7

The registration and display of functional data

7.1 Introduction

We can now assume that our observations are in functional form, and want to proceed to consider methods for their analysis. We are not quite ready, however; a problem of critical importance to functional data needs a solution. We see often that variation in functional observations involves both phase and amplitude, and that confounding these two leads to many problems. Our main emphasis is on *registration* of the data, involving transformations of the argument t rather than the values $x(t)$.

Figure 1.2 illustrates a problem that can frustrate even the simplest analyses of replicated curves. Ten records of the acceleration in children's height show individually the salient features of growth: the large deceleration during infancy is followed by a rather complex but small-sized acceleration phase during late childhood. Then the dramatic acceleration-deceleration pulses of the pubertal growth spurt finally give way to zero acceleration in adulthood. But the timing of these salient features obviously varies from child to child, and ignoring this timing variation in computing a cross-sectional mean function, shown by the heavy dashed line in Figure 1.2, can result in a estimate of average acceleration that does not resemble any of the observed curves. In this case, the mean curve has less variation during the pubertal phase than any single curve, and the duration of the mean pubertal growth spurt is rather larger than that of any individual curve.

The problem is that the growth curves exhibit two types of variability. *Amplitude variability* pertains to the sizes of particular features such as the

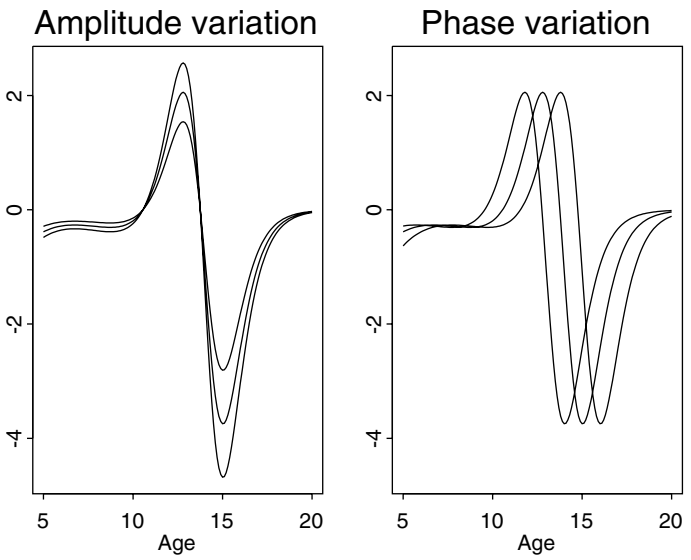

Figure 7.1. The left panel shows three height acceleration curves varying only in amplitude. The right panel shows three curves varying only in phase.

velocity peak in the pubertal growth spurt, ignoring their timings. *Phase variability* is variation in the timings of the features without considering their sizes. Before we can get a useful measure of a typical growth curve, we must separate these two types of variation, so that features such as the pubertal spurt occur at roughly the same "times" for all girls. The problem is expressed in schematic terms in Figure 7.1, where we see in the left panel two acceleration curves that differ only in amplitude, and in the right panel two curves with the same amplitude, but differing in phase.

The need to transform curves by transforming their arguments, which we call *curve registration*, can be motivated as follows. The rigid metric of physical time may not be directly relevant to the internal dynamics of many real-life systems. Rather, there can be a sort of biological or meteorological time scale that can be nonlinearly related to physical time, and can vary from case to case.

Human growth, for example, is the consequence of a complex sequence of hormonal events that do not happen at the same rate for every child. The intensity of the pubertal growth spurts of two children should be compared at their respective ages of peak velocity rather than at any fixed age. A colleague with a musical turn of mind refers to this as differences in the *tempo* of growth.

Similarly, weather is driven by ocean currents, reflectance changes for land surfaces, and other factors that are timed differently for different spatial locations and different years. Winter comes early in some years, and

late in others, and typically arrives later at some weather stations than others. We need to assess how cold the average winter is at the time the average temperature bottoms out rather than at any fixed time.

Put more abstractly, the values of two or more function values $x_i(t_i)$ can in principle differ because of two types of variation. The first is the more familiar vertical variation, or *amplitude variation*, due to the fact that $x_1(t)$ and $x_2(t)$ may simply differ at points of time t at which they are compared, but otherwise exhibit the same shape features at that time. But they may also exhibit *phase variation* in the sense that functions x_1 and x_2 should not be compared at the same time t because they are not exhibiting the same behavior. Instead, in order to compare the two functions, the time scale itself has to be distorted or transformed.

We now look at several types of curve registration problems, beginning first with the problem of simply translating or shifting the values of t by a constant amount δ. Then we discuss landmark registration, which involves transforming t nonlinearly in order to line up important features or landmarks for all curves. Finally, we look at a more general method for curve registration.

7.2 Shift registration

Many of the issues involved in registration can be illustrated by considering the simplest case, a simple shift in the time scale. The pinch force data illustrated in Figure 1.11 are an example of a set of functional observations that must be aligned by moving each curve horizontally before any meaningful cross-curve analysis is possible. This often happens because the time at which the recording process begins is arbitrary, and is unrelated to the beginning of the interesting segment of the data, in this case the period over which the measured squeeze actually takes place.

Let the interval $\mathcal{T}$ over which the functions are to be registered be $[T_1, T_2]$. We also need to assume that each sample function x_i is available for some region beyond each end of $\mathcal{T}$. The pinch force data, for example, are observed for substantial periods both before and after the force pulse that we wish to study. In the case of periodic data such as the Canadian temperature records, this requirement is easily met since one can wrap the function around by using the function's behavior at the opposing end of the interval.

We are actually interested in the values

$$x_i^*(t) = x_i(t + \delta_i),$$

where the shift parameter δ_i is chosen in order to appropriately align the curves. For the pinch force data, the size of δ_i is of no real interest, since it merely measures the gap between the initialization of recording and the beginning of a squeeze. Silverman (1995) refers to this situation, in which

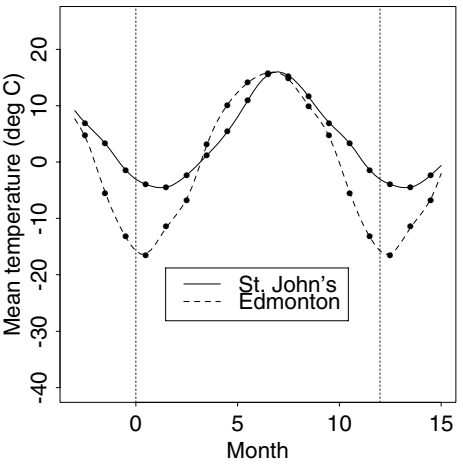

Figure 7.2. Temperature records for two weather stations where in the timing of the seasons differs by a roughly constant shift.

a shift parameter must be accounted for but is of no real interest, as a *nuisance effects* problem.

The Canadian temperature data present a curve alignment problem of a somewhat different nature. As Figure 7.2 indicates, two temperature records, such as those for St. John's, Newfoundland, and Edmonton, Alberta, can differ noticeably in terms of the phase or timing of key events, such as the lowest mean temperature and the timing of spring and autumn. In this case, the shifts that would align these two curves vertically are of intrinsic interest, and should be viewed as a component of variation that needs careful description. It turns out that continental stations such as Edmonton have earlier seasons than marine stations such as St. John's, because of the capacity of oceans to store heat and to release it slowly. In fact, either station's weather would have to be shifted by about three weeks to align the two.

When, as in the temperature data case, the shift is an important feature of each curve, we characterize its estimation as a *random effects* problem. Silverman (1995) also distinguishes a third and intermediate *fixed effects* case in which the shift must be carried out initially, and while not being discarded completely once the functions x_i^* have been constructed, is nevertheless only of tangential interest.

7.2.1 The least squares criterion for shift alignment

The basic mechanics of estimating the shifts δ_i are the same, whether they are considered as nuisance or random effects. The differences become important when we consider the analysis in subsequent chapters, because in the random effects case (and, to some extent, the fixed effects case) the δ_i enter the analysis. However, for present purposes we concentrate on the pinch force data as an example.

The estimation of a shift or an alignment requires a criterion that defines when several curves are properly registered. One possibility is to identify a specific feature or *landmark* for a curve, and shift each curve so that this feature occurs at a fixed point in time. The time of the maximum of the smoothed pinch force is an obvious landmark. Note that this might also be expressed as the time at which the first derivative crosses zero with negative slope, and landmarks are often more easily identifiable at the level of some derivative.

However, the registration by landmark or feature alignment has some potentially undesirable aspects: The location of the feature may be ambiguous for certain curves, and if the alignment is only of a single point, variations in other regions may be ignored. If, for example, we were to register the two temperature curves by aligning the midsummers, the midwinters might still remain seriously out of phase.

Instead, we can define a global registration criterion for identifying a shift δ_i for curve i as follows. First we estimate an overall mean function $\hat{\mu}(t)$ for t in $\mathcal{T}$. If the individual functional observations x_i are smooth, it usually suffices to estimate $\hat{\mu}$ by the sample average $\bar{x}$. However, we wish to be able to evaluate derivatives of $\hat{\mu}$, and so more generally we want to smooth the overall estimate using one of the methods described in Chapters 4 and 5. We can now define our global registration criterion by

$$\begin{aligned} \texttt{REGSSE} &= \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i(t+\delta_i) - \hat{\mu}(t)]^2 \, ds \\ &= \sum_{i=1}^{N} \int_{\mathcal{T}} [x_i^*(t) - \hat{\mu}(t)]^2 \, ds. \end{aligned} \tag{7.1}$$

Thus, our measure of curve alignment is the integrated or global sum of squared vertical discrepancies between the shifted curves and the sample mean curve.

The target function for transformation in (7.1) is the unregistered cross-sectional estimated mean $\hat{\mu}$. But of course one of the goals of registration is to produce a better estimate of this same mean function. We therefore expect to proceed iteratively: beginning with the unregistered cross-sectional estimated mean, argument values for each curve are shifted so as to minimize `REGSSE`, then the estimated mean $\hat{\mu}$ is updated by re-estimating it from the *registered* curves x_i^*, and a new iteration is then undertaken us-

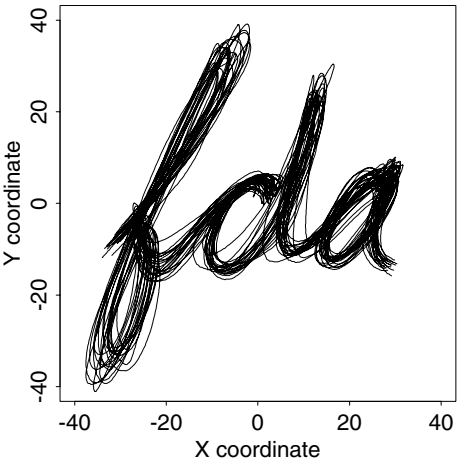

Figure 7.3. Twenty replications of "fda" written by one of the authors.

ing this revised target. This procedure of estimating a transformation by transforming to an iteratively updated average is often referred to as the *Procrustes method*. In practice, we have found that the process usually converges within one or two iterations.

7.3 Feature or landmark registration

A landmark or a feature of a curve is some characteristic that one can associate with a specific argument value t. These are typically maxima, minima, or zero crossings of curves, and may be identified at the level of some derivatives as well as at the level of the curves themselves.

We now turn to the more general problem of estimating a possibly nonlinear transformation h_i of t, and indicate how we can use landmarks to estimate this transformation. Coincidentally, the illustrative example we use shows how vector-valued functional data can be handled by obvious extensions of methods for scalar-valued functions.

The landmark registration process requires for each curve x_i the identification of the argument values $t_{if}, f = 1, \ldots, F$ associated with each of F features. The goal is to construct a transformation h_i for each curve such that the registered curves with values

$$x^*(t) = x_i[h_i(t)]$$

have more or less identical argument values for any given landmark.

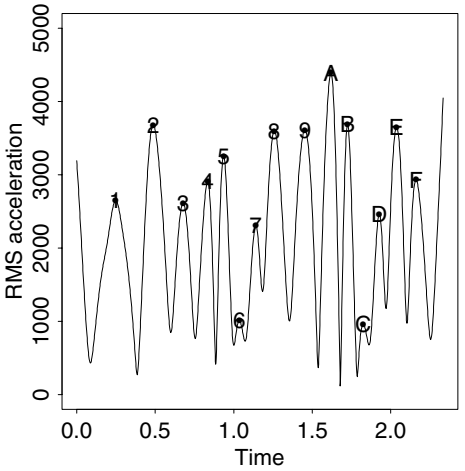

Figure 7.4. The average length of the acceleration vector for the 20 handwriting samples. The characters identify the 15 features used for landmark registration.

Consider, for example, the 20 replications of the letters "fda" in Figure 7.3. Each sample of handwriting was obtained by recording the position of a pen at a sampling rate of 600 times per second. There was some preprocessing to make each script begin and end at times 0 and 2.3 seconds, and to compute coordinates at the same 1,401 equally-spaced time-values. Each curve x_i in this situation is vector-valued, since two spatial coordinates are involved, and we use $\texttt{ScriptX}_i$ and $\texttt{ScriptY}_i$ to designate the X- and Y-coordinates, respectively.

Not surprisingly, there is some variation from observation to observation, and one goal is to explore the nature of this variation. But we want to take into account that, for example, variation in the "f" can be of two sorts. There is temporal variation due to the fact that timing of the top of the upper loop, for example, is variable. While this type of variation would not show up in the plots in Figure 7.3, it may still be an important aspect of how these curves vary. On the other hand, there is variation in the way the shape of each letter is formed, and this is obvious in the figure.

We estimated the accelerations or second derivatives of the two coordinate functions $D^2\texttt{ScriptX}_i$ and $D^2\texttt{ScriptY}_i$ by the local polynomial method described in Chapter 4. Figure 7.4 displays the average length of the acceleration vector

$$\sqrt{(D^2\texttt{ScriptX}_i)^2 + (D^2\texttt{ScriptY}_i)^2}$$

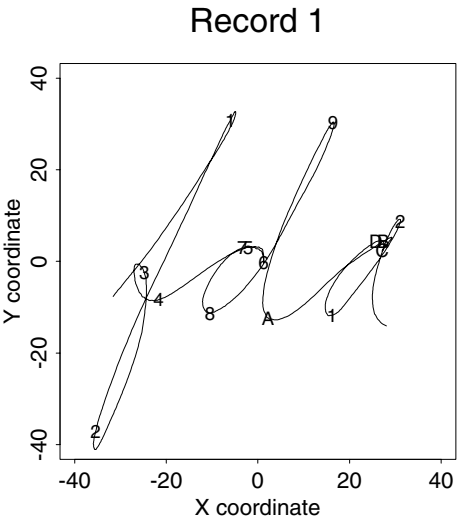

Figure 7.5. The first handwriting curve with the location of the 15 landmarks indicated by the characters used in Figure 7.4.

and we note that there are 15 clearly identified maxima, indicating points where the pen is changing direction. We also found that these maxima were easily identifiable in each record, and we were able to determine the values of t_{if} corresponding to them by just clicking on the appropriate points in a plot. Figure 7.5 shows the first curve with these 15 features labelled, and we can see that landmarks labelled "4" and "A" mark the boundaries between letters. Figure 7.6 plots the values of the landmark timings t_{if} against the corresponding timings for the mean function, t_{0f}. We were interested to see that the variability of the landmark timings was rather larger for the initial landmarks than for the later ones, and we were surprised by how small the variability was for all of them.

The identification of landmarks enabled us to compare the X- and Y-coordinate values for the 20 curves at the landmark times, but of course we also wanted to make comparisons at arbitrary points between landmarks. This required the computation of a function h_i for each curve, called a *time-warping function* in the engineering literature, with the properties

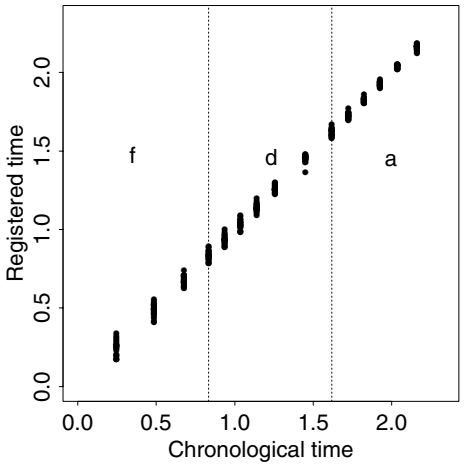

Figure 7.6. The timings of the landmarks for all 20 scripts plotted against the corresponding timings for the mean curve.

- $h_i(0) = 0$
- $h_i(2.3) = 2.3$
- $h_i(t_{0f}) = t_{if}, f = 1, \ldots, 15$
- h_i is strictly monotonic: $s < t$ implies that $h_i(t) < h_i(t)$.

The values of the adjusted curves at time t are $\texttt{ScriptX}[h_i(t)]$ and $\texttt{ScriptY}[h_i(t)]$. In all the adjusted curves, the landmarks each occur at the same time as in the mean function. In addition, the adjusted curves are also more or less aligned between landmarks. In this application, we merely used linear interpolation for time values between the points (t_{0f}, t_{if}) (as well as (0,0) and (2.3,2.3)) to define the time warping function h_i for each curve. We introduce more sophisticated notions in the next section. Figure 7.7 shows the warping function computed in this manner for the first script record. Because h_1 is below the diagonal line in the region of "f," the aligned time $h_1(t)$ is earlier than the actual time of features, and hence the actual times for curve 1 are retarded with respect to the mean curve.

We can now re-compute the mean curve by averaging the registered curves. The result is in Figure 7.8, shown along with the mean for the unregistered data. Although the differences are not dramatic, as we might expect given the mild curvature in h_1, we do see that the upper and lower loops of the "f" are now more pronounced, and in fact do represent the original curves substantially better.

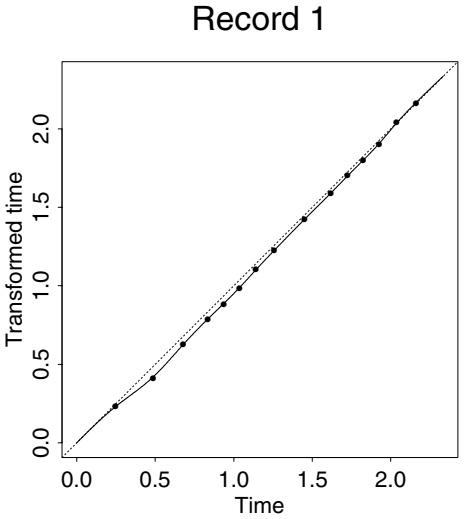

Figure 7.7. The time warping function h_1 estimated for the first record that registers its features with respect to the mean curve.

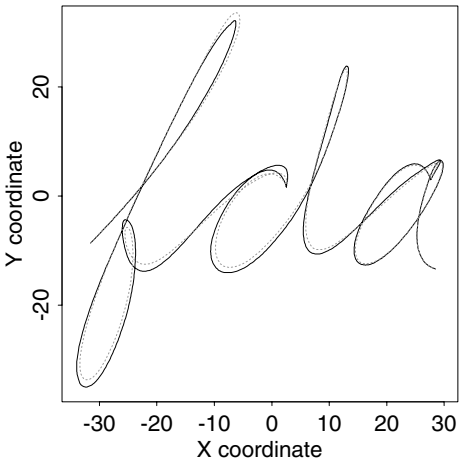

Figure 7.8. The solid line is the mean of the registered "fda" curves, and the dashed line is the mean of the unregistered curves.

7.4 Using the warping function h to register x

Now that a warping function h has been estimated from landmark registration, or by using the continuous method described in a later section, you will want to calculate the registered function values $x^*(t) = x[h(t)]$. This requires two steps.

First, estimate the *inverse warping function* $h^{-1}(t)$ with the property $h^{-1}[h(t)] = t$. Note that this is not an inverse in the sense of the reciprocal. Instead, $h^{-1}(t)$ is computed by smoothing or interpolating the relationship between $h(t)$ plotted on the horizontal axis and t plotted on the vertical axis. You can then use simple interpolation to get the values of this inverse function at an equally spaced set of values of t if required. Note that it will be essential that this smoothing or interpolation function be strictly monotonic, so you may have to use lots of values of t and/or employ monotone smoothing described in Chapter 6.

The second step is to smooth or interpolate the relationship between $h^{-1}(t)$ plotted on the abscissa and $x(t)$ plotted on the ordinate. You can then use simple interpolation to get the values of this registered function at an equally spaced set of values of t if required.

7.5 A more general warping function h

The linear interpolation scheme that we used on the handwriting data to estimate the time-warping function h has two limitations. First, if we want to compute higher order derivatives of the curves with respect to warped time, the warping function must also be differentiable to the same order, a linear interpolation would not carry us beyond the first derivative. Secondly, we will shortly consider *continuous registration* methods that do not use landmarks and where the idea of interpolating a sequence of points will not be helpful.

Time is itself a growth process, and thus can be linked to our discussion in Chapter 6 on how to model the children's growth curves. That is, we can use the formulation

$$h(t) = C_0 + C_1 \int_0^t \exp W(u)\, du \tag{7.2}$$

that we used in (6.9). Here the constants C_0 and C_1 are fixed by the requirement that $h(t) = t$ at the lower and upper limits of the interval over which we model the data. Or, if shift registration is a possibility, the constant term C_0 can be allowed to pick any constant phase shift that is required.

Physical or clock time grows linearly, of course, and thus corresponds to $W(u) = 0$. If $W(u)$ is *positive*, then $h(t) > t$, warped time is growing faster than clock time, and this is what we want if our observed process

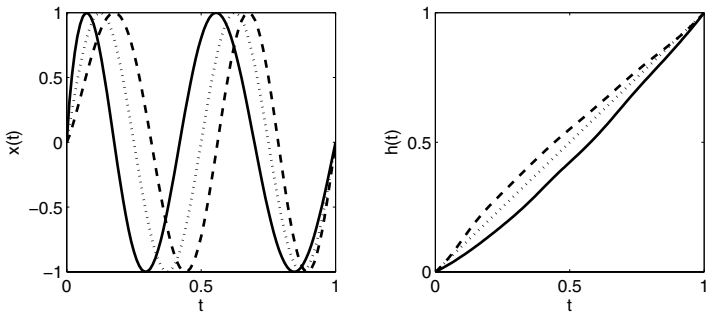

Figure 7.9. The left panel shows the target function, $x_0(t) = \sin(4\pi t)$, as a dotted line; an early function, $x_E(t) = \sin(4\pi t^{0.8})$, as a solid line; and a late function, $x_L(t) = \sin(4\pi t^{1.2})$, as a dashed line. The corresponding warping functions that register the early and late curves to the target are shown in the right panel.

is running *late*. Similarly, for negative values of $W(u)$, $h(t) < t$, and clock time is being slowed down for a process that is running ahead of some target.

The left panel of Figure 7.9 displays two examples. Here the target or standard function is $x_0(t) = \sin(4\pi t)$, the early function is $x_E(t) = \sin(4\pi t^{0.8})$ and the late function is $x_L(t) = \sin(4\pi t^{1.2})$. Warping $h_E(t) = t^{0.125}$ will register the first example since $\sin[4\pi(t^{0.8})^{1.25}] = \sin(4\pi t)$, and similarly $h_L(t) = t^{0.833}$. Approximations to the two warping functions by a method to be described below are presented in the right panel, and we can see there how early functions are associated with time-decelerating warpings, and late functions with time-accelerating warpings.

The use of (7.2) as a representation of a warping function has a very handy bonus. Providing that the warp h is reasonably smooth and mild, the inverse warp h^{-1} is achieved to a close approximation by merely replacing W in the equation by $-W$.

7.6 A continuous fitting criterion for registration

The least squares criterion (7.1) worked well for simple shift registration, but gets us into trouble for more general warping functions. The lower panel in Figure 7.10 shows why. When two functions differ in terms of amplitude as well as phase, the least squares criterion uses time warping to also minimize amplitude differences by trying to squeeze out of existence regions where amplitudes differ. Put another way, the least squares fitting criterion is intrinsically designed to assess differences in amplitude rather

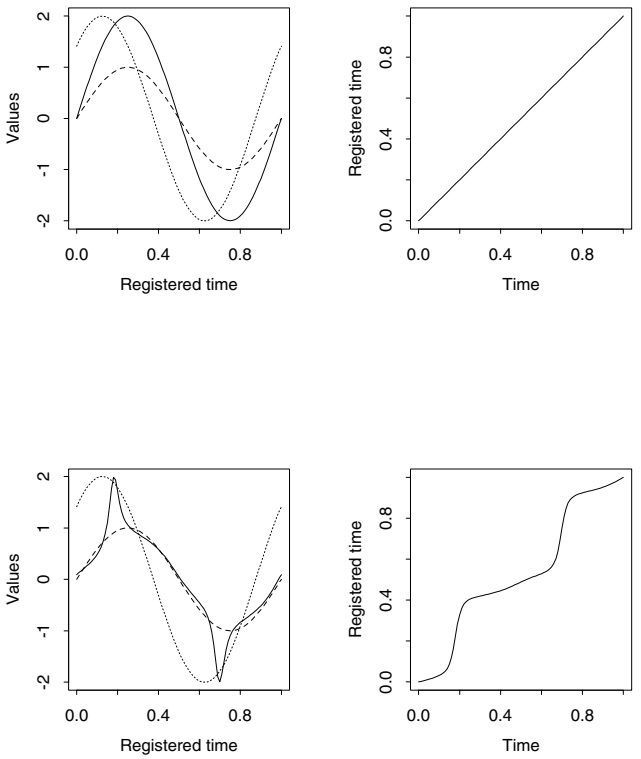

Figure 7.10. The upper two panels show results for an artificial registration problem using the minimum eigenvalue criterion. The dotted curve in the upper-left panel is the curve to be registered to the curve indicated by the dashed line. The solid line is the registered curve. The upper-right panel contains the warping function for this case, $h(t) = t$. The lower panels show the same results using the least squares criterion.

than phase. This wasn't a problem when only time shifts were involved since such simple time warps cannot affect amplitude differences.

Suppose two curves x_0 and x_1 differ only in amplitude but not in phase, such as in the left panel of Figure 7.10. Then, if we plot the function values $x_0(t)$ and $x_1(t)$ against each other, we will see a straight line. Amplitude differences will then be reflected in the slope of the line, a line at 45^o corresponding to no amplitude differences.

Now thinking about a line as a one-dimensional set of points on a plane, we can turn to principal components analysis as just the right technique for assessing how many dimensions are required to represent the distribution of these points. This technique will yield only one positive eigenvalue if the

point spread is, in fact, one-dimensional. That is, the size of the smallest eigenvalue measures departures from unidimensionality.

Let us consider now evaluating both the target function x_0 and the registered function x^* at a fine mesh of n values of t to obtain the pairs of values $(x_0(t), x[h(t)])$. Let the n by two matrix $\mathbf{X}$ contain these pairs of values. Then the two-by-two cross-product matrix $\mathbf{X}'\mathbf{X}$ would be what we would analyze by principal components.

The following order two matrix is the functional analogue of the cross-product matrix $\mathbf{X}'\mathbf{X}$.

$$\mathbf{T}(h) = \begin{bmatrix} \int \{x_0(t)\}^2\, dt & \int x_0(t)x[h(t)]\, dt \\ \int x_0(t)x[h(t)]\, dt & \int \{x[h(t)]\}^2\, dt \end{bmatrix} \tag{7.3}$$

We see that the summations over points implied by the expression $\mathbf{X}'\mathbf{X}$ have here been replaced by integrals. Otherwise this is the same matrix. We have expressed the matrix as a function of warping function h to remind ourselves that it does depend on h.

Consequently, we can now express our fitting criterion for assessing the degree to which two functions are registered as follows:

$$\texttt{MINEIG}(h) = \mu_2[\mathbf{T}(h)], \tag{7.4}$$

where the function μ_2 is the size of the second eigenvalue of its argument, which is an order two symmetric matrix. When $\texttt{MINEIG}(h) = 0$, we have achieved registration, and h is the warping function that does the job.

As is now routine, we will want to apply some regularization now and then to impose smoothness on h, so we extend our criterion to

$$\texttt{MINEIG}_\lambda(h) = \texttt{MINEIG}(h) + \lambda \int \{W^{(m)}(t)\}^2\, dt. \tag{7.5}$$

Here we are assuming that h is of the form (7.2), and that we achieve smoothness in h by smoothing the function W that defines it.

The results in Figure 7.9 were achieved by expanding W in terms of 13 B-splines with equally spaced knots, and penalizing the size of its second derivative using a smoothing parameter of $\lambda = 10^6$.

7.7 Registering the height acceleration curves

The 10 acceleration functions in Figure 1.2 were registered by the Procrustes method and the regularized basis expansion method set out in Section 7.6. The interval $\mathcal{T}$ was taken to be $[4, 18]$ with time measured in years. The break-values τ_k defining the monotone transformation family (7.2) were 4, 7, 10, 12, 14, 16 and 18 years, and the curves were registered over the interval [4,18] using criterion (7.5) with $\lambda = 0.001$. A single Procrustes iteration produced the results displayed in Figure 7.11. The left panel displays the 10 warping functions h_i, and the right panel shows

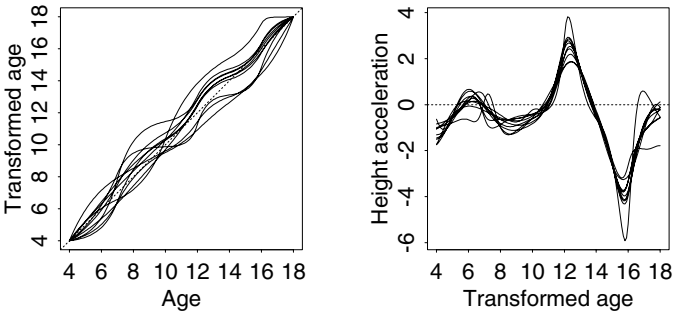

Figure 7.11. The left panel contains the estimated time warping functions h_i for the 10 height acceleration curves in Figure 1.2. The right panel displays the registered curves.

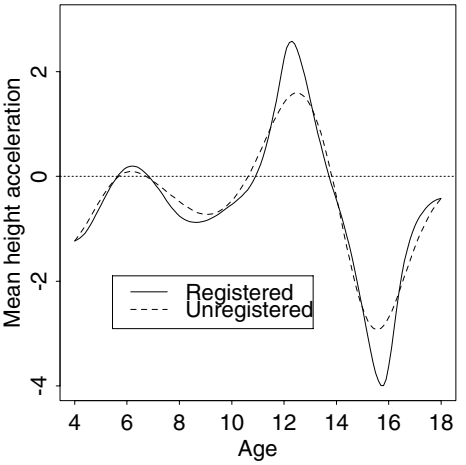

Figure 7.12. The cross-sectional means of the registered and unregistered height acceleration curves displayed in Figure 1.2.

the curve values $x_i[h_i(t)]$. Figure 7.12 compares the unregistered and registered cross-sectional means. We see that the differences are substantial, and moreover that the mean of the registered function tends to resemble much more closely most of the sample curves.

7.8 Some practical advice

Before registration, remove amplitude effects that can be accounted for by vertical shifts or scale changes, by centering and possibly rescaling the curves. This is standard advice in data analysis; deal with obvious effects in a simple way before moving on to more sophisticated procedures.

In general it is not clear that variation in the amplitude of curves can be cleanly separated from the variation that the registration process aims to account for. It is easy to construct examples where a registration function h that is allowed to be highly nonlinear can remove variation that is clearly of an amplitude nature, and the lower panels of Figure 7.10.

This problem of lack of identifiability of the two types of variation, horizontal and vertical, is perhaps less of a concern if only linear transformations are permitted, and is also not acute for landmark registration, where the role of the transformation is to only align curve features.

However, there is one situation that implies relatively unambiguous separation of the two types of variation. This happens with curves that cross zero at a number of points. At and near these zero crossings, only phase variation is possible. In effect, zero crossings are landmarks that should be aligned. Consequently, it may be wise to consider registering a derivative of a curve rather than the curve itself, since derivatives often cross zero. This is why we registered the acceleration curves above rather than the height or velocity curves.

If flexible families of monotone transformations such as those described above are used in conjunction with a global fitting criterion such as `MINEIG`, allow transformations to differ from linear only with caution by careful application of regularization.

In general, we have found it wise to first register on any landmarks that are clearly identifiable before using the continuous registration procedure. For example, in our work with the growth data we first register the curves using the zero-crossing in the middle of the pubertal growth spurt as a single landmark. Then we use the curves resulting from this preliminary registration as inputs to a continuous registration. If we use the notation h_L and $h_{C|L}$ to refer to the landmark warps and the continuous warps after landmark registration, respectively, then the final composite warping function is $h(t) = h_{C|L}[h_L(t)]$ or $h = h_{C|L} \circ h_L$.

7.9 Computational details

7.9.1 Shift registration by the Newton-Raphson algorithm

We can estimate a specific shift parameter δ_i iteratively by using a modified Newton-Raphson algorithm for minimizing `REGSSE`. This procedure requires derivatives of `REGSSE` with respect to the δ_i. If we assume that the

differences between x_i^* and $\hat{\mu}$ at the ends of the interval can be ignored (this is exactly true in the periodic case, and often approximately true in the non-periodic case if the effects of real interest are concentrated in the middle of the interval), then we have

$$\begin{aligned}\frac{\partial}{\partial\delta_i}\texttt{REGSSE} &= 2\int_{\mathcal{T}}\{x_i(t+\delta_i)-\hat{\mu}(t)\}Dx_i(t)\,dt \\ \frac{\partial^2}{\partial\delta_i^2}\texttt{REGSSE} &= 2\int_{\mathcal{T}}\{x_i(t+\delta_i)-\hat{\mu}(t)\}D^2x_i(t)\,dt \\ &+ 2\int_{\mathcal{T}}\{Dx_i(t)\}^2\,dt. \qquad (7.6)\end{aligned}$$

The modified Newton-Raphson algorithm works as follows:

Step 0: Begin with some initial shift estimates $\delta_i^{(0)}$, perhaps by aligning with respect to some feature, or even $\delta_i^{(0)} = 0$. But the better the initial estimate, the faster and more reliably the algorithm converges. Complete this step by estimating the average $\hat{\mu}$ of the shifted curves, using a method that allows the first two derivatives of $\hat{\mu}$ to give good estimates of the corresponding derivatives of the population mean, such as local polynomial regression of degree 4, or roughness penalty smoothing with an integrated squared fourth derivative penalty.

Step ν, for $\nu = 1, 2, \ldots$: Modify the estimate $\delta_i^{(\nu-1)}$ on the previous iteration by

$$\delta_i^{(\nu)} = \delta_i^{(\nu-1)} - \alpha\frac{(\partial/\partial\delta_i)\texttt{REGSSE}}{(\partial^2/\partial\delta_i^2)\texttt{REGSSE}},$$

where α is a step-size parameter that can sometimes simply be set to one. It is usual to drop the first term (7.6) in the second derivative of REGSSE since it vanishes at the minimizing values, and convergence without this term tends to be more reliable when current estimates are substantially far from the minimizing values. Once the new shifts are estimated, recompute the estimated average $\hat{\mu}$ of the shifted curves.

Although the algorithm can in principle be iterated to convergence, and although convergence is generally fast, we have found that a single iteration is often sufficient with reasonable initial estimates. For the pinch force data, we began by aligning the smoothed curves by setting the location of the maximum of each curve at 0.1 seconds. The shifts involved ranged from -20 to 50 milliseconds. We then carried out a single Newton-Raphson update ($\nu = 1$ above) where the range $\mathcal{T}$ of integration was from 23 to 251 milliseconds. The changes in the δ_i ranged from -3 to 2 milliseconds, and after this update, a second iteration did not yield any changes larger than a millisecond. The aligned curves are shown in Figure 7.13.

As part of a technique that they call *self-modelling nonlinear regression*, which attempts to estimate both parametric and nonparametric components of variation among several curves, Kneip and Gasser (1988) use linear

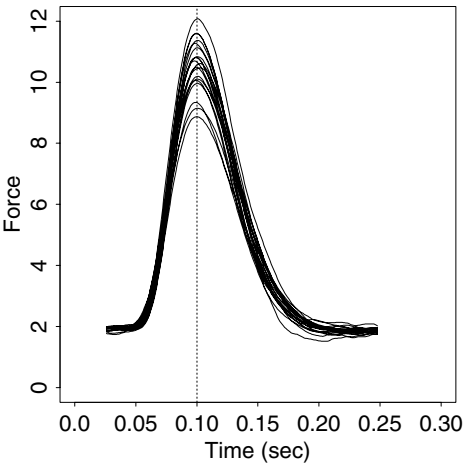

Figure 7.13. The pinch force curves aligned by minimizing the Procrustes criterion `REGSSE`.

transformations of t, that is both shift and scale changes. Kneip and Engel (1995) use such shift-scale transformations to identify "shape invariant features" of curves, which remain unaltered by these changes in t.

7.10 Further reading and notes

The classic paper on the estimation of time warping functions is Sakoe and Chiba (1978), who used dynamic programming to estimate the warping function in a context where there was no need for the warping function to be smooth.

Landmark registration has been studied in depth by Kneip and Gasser (1992) and Gasser and Kneip (1995), who refer to a landmark as a *structural feature*, its location as a *structural point*, to the distribution of landmark locations along the t axis as *structural intensity*, and to the process of averaging a set of curves after registration as *structural averaging*. Their papers contain various technical details on the asymptotic behavior of landmark estimates and warping functions estimated from them. Their papers on growth curves (Gasser et al., 1990, 1991a,b) are applications of this process. Another source of much information on the study of landmarks and their use in registration is Bookstein (1991).

Ramsay (1996b) and Ramsay and Li (1996) developed the fitting of a general and flexible family of warping functions h_i making use of a regular-

ization technique. Their work used a piecewise linear basis for function W in order to avoid numerical integration, but our subsequent work has found numerical integration to be easy to apply here as well as elsewhere, and consequently W may now be expanded in terms of any basis. Kneip, Li, MacGibbon and Ramsay (2000) developed a method that is rather analogous to local polynomial smoothing for identifying warping functions that register a sample of curves.

Wang and Gasser (1997, 1998, 1999) and Gervini and Gasser (2004) have evolved registration technology that does not use landmarks in a number of useful ways, and consider some important theoretical issues. Liu and Müller (2004) advanced their theoretical framework by discussing curve registration in the context of a model for random or stochastic functions where time is itself transformed in a random manner. They propose the operation of taking a *functional convex sum* as a way of computing convex sums of unregistered functions. This operation defines a type of mean that preserves the locations and shapes of features. See also Rønn (2001) for a model-based approach to shift registration.

The functional two-sample functional testing problem considered by Munoz, Maldonado, Staniswalis, Irwin and Byers (2002) uses landmark registration of some image density curves as a pre-processing step.