

## SYNCHRONIZING SAMPLE CURVES NONPARAMETRICALLY<sup>1</sup>

BY KONGMING WANG<sup>2</sup> AND THEO GASSER

*SUNY Brooklyn and University of Zürich*

More and more often, the outcome of a study is not a random variable but a noisy function for each experimental unit, resulting in a sample of curves. Typically, the individual curves vary not only in amplitude or intensity, but also with respect to the time axis: different subjects experience certain events sooner or later. Analyzing such data involves finding out the time changes (or curve registration) among curves. Following our previous work where modified dynamic time warping is applied to align two curves, we formulate a global minimization problem to align all curves in a sample and to compute the aligned average curve. Algorithms for solving the minimization problem are presented and tested with simulated and real data. The test results are promising. The method, which involves kernel smoothing of regression functions, estimates the time changes and the average of the aligned curves from noisy data. Large sample asymptotics is derived.

**1. Introduction.** More and more often the outcome of a study is not a random variable but a noisy function for each experimental unit, resulting in a sample of curves. Examples are growth curves and brain potentials in medicine, speech signals in engineering and expenditure curves for various goods and countries in economics. The data  $y_{ij}$  are usually obtained at discrete time points  $t_{ij}$  (or some other discrete grid) and can be modeled as follows:

$$(1) \quad y_{ij} = f_i(t_{ij}) + \varepsilon_{ij}, \quad j = 1, \dots, n_i; \quad i = 1, \dots, m.$$

Here,  $f_i$  denotes the true individual regression function for subject  $i$ , and  $\varepsilon_{ij}$  denote independent zero mean random errors with variance  $V(\varepsilon_{ij}) = \sigma_i^2 > 0$ . We assume  $t_{ij} \in J \equiv [0, 1]$  without loss of generality. Typically, the individual  $f_i$  vary not only in amplitude or intensity, but also with respect to the time axis: different subjects experience certain events sooner or later. Classical methods like repeated measures ANOVA or principal component analysis of curves [Rao (1958), Rice and Silverman (1991)] ignore the second type of variability. This may severely distort the analysis and in particular the pattern of an average curve (“interference bias”).

These problems led Kneip and Gasser (1992) to apply a strictly monotone time transformation  $g_i$  from a physical to a “biological” time scale (in a biological application) in order to eliminate (nonlinear) shifts between curves (the functions  $g_i$  are also called warping or alignment or registration functions).

---

Received April 1997; revised October 1998.

<sup>1</sup>Supported by Swiss NSF Grants 20-42253.94 and B-2006.295.5942.

<sup>2</sup>Also supported by NIH Grants AA 05524 and AA 02686.

AMS 1991 subject classifications. Primary 62G07; secondary 62H05.

Key words and phrases. Curves, dynamic time warping, kernel estimation, structural analysis, warping functions.

Analyzing subsequently the  $f_i(g_i(t))$  instead of the  $f_i(t)$  leads to more meaningful results, in particular to a more meaningful “structural” average curve, reflecting the average dynamic and the average intensity. The determination of the transformation  $g_i$  is, in this approach, based on features common to the sample of curves, followed by monotone interpolation.

Figure 1 illustrates the structural averaging procedure, computed by dynamic time warping. The subgraph (d) of Figure 1 shows that small details of the sample curves are preserved in the average of the aligned curves. The application of structural analysis to longitudinal growth curves led to a wealth of new information on the many facets of growth and some underlying mechanisms [Gasser, Kneip, Binding, Prader and Molinari (1991), Gasser, Kneip, Zieger, Molinari, Prader and Largo (1994)]. However, the definition of features and their unequivocal identification in individual curves might pose problems in some applications [Gasser and Kneip (1995) have a proposal for defining common features]. As an alternative, we studied dynamic time warping (DTW) for computing the time transformation between two curves in a nonparametric way [Wang and Gasser (1997)]. DTW has been developed in engineering, in particular for speech analysis [Sakoe and Chiba (1978)].

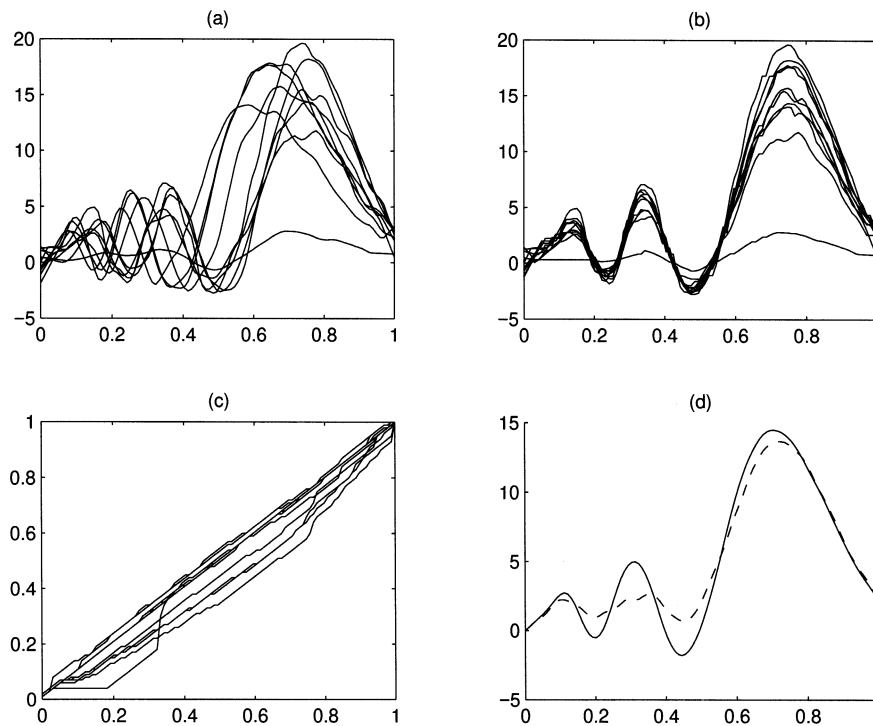


FIG. 1. (a) 10 sample curves; (b) aligned sample curves; (c) the estimated time transformations used for the alignment; (d) the average curves [solid: average of the aligned curves in (b); dashed: average of the sample curves in (a)].

One problem that arises in the application of DTW to a sample of curves is that it applies only to a pair of curves, typically an individual curve and some template or master curve. This problem will be the focus of this paper. We define a global minimization problem for samples of curves rendering the unknown warping functions as well as the average of the aligned curves. We propose an iterative algorithm for estimation, and we will study convergence and statistical properties of this procedure. The incorporation of shift functions into a spline smoothing framework has been studied by Ramsay and Li (1998). In a rather different approach to the problem, Silverman (1995) models the shift functions parametrically and estimates the parameters iteratively. A recent book by Ramsay and Silverman (1997) gives a good introduction to functional data analysis. The advantage of the DTW is its generality in choosing the time transformations which are assumed only to be strictly increasing.

The paper is organized as follows: Section 2 describes our way of performing DTW to make this paper self-contained. In Section 3 we form a global variational problem whose solution gives the time transformations and the average of the aligned curves. An iterated estimation procedure for solving the variational problem is also described in Section 3. Computational details are discussed in Section 4. Section 5 is devoted to simulations and applications. Statistical asymptotics is derived in Section 6.

**2. Dynamic time warping (DTW).** Suppose that two sequences  $\{f(l), l = 1, \dots, M\}$  and  $\{h(j), j = 1, \dots, N\}$  characterize two signals  $f$  and  $h$ . The best match between  $f$  and  $h$  by some alignment  $w$  is achieved by minimizing some cost function, classically

$$\inf_w \sum_{(l, j) \in w} (f(l) - h(j))^2.$$

The warping path  $w = \{(l, j)\}$  connects  $(1, 1)$  and  $(M, N)$  in a two-dimensional square lattice, has to be monotone and has in addition to satisfy some side conditions penalizing a too irregular path [Parsons (1986); Qi (1992)]. The method of choice for estimating  $w = ((l(1), j(1)), \dots, (l(K), j(K)))$  is dynamic programming, and this is also true for general cost functions  $C$  of the form

$$C(f, h, w) \equiv \sum_{k=1}^K d(f(l(k)), h(j(k))),$$

where  $d(\cdot, \cdot)$  is a distance measure and thus symmetric in  $f$  and  $h$ . The length  $K$  of the warping path is determined as well while minimizing the cost function  $C$ . In the framework of Wang and Gasser (1997) we are moving from discrete signals to functions  $f$  and  $h$  and from a warping path to a monotone warping function  $g$ . The cost function introduced in Wang and Gasser (1997) is motivated by the Sobolev norm and by the goal of aligning important common features such as extrema and inflection points,

$$(2) \quad C(f, f', h, h', g, \alpha) \equiv \int_0^1 [F(f, f', h, h', g, \alpha)(t) + 2\phi(g'(t))] dt,$$

where

$$F(f, f', h, h', g, \alpha)(t) \equiv \alpha^2 \left( \frac{f(t)}{\|f\|} - \frac{h(g(t))}{\|h\|} \right)^2 + (1 - \alpha)^2 \left( \frac{f'(t)}{\|f'\|} - \frac{h'(g(t))}{\|h'\|} \right)^2$$

with  $\|f\|$  being the sup-norm of  $f$ .

The variational problem

$$(3) \quad \inf_{\alpha} \inf_g C(f, f', h, h', g, \alpha)$$

can be solved by dynamic programming for fixed  $\alpha$  and by grid search for  $\alpha \in [0, 1]$ .

REMARK 2.1. For  $\alpha = 1$ , the alignment is completely determined by the amplitudes of the two curves, and for  $\alpha = 0$  by the amplitudes of their derivatives. In general, the alignment depends on both the amplitudes of the curves and the amplitudes of their derivatives, and the parameter  $\alpha$  tunes the compromise. The sum of the weights,  $\alpha^2 + (1 - \alpha)^2$ , has a minimum at  $\alpha = 0.5$ . This means that we would weight the alignment of the amplitudes of the two curves and the alignment of the amplitudes of their derivatives equally importantly unless the data clearly suggest otherwise. In all computations, we solve the variational problem (3) for three values  $\alpha = 0.3, 0.5, 0.7$  and then choose the best alignment which yields smallest cost  $C(f, f', h, h', g, \alpha)$ .

REMARK 2.2. Standardizing the functional  $F$  using division by the respective sup-norm makes the asymptotic analysis much more complicated but offers definite advantages: it helps to avoid explaining differences in amplitude wrongly by warping functions. Ramsay and Li (1998), who do not use such a normalization, also pointed out this problem.

REMARK 2.3. The function  $\phi$  penalizes too irregular warping functions allowed to before, in particular too steep warping functions. A simple choice is  $\phi(x) = a(x - 1)^2$  for some  $a > 0$ . A general  $\phi$  can be defined by the following assumption.

ASSUMPTION 0. The function  $\phi$  satisfies the following conditions:  $\phi$  is convex;  $\phi(x) = 0$  for  $x \in [\delta + r, \Delta - r]$  with some given constants  $\Delta > \delta > 0$  and  $(\Delta - \delta)/2 \geq r > 0$ ;  $\phi(\delta^+) = \phi(\Delta^-) = \infty$ ;  $\phi(x) = \infty$  for  $x \in (\delta, \Delta)^c$  and  $\phi \in C^4[\delta, \Delta]$ .

REMARK 2.4. In our formalization we have gone from a (discrete) warping path  $w$ , as suggested in the engineering literature, to a smooth warping function  $g$  (at least  $g \in C^1$ ). This offers substantial advantages in formal terms. Frequently, a common starting point is known which leads to  $g(0) = 0$ .

REMARK 2.5. An important advantage of the new cost function compared to established ones is the following: in quite general classes of regression models,

dynamic time warping is able to identify the correct warping function [Wang and Gasser (1997)], which is not true for other cost functions. It should also be pointed out that important parametric models such as the logistic or Gompertz model have linear warping functions as an inherent property to characterize individual differences.

REMARK 2.6. It makes sense to add terms involving derivatives of higher orders to the cost function  $C$ . However, these higher order derivatives have to be estimated from noisy data. While many methods, for example, the kernel method [Gasser and Müller (1984)] and local polynomial procedures [Fan (1992)] have been proposed for derivative estimation, the estimation error always increases sharply as the order of the derivative increases. The cost function used in this paper involves only first-order derivatives, which are estimated by the kernel method (see Section 4.1). In our experience this is feasible even for small to moderate sample size despite the unattractive asymptotic performance for derivatives.

**3. Estimating warping functions and a structural average curve.**

3.1. *Formalization of the problem.* Dynamic time warping produces a relative warping function between two curves. With a set of functions  $\{f_i\}$ , we want to estimate a set of warping functions  $\{g_i\}$  in order to align all the sample curves to an average dynamic. To this end, we define a cost functional,

$$\begin{aligned}
 & \mathcal{F}(g_1, \dots, g_m; f_1, \dots, f_m)(t) \\
 (4) \quad &= \sum_{i=1}^m \left[ \alpha^2 \left( \frac{f_i(g_i(t))}{\|f_i\|} - \frac{1}{m} \sum_{j=1}^m \frac{f_j(g_j(t))}{\|f_j\|} \right)^2 \right. \\
 & \quad \left. + (1 - \alpha)^2 \left( \frac{f'_i(g_i(t))}{\|f'_i\|} - \frac{1}{m} \sum_{j=1}^m \frac{f'_j(g_j(t))}{\|f'_j\|} \right)^2 + 2\phi(g'_i(t)) \right].
 \end{aligned}$$

As in the last section, the parameter  $\alpha$  weights the alignment of curves and the alignment of their derivatives, and  $\phi$  is a penalty function to regularize the warping functions. The best warping functions are given as the solution of the variational problem

$$(5) \quad \inf_{\{g_i\}} \int_0^1 \mathcal{F}(g_1, \dots, g_m; f_1, \dots, f_m)(t) dt,$$

where  $g_i \in C^1[0, 1]$ ,  $g_i(0) = 0$  and  $g_i(1) = 1$ .

Once the warping functions have been obtained, the structural average  $\mu$  is then computed as a smooth fitting to the aligned sample curves,

$$(6) \quad \inf_{\mu} \int_0^1 \left[ \frac{1}{m} \sum_{i=1}^m (f_i(g_i(t)) - \mu(t))^2 + \lambda(\mu''(t))^2 \right] dt,$$

with  $\mu \in C^2[0, 1]$  and smoothing parameter  $\lambda \geq 0$ . Note that the average warping function, or the average development of individuals, is given by

$$\bar{g}(t) = \frac{1}{m} \sum_{i=1}^m g_i(t).$$

Therefore, the average  $\mu(t)$  is the amplitude of the structural average at time  $\bar{g}(t)$ . That is,  $\{(\bar{g}(t), \mu(t)): t \in [0, 1]\}$  defines the structural average.

REMARK 3.1. From the cost function (4) and (5), it is clear that the best warping functions will align the sample curves to the average of the aligned curves. There always exists a solution to (5) because the objective function to be minimized is bounded below and because the minimization is performed in a compact set of functions. When  $\phi$  is strictly convex, the solution is unique.

REMARK 3.2. The conditions  $g_i(0) = 0$  and  $g_i(1) = 1$  correspond to known start and end points of the underlying process. An example is the growth from birth to an adult age (say 20 years old). The case of unknown start and end points can be easily dealt with. Details are given in Section 4.

REMARK 3.3. Rice and Silverman (1991) proposed a penalized smoothing estimation of a representative curve for a set of curves, using unaligned data. We have replaced their term  $\sum_{i=1}^m (1/n_i) \sum_{j=1}^{n_i} (f_i(t_{ij}) - \mu(t_{ij}))^2$  by  $\int_0^1 \sum_{i=1}^m (f_i(g_i(t)) - \mu(t))^2 dt$  so that the penalized smoothing estimation is based on aligned data.

Any solution of (5) has to satisfy the Euler equations

$$(7) \quad \frac{\partial \mathcal{F}}{\partial g_i} - \frac{d}{dt} \frac{\partial \mathcal{F}}{\partial g_i'} = 0, \quad i = 1, \dots, m.$$

These equations lead to a system of nonlinear equations,

$$(8) \quad \begin{aligned} 0 = & \alpha^2 \left( \frac{f_i(g_i(t))}{\|f_i\|} - \frac{1}{m} \sum_{j=1}^m \frac{f_j(g_j(t))}{\|f_j\|} \right) \frac{f_i'(g_i(t))}{\|f_i\|} \\ & + (1 - \alpha)^2 \left( \frac{f_i'(g_i(t))}{\|f_i'\|} - \frac{1}{m} \sum_{j=1}^m \frac{f_j'(g_j(t))}{\|f_j'\|} \right) \frac{f_i''(g_i(t))}{\|f_i'\|} \\ & - \phi''(g_i'(t)) g_i''(t). \end{aligned}$$

While these equations are useful for the theoretical analysis (Section 6), they are not an attractive way, computationally, to estimate the warping functions  $g_i$ .

Similarly, a solution of (6) has to satisfy the Euler equation

$$(9) \quad \frac{\partial \mathcal{F}}{\partial \mu} + \frac{d}{dt^2} \frac{\partial \mathcal{F}}{\partial \mu''} = 0.$$

This leads to the spline smoothing of the average of the aligned curves,

$$(10) \quad \lambda\mu^{(4)}(t) + \mu(t) = \frac{1}{m} \sum_{i=1}^m f_i(g_i(t)).$$

3.2. *An iterative algorithm.* Since spline smoothing is well understood, we will focus on (5) and propose an iterative algorithm, based on dynamic time warping (as described in Section 2), for solving (5). A discussion of (10) will be given in Section 4. For given  $k \geq 1$ , let

$$\eta_i^k(t) = \frac{1}{m} \sum_{j=1}^m \frac{f_j(g_j^{k-1}(t))}{\|f_j\|}, \quad \tau_i^k(t) = \frac{1}{m} \sum_{j=1}^m \frac{f'_j(g_j^{k-1}(t))}{\|f'_j\|}.$$

The iterative algorithm works as follows: first, set the initial warping functions  $g_i^0(t) = t$  (no alignment). Then update the warping functions as follows: for  $k \geq 1$ , find  $g_i^k(t)$  as the solution of the minimization problem

$$(11) \quad \inf_{g_i} \int_0^1 \left[ \alpha^2 \left( \frac{f_i(g_i(t))}{\|f_i\|} - \eta_i^k(t) \right) + (1 - \alpha)^2 \left( \frac{f'_i(g_i(t))}{\|f'_i\|} - \tau_i^k(t) \right)^2 + 2\phi(g'_i(t)) \right] dt$$

for  $i = 1, \dots, m$ . Note that (11) is of the form (3) and can be solved by dynamic time warping. Further computation details such as the number of iterations and the choice of the weight parameter  $\alpha$  will be discussed in Section 4.

Now we discuss convergence of the above algorithm. To simplify the analysis, we will assume that  $\phi(x) = a(x - 1)^2/2$  for some  $a > 0$ . This guarantees that the solution to (5) is unique. For a general  $\phi$  defined by Assumption 0 of Section 2, the discussion becomes more complicated because of possible nonuniqueness of the solution. Write (8) in a compact form

$$ag''_i(t) = H_i(F_1(g_1(t)), \dots, F_m(g_m(t)))$$

for some functionals  $H_i$  ( $i = 1, \dots, m$ ), where

$$F_i(t) \equiv (f_i(t), \|f_i\|, f'_i(t), \|f'_i\|, f''_i(t))^T.$$

In matrix form it is

$$ag''(t) = H(F_1(g_1(t)), \dots, F_m(g_m(t)))$$

with  $g = (g_1, \dots, g_m)^T$  and  $H = (H_1, \dots, H_m)^T$ . The solution of this equation can be written in integral form as

$$g(t) = \frac{1}{\alpha} \int_0^t du \int_0^u H(F_1(g_1(s)), \dots, F_m(g_m(s))) ds + (C_1 t, \dots, C_m t)^T,$$

where  $C_1, \dots, C_m$  are constants determined by the boundary condition  $g(1) = (1, \dots, 1)^T$ . It follows that

$$(12) \quad g(t) = \begin{pmatrix} t \\ \vdots \\ t \end{pmatrix} - \frac{1}{\alpha} \int_0^1 \min(t, s)[1 - \max(t, s)] \\ \times H(F_1(g_1(s)), \dots, F_m(g_m(s))) ds.$$

This represents the warping functions as the identity function plus fluctuations (or deformation functions). With a relatively large constant  $\alpha$ , the right-hand side of (12) is a contraction mapping because  $H$  has bounded second-order partial derivatives. The contraction mapping has a fixed point. The sequence  $\{g^k\}$  produced by the iterative algorithm converges to the fixed point. This proves the convergence of the above iterated algorithm.

**3.3. Relations to models.** To end this section, we discuss another important problem: if we have a set of curves generated by some model, what does the algorithm retrieve as aligned average curve  $\mu$ ? We look at two classes of models which are important in applications. We will assume that  $\phi''(g'_i(t)) = 0$  holds for all  $i = 1, \dots, m$ . This is equivalent to assuming "regularity" for the alignments defined by Assumption 0.

Our first example is the shape invariant model [Lawton, Sylvestre and Maggo (1972), Stützle, Gasser, Molinari, Largo, Prader and Huber (1980)]. Here the individual curves are generated from a common shape function  $\psi$  by linear time transformations and linear amplitude variation,

$$(13) \quad f_i(t) = a_i \psi\left(\frac{t - b_i}{c_i}\right) + d_i, \quad i = 1, \dots, m.$$

Note that this class of models includes the logistic and the Gompertz models as important special cases. A relatively demanding estimation procedure for estimating the individual parameters  $\{a_i, b_i, c_i, d_i: i = 1, \dots, m\}$  together with the model  $\psi$  has been described and studied by Kneip and Gasser (1988). Note that identifiability conditions on parameters such as  $a_i > 0$ ,  $c_i > 0$ ,  $a_1 + \dots + a_m = 1$  and  $d_1 + \dots + d_m = 0$  have to be imposed.

When we take  $\alpha = 0$  and when ignoring the boundary condition  $g_i(0) = 0$  and  $g_i(1) = 1$ , the solution of (5) is given by  $g_i(t) = b_i + c_i t$ . Different boundary conditions can be incorporated into the algorithm (Section 4). With these warping functions, we get

$$\mu(t) = \frac{1}{m} \sum_{i=1}^m f_i(g_i(t)) = \psi(t).$$

Thus, the true shape function can be recovered and we have a possibility of estimating a model nonparametrically [see also Kneip and Engel (1995)].



The second, rather general, example is the nonlinear shift model where we have

$$f_i(t) = a_i \psi(h_i(t))$$

for some strictly increasing functions  $h_i$ . It is easy to see that a solution of (5) is given by  $g_i(t) = h_i^{-1}(t)$  when setting  $\alpha = 1$ . If we define  $(a_1 + \dots + a_m)/m = 1$ , then we get

$$\mu(t) = \frac{1}{m} \sum_{i=1}^m f_i(g_i(t)) = \psi(t).$$

Again, the true shape function could be recovered, which is quite remarkable for a model as general as this. Note that this model is not uniquely defined unless some restrictions are imposed on the time transformations  $h_i$ . For strictly increasing one-to-one maps  $\gamma: [0, 1] \mapsto [0, 1]$ , we can rewrite the model as  $f_i(t) = a_i \nu(\xi_i(t))$  with  $\xi_i(t) = \gamma(h_i(t))$  and  $\nu(t) = \psi(\gamma^{-1}(t))$ . This would result in the nonuniqueness of the solution of (5) if  $\phi$  is not strictly convex. This means that the basic shape of  $\psi$  will be recovered, but the positions of the peaks of  $\psi$  cannot be uniquely determined.

For more general cases, where recovering the model may not be possible in full generality, the algorithm still provides good approximations (see Section 5).

**4. Some computational considerations.** This section discusses computational details of the iterative algorithm of the last section.

4.1. *Estimating derivatives.* The most frequently used methods for estimating regression functions and their derivatives are kernel smoothing, spline smoothing and local polynomial fitting. These three methods are asymptotically equivalent for fixed design as dealt with here. We will apply the kernel method of Gasser and Müller (1984) to the data of the form (1),

$$(14) \quad \hat{f}_i^{(k)}(t) = \hat{f}_i^{(k)}(t; b) = \frac{1}{b^{k+1}} \sum_{j=1}^{n_i} \int_{s_{ij-1}}^{s_{ij}} y_{ij} K_k\left(\frac{t-u}{b}\right) du.$$

Here  $k$  is the order of derivatives,  $K_k$  is a kernel of order  $k + 2$  and  $b$  is a global bandwidth which is independent of  $t$  but depends on  $k$ . The  $s$ -sequence is defined by  $s_{ij} = (t_{ij} + t_{i,j+1})/2$ ,  $j = 1, \dots, n_i - 1$ ,  $s_{i0} = 0$  and  $s_{i n_i} = 1$ .

The selection of the bandwidth  $b$  is very important for the quality of the estimate. Here the bandwidth is chosen by the plug-in method: estimate the optimal bandwidth which minimizes the estimated asymptotically optimal mean (integrated) squared error [Gasser, Kneip and Kohler (1991)]. The subroutine used in our computation is `glkern.f`, which was written by Eva Herrmann and can be obtained from the Web site [www.unizh.ch/biostat/](http://www.unizh.ch/biostat/).

The optimal bandwidth increases as the order  $k$  of the derivative increases, and a larger bandwidth produces smoother estimates. To reduce the amount of irregular alignments due to larger errors when estimating derivatives, the

derivatives of the sample curves are estimated with oversmoothing. First, the optimal bandwidth  $b_2$  for estimating  $f_i''$  ( $k = 2$ ) is estimated. Then we use this bandwidth  $b_2$  in (14) with  $k = 1$  to estimate the first derivatives of  $f_i$ . This oversmoothing reduces the peak amplitudes of  $f_i'$  but it preserves the shape of  $f_i'$ . For the alignment, the shapes of  $f_i$  and  $f_i'$  are more important than the peak amplitudes because of the normalization by the supremum norm.

4.2. *The parameter  $\alpha$ .* The parameter  $\alpha$  determines whether the alignment should rely more on the function itself or on its derivative. The best  $\alpha$  can be obtained by a grid search. For a finite number of  $\alpha \in [0, 1]$ , solve the variational problem (5) and compute the cost  $\mathcal{F}_\alpha$  (the minimum value of  $\mathcal{F}$  given  $\alpha$ ). Then the best weight parameter  $\alpha^*$  is given by the solution of

$$\min_{\alpha} \mathcal{F}_\alpha.$$

To speed up computation, we have chosen the best weight parameter  $\alpha^*$  from three values  $\{0.3, 0.5, 0.7\}$  (see also the Remark 2.1). The simulations of Section 5 show that this choice works well.

4.3. *The case of unknown starting and ending times.* The start and end points of the underlying process may be unknown. An example is the recording of evoked potentials of the human brain. The effect of a stimulus on the human brain (e.g., subjects push a button with a finger as soon as possible after they see a specified graph on a computer screen) lasts for different periods for different subjects, and the length of the period is unknown. This means that the end points of the potentials are unknown.

It is easy to deal with unknown start and end points in programming. In each iteration of the algorithm of Section 3, every curve is aligned to the average of the aligned curves from previous iterations. If  $g_i^k(t) = 0$  for  $0 < t \leq t_0 < 1$ , then the starting time of the  $i$ th curve is later than the average starting time, and one can omit the  $i$ th curve when computing the average of the aligned curves for the time interval  $[0, t_0]$  as if there were only  $m - 1$  curves. On the other hand, if  $g_i^k(0) > 0$ , then the starting time of the  $i$ th curve is earlier than the average starting time, and that part of the  $i$ th curve,  $\{f_i(t): 0 \leq t \leq g_i^k(0)\}$ , can be omitted. End times can be treated in the same way.

4.4. *The penalty function  $\phi$ .* On one hand, the penalty function  $\phi$  should be strictly convex such that the solution of the minimization problem (5) is unique. On the other hand,  $\phi$  should be flat so that it does not produce artificial warping functions by pushing toward the identity function  $g(t) = t$ . To avoid estimating derivatives of the warping functions, we have replaced the single penalty term  $\phi(g_i'(t))$  with three conditions in programming.

(A) The warping functions  $g_i(t)$  are nondecreasing. This is automatically obtained by dynamic time warping.

- (B) The warping functions  $g_i(t)$  do not jump steeply:  $|g_i(t) - t| \leq \theta$  for some  $\theta > 0$ . We have used  $\theta = 1/3$  and  $\theta = 1/2$  in the simulations and application of Section 5. Small  $\theta$  means a narrow search by dynamic time warping speeding up the computation. If one has no information about the time transformations between sample curves, a relatively large  $\theta$  should be specified, say  $\theta = 1/2$ .
- (C) Within the boundary of (B), a strictly convex penalty function on the warping functions is defined by

$$\phi(g_i(t)) = \delta_p \frac{(g_i(t) - t)^2}{1 + (g_i(t) - t)^2}.$$

We choose  $\delta_p = 0.001$ , resulting in a flat penalty function.

While these conditions are easy to handle in programming, the penalty term  $\phi(g'_i(t))$  of (5) is easier for a theoretical analysis.

4.5. *Number of iterations.* We are unable to perform a theoretical analysis on the convergence rate of the iterative algorithm of the last section. Our simulations show that a few iterations (less than or equal to five iterations) produce nice results. In all computations in this paper, we use three iterations.

4.6. *The initial warping functions.* The initial warping functions,  $g_i^0(t) = t$  of the iterative algorithm of Section 3, work well in most cases. However, in some extreme cases where large time transformations occur, these initial warping functions could slow down the convergence of the iterative algorithm and even lead to misalignment of some components with relatively small amplitudes. If we had a template curve, then we could obtain initial warping functions by warping each sample curve to the template (Section 2). The template can be obtained as follows. Choose a few sample curves (say  $2^q$  curves,  $q = 2$  or  $3$ ) such that these curves have the smallest distance from the average of the sample curves. Divide these curves into  $2^{q-1}$  pairs. Warp one curve to another for each pair and compute an aligned average of each pair. Then divide the  $2^{q-1}$  averages into  $2^{q-2}$  pairs. Repeat this procedure until there is only one average curve left, which is the template.

4.7. *The parameter  $\lambda$ .* There are well-known methods such as GCV for choosing the smoothing parameter  $\lambda$ . If one uses the smoothed curves in the computation of the right-hand side of (10), then one can take  $\lambda = 0$ . This produces a visually smooth average of the aligned curves in most cases, and this is what we do in the simulations and applications.

**5. Simulations and applications.** This section presents simulation results of the algorithm of Section 3 and an application to growth curves. An application to the well-known gait data (for the knee) has also led to very satisfactory results. Fortran subroutines used for these computations can be

obtained by sending a request via E-mail to Wang (kmw@mendel.neurodyn.hscbklyn.edu).

5.1. *Simulations.* The shape function is defined by

$$s(t) = 15(\exp(-20(t - 0.7)^2) - 0.5 \exp(-50(t - 0.45)^2) \\ + 0.6 \exp(-100(t - 0.3)^2) - 0.6 \exp(-150(t - 0.2)^2) \\ + 0.5 \exp(-200(t - 0.15)^2)).$$

This function has several positive and negative components, simulating the basic pattern of visually evoked brain potentials. Figure 2 plots the function  $s(t)$ . The simulated data are generated from the model

$$f_i(t) = \alpha_i s(h_i(t)) + \varepsilon_i(t).$$

Here  $\alpha_i$  is an individual constant and  $\varepsilon_i(t)$  is white noise. We will assume an equally spaced design with  $n_i = 100$  observations per curve. The number of curves is set to  $m = 30$ . We consider two forms of the time transformations  $h_i$ .

- (F1) Quadratic transformations  $h_i(t) = t + b_i t(1 - t)$  for constants  $b_i$ . To insure that  $h_i$  is strictly increasing, the condition  $|b_i| < 1$  has to be satisfied. It is obvious that  $h_i(0) = 0$  and  $h_i(1) = 1$ .
- (F2) More complicated transformations  $h_i(t) = t + b_i \sin(2\pi c_i t)$  for integers  $c_i \in \{0, 1, 2, 3\}$ . To insure that  $h_i$  is strictly increasing, the condition  $|2\pi c_i b_i| < 1$  has to be satisfied. Again we have  $h_i(0) = 0$ ,  $h_i(1) = 1$  and in addition  $h_i(1/2) = 1/2$ .

For comparisons between the estimated average of the aligned curves and the shape function, we will require that

$$\frac{1}{m} \sum_{i=1}^m \alpha_i = 1, \quad \frac{1}{m} \sum_{i=1}^m h_i(t) = t.$$

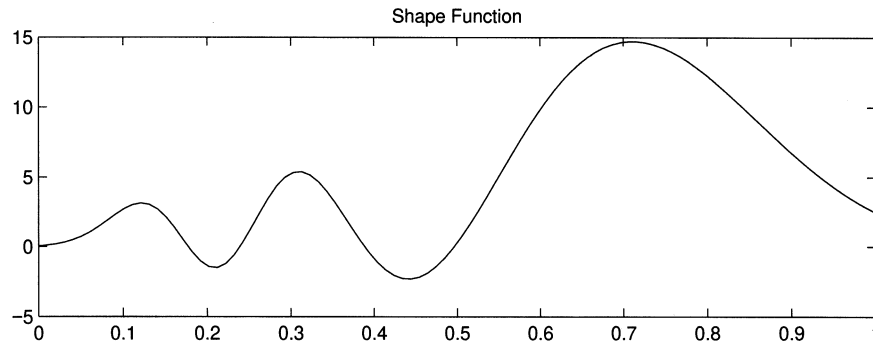


FIG. 2. The shape function used in the simulations of Section 5.

It follows that the shape function becomes the average of the aligned curves. For (F1), the latter condition implies that  $\sum_{i=1}^m b_i = 0$ . It is a bit more complicated for the model (F2), but a sufficient condition is

$$(15) \quad \sum_{\{i: c_i=j\}} b_i = 0 \quad \text{for } j = 0, 1, 2, 3.$$

The parameters are generated as follows. For  $\alpha_i$ , first generate  $m$  normal random variables  $N_i, i = 1, \dots, m$ . Then normalize them as  $A_i = 0.5N_i/N$  where  $N = \max_i |N_i|$ . Set

$$\alpha_i = 1 + A_i - \frac{1}{m} \sum_{j=1}^m A_j.$$

The generation of  $c_i$  is simple:

$$c_i = \text{round}(3U(0, 1)),$$

where  $U(0, 1)$  is a uniform random variable on  $(0,1)$  and the function  $\text{round}(x)$  rounds up a number  $x$  to its nearest integer. The generation of the parameters  $b_i$  depends on the model for time transformations. For (F1), again generate  $m$  normal random variables  $M_i, i = 1, \dots, m$  and center them as in

$$B_i = M_i - \frac{1}{m} \sum_{j=1}^m M_j.$$

Then let

$$b_i = \frac{0.5B_i}{\max_j |B_j|}.$$

For (F2), we start from  $m$  normal random variables and repeat the following two steps:

- (i) Center the variables according to (15).
- (ii) If  $|2\pi c_i b_i| \geq 1$  for some  $i$ , replace  $b_i$  by  $0.95\text{sign}(b(i))/|2\pi c_i|$ .

These two steps are repeated alternatively until  $|2\pi c_i b_i| < 1$  is satisfied and the parameters  $b_i$  are centered according to (15).

For model (F1), it is easy to see that  $|h_i(t) - t| = |b_i t(1 - t)| \leq 1/2$  but it may be  $|h_i(t) - t| = 1/2$  for some  $i$ . So the search band parameter  $\theta$  of the condition (B) of Section 4.4 should be at least  $\theta = 1/2$ ; otherwise the optimal warping functions could not be found for some sample curves. We will run the simulation with both  $\theta = 1/2$  and  $\theta = 1/3$  to show possible wrong alignment with a too small band parameter  $\theta$ . For model (F2), we have  $|h_i(t) - t| = |b_i \sin(2\pi c_i t)| \leq 1/(2\pi) < 1/3$ . So  $\theta = 1/3$  is appropriate for the model (F2).

Two error measurements are computed for each replication. First, the mean squared error of the estimated structural average is given by

$$\text{MSEA} = \int_0^1 [s(t) - \mu(\bar{g}(t))]^2 dt.$$

Second, the error of estimating the warping functions  $\{g_i\}$  (the inverse of the time transformations  $h_i$ ) is given by

$$\text{MSEW} = \frac{1}{m} \sum_{i=1}^m \int_0^1 [h_i(g_i(t)) - t]^2 dt.$$

The simulation results with 500 runs are plotted in Figures 3 and 4.

The average MSEA of 500 runs is 0.5906 (standard deviation 0.3051), 6.0194 (standard deviation 3.1141) and 0.1428 (standard deviation 0.0377) for the time transformation models (F2), (F1) with the search band parameter  $\theta = 1/3$  and (F1) with  $\theta = 1/2$ , respectively. The average MSEW of estimating warping functions in 500 runs is 0.0002 (standard deviation 0.0005), 0.0101 (standard deviation 0.0041) and  $2.2667e-06$  (standard deviation  $4.4808e-05$ ) for the models (F2), (F1) with  $\theta = 1/3$  and (F1) with  $\theta = 1/2$ , respectively.

For the simple time transformations (F1), the estimation of the warping functions is quite accurate with a large enough  $\theta$ . When  $\theta$  is too small, the estimation error could be large. For the more complicated time transformations (F2), it is more difficult to estimate the warping functions even though the

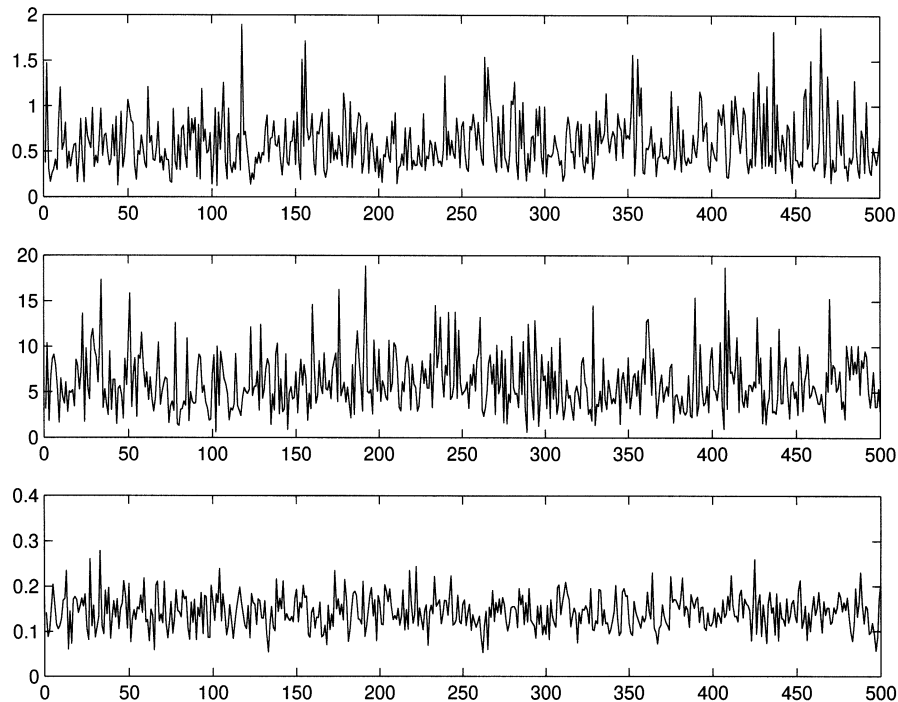


FIG. 3. The errors MSEA of 500 runs. Top: model (F2); middle: model (F1) with  $\delta = 1/3$ ; bottom: model (F1) with  $\delta = 1/2$ .

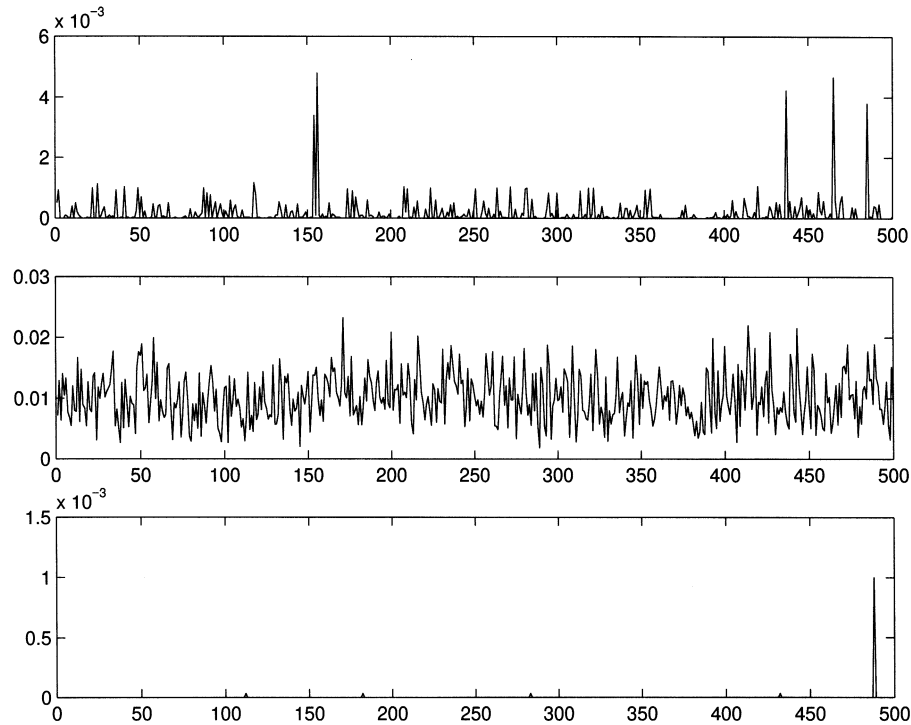


FIG. 4. The errors MSEW of 500 runs. Top: model (F2); middle: model (F1) with  $\delta = 1/3$ ; bottom: model (F1) with  $\delta = 1/2$ .

amount of time transformations ( $\sup_t |h_i(t) - t|$ ) is relatively small. Figure 5 plots a typical run with the time transformations (F2).

5.2. *Application to growth data.* The Zürich growth data is analyzed here. The data are the measurements of the growth (height, shoulder width, etc.) of children from birth to adulthood. There are 120 boys and 112 girls. For each person, the measurements are taken quarterly from birth to one year old, semiannually from one to two years old, annually from two to ten years old and then semiannually from ten until pubertal growth has stopped. This results in at most 36 measurements per person. The actual number of measurements per person ranges from 27 to 35, with an average of 32 measurements per person.

Here we discuss growth velocity of the shoulder widths of the children. The growth velocity is more informative since growth curves are simply increasing functions of age. From birth to two years old, the growth speed is high but it is simply decreasing. Therefore, we will analyze the growth speed of the children's shoulder widths from 2 to 20 years old. The measurement unit is in centimeters. Figures 6 and 7 plot the results for boys and girls, respectively.

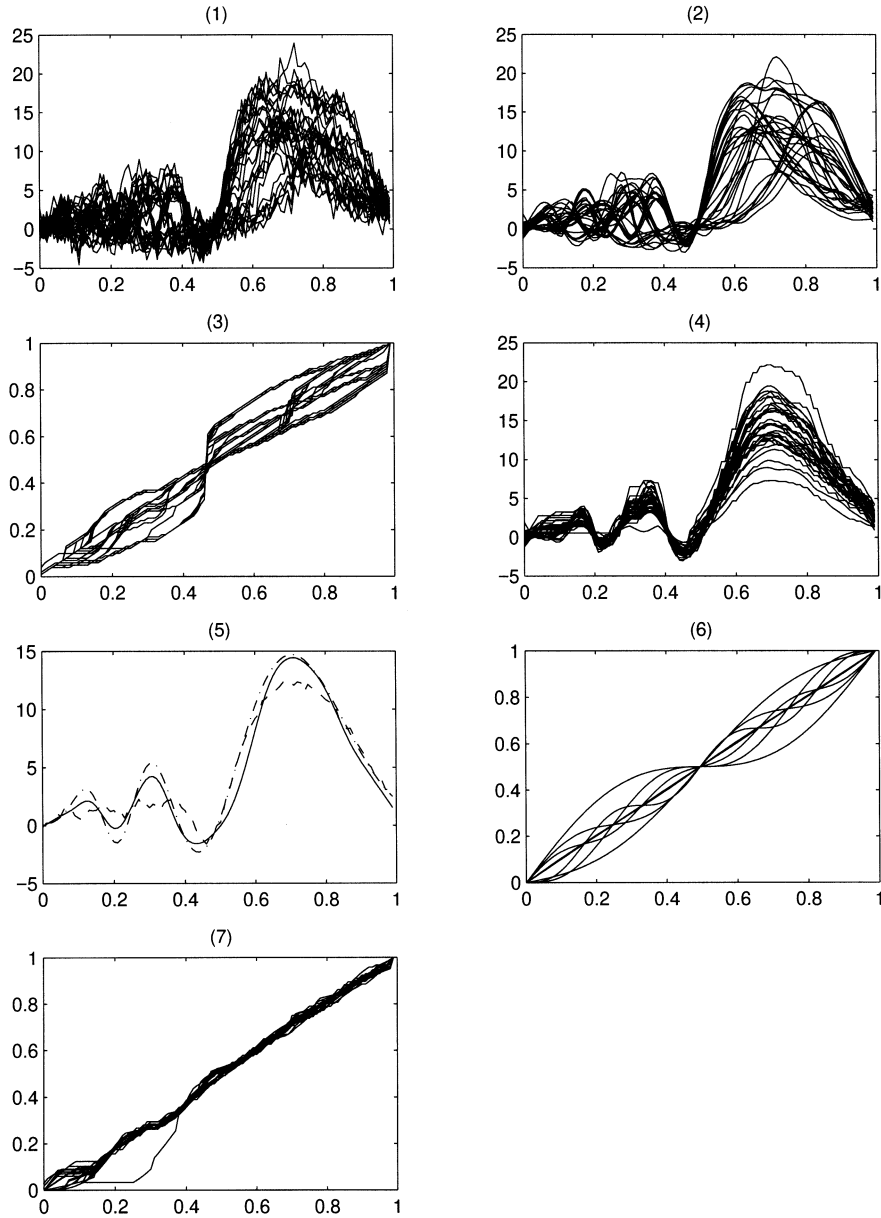


FIG. 5. A typical run of the simulations with model (F2): (1) the 30 sample curves; (2) the smoothed curves; (3) the estimated warping functions  $g_i(t)$ ; (4) the aligned curves; (5) the shape function (dash-dot), the average of the aligned curves (solid), and the average of the sample curves (dash); (6) the true time transformations  $h_i(t)$ ; (7) the plot of the functions  $h_i(g_i(t))$ , which should be  $h_i(g_i(t)) = t$  for all  $i$ .



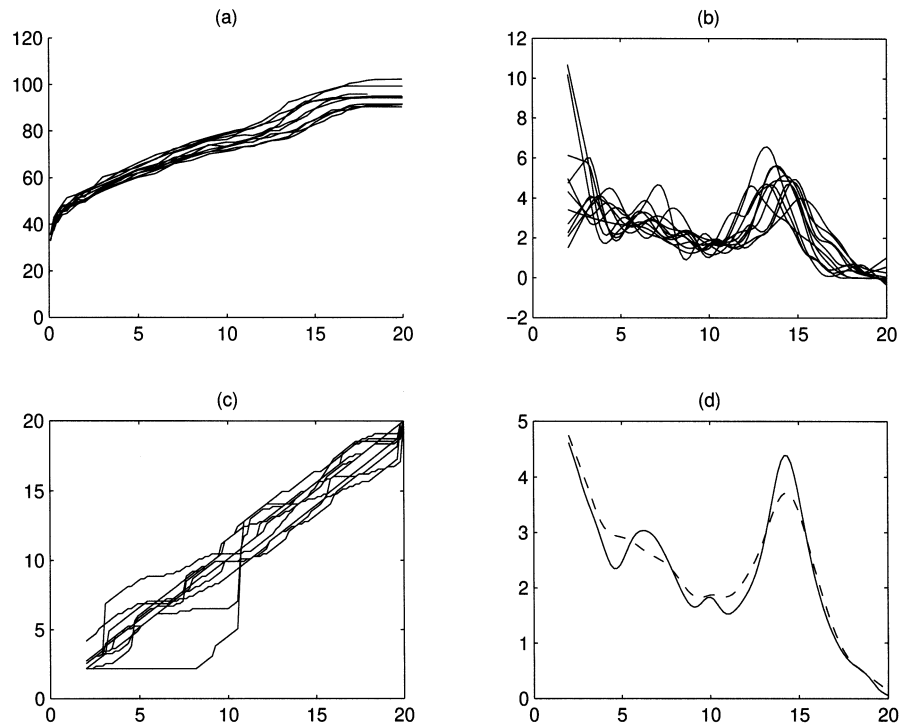


FIG. 6. Growth velocity of shoulder widths of 120 boys. (a) measurements of shoulder widths from 10 boys. (b) estimated growth speed of the 10 sample curves in (a). (c) the 10 warping functions to align the 10 curves of (b) to the average development pace of boys. (d) the average growth speed of boys: average of aligned speeds (solid) and average of unaligned speeds (dash).

From subgraph (b) of Figures 6 and 7, we see that the growth speed of a child's shoulder width changes quite a bit between individuals. The simple average of the growth speed of more than 100 children does catch the growth spurt at about age 14, but the amplitude of this peak is reduced. Furthermore, it misses the early spurt at about age 7 [see subgraph (d) of Figures 6 and 7]. For boys, there is a small peak at about age 10. This phenomenon has been seen before but its biological relevance is unclear. The midgrowth spurt occurs at about the same age for boys and girls, but the major spurt occurs at about age 13 for girls and at about age 15 for boys. These results agree with the findings of Gasser, Kneip, Binding, Prader and Molinari (1991), Gasser, Kneip, Zieger, Molinari, Prader and Largo (1994).

These and further simulations and applications illustrate that the algorithm of Section 3 works well indeed.

**6. Statistical analysis.** In many applications data are of the form (1), and the functions  $f_i$  and their derivatives can, for example, be estimated by the kernel method described in Section 4. Then the solution  $\{g_i^*\}$  of (5) is

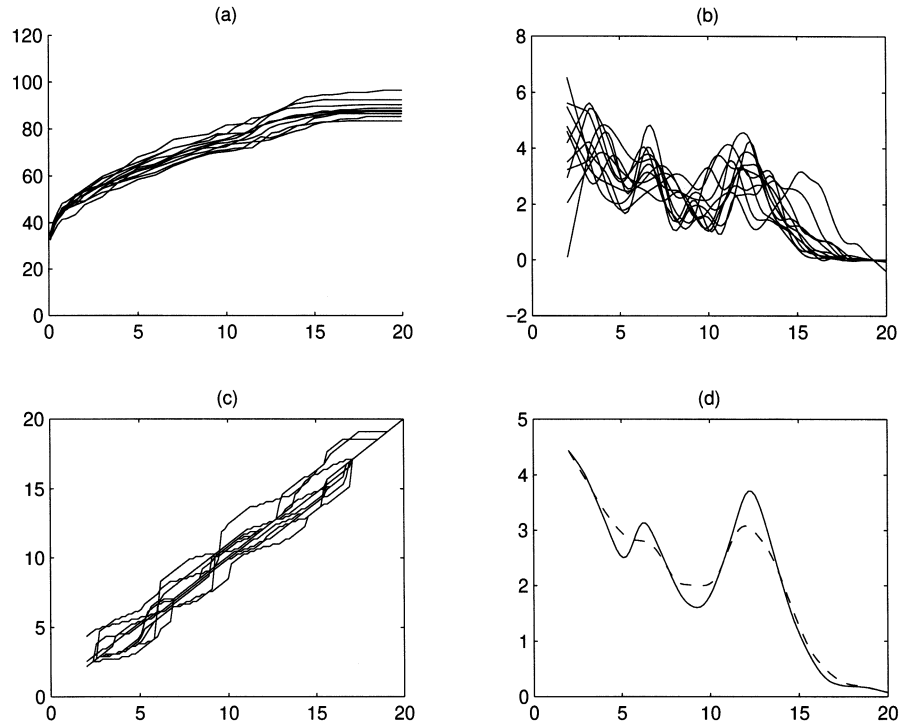


FIG. 7. Growth speed of shoulder widths of 112 girls. (a) measurements of shoulder widths from 10 girls. (b) estimated growth speed of the 10 sample curves in (a). (c) the 10 warping functions to align the 10 curves of (b) to the average development pace of girls. (d) the average growth speed of girls: average of aligned speeds (solid) and average of unaligned speeds (dash).

estimated by  $\{\hat{g}_i\}$ , which solve the minimization problem

$$(16) \quad \inf_{\{g_i\}} \int_0^1 \mathcal{F}(g_1, \dots, g_m; \hat{f}_1, \dots, \hat{f}_m)(t) dt.$$

Similarly, the solution  $\mu^*$  of (6) is estimated by  $\hat{\mu}$ , which solves the minimization problem (6) in which  $f_i$  is replaced by  $\hat{f}_i$  and  $g_i$  by  $\hat{g}_i$ . We will derive the order of the bias and variance of these estimators based on the following assumptions.

ASSUMPTION 1. There exists a constant  $\rho > 0$  such that

$$\delta + r + \rho \leq \inf_i \inf_t g'_i(t) \leq \sup_i \sup_t g'_i(t) \leq (\Delta - r) - \rho.$$

As to parameters  $\delta$ ,  $r$  and  $\Delta$ , see Assumption 0.

ASSUMPTION 2. For all  $i$ ,  $f_i \in \Omega \equiv \{f \in C^4[0, 1]: \|f^{(k)}\| \geq \delta_0 > 0, k = 0, 1\}$  for a positive number  $\delta_0$ .

ASSUMPTION 3. For a bounded continuous functional  $Z: \Omega \mapsto \mathbf{R}$ , there exists a continuous function  $Z^\infty \in C[0, 1]$  such that

$$(17) \quad \frac{1}{m} \sum_{i=1}^m Z(f_i)(g_i(t)) - Z^\infty(t) = O_p(m^{-1/2}).$$

ASSUMPTION 4. The gradient matrix

$$A(t) = \nabla H(F_1(g_1(t)), \dots, F_m(g_m(t))) \begin{pmatrix} \nabla_g F_1(g_1(t)) \\ \vdots \\ \nabla_g F_m(g_m(t)) \end{pmatrix}$$

has an inverse  $A^{-1}(t)$  in a neighborhood of  $t$ .

REMARK 6.1. Assumption 1 imposes a regularity condition on  $\{g_i\}$ . Since the parameters  $\delta$ ,  $\Delta^{-1}$ ,  $r$ , and  $\rho$  can be small and since we expect regular warping functions  $g_i$  in applications, Assumption 1 does not restrict the applicability. It simplifies, however, the statistical analysis because (8) is reduced to a system of algebraic equations. In principle, the differential equations can be treated in a similar way and we will make a remark on it [Remark 6.3(ii)]. Assumption 2 is the usual smoothness condition and avoids also linear curves which are degenerate cases in kernel fitting. Assumption 3 is satisfied if the sample  $\{f_i\}$  follows some probabilistic model. It simply says that for bounded continuous functionals on a proper Banach space, the central limit theorem holds. Finally, Assumption 4 makes the solution of (8) unique in a neighborhood of  $t$  and hence makes it possible to derive asymptotics of  $\hat{\mu}$  and  $\{\hat{g}_i\}$  at  $t$ .

THEOREM 6.1. Under Assumptions 0–4, the following conclusions hold.

(i) The order of bias is given by

$$E(\hat{g}_i(t)) - g_i(t) = O\left(b^2 + \left(\frac{\log(n)}{nb^3}\right)^{1/2}\right) = E(\hat{\mu}(t)) - \mu(t).$$

(ii) The stochastic terms  $(nb^5)^{1/2}[\hat{g}_i(t) - E(\hat{g}_i(t))]$  ( $i = 1, 2, \dots, m$ ) and  $(mnb^5)^{1/2}[\hat{\mu}(t) - E(\hat{\mu}(t))]$  of the estimators are asymptotically normal.

(iii) The estimators  $\{\hat{g}_i\}$  for the warping functions are asymptotically independent.

REMARK 6.2. (i) The true shift functions  $g_i$  depend on the sample functions  $f_i$  and the pattern of the underlying process, which is estimated by the average of the aligned sample curves. The asymptotic independence of the  $\{\hat{g}_i\}$  simply says that the individuals are developing at their own pace, respectively, and the development of one individual is not influenced by others.

(ii) It would be possible to write down an explicit formula for the leading bias term of  $\hat{\mu}$  and  $\hat{g}_i$ , if we knew a formula for the leading bias term of  $\|f_i^{\hat{(k)}}\|$  as an estimator of  $\|f_i^{(k)}\|$ ,  $k = 0, 1$ .

PROOF OF THE THEOREM. The proof goes as follows. The differential equations (8) are reduced to algebraic equations because of Assumption 1. Then use a Taylor expansion to express  $\hat{g}_i(t) - g_i(t)$  as a linear function of the statistics  $\{\hat{f}_i(g_i(t)) - f_i(g_i(t))\}, \dots, \{\hat{f}_i''(g_i(t)) - f_i''(g_i(t))\}$  and higher order error terms. The theorem follows from known results for kernel estimation of regression functions. Therefore, we will keep the proof short and present only the ideas of the proof.

From (8) and Assumption 1, we get

$$0 = H_i(F_1(g_1(t)), \dots, F_m(g_m(t)))$$

for a functional  $H_i$ . In matrix form it is

$$0 = H(F_1(g_1(t)), \dots, F_m(g_m(t))) \quad \text{and} \quad H = (H_1, \dots, H_m)^T.$$

Recall that  $F_i(t) \equiv (f_i(t), \|f_i\|, f_i'(t), \|f_i'\|, f_i''(t))^T$ . As sample size per subject increases to infinity,  $\hat{F}_i(t)$  converges to  $F_i(t)$  almost surely. Again, Assumption 1 implies that the estimated shift functions  $\{\hat{g}_i\}$  satisfy the same equation

$$0 = H(\hat{F}_1(\hat{g}_1(t)), \dots, \hat{F}_m(\hat{g}_m(t))).$$

Combining the last two equations and applying the Taylor expansion, we derive that

$$\begin{aligned} (18) \quad 0 &= H(\hat{F}_1(\hat{g}_1(t)), \dots, \hat{F}_m(\hat{g}_m(t))) - H(F_1(g_1(t)), \dots, F_m(g_m(t))) \\ &= A(t) \begin{pmatrix} \hat{g}_1(t) - g_1(t) \\ \vdots \\ \hat{g}_m(t) - g_m(t) \end{pmatrix} + B(t) + \text{higher order error term.} \end{aligned}$$

The matrix  $A(t)$  is defined in Assumption 4 and the term  $B(t)$  is defined by

$$\begin{aligned} B(t) &= \begin{pmatrix} B_1(t) \\ \vdots \\ B_m(t) \end{pmatrix} \\ &= \nabla H(F_1(g_1(t)), \dots, F_m(g_m(t))) \begin{pmatrix} \hat{F}_1(g_1(t)) - F_1(g_1(t)) \\ \vdots \\ \hat{F}_m(g_m(t)) - F_m(g_m(t)) \end{pmatrix}. \end{aligned}$$

The matrices  $A(t)$  and  $B(t)$  can be calculated and they are given by complicated yet explicit formulas.

From well known results for kernel smoothing [Gasser and Müller (1984)], one can derive that

$$E(B_i(t)) = O\left(b^2 + \left(\frac{\log(n)}{nb^3}\right)^{1/2}\right),$$

and that the functions  $\{B_i: i = 1, \dots, m\}$  are asymptotically independent.

Based on Assumptions 3 and 4, the inverse of the matrix  $A(t)$  exists. Denote this inverse by  $A^{-1}(t)$ . Then

$$(19) \quad \begin{pmatrix} \hat{g}_1(t) - g_1(t) \\ \vdots \\ \hat{g}_m(t) - g_m(t) \end{pmatrix} = A^{-1}(t) \begin{pmatrix} B_1(t) \\ \vdots \\ B_m(t) \end{pmatrix} + \text{higher order error terms.}$$

This leads to the bias estimation

$$E(\hat{g}_i(t)) - g_i(t) = O\left(b^2 + \left(\frac{\log(n)}{nb^3}\right)^{1/2}\right).$$

The leading stochastic term of  $\hat{g}_i(t)$  is obtained similarly. First, the results for kernel smoothing [Gasser and Müller (1984)] lead to  $(B_i(t) - E[B_i(t)]) = O_p((nb^5)^{-1/2})$  for all  $i$ . Second, the asymptotic independence and normality of  $\{B_i(t): i = 1, \dots, m\}$  leads to the asymptotic independence and normality of the warping functions  $\hat{g}_i(t)$ .

The statistical analysis of the estimator  $\hat{\mu}$  follows from the asymptotics of  $\hat{g}_i(t)$  and (10). This completes the proof of the theorem.  $\square$

REMARK 6.3. (i) In Theorem 6.1 and its proof, we assume that the weight parameter  $\alpha$  is fixed. The best choice of the parameter  $\alpha$  (see Section 4.2) can be estimated from the data. For the case of two curves, this is treated in Wang and Gasser (1997). We will not deal with the statistical properties of the estimator of  $\alpha$  in this paper.

(ii) The convergence rates of the estimators are still valid without Assumption 1, but the proof is more complicated. As an example we take the special penalty function  $\phi(x) = \alpha(x - 1)^2/2$  for a constant  $\alpha > 0$ . Apply the derivation of (12) again to get

$$\hat{g}(t) = \begin{pmatrix} t \\ \vdots \\ t \end{pmatrix} - \frac{1}{\alpha} \int_0^1 \min(t, s)[1 - \max(t, s)]H(\hat{F}_1(\hat{g}_1(s)), \dots, \hat{F}_m(\hat{g}_m(s))) ds.$$

This shows that

$$\begin{aligned} \hat{g}(t) - g(t) &= -\frac{1}{\alpha} \int_0^1 \min(t, s)[1 - \max(t, s)] \\ &\quad \times [H(\hat{F}_1(\hat{g}_1(s)), \dots, \hat{F}_m(\hat{g}_m(s))) \\ &\quad - H(F_1(g_1(s)), \dots, F_m(g_m(s)))] ds. \end{aligned}$$

By expanding the difference

$$H(\hat{F}_1(\hat{g}_1(s)), \dots, \hat{F}_m(\hat{g}_m(s))) - H(F_1(g_1(s)), \dots, F_m(g_m(s)))$$

as we do in the proof of Theorem 6.1 and by approximating the nonlinear integral equations with linear equations, we can derive the same convergence rates in this case.

**Acknowledgments.** We thank the referees and an Associate Editor for very helpful comments.

## REFERENCES

- FAN, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.* **87** 998–1004.
- GASSER, T. and KNEIP, A. (1995). Searching for structure in curve samples. *J. Amer. Statist. Assoc.* **90** 1179–1188.
- GASSER, T., KNEIP, A., BINDING, A., PRADER, A. and MOLINARI, L. (1991). The dynamics of linear growth in distance, velocity and acceleration. *Ann. Human Biology* **18** 187–205.
- GASSER, T., KNEIP, A. and KÖHLER, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86** 643–652.
- GASSER, T., KNEIP, A., ZIEGER, P., MOLINARI, L., PRADER, A. and LARGO, R. (1994). Development and outcome of indices of obesity in normal children. *Ann. Human Biology* **21** 275–286.
- GASSER, T. and MÜLLER, H. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.
- KNEIP, A. and ENGEL, J. (1995). Model estimation in nonlinear regression under shape invariance. *Ann. Statist.* **23** 551–570.
- KNEIP, A. and GASSER, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–112.
- KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze data representing a sample of curves. *Ann. Statist.* **20** 1266–1305.
- LAWTON, W. H., SYLVESTRE, E. A. and MAGGO, M. S. (1972). Self-modeling regression. *Technometrics* **14** 513–532.
- PARSONS, T. (1986). *Voice and Speech Processing*. McGraw-Hill, New York.
- QI, Y. (1992). Time normalization in voice analysis. *J. Acoust. Soc. Amer.* **92** 2569–2576.
- RAMSAY, J. and LI, X. (1998). Curve registration. *J. Roy. Statist. Soc. Ser. B* **60** 351–363.
- RAMSAY, J. and SILVERMAN, B. (1997). *Functional Data Analysis*. Springer, New York.
- RAO, C. (1958). Some statistical methods for the comparison of growth curves. *Biometrics* **14** 1–17.
- RICE, J. and SILVERMAN, B. (1991). Estimating the mean and covariance structure nonparametrically when data are curves. *J. Roy. Statist. Soc. Ser. B* **53** 233–243.
- SAKOE, H. and CHIBA, S. (1978). Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoustics, Speech, and Signal Processing* **26** 43–49.
- SILVERMAN, B. (1995). Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B* **57** 673–689.
- STÜTZLE, W., GASSER, T., MOLINARI, L., LARGO, R., PRADER, A. and HUBER, P. (1980). Shape-invariant modeling of human growth. *Ann. Human Biology* **7** 507–528.
- WANG, K. and GASSER, T. (1997). Alignment of curves by dynamic time warping. *Ann. Statist.* **25** 1251–1276.

DEPARTMENT OF PSYCHIATRY  
 BOX 1203  
 SUNY HSCB  
 450 CLARKSON AVE  
 BROOKLYN, NY 11203  
 E-MAIL: kmw@mendel.neurodyn.hscbklyn.edu

DEPARTMENT OF BIostatISTICS, ISPM  
 UNIVERSITY OF ZÜRICH  
 SUMATRASTRASSE 30  
 CH-8006 ZÜRICH  
 SWITZERLAND  
 E-MAIL: tgasser@biostat9.unizh.ch