8

Principal components analysis for functional data

8.1 Introduction

For many reasons, principal components analysis (PCA) of functional data is a key technique to consider. First, our own experience is that, after the preliminary steps of registering and displaying the data, the user wants to explore that data to see the features characterizing typical functions. Some of these features are expected to be there, for example the sinusoidal nature of temperature curves, but other aspects may be surprising. Some indication of the complexity of the data is also required, in the sense of how many types of curves and characteristics are to be found. Principal components analysis serves these ends admirably, and it is perhaps also for these reasons that it was the first method to be considered in the early literature on FDA.

Just as for the corresponding matrices in the classical multivariate case, the variance-covariance and correlation functions can be difficult to interpret, and do not always give a fully comprehensible presentation of the structure of the variability in the observed data directly. The same is true, of course, for variance-covariance and correlation matrices in classical multivariate analysis. A principal components analysis provides a way of looking at covariance structure that can be much more informative and can complement, or even replace altogether, a direct examination of the variance-covariance function.

PCA also offers an opportunity to consider some issues that reappear in subsequent chapters. For example, we consider immediately how PCA is

defined by the notion of a linear combination of function values, and why this notion, at the heart of most of multivariate data analysis, requires some care in a functional context. A second issue is that of *regularization*; for many data sets, PCA of functional data is more revealing if some type of smoothness is required of the principal components themselves. We consider this topic in detail in Chapter 9.

8.2 Defining functional PCA

8.2.1 PCA for multivariate data

The central concept exploited over and over again in multivariate statistics is that of taking a linear combination of variable values,

$$f_i = \sum_{j=1}^{p} \beta_j x_{ij},\ i = 1, \ldots, N, \tag{8.1}$$

where β_j is a weighting coefficient applied to the observed values x_{ij} of the jth variable. We can express (8.1) as

$$f_i = \beta' x_i,\ i = 1, \ldots, N, \tag{8.2}$$

where β is the vector $(\beta_1, \ldots, \beta_p)'$ and x_i is the vector $(x_{i1}, \ldots, x_{ip})'$.

In the multivariate situation, we choose the weights so as to highlight or display types of variation that are very strongly represented in the data. Principal components analysis can be defined in terms of the following stepwise procedure, which defines sets of normalized weights that maximize variation in the f_i's:

1. Find the weight vector $\boldsymbol{\xi}_1 = (\xi_{11}, \ldots, \xi_{p1})'$ for which the linear combination values
$$f_{i1} = \sum_j \xi_{j1} x_{ij} = \boldsymbol{\xi}_1' x_i$$
have the largest possible mean square $N^{-1} \sum_i f_{i1}^2$ subject to the constraint
$$\sum_j \xi_{j1}^2 = \|\boldsymbol{\xi}_1\|^2 = 1.$$

2. Carry out second and subsequent steps, possibly up to a limit of the number of variables p. On the mth step, compute a new weight vector $\boldsymbol{\xi}_m$ with components ξ_{jm} and new values $f_{im} = \boldsymbol{\xi}_m' x_i$. Thus, the values f_{im} have maximum mean square, subject to the constraint $\|\boldsymbol{\xi}_m\|^2 = 1$ and the $m - 1$ additional constraint(s)
$$\sum_j \xi_{jk} \xi_{jm} = \boldsymbol{\xi}_k' \boldsymbol{\xi}_m = 0,\ k < m.$$

The motivation for the first step is that by maximizing the mean square, we are identifying the strongest and most important mode of variation in the variables. The unit sum of squares constraint on the weights is essential to make the problem well defined; without it, the mean squares of the linear combination values could be made arbitrarily large. On second and subsequent steps, we seek the most important modes of variation again, but require the weights defining them to be orthogonal to those identified previously, so that they are indicating something new. Of course, the amount of variation measured in terms of $N^{-1}\sum_i f_{im}^2$ will decline on each step. At some point, usually well short of the maximum index p, we expect to lose interest in modes of variation thus defined.

The definition of principal components analysis does not actually specify the weights uniquely; for example, it is always possible to change the signs of all the values in any vector $\boldsymbol{\xi}_m$ without changing the value of the variance that it defines.

The values of the linear combinations f_{im} are called *principal component scores* and are often of great help in describing what these important components of variation mean in terms of the characteristics of specific cases or replicates.

To be sure, the mean is a very important aspect of the data, but we already have an easy technique for identifying it. Therefore, we usually subtract the mean for each variable from corresponding variable values before doing PCA. When this is done, maximizing the mean square of the principal component scores corresponds to maximizing their sample variance.

8.2.2 Defining PCA for functional data

How does PCA work in the functional context? The counterparts of variable values are function values $x_i(s)$, so that the discrete index j in the multivariate context has been replaced by the continuous index s. When we were considering vectors, the appropriate way of combining a weight vector β with a data vector x was to calculate the inner product

$$\beta' x = \sum_j \beta_j x_j.$$

When β and x are functions $\beta(s)$ and $x(s)$, summations over j are replaced by integrations over s to define the inner product

$$\int \beta x = \int \beta(s)x(s)\,ds. \tag{8.3}$$

Within the principal components analysis, the weights β_j now become functions with values $\beta_j(s)$. Using the notation (8.3), the principal

component scores corresponding to weight β are now

$$f_i = \int \beta x_i = \int \beta(s) x_i(s)\, ds. \tag{8.4}$$

For the rest of our discussion, we will often use the short form $\int \beta x_i$ for integrals in order to minimize notational clutter.

In the first functional PCA step, the weight function $\xi_1(s)$ is chosen to maximize $N^{-1}\sum_i f_{i1}^2 = N^{-1}\sum_i (\int \xi_1 x_i)^2$ subject to the continuous analogue $\int \xi_1(s)^2 ds = 1$ of the unit sum of squares constraint. This time, the notation $\|\xi_1\|^2$ is used to mean the squared norm $\int \xi_1(s)^2 ds = \int \xi_1^2$ of the *function* ξ_1.

Postponing computational details until Section 8.4, now consider as an illustration in the upper left panel in Figure 8.1. This displays the weight function ξ_1 for the Canadian temperature data after the mean across all 35 weather stations has been removed from each station's monthly temperature record. Although ξ_1 is positive throughout the year, the weight placed on the winter temperatures is about four times that placed on summer temperatures. This means that the greatest variability between weather stations will be found by heavily weighting winter temperatures, with only a light contribution from the summer months; Canadian weather is most variable in the wintertime, in short. Moreover, the percentage 89.3% at the top of the panel indicates that this type of variation strongly dominates all other types of variation. Weather stations for which the score f_{i1} is high will have much warmer than average winters combined with warm summers, and the two highest scores are in fact assigned to Vancouver and Victoria on the Pacific Coast. To no one's surprise, the largest negative score goes to Resolute in the High Arctic.

As for multivariate PCA, the weight function ξ_m is also required to satisfy the orthogonality constraint(s) $\int \xi_k \xi_m = 0,\ k < m$ on subsequent steps. Each weight function has the task of defining the most important mode of variation in the curves subject to each mode being orthogonal to all modes defined on previous steps. Note again that the weight functions are defined only to within a sign change.

The weight function ξ_2 for the temperature data is displayed in the upper right panel of Figure 8.1. Because it must be orthogonal to ξ_1, we cannot expect that it will define a mode of variation in the temperature functions that will be as important as the first. In fact, this second mode accounts for only 8.3% of the total variation, and consists of a positive contribution for the winter months and a negative contribution for the summer months, therefore corresponding to a measure of uniformity of temperature through the year. On this component, one of the highest scores f_{i2} goes to Prince Rupert, also on the Pacific coast, for which there is comparatively low discrepancy between winter and summer. Prairie stations such as Winnipeg, on the other hand, have hot summers and very cold winters, and receive large negative second component scores.

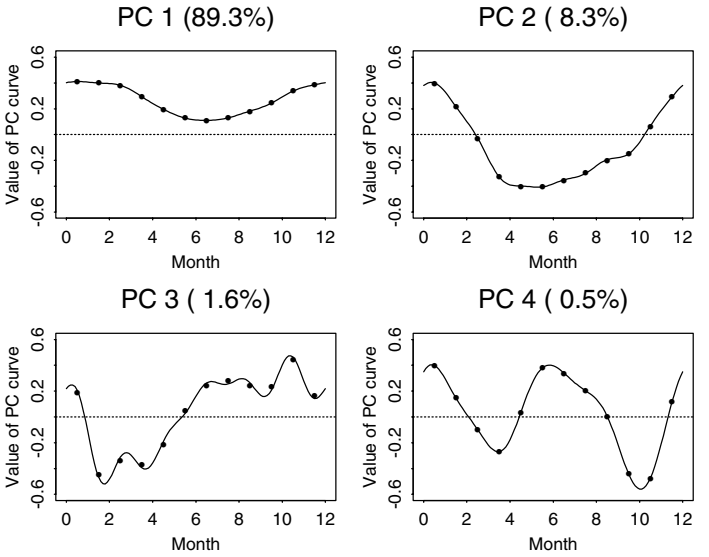

Figure 8.1. The first four principal component curves of the Canadian temperature data estimated by two techniques. The points are the estimates from the discretization approach, and the curves are the estimates from the expansion of the data in terms of a 12-term Fourier series. The percentages indicate the amount of total variation accounted for by each principal component.

The third and fourth components account for small proportions of the variation, since they are required to be orthogonal to the first two as well as to each other. At this point they are difficult to interpret, but we look at techniques for understanding them in Section 8.3.

Displays such as Figure 8.1 can remind one of the diagrams of modes of vibration in a string fixed at both ends always found in introductory physics texts. The first and dominant type is simple in structure and resembles a single cycle of a sine wave. Subdominant or higher order components are also roughly sinusoidal, but with more and more cycles. With this analogy in mind, we find the term *harmonics* evocative in referring to principal components of variation in curves in general.

8.2.3 Defining an optimal empirical orthonormal basis

There are several other ways to motivate PCA, and one is to define the following problem: We want to find a set of exactly K orthonormal functions ξ_m so that the expansion of each curve in terms of these basis functions approximates the curve as closely as possible. Since these basis functions

are orthonormal, it follows that the expansion will be of the form

$$\hat{x}_i(t) = \sum_{k=1}^{K} f_{ik}\xi_k(t),$$

where f_{ik} is the principal component value $\int x_i\xi_k$. As a fitting criterion for an individual curve, consider the integrated squared error

$$\|x_i - \hat{x}_i\|^2 = \int [x(s) - \hat{x}(s)]^2\, ds$$

and as a global measure of approximation,

$$\texttt{PCASSE} = \sum_{i=1}^{N} \|x_i - \hat{x}_i\|^2. \tag{8.5}$$

The problem is then, more precisely, what choice of basis will minimize the error criterion (8.5)?

The answer, it turns out, is precisely the same set of principal component weight functions that maximize variance components as defined above. For this reason, these functions ξ_m are referred to in some fields as *empirical orthonormal functions*, because they are determined by the data they are used to expand.

8.2.4 PCA and eigenanalysis

In this section, we investigate another characterization of PCA, in terms of the eigenanalysis of the variance-covariance function or operator.

Assume for this section that our observed values, x_{ij} in the multivariate context and $x_i(t)$ in the functional situation, result from subtracting the mean variable or function values, so that their sample means $N^{-1}\sum_i x_{ij}$, or cross-sectional means $N^{-1}\sum_i x_i(t)$, respectively, are zero.

Texts on multivariate data analysis tend to define principal components analysis as the task of finding the eigenvalues and eigenvectors of the covariance or correlation matrix. The logic for this is as follows. Let the $N \times p$ matrix $\mathbf{X}$ contain the values x_{ij} and the vector $\boldsymbol{\xi}$ of length p contain the weights for a linear combination. Then the mean square criterion for finding the first principal component weight vector can be written as

$$\max_{\boldsymbol{\xi}'\boldsymbol{\xi}=1} N^{-1}\boldsymbol{\xi}'\mathbf{X}'\mathbf{X}\boldsymbol{\xi} \tag{8.6}$$

since the vector of principal component scores f_i can be written as $\mathbf{X}\boldsymbol{\xi}$.

Use the $p \times p$ matrix $\mathbf{V}$ to indicate the sample variance-covariance matrix $\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$. (One may prefer to use a divisor of $N-1$ to N since the means have been estimated, but it makes no essential difference to the principal components analysis.) The criterion (8.6) can now be expressed

as

$$\max_{\boldsymbol{\xi}'\boldsymbol{\xi}=1} \boldsymbol{\xi}'\mathbf{V}\boldsymbol{\xi}.$$

As explained in Section A.5, this maximization problem is now solved by finding the solution with largest eigenvalue ρ of the eigenvector problem or *eigenequation*

$$\mathbf{V}\boldsymbol{\xi} = \rho\boldsymbol{\xi}. \tag{8.7}$$

There is a sequence of different eigenvalue-eigenvector pairs $(\rho_j, \boldsymbol{\xi}_j)$ satisfying this equation, and the eigenvectors $\boldsymbol{\xi}_j$ are orthogonal. Because the mean of each column of $\mathbf{X}$ is usually subtracted from all values in that column as a preliminary to principal components analysis, the rank of $\mathbf{X}$ is $N-1$ at most, and hence the $p \times p$ matrix $\mathbf{V}$ has, at most, $\min\{p, N-1\}$ nonzero eigenvalues ρ_j. For each j, the eigenvector $\boldsymbol{\xi}_j$ satisfies the maximization problem (8.6) subject to the additional constraint of being orthogonal to all the eigenvectors $\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, \ldots, \boldsymbol{\xi}_{j-1}$ found so far. This is precisely what was required of the principal components in the second step laid out in Section 8.2.1. Therefore, as we have defined it, the multivariate PCA problem is equivalent to the algebraic and numerical problem of solving the eigenequation (8.7). Of course, there are standard computer algorithms for doing this.

Now consider the functional version of PCA. Define the covariance function $v(s,t)$ by

$$v(s,t) = N^{-1}\sum_{i=1}^{N} x_i(s)x_i(t). \tag{8.8}$$

Again, note that we may prefer to use $N-1$ to define the variance-covariance function v; nothing discussed here changes in any essential way.

The more general results set out in Section A.5.2 can be applied, to find the principal component weight functions $\xi_j(s)$. Each of these satisfies the equation

$$\int v(s,t)\xi(t)\,dt = \rho\xi(s) \tag{8.9}$$

for an appropriate eigenvalue ρ. The left side of (8.9) is an *integral transform* V of the weight function ξ defined by

$$V\xi = \int v(\cdot,t)\xi(t)\,dt. \tag{8.10}$$

This integral transform is called the *covariance operator* V. Therefore we may also express the eigenequation directly as

$$V\xi = \rho\xi, \tag{8.11}$$

where ξ is now an eigenfunction rather than an eigenvector. By suitable choice of notation, the equation (8.11) for functional PCA now looks the same as the eigenequation (8.7) relevant to conventional PCA.

There is an important difference between the multivariate and functional eigenanalysis problems, concerning the maximum number of different eigenvalue-eigenfunction pairs. The counterpart of the number of variables p in the multivariate case is the number of function values in the functional case, and thus infinity. However, provided the functions x_i are not linearly dependent, the operator V will have rank $N-1$, and there will be only $N-1$ nonzero eigenvalues.

To summarize, in this section we find that principal components analysis is defined as the search for a set of mutually orthogonal and normalized weight functions ξ_m. Functional PCA can be expressed as the problem of the eigenanalysis of the covariance operator V. By suitable choice of notation, the formal steps to be carried out are the same, whether the data are multivariate or functional.

In Section 8.4 we discuss practical methods for actually computing the eigenfunctions ξ_m, but first we consider some aspects of the display of principal components once they have been found.

8.3 Visualizing the results

The fact that interpreting the components is not always an entirely straightforward matter is common to most functional PCA problems. We now consider some techniques that may aid their interpretation.

8.3.1 Plotting components as perturbations of the mean

A method found to be helpful is to examine plots of the overall mean function and the functions obtained by adding and subtracting a suitable multiple of the principal component function in question. Figure 8.2 shows such a plot for the temperature data. In each case, the solid curve is the overall mean temperature, and the dotted and dashed curves show the effects of adding and subtracting a multiple of each principal component curve. This considerably clarifies the effects of the first two components. We can now see that the third principal component corresponds to a time shift effect combined with an overall increase in temperature and in range between winter and summer. The fourth corresponds to an effect whereby the onset of spring is later and autumn ends earlier.

In constructing this plot, it is necessary to choose which multiple of the principal component function to use. Define a constant C to be the root-mean-square difference between $\hat{\mu}$ and its overall time average,

$$C^2 = T^{-1}\|\hat{\mu} - \bar{\mu}\|^2, \tag{8.12}$$

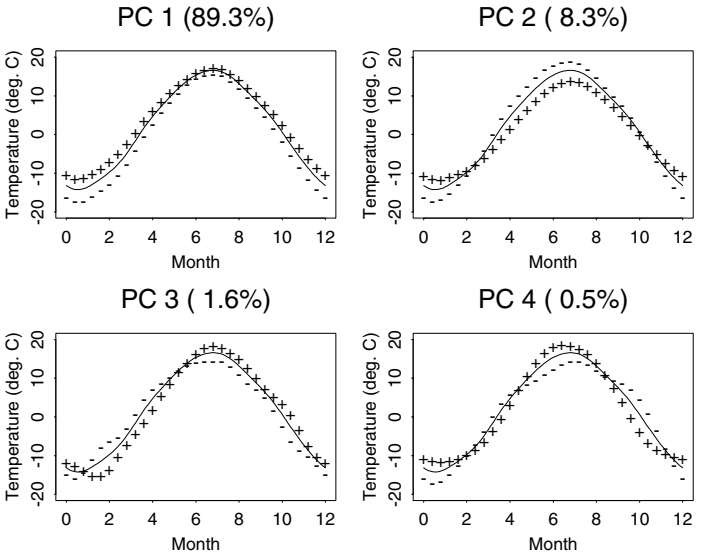

Figure 8.2. The mean temperature curves and the effects of adding (+) and subtracting (−) a suitable multiple of each PC curve.

where

$$\bar{\mu} = T^{-1} \int \hat{\mu}(t)\, dt.$$

It is then appropriate to plot $\hat{\mu}$ and $\hat{\mu} \pm 0.2C\hat{\gamma}_j$, where we have chosen the constant 0.2 to give easily interpretable results. Depending on the overall behavior of $\hat{\mu}$, it may be helpful to adjust the value 0.2 subjectively. But for ease of comparison between the various modes of variability, it is best to use the same constant for all the principal component functions plotted in any particular case.

In Figure 8.3, we consider the hip angles observed during the gait of 39 children, as plotted in Figure 1.8. The angles for a single cycle are shown, along with the results of a functional PCA of these data. The effect of the first principal component of variation is approximately to add or subtract a constant to the angle throughout the gait cycle. The second component corresponds roughly to a time shift effect, which is not constant through the cycle. The third component corresponds to a variation in the overall amplitude of the angle traced out during the cycle.

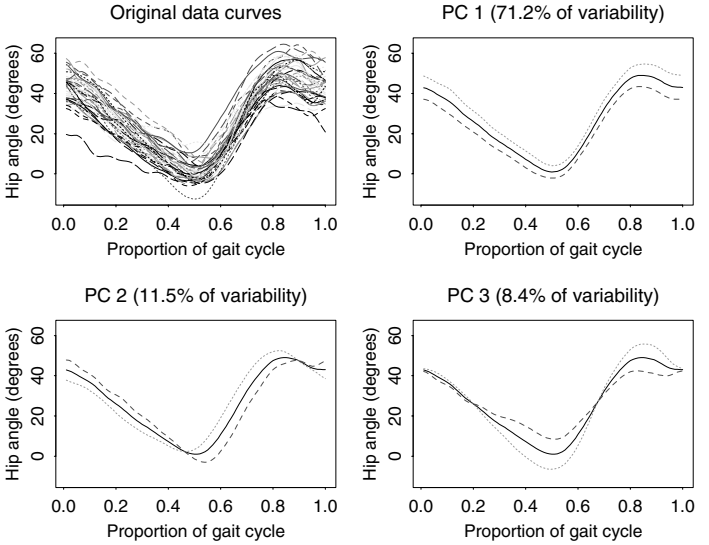

Figure 8.3. The hip angle observed in the gait cycles of 39 children, and the effect on the overall mean of adding and subtracting a suitable multiple of each of the first three principal component functions.

8.3.2 Plotting principal component scores

An important aspect of PCA is the examination of the scores f_{im} of each curve on each component. In Figure 8.4, each weather station is identified by a four-letter abbreviation of its name given in Table 8.1. The strings are positioned roughly according to the scores on the first two principal components; some positions have been adjusted slightly to improve legibility. The West Coast stations Vancouver (VANC), Victoria (VICT) and Prince Rupert (PRUP) are in the upper right corner because they have warmer winters than most stations (high on PC 1) and less summer-winter temperature variation (high on PC 2). Resolute (RESO), on the other hand, has an extremely cold winter, but does resemble the Pacific weather stations in having less summer/winter variation than some Arctic cousins, such as Inuvik (INUV).

8.3.3 Rotating principal components

In Section 8.2 we observed that the weight functions ξ_m can be viewed as defining an orthonormal set of K functions for expanding the curves to minimize a summed integrated squared error criterion (8.5). For the

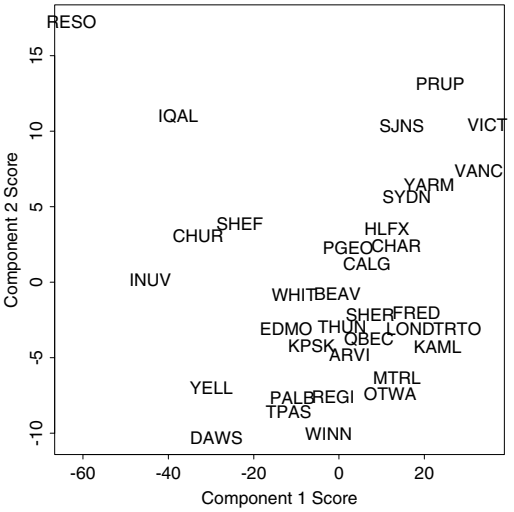

Figure 8.4. The scores of the weather stations on the first two principal components of temperature variation. The location of each weather station is shown by the four-letter abbreviation of its name assigned in Table 8.1.

Table 8.1. The Canadian Weather Stations

Arvida, Que.	Kapuskasing, Ont.	St. John's, Nfld
Beaverlodge, B.C.	London, Ont.	Sydney, N.S.
Calgary, Alta.	Montreal, Que.	The Pas, Man.
Charlottetown, P.E.I.	Ottawa, Ont.	Thunder Bay, Ont.
Churchill, Man.	Prince Albert, Sask.	Toronto, Ont.
Dawson, Yukon	Prince George, B.C.	Vancouver, B.C.
Edmonton, Alta.	Prince Rupert, B.C.	Victoria, B.C.
Fredericton, N.B.	Quebec City, Que.	Whitehorse, Yukon
Halifax, N.S.	Regina, Sask.	Winnipeg, Man.
Inuvik, N.W.T.	Resolute, N.W.T.	Yarmouth, N.S.
Iqualuit, N.W.T.	Schefferville, Que.	Yellowknife, N.W.T.
Kamloops, B.C.	Sherbrooke, Que.	

temperature data, for example, no set of four orthonormal functions will do a better job of approximating the curves than those displayed in Figure 8.1.

This does not mean, however, that there aren't other orthonormal sets that will do just as well. In fact, if we now use ξ to refer to the vector-valued

function $(\xi_1, \ldots, \xi_K)'$, then an equally good orthonormal set is defined by

$$\psi = \mathbf{T}\xi, \tag{8.13}$$

where $\mathbf{T}$ is any orthonormal matrix of order K, meaning that $\mathbf{T}'\mathbf{T} = \mathbf{T}\mathbf{T}' = \mathbf{I}$. From a geometrical perspective, the vector of functions ψ is a rigid rotation of ξ. Of course, after rotation, we can no longer expect that ψ_1 will define the largest component of variation. But the point is that the orthonormal basis functions $\psi_1, \ldots, \psi_K$ are just as effective at approximating the original curves in K dimensions as their unrotated counterparts.

Can we find some rotated functions that are perhaps a little easier to interpret? Here again, we can borrow a tool that has been invaluable in multivariate analysis, VARIMAX rotation. Let $\mathbf{B}$ be a $K \times n$ matrix representing the first K principal component functions $\xi_1, \ldots, \xi_K$. For the moment, suppose that $\mathbf{B}$ has, as row m, the values $\xi_m(t_1), \ldots, \xi_m(t_n)$ for n equally spaced argument values in the interval $\mathcal{T}$. The corresponding matrix $\mathbf{A}$ of values of the rotated basis functions $\psi = \mathbf{T}\xi$ will be given by

$$\mathbf{A} = \mathbf{T}\mathbf{B}. \tag{8.14}$$

The VARIMAX strategy for choosing the orthonormal rotation matrix $\mathbf{T}$ is to maximize the variation in the values a_{mj}^2 strung out as a single vector. Since $\mathbf{T}$ is a rotation matrix, the overall sum of these squared values will remain the same no matter what rotation we perform. In algebraic terms,

$$\sum_m \sum_j a_{mj}^2 = \operatorname{trace} \mathbf{A}'\mathbf{A} = \operatorname{trace} \mathbf{B}'\mathbf{T}'\mathbf{T}\mathbf{B} = \operatorname{trace} \mathbf{B}'\mathbf{B}.$$

Therefore, maximizing the variance of the a_{mj}^2 can happen only if these values tend either to be relatively large or relatively near zero. The values a_{mj} themselves are encouraged to be either strongly positive, near zero, or strongly negative; in-between values are suppressed. This clustering of information tends to make the components of variation easier to interpret.

There are fast and stable computational techniques for computing the rotation matrix $\mathbf{T}$ that maximizes the VARIMAX criterion. A C function for computing the VARIMAX rotation can be found through the book's world-wide web page described in Section 1.9.

Figure 8.5 displays the VARIMAX rotation of the four principal components for the temperature data. There, $n = 12$ equally spaced time points t_j were used, and the variance of the squared values $\psi_m^2(t_j)$ was maximized with respect to $\mathbf{T}$. The resulting rotated functions ψ_m, along with the percentages of variances that they account for, are now quite different.

Collectively, the rotated functions ψ_m still account for a total of 99.7% of the variation, but they divide this variation in different proportions. The VARIMAX rotation has suppressed medium-sized values of ψ_m while preserving orthonormality. (Note that the rotated component scores are no longer uncorrelated; however, the sum of their variances is still the same, because $\mathbf{T}$ is a rotation matrix, and so they may still be considered to

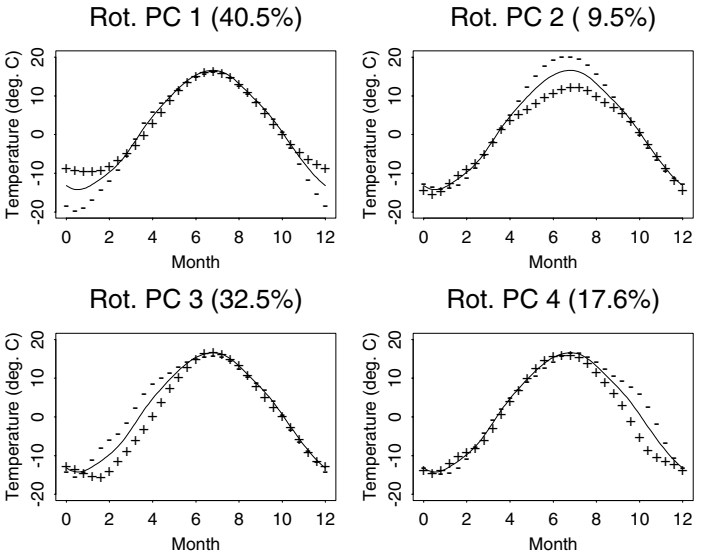

Figure 8.5. Weight functions rotated by applying the VARIMAX rotation criterion to weight function values, and plotted as positive and negative perturbations of the mean function.

partition the variability in the original data.) The result is four functions that account for local variation in the winter, summer, spring and autumn, respectively. Not only are these functions much easier to interpret, but we see something new: although winter variation remains extremely important, now spring variation is clearly almost as important, about twice as important as autumn variation and over three times as important as summer variation.

Another way of using the VARIMAX idea is to let $\mathbf{B}$ contain the coefficients for the expansion of each ξ_m in terms of a basis ϕ of n functions. Thus we rotate the coefficients of the basis expansion of each ξ_m rather than rotating the values of the ξ_m themselves. Figure 8.6 shows the results using a Fourier series expansion of the principal components. The results are much more similar to the original principal components displayed in Figure 8.2. The main difference is in the first two components. The first rotated component function in Figure 8.6 is much more constant than the original first principal component, and corresponds almost entirely to a constant temperature effect throughout the year. The general shape of the second component is not changed very much, but it accounts for more of the variability, having essentially taken on part of the variability in the first unrotated component. Because the first component originally accounted for such a large proportion, 89.3%, of the variability, it is not surprising that a

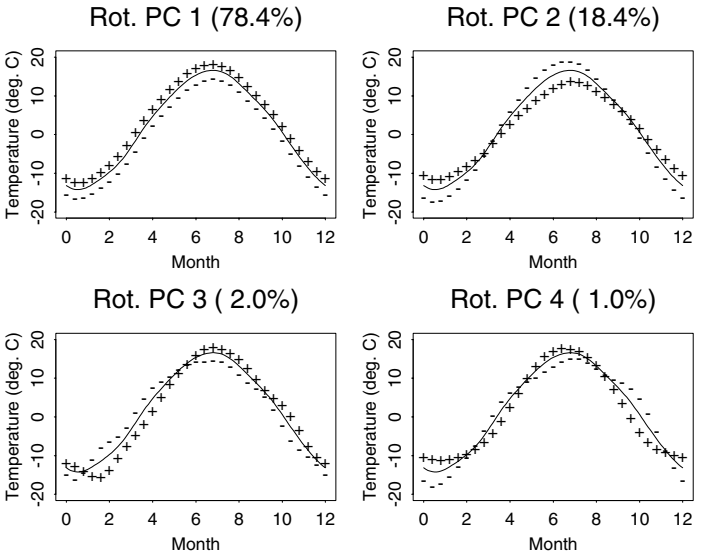

Figure 8.6. Weight functions rotated by applying the VARIMAX rotation criterion to weight function coefficients, and plotted as positive and negative perturbations of the mean function.

fairly small change in the shape of the second component results in moving about 10% of the total variability from the first to the second component. The third and fourth components are not enormously affected by the VARIMAX rotation in the Fourier domain.

By no means is the VARIMAX criterion the only rotation criterion available. References on factor analysis and multivariate statistics such as Basilevsky (1994), Johnson and Wichern (1988), Mulaik (1972) and Seber (1984) offer a number of other possibilities. Even from the relatively brief discussion in this section, it is clear that much research remains to be done on rotation schemes tailored more directly to the functional context.

8.4 Computational methods for functional PCA

Now suppose that we have a set of N curves x_i, and that preliminary steps such as curve registration and the possible subtraction of the mean curve from each (curve centering) have been completed. Let $v(s,t)$ be the sample covariance function of the observed data. In this section, we consider possible strategies for approaching the eigenanalysis problem in (8.9). In all cases, we convert the continuous functional eigenanalysis problem to an approximately equivalent matrix eigenanalysis task.

8.4.1 Discretizing the functions

A simple approach is to discretize the observed functions x_i to a fine grid of n equally spaced values s_j that span the interval $\mathcal{T}$. This yields an $N \times n$ data matrix $\mathbf{X}$ that can be fed into a standard multivariate principal components analysis program such as the S-PLUS routine `prcomp`. This produces eigenvalues and eigenvectors satisfying

$$\mathbf{V}\mathbf{u} = \lambda\mathbf{u} \tag{8.15}$$

for n-vectors $\mathbf{u}$.

Notice that we may well have n much larger than N. Rather than working with the $n \times n$ matrix $\mathbf{V}$, one possible approach to finding the solutions of the eigenequation (8.15) is to work in terms of the SVD $\mathbf{UDW}'$ of $\mathbf{X}$. The variance matrix satisfies $N\mathbf{V} = \mathbf{WD}^2\mathbf{W}'$, and hence the nonzero eigenvalues of $\mathbf{V}$ are the squares of the singular values of $\mathbf{X}$, and the corresponding eigenvectors are the columns of $\mathbf{U}$. If we use a standard PCA package, these steps, or corresponding ones, will be carried out automatically in any case.

How do we transform the vector principal components back into functional terms? The sample variance-covariance matrix $\mathbf{V} = N^{-1}\mathbf{X}'\mathbf{X}$ will have elements $v(s_j, s_k)$ where $v(s,t)$ is the sample covariance function. Given any function ξ, let $\tilde{\boldsymbol{\xi}}$ be the n-vector of values $\xi(s_j)$. Let $w = T/n$ where T is the length of the interval $\mathcal{T}$. Then, for each s_j,

$$V\xi(s_j) = \int v(s_j, s)\xi(s)ds \approx w \sum v(s_j, s_k)\tilde{\boldsymbol{\xi}}_k,$$

so the functional eigenequation $V\xi = \rho\xi$ has the approximate discrete form

$$w\mathbf{V}\tilde{\boldsymbol{\xi}} = \rho\tilde{\boldsymbol{\xi}}.$$

The solutions of this equation will correspond to those of (8.15), with eigenvalues $\rho = w\lambda$. The discrete approximation to the normalization $\int \xi(s)^2 ds = 1$ is $w\|\tilde{\boldsymbol{\xi}}\|^2 = 1$, so that we set $\tilde{\boldsymbol{\xi}} = w^{-1/2}\mathbf{u}$ if $\mathbf{u}$ is a normalized eigenvector of $\mathbf{V}$. Finally, to obtain an approximate eigenfunction ξ from the discrete values $\tilde{\boldsymbol{\xi}}$, we can use any convenient interpolation method. If the discretization values s_j are closely spaced, the choice of interpolation method will not usually have a great effect.

The discretization approach is the earliest approach to functional principal components analysis, used by Rao (1958, 1987) and Tucker (1958), who applied multivariate principal components analysis without modification to observed function values. We discuss the idea of discretizing the integral in more detail in Section 8.4.3, but first we consider an alternative approach that makes use of basis expansions.

8.4.2 Basis function expansion of the functions

One way of reducing the eigenequation (8.9) to discrete or matrix form is to express each function x_i as a linear combination of known basis functions

ϕ_k. The number K of basis functions used depends on many considerations: how many discrete sampling points n were in the original data, whether some level of smoothing was to be imposed by using $K < n$, how efficient or powerful the basis functions are in reproducing the behavior of the original functions, and so forth. For the monthly temperature data, for example, it would be logical to use a Fourier series basis orthonormal over the interval [0,12], with $K = 12$ the maximum possible dimension of the basis for the monthly temperature data, because only 12 sampling points are available per curve. Actually, for these data, a value of K as small as 7 would capture most of the interesting variation in the original data, but there is little point in reducing K below the value of 12.

Now suppose that each function has basis expansion

$$x_i(t) = \sum_{k=1}^{K} c_{ik}\phi_k(t). \tag{8.16}$$

We may write this more compactly by defining the vector-valued function $\mathbf{x}$ to have components $x_1, \ldots, x_N$, and the vector-valued function $\boldsymbol{\phi}$ to have components $\phi_1, \ldots, \phi_K$. We may then express the simultaneous expansion of all N curves as

$$\mathbf{x} = \mathbf{C}\boldsymbol{\phi},$$

where the coefficient matrix $\mathbf{C}$ is $N \times K$. In matrix terms the variance-covariance function is

$$v(s,t) = N^{-1}\boldsymbol{\phi}(s)'\mathbf{C}'\mathbf{C}\boldsymbol{\phi}(t),$$

remembering that $\boldsymbol{\phi}(s)'$ denotes the transpose of the vector $\boldsymbol{\phi}(s)$ and has nothing to do with differentiation.

Define the order K symmetric matrix $\mathbf{W}$ to have entries

$$w_{k_1,k_2} = \int \phi_{k_1}\phi_{k_2}$$

or $\mathbf{W} = \int \boldsymbol{\phi}\boldsymbol{\phi}'$. For some choices of bases, $\mathbf{W}$ will be readily available. For example, for the orthonormal Fourier series that we might use for the temperature data, $\mathbf{W} = \mathbf{I}$, the order K identity matrix. In other cases, we may have to resort to numerical integration to evaluate $\mathbf{W}$.

Now suppose that an eigenfunction ξ for the eigenequation (8.9) has an expansion

$$\xi(s) = \sum_{k=1}^{K} b_k\phi_k(s)$$

or, in matrix notation, $\xi(s) = \boldsymbol{\phi}(s)'\mathbf{b}$. This yields

$$\begin{aligned}\int v(s,t)\xi(t)\,dt &= \int N^{-1}\boldsymbol{\phi}(s)'\mathbf{C}'\mathbf{C}\boldsymbol{\phi}(t)\boldsymbol{\phi}(t)'\mathbf{b}\,dt \\ &= \boldsymbol{\phi}(s)'N^{-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b}.\end{aligned}$$

Therefore the eigenequation (8.9) can be expressed as

$$\boldsymbol{\phi}(s)'N^{-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b} = \rho\boldsymbol{\phi}(s)'\mathbf{b}.$$

Since this equation must hold for all s, this implies the purely matrix equation

$$N^{-1}\mathbf{C}'\mathbf{C}\mathbf{W}\mathbf{b} = \rho\mathbf{b}.$$

But note that $\|\xi\| = 1$ implies that $\mathbf{b}'\mathbf{W}\mathbf{b} = 1$ and, similarly, two functions ξ_1 and ξ_2 will be orthogonal if and only if the corresponding vectors of coefficients satisfy $\mathbf{b}_1'\mathbf{W}\mathbf{b}_2 = 0$. To get the required principal components, we define $\mathbf{u} = \mathbf{W}^{1/2}\mathbf{b}$, solve the equivalent symmetric eigenvalue problem

$$N^{-1}\mathbf{W}^{1/2}\mathbf{C}'\mathbf{C}\mathbf{W}^{1/2}\mathbf{u} = \rho\mathbf{u}$$

and compute $\mathbf{b} = \mathbf{W}^{-1/2}\mathbf{u}$ for each eigenvector.

Two special cases deserve particular attention. As already mentioned, if the basis is orthonormal, meaning that $\mathbf{W} = \mathbf{I}$, the functional PCA problem finally reduces to the standard multivariate PCA of the coefficient array $\mathbf{C}$, and we need only carry out the eigenanalysis of the order K symmetric array $N^{-1}\mathbf{C}'\mathbf{C}$.

As a rather different special case, particularly appropriate if the number of observed functions is not enormous, we may also view the observed functions x_i as their *own* basis expansions. In other words, there are N basis functions, and they happen to be the observed functions. This implies, of course, that $\mathbf{C} = \mathbf{I}$, and now the problem becomes one of the eigenanalysis of the symmetric matrix $N^{-1}\mathbf{W}$, which has entries

$$w_{ij} = \int x_i x_j.$$

As a rule, these entries will have to be computed by some quadrature technique.

In every case, the maximum number of eigenfunctions that can in principle be computed by the basis function approach is K, the dimension of the basis. However, if the basis expansions have involved any approximation of the observed functions, then it is not advisable to use a basis expansion to K terms to calculate more than a fairly small proportion of K eigenfunctions.

The results of both strategies that we have discussed are illustrated in Figure 8.1, which shows the first four estimated eigenfunctions ξ_m of the centered temperature functions

$$x_i = \texttt{Temp}_i - \frac{1}{35}\sum_j \texttt{Temp}_j.$$

The smooth curves give the estimated eigenfunctions using the complete 12-term Fourier series expansion. For comparison purposes, the results of applying the discretization approach to the data are also displayed as points

indicating the values of the eigenvectors. There is little discrepancy between the two sets of results. The proportions of variances for the basis function analysis turn out to be identical to those computed for the discretization approach. No attempt has been made to interpolate the discretized values to give continuous eigenfunctions, but if the Fourier series interpolation method were used, the results would be identical to the results obtained by the basis method; this is a consequence of special properties of Fourier series.

8.4.3 More general numerical quadrature

The eigenequation (8.9) involves the integral $\int x_i(s)\xi(s)\,ds$, and the discretization strategy is to approximate this integral by a sum of discrete values. Most schemes for numerical integration or quadrature (Stoer and Bulirsch, 2002, is a good reference) involve an approximation of the form

$$\int f(s)\,ds \approx \sum_{j=1}^{n} w_j f(s_j) \tag{8.17}$$

and the method set out in Section 8.4.1 is a fairly crude special case. We restrict our attention to linear quadrature schemes of the form (8.17). There are three aspects of the approximation that can be manipulated to meet various objectives:

- n, the number of discrete argument values s_j
- s_j, the argument values, called *quadrature points*
- w_j, the weights, called *quadrature weights*, attached to each function value in the sum.

A simple example is the *trapezoidal rule*, in which the interval of integration is divided into $n-1$ equal intervals, each of width h. The s_j are the boundaries of the interval with s_1 and s_n the lower and upper limits of integration, respectively, and the approximation is

$$\int f(s)\,ds \approx h[f(s_1)/2 + \sum_{j=2}^{n-1} f(s_j) + f(s_n)/2]. \tag{8.18}$$

Note that the weights w_j are $h/2, h, \ldots, h, h/2$ and that accuracy is controlled simply by the choice of n. The trapezoidal rule has some important advantages: the original raw data are often collected for equally spaced argument values, the weights are trivial, and although the accuracy of the method is modest relative to other more sophisticated schemes, it is often entirely sufficient for the objectives at hand. The method we set out in Section 8.4.1 is similar to the trapezoidal rule, and indeed if we use periodic boundary conditions, the methods are the same, since the values $f(s_n)$ and $f(s_1)$ are identical.

Other techniques, Gaussian quadrature schemes for example, define quadrature weights and points that yield much higher accuracy for fixed n under suitable additional conditions on the integrand. Another class of procedures chooses the quadrature points adaptively to provide more resolution in regions of high integrand curvature; for these to be relevant to the present discussion, we must choose the quadrature points once for all the functions considered in the analysis.

Applying quadrature schemes of the type (8.17) to the operator V in (8.10), yields the discrete approximation

$$V\xi \approx \mathbf{V}\mathbf{W}\tilde{\boldsymbol{\xi}}, \tag{8.19}$$

where, as in Section 8.4.1, the matrix $\mathbf{V}$ contains the values $v(s_j, s_k)$ of the covariance function at the quadrature points, and $\tilde{\boldsymbol{\xi}}$ is an order n vector containing values $\xi(s_j)$. The matrix $\mathbf{W}$ is a diagonal matrix with diagonal values being the quadrature weights w_j.

The approximately equivalent matrix eigenanalysis problem is then

$$\mathbf{V}\mathbf{W}\tilde{\boldsymbol{\xi}} = \rho\tilde{\boldsymbol{\xi}},$$

where the orthonormality requirement is now

$$\tilde{\boldsymbol{\xi}}_m'\mathbf{W}\tilde{\boldsymbol{\xi}}_m = 1 \text{ and } \tilde{\boldsymbol{\xi}}_{m_1}'\mathbf{W}\tilde{\boldsymbol{\xi}}_{m_2} = 0,\ m_1 \neq m_2.$$

Since most quadrature schemes use positive weights, we can put the approximate eigenequation in more standard form, analogous to the calculations carried out in Section 8.4.2:

$$\mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}\mathbf{u} = \rho\mathbf{u},$$

where $\mathbf{u} = \mathbf{W}^{1/2}\tilde{\boldsymbol{\xi}}$ and $\mathbf{u}'\mathbf{u} = 1$. Then the whole procedure is as follows:

1. Choose n, the w_j's, and the s_j's.
2. Compute the eigenvalues ρ_m and eigenvectors $\mathbf{u}_m$ of $\mathbf{W}^{1/2}\mathbf{V}\mathbf{W}^{1/2}$.
3. Compute
$$\tilde{\boldsymbol{\xi}}_m = \mathbf{W}^{-1/2}\mathbf{u}_m.$$
4. If needed, use an interpolation technique to convert each vector $\tilde{\boldsymbol{\xi}}_m$ to a function ξ_m.

If the number n of quadrature points is less than the number of curves N, we cannot recover more than n approximate eigenfunctions. However, many applications of PCA require only a small number of the leading eigenfunctions, and any reasonably large n will serve.

To illustrate the application of this discretizing approach, we analyze the acceleration in human growth described in Chapter 1. Each curve consists of 141 equally spaced values of acceleration in height estimated for ages from 14 to 18 years, after spline smoothing and registration by certain

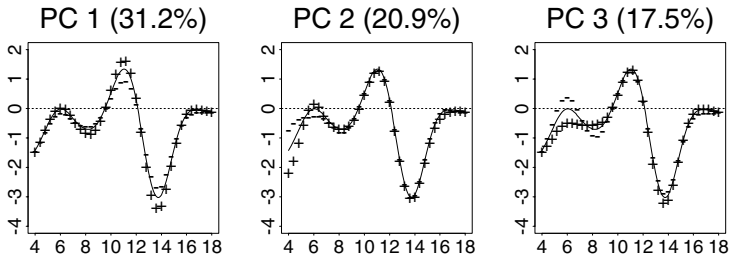

Figure 8.7. The solid curve in each panel is the mean acceleration in height in cm/year^2 for girls in the Zurich growth study. Each principal component is plotted in terms of its effect when added (+) and subtracted (−) from the mean curve.

marker events. Full details of this process can be found in Ramsay, Bock and Gasser (1995). The curves are for 112 girls who took part in the Zurich growth study (Falkner, 1960).

Figure 8.7 shows the first three eigenfunctions or harmonics plotted as perturbations of the mean function. Essentially, the first principal component reflects a general variation in the amplitude of the variation in acceleration that is spread across the entire curve, but is particularly marked during the pubertal growth spurt lasting from 10 to 16 years of age. The second component indicates variation in the size of acceleration only from ages 4 to 6, and the third component, of great interest to growth researchers, shows a variation in intensity of acceleration in the prepubertal period around ages 5 to 9 years.

8.5 Bivariate and multivariate PCA

We often wish to study the simultaneous variation of more than one function. The hip and knee angles described in Chapter 1 are an example; to understand the total system, we want to know how hip and knee angles vary jointly. Similarly, the handwriting data require the study of the simultaneous variation of the X and Y coordinates; there would be little point in studying one coordinate at a time. In both these cases, the two variables being considered are measured relative to the same argument, time in both cases. Furthermore, they are measuring quantities in the same units (degrees in the first case and cm in the second). The discussion in this section is particularly aimed towards problems of this kind.

8.5.1 Defining multivariate functional PCA

For clarity of exposition, we discuss the extension of the PCA idea to deal with bivariate functional data in the specific context of the hip and knee data. Suppose that the observed hip angle curves are $\texttt{Hip}_1, \texttt{Hip}_2, \ldots, \texttt{Hip}_n$ and the observed knee angles are $\texttt{Knee}_1, \texttt{Knee}_2, \ldots, \texttt{Knee}_n$. Let `Hipmn` and `Kneemn` be estimates of the mean functions of the `Hip` and `Knee` processes. Define $v_{\texttt{HH}}$ to be the covariance operator of the $\texttt{Hip}_i$, $v_{\texttt{KK}}$ that of the $\texttt{Knee}_i$, $v_{\texttt{HK}}$ to be the cross-covariance function, and $v_{\texttt{KH}}(t,s) = v_{\texttt{HK}}(s,t)$.

A typical principal component is now defined by a 2-vector $\xi = (\xi^{\texttt{H}}, \xi^{\texttt{K}})'$ of weight functions, with $\xi^{\texttt{H}}$ denoting the variation in the `Hip` curve and $\xi^{\texttt{K}}$ that in the `Knee` curve. To proceed, we need to define an inner product on the space of vector functions of this kind. Once this has been defined, the principal components analysis can be formally set out in exactly the same way as previously.

The most straightforward definition of an inner product between bivariate functions is simply to sum the inner products of the two components. Suppose ξ_1 and ξ_2 are both bivariate functions each with hip and knee components. We then define the inner product of ξ_1 and ξ_2 to be

$$\langle \xi_1, \xi_2 \rangle = \int \xi_1^{\texttt{H}} \xi_2^{\texttt{H}} + \int \xi_1^{\texttt{K}} \xi_2^{\texttt{K}}. \tag{8.20}$$

The corresponding squared norm $\|\xi\|^2$ of a bivariate function ξ is simply the sum of the squared norms of the two component functions $\xi^{\texttt{H}}$ and $\xi^{\texttt{K}}$.

What all this amounts to, in effect, is stringing two (or more) functions together to form a composite function. We do the same thing with the data themselves: define $\texttt{Angles}_i = (\texttt{Hip}_i, \texttt{Knee}_i)$. The weighted linear combination (8.4) becomes

$$f_i = \langle \xi, \texttt{Angles}_i \rangle = \int \xi^{\texttt{H}} \texttt{Hip}_i + \int \xi^{\texttt{K}} \texttt{Knee}_i. \tag{8.21}$$

We now proceed exactly as in the univariate case, extracting solutions of the eigenequation system $V\xi = \rho\xi$, which can be written out in full detail as

$$\begin{aligned} \int v_{\texttt{HH}}(s,t)\xi^{\texttt{H}}(t)\,dt + \int v_{\texttt{HK}}(s,t)\xi^{\texttt{K}}(t)\,dt &= \rho\xi^{\texttt{H}}(s) \\ \int v_{\texttt{KH}}(s,t)\xi^{\texttt{H}}(t)\,dt + \int v_{\texttt{KK}}(s,t)\xi^{\texttt{K}}(t)\,dt &= \rho\xi^{\texttt{K}}(s). \end{aligned} \tag{8.22}$$

In practice, we carry out this calculation by replacing each function $\texttt{Hip}_i$ and $\texttt{Knee}_i$ with a vector of values at a fine grid of points or coefficients in a suitable expansion. For each i these vectors are concatenated into a single long vector Z_i; the covariance matrix of the Z_i is a discretized version of the operator V as defined in (8.7). We carry out a standard principal components analysis on the vectors Z_i, and separate the resulting principal component vectors into the parts corresponding to `Hip` and to `Knee`. The

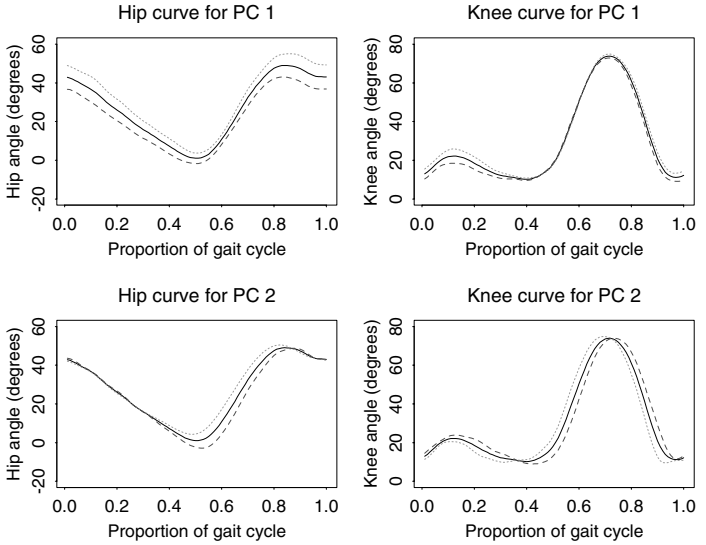

Figure 8.8. The mean hip and knee angle curves and the effects of adding and subtracting a multiple of each of the first two vector principal components.

analysis is completed by applying a suitable inverse transform to each of these parts if necessary.

If the variability in one of the sets of curves is substantially greater than that in the other, then it is advisable to consider down-weighting the corresponding term in the inner product (8.20), and making the consequent changes in the remainder of the procedure. In the case of the hip and knee data, however, both sets of curves have similar amounts of variability and are measured in the same units (degrees) and so there is no need to modify the inner product.

8.5.2 Visualizing the results

In the bivariate case, the best way to display the result depends on the particular context. In some cases it is sufficient to consider the individual parts $\xi_m^{\mathtt{H}}$ and $\xi_m^{\mathtt{K}}$ separately. An example of this is given in Figure 8.8, which displays the first two principal components. Because $\|\xi_m^{\mathtt{H}}\|^2 + \|\xi_m^{\mathtt{K}}\|^2 = 1$ by definition, calculating $\|\xi_m^{\mathtt{H}}\|^2$ gives the proportion of the variability in the mth principal component accounted for by variation in the hip curves.

For the first principal components, this measure indicates that 85% of the variation is due to the hip curves, and this is borne out by the presentation in Figure 8.8. The effect on the hip curves of the first combined principal component of variation is virtually identical to the first principal

component curve extracted from the hip curves considered alone. There is also little associated variation in the knee curves, apart from a small associated increase in the bend of the knee during the part of the cycle where all the weight is on the observed leg. The main effect of the first principal component remains an overall shift in the hip angle. This could be caused by an overall difference in stance; some people stand up more straight than others and therefore hold their trunks at a different angle from the legs through the gait cycle. Alternatively, there may simply be variation in the angle of the marker placed on the trunk.

For the second principal component, the contributions of both hip and knee are important, with somewhat more of the variability (65%) due to the knee than to the hip. We see that this principal component is mainly a distortion in the timing of the cycle, again correlated with the way in which the initial slight bend of the knee takes place. There is some similarity to the second principal component found for the hip alone, but this time there is very substantial interaction between the two joints.

A particularly effective method for displaying principal components in the bivariate case is to construct plots of one variable against the other. Suppose we are interested in displaying the mth principal component function. For equally spaced points t in the time interval on which the observations are taken, we indicate the position of the mean function values $(\texttt{Hipmn}(t), \texttt{Kneemn}(t))$ by a dot in the (x, y) plane, and we join this dot by an arrow to the point $(\texttt{Hipmn}(t) + C\xi_m^{\texttt{H}}(t),\ \texttt{Kneemn}(t) + C\xi_m^{\texttt{K}}(t))$. We choose the constant C to give clarity. Of course, the sign of the principal component functions, and hence the sense of the arrows, is arbitrary, and plots with all the arrows reversed convey the same information.

This technique is displayed in Figure 8.9. The plot of the mean cycle alone demonstrates the overall shape of the gait cycle in the hip-knee plane. The portion of the plot between time points 11 and 19 (roughly the part where the foot is off the ground) is approximately half an ellipse with axes inclined to the coordinate axes. The points on the ellipse are roughly at equal angular coordinates — somewhat closer together near the more highly curved part of the ellipse. This demonstrates that in this part of the cycle, the joints are moving roughly in simple harmonic motion but with different phases. During the other part of the cycle, the hip angle is changing at a approximately constant rate as the body moves forward with the leg approximately straight, and the knee bends slightly in the middle.

Now consider the effect of the first principal component of variation. As we have already seen, this has little effect on the knee angle, and all the arrows are approximately in the x-direction. The increase in the hip angle due to this mode of variation is somewhat larger when the angle itself is larger. This indicates that the effect contains an exaggeration (or diminution) in the amount by which the hip joint is bent during the cycle, and is also related to the overall angle between the trunk and the legs.

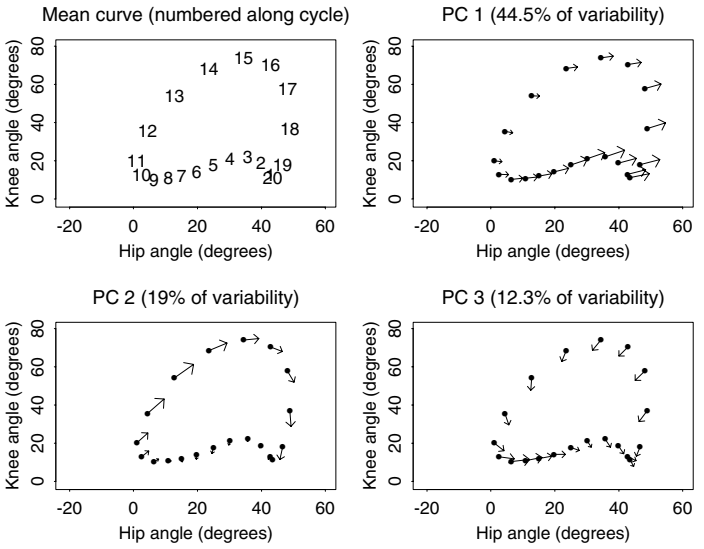

Figure 8.9. A plot of 20 equally spaced points in the average gait cycle, and the effects of adding a multiple of each of the first three principal component cycles in turn.

The second principal component demonstrates an interesting effect. There is little change during the first half of the cycle. However, during the second half, individuals with high values of this principal component would traverse roughly the same cycle but at a roughly constant time ahead. Thus this component represents a uniform time shift during the part of the cycle when the foot is off the ground.

A high score on the third component indicates two effects. There is some time distortion in the first half of the cycle, and then a shrinking of the overall cycle; an individual with a high score would move slowly through the first part of the cycle, and then perform simple harmonic motion of knee and hip joints with somewhat less than average amplitude.

8.5.3 Inner product notation: Concluding remarks

One of the features of the functional data analysis approach to principal components analysis is that, once the inner product has been defined appropriately, principal components analysis looks formally the same, whether the data are the conventional vectors of multivariate analysis, scalar functions as considered in Section 8.2.2, or vector-valued functions as in Section 8.5.1. Indeed, principal component analyses for other possible forms of functional data can be constructed similarly; all that is needed

is a suitable inner product, and in most contexts the definition of such an inner product will be a natural one. For example, if our data are functions defined over a region $\mathcal{S}$ in two-dimensional space, for example temperature profiles over a geographical region, then the natural inner product will be given by

$$\int_{\mathcal{S}} f(\mathbf{s})g(\mathbf{s})d\mathbf{s},$$

and the principal component weight functions will also be functions defined over $\mathbf{s}$ in $\mathcal{S}$.

Much of our subsequent discussion of PCA, and of other functional data analysis methods, will use univariate functions of a single variable as the standard example. This choice simplifies the exposition, but in most or all cases the methods generalize immediately to other forms of functional data, simply by substituting an appropriate definition of inner product.

8.6 Further readings and notes

An especially fascinating and comprehensive application of functional principal components analysis can be found in Locantore, Marron, Simpson, Tripoli, Zhang and Cohen (1999). These authors explore abnormalities in the curvature of the cornea in the human eye, and along the way extend functional principal components methodology in useful ways. Since the variation is over the spherical or elliptical shape of the cornea, they use Zernicke orthogonal basis functions. Their color graphical displays and the importance of the problem make this a showcase paper.

Viviani, Grön and Spitzer (2005) apply PCA to repeated fMRI scans of areas in the human brain, where each curve is associated with a specific voxel. They compare the functional and multivariate versions, and find that the functional approach offers a rather better image of experimental manipulations underlying the data. They also find that the use of the GCV criterion is particularly effective in choosing the smoothing parameter prior to applying functional PCA.

While most of our examples have time as the argument, there are many important problems in the physical and engineering sciences where spectral analysis is involved. An example involving elements of both registration and principal components analysis is reported in Liggett, Cazares and Semmes (2003). Kneip and Utikal (2001) apply functional principal components analysis to the problem of describing a set of density curves where the argument variable is log income.

Besse, Cardot and Ferraty (1997) studied the properties of estimates of curves where these are assumed to lie within a finite-dimensional subspace, and where principal components analysis is used in the estimation process, and Cardot (2004) extended this work.

Valderrama, Aguilera and Ocaña (2000) is a monograph in Spanish that contains many interesting applications of principal components analysis to functional data at the University of Granada, some of which precede the publication of our first edition. Ocaña, Aguilera and Valderrama (1999) discuss the role of the norm used to define functional principal components analysis.

James, Hastie and Sugar (2000) have developed a useful extension of functional principal components analysis that permits the estimation of harmonics from fragments of curves. They analyze measurements of spinal bone mineral density in females children and young adults taken at various ages. Yao, Müller and Wang (2004) is a more recent reference on this important problem.

There is a considerable literature on cluster analysis of samples of curves, a topic not far removed from principal components analysis. Abraham, Cornillion, Matzner-Lober and Molinari (2003) and Tarpey and Kinateder (2003) are recent references. James and Sugar (2003) adapt their functional principal components approach to this problem. Tarpey, Petkova and Ogden (2003) use functional cluster analysis to profile placebo responders.