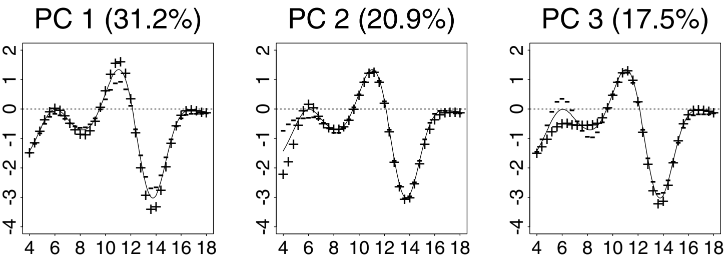

Figure 8.7. The solid curve in each panel is the mean acceleration in height in cm/year^2 for girls in the Zurich growth study. Each principal component is plotted in terms of its effect when added (+) and subtracted (−) from the mean curve.

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

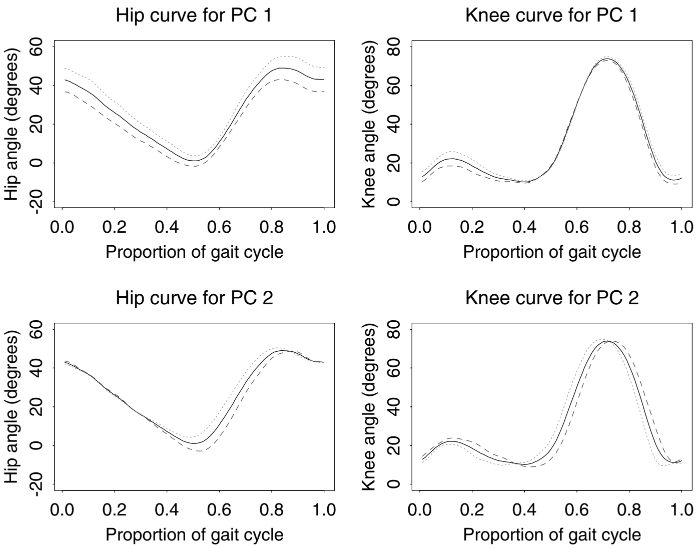

Figure 8.8. The mean hip and knee angle curves and the effects of adding and subtracting a multiple of each of the first two vector principal components.

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