (1-\kappa) f_2$ *convex combinations for* $f_j(x) = \beta_j \frac{1}{\sqrt{2\pi}\sigma_j} \exp\{-\frac{1}{2\sigma_j^2}(x-\mu_j)^2\},\ x \in [0,T],\ j = 1,2,$ *where* $\mu_1 = .5,\ \mu_2 = 1.5,\ \sigma_1 = .15,\ \sigma_2 = .25,\ \beta_1 = 2,\ \beta_2 = 6,\ T = 3,\ \kappa = 0, .25, .5, .75, 1.$ *Solid thin curves in both panels denote* f_1 *(left curve) and solid bold curves denote* f_2 *(right curve). Upper panel: Conventional convex combinations. Lower panel: Functional convex combinations.*

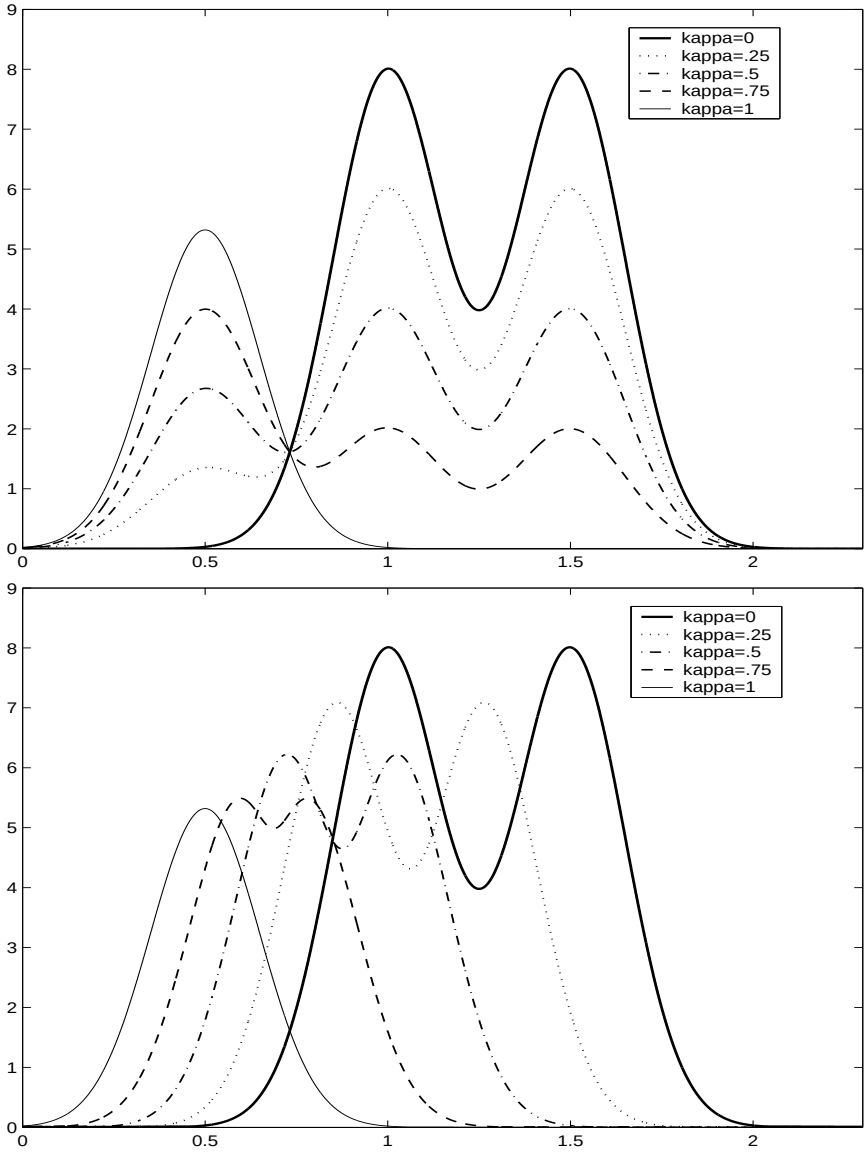


Figure 2: *Conventional* $\kappa f_1 + (1-\kappa) f_2$ *and functional* $\kappa f_1 \oplus (1-\kappa) f_2$ *convex combinations for* $f_1(x)$ *and* $f_2(x) = \frac{\beta_2}{\sqrt{2\pi}\sigma_2} \exp\{-\frac{1}{2\sigma_2^2}(x-\mu_2)^2\} + \frac{\beta_3}{\sqrt{2\pi}\sigma_3} \exp\{-\frac{1}{2\sigma_3^2}(x-\mu_3)^2\}$ *where* $\mu_1 = .5,\ \mu_2 = 1,\ \mu_3 = 1.5,\ \beta_2 = \beta_3 = 3,\ \sigma_1 = \sigma_2 = \sigma_3 = .15,\ T = 3,\ \kappa = 0, .25, .5, .75, 1$. *Solid thin curves in both panels denote* f_1 *(left curve) and solid bold curves denote* f_2 *(right curve). Upper panel: Conventional convex combinations. Lower panel: Functional convex combinations.*

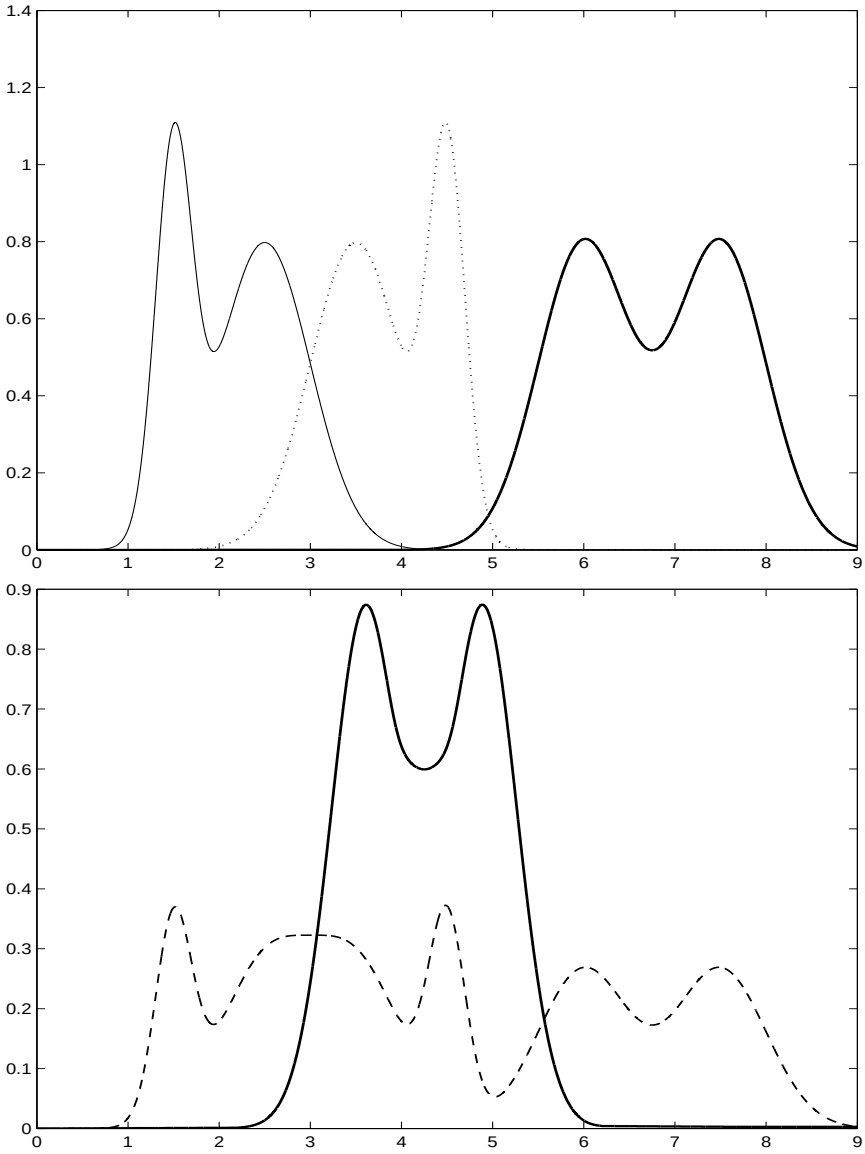


Figure 3: *Upper panel: Three curves generated as bimodal Gaussian mixtures. Lower panel: The dashed curve is the conventional cross-sectional average of the three curves, while the solid curve is their functional convex average.*

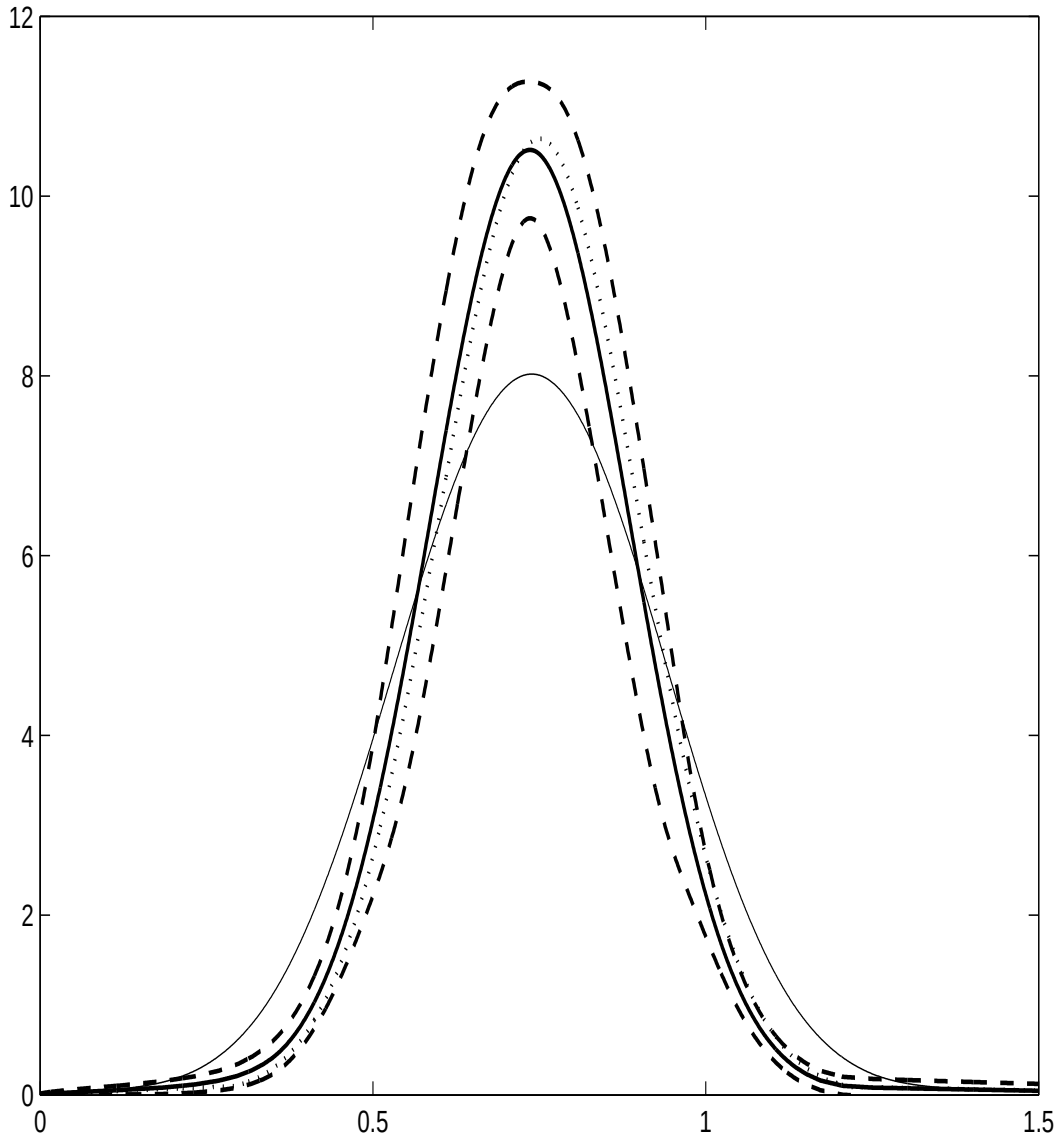


Figure 4: 95%-*confidence bands (36) for the functional convex mean curve obtained from* 50 *unimodal sample curves. Shown are target function (dotted), functional convex mean function (solid bold), conventional cross-sectional mean function (solid thin lower curve), and upper and lower confidence limits (dashed).*

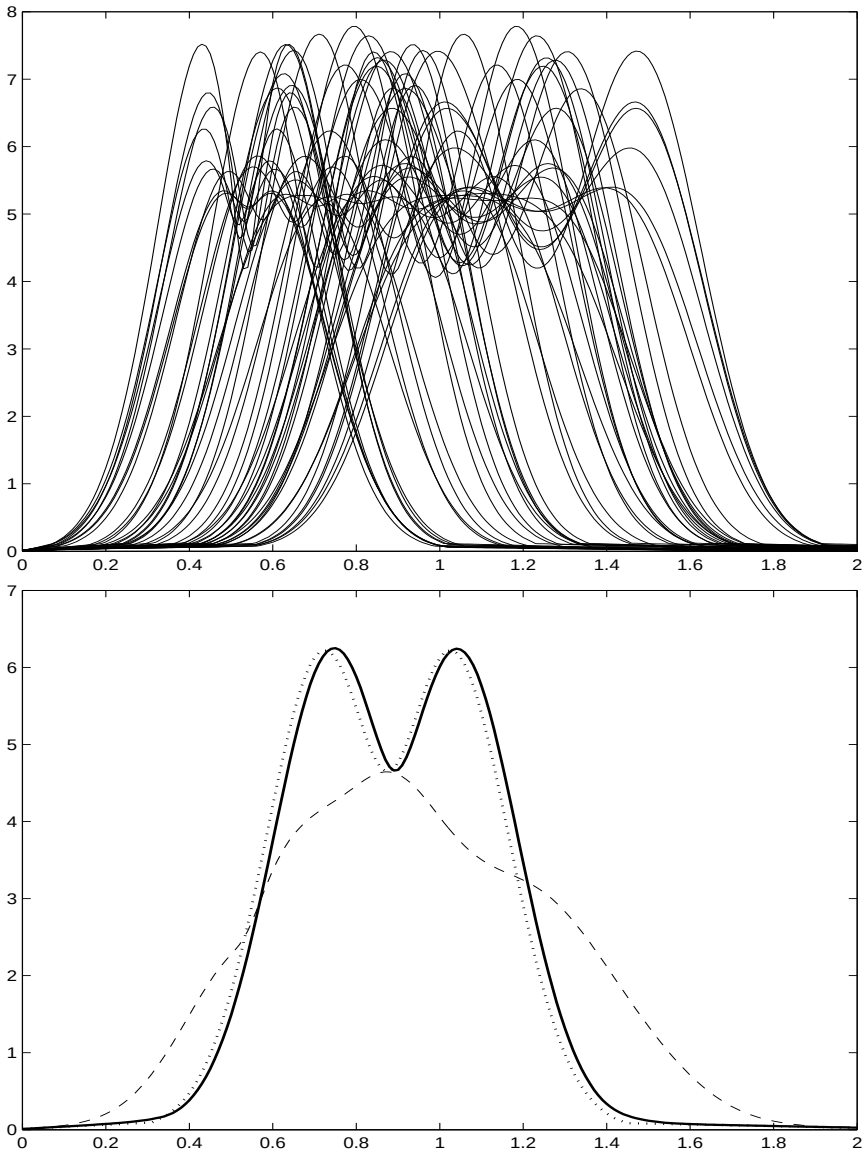


Figure 5: *Simulation for bimodal curves. Upper panel:* 50 *observed sample curves. Lower panel: Functional convex average (solid), conventional cross-sectional average (dashed), and target function (dotted).*

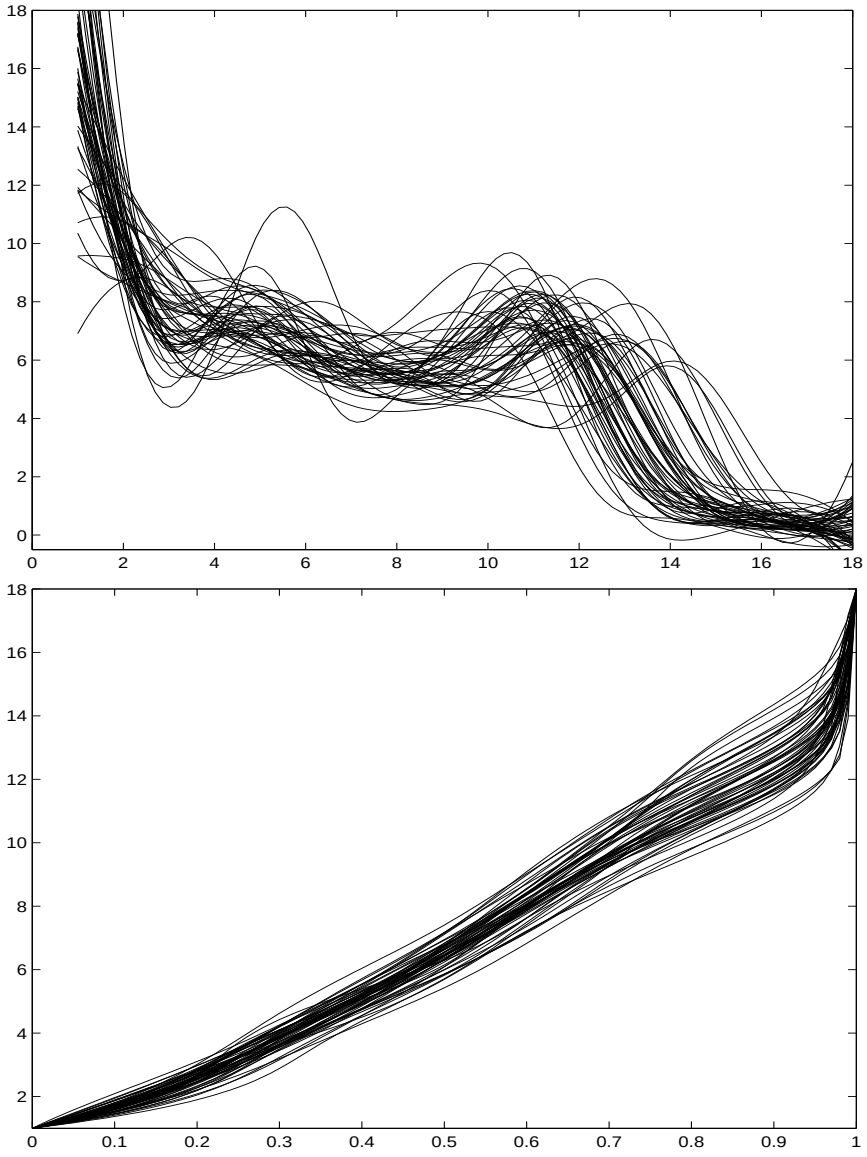


Figure 6: *The Berkeley growth data. Upper panel: Smoothed growth velocity curves for* 54 *girls. The* $x-axis$ *corresponds to time (time unit is years) and the* $y-axis$ *corresponds to the growth velocity in cm/year. Lower panel: Time-warping mapping functions* $\varphi_{\tilde{Y}}^{-1}(\cdot)$ *for these* 54 *girls.*

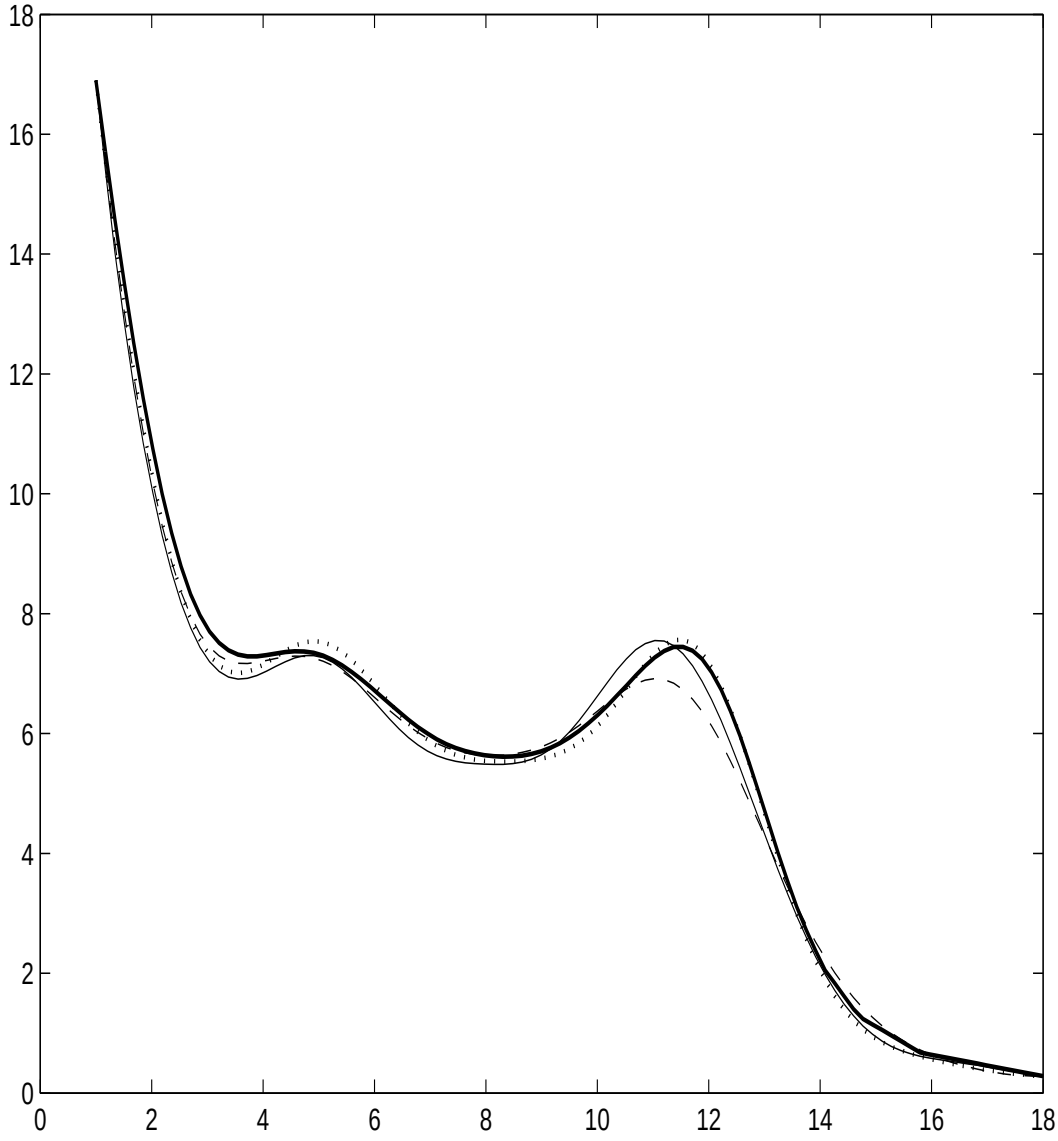


Figure 7: *Comparison of mean functions from various methods for the Berkeley growth data shown in Figure 6: Functional convex mean (solid bold), continuous monotone registration (solid), landmark registration (dotted), and cross-sectional average (dashed).*