Supplement to “Outlier detection and trimmed estimation for general functional data”

Daniel Gervini
Department of Mathematical Sciences
University of Wisconsin–Milwaukee
gervini@uwm.edu

July 4, 2011

Abstract

This Technical Supplement contains proofs of the theoretical results in the paper "Outlier detection and trimmed estimation for general functional data", and an additional real-data example.

1 Proofs

Proof of Proposition 1

If $\tilde{X}_i = a\mathfrak{U}X_i + b$ with $\mathfrak{U}$ unitary, then $\tilde{d}_{ij} = \|(a\mathfrak{U}X_i + b) - (a\mathfrak{U}X_j + b)\| = |a|\,\|\mathfrak{U}(X_i - X_j)\| = |a|\,d_{ij}$. This implies that $\tilde{r}_i = |a|\,r_i$ and then $w(\tilde{X}_i) = w(X_i)$, because $\mathrm{rank}(\tilde{r}_i) = \mathrm{rank}(r_i)$. Therefore

$$\begin{aligned}
\widehat{\tilde{\mu}} &= \frac{\sum_{i=1}^n w(\tilde{X}_i)\tilde{X}_i}{\sum_{i=1}^n w(\tilde{X}_i)} = \frac{\sum_{i=1}^n w(X_i)(a\mathfrak{U}X_i + b)}{\sum_{i=1}^n w(X_i)} \\
&= a\mathfrak{U}\hat{\mu} + b,
\end{aligned}$$

and for every f and g in $\mathscr{H}$ we have

$$\begin{aligned}
\widehat{\tilde{\mathfrak{C}}}(f,g) &= \frac{1}{\sum_{i=1}^n w(\tilde{X}_i)} \sum_{i=1}^n w(\tilde{X}_i)\langle \tilde{X}_i - \widehat{\tilde{\mu}}, f\rangle\langle \tilde{X}_i - \widehat{\tilde{\mu}}, g\rangle \\
&= \frac{1}{\sum_{i=1}^n w(X_i)} \sum_{i=1}^n w(X_i)\langle a\mathfrak{U}(X_i - \hat{\mu}), f\rangle\langle a\mathfrak{U}(X_i - \hat{\mu}), g\rangle \\
&= \frac{a^2}{\sum_{i=1}^n w(X_i)} \sum_{i=1}^n w(X_i)\langle X_i - \hat{\mu}, \mathfrak{U}^* f\rangle\langle X_i - \hat{\mu}, \mathfrak{U}^* g\rangle \\
&= a^2\widehat{\mathfrak{C}}(\mathfrak{U}^* f, \mathfrak{U}^* g).
\end{aligned}$$

To complete the proof note that for every $g \in \mathscr{H}$,

$$\begin{aligned}
\widehat{\tilde{\mathfrak{C}}}(\mathfrak{U}\hat{\phi}_k, g) &= a^2\widehat{\mathfrak{C}}(\mathfrak{U}^*\mathfrak{U}\hat{\phi}_k, \mathfrak{U}^* g) \\
&= a^2\widehat{\mathfrak{C}}(\hat{\phi}_k, \mathfrak{U}^* g) \\
&= a^2\hat{\lambda}_k\langle\hat{\phi}_k, \mathfrak{U}^* g\rangle \\
&= a^2\hat{\lambda}_k\langle\mathfrak{U}\hat{\phi}_k, g\rangle.
\end{aligned}$$

Then $\mathfrak{U}\hat{\phi}_k$ is an eigenfunction of $\widehat{\tilde{\mathfrak{C}}}$ with eigenvalue $a^2\hat{\lambda}_k$, and since $a^2 > 0$, the order is preserved (i.e. if $\hat{\phi}_k$ is the kth eigenfunction of $\widehat{\mathfrak{C}}$ then $\mathfrak{U}\hat{\phi}_k$ is the kth eigenfunction of $\widehat{\tilde{\mathfrak{C}}}$.)

Proof of Proposition 2

We will first show that $\varepsilon_n^*(\hat{\mu}) \geq \min(\lceil \alpha n \rceil, \lfloor \beta n \rfloor + 2)/n$, and then proceed to show that the equality holds by exhibiting a particular sequence of contaminations that makes $\|\hat{\mu}\|$ go to infinity.

Suppose, then, that $\{\tilde{\mathscr{X}}^{(m)}\}_{m\geq 1}$ is a sequence of contaminated samples obtained from $\mathscr{X}$ by replacing k observations, and such that $\|\hat{\mu}^{(m)}\| \to \infty$ as $m \to \infty$. Since $\|\hat{\mu}^{(m)}\| \leq \sum_{i=1}^n w(\tilde{X}_i^{(m)})\|\tilde{X}_i^{(m)}\| / \sum_{i=1}^n w(\tilde{X}_i^{(m)})$, one can choose a sequence of points $\{\tilde{X}_{i_m}^{(m)}\}_{m\geq 1}$ such that $\|\tilde{X}_{i_m}^{(m)}\| \to \infty$ when $m \to \infty$ and $w(\tilde{X}_{i_m}^{(m)}) > 0$ for all m. The latter implies that the rank of the corresponding radii $\tilde{r}_{i_m}^{(m)}$ is strictly less than $(1-\beta)n$, because $g(t) = 0$ for any $t \geq 1-\beta$; therefore $\tilde{r}_{i_m}^{(m)} < \tilde{r}_{(\lceil(1-\beta)n\rceil)}^{(m)}$.

Since $\tilde{r}_{i_m}^{(m)}$ is defined as the distance between $\tilde{X}_{i_m}^{(m)}$ and its $\lceil \alpha n \rceil$-th closest point in $\tilde{\mathscr{X}}^{(m)}$, and $\tilde{\mathscr{X}}^{(m)}$ has only k outliers, if $k < \lceil \alpha n \rceil$ there has to be at least one point X_{j_m} of the original sample $\mathscr{X}$ such that $\|\tilde{X}_{i_m}^{(m)} - X_{j_m}\| \leq r_{i_m}^{(m)}$. Since $\|\tilde{X}_{i_m}^{(m)}\| \to \infty$, it follows that $\|\tilde{X}_{i_m}^{(m)} - X_{j_m}\| \to \infty$ when $m \to \infty$ (the X_{j_m}s are bounded), so $\tilde{r}_{i_m}^{(m)} \to \infty$ as well. But $\tilde{r}_{i_m}^{(m)} < \tilde{r}_{(\lceil(1-\beta)n\rceil)}^{(m)}$, so a total of at least $1 + (n - \lceil(1-\beta)n\rceil + 1) = \lfloor \beta n \rfloor + 2$ radii $\tilde{r}_i^{(m)}$ go to infinity when $m \to \infty$.

Now, if $k < \lceil \alpha n \rceil$ then the number of non-outliers in $\tilde{\mathscr{X}}^{(m)}$ is $n - k \geq \lceil \alpha n \rceil$ (because $\alpha \leq .50$), so for each non-outlier X_i in $\tilde{\mathscr{X}}^{(m)}$ there are at least $\lceil \alpha n \rceil$ non-outliers X_j such that $\tilde{d}_{ij}^{(m)} = \|X_i - X_j\|$ remains bounded regardless of m; therefore the corresponding $\tilde{r}_i^{(m)}$s cannot go to infinity. This means that the $\lfloor \beta n \rfloor + 2$ radii that go to infinity correspond to observations in $\tilde{\mathscr{X}}^{(m)}$ that *are* outliers, so the number of outliers k cannot be less that $\lfloor \beta n \rfloor + 2$. Then $k \geq \lfloor \beta n \rfloor + 2$ when $k < \lceil \alpha n \rceil$. The other possibility is that $k \geq \lceil \alpha n \rceil$, so $k \geq \min(\lceil \alpha n \rceil, \lfloor \beta n \rfloor + 2)$. This proves that $\varepsilon_n^*(\hat{\mu}) \geq \min(\lceil \alpha n \rceil, \lfloor \beta n \rfloor + 2)/n$.

To see that $\varepsilon_n^*(\hat{\mu}) = \min(\lceil \alpha n \rceil, \lfloor \beta n \rfloor + 2)/n$, take $k = \min(\lceil \alpha n \rceil, \lfloor \beta n \rfloor + 2)$ and consider the two possibilities: $\lfloor \beta n \rfloor + 2 \leq \lceil \alpha n \rceil$ or $\lfloor \beta n \rfloor + 2 > \lceil \alpha n \rceil$. If $\lfloor \beta n \rfloor + 2 \leq \lceil \alpha n \rceil$, take any $X_0 \in \mathscr{H}$ with norm one and define the outliers $\tilde{X}_i^{(m)} = m^i X_0$, for $i = 1, \ldots, k$; then the distance between each $\tilde{X}_i^{(m)}$ and any other point in $\tilde{\mathscr{X}}^{(m)}$ (including other outliers) goes to infinity when $m \to \infty$, so $\tilde{r}_i^{(m)} \to \infty$ for $i = 1, \ldots, k$; since $k = \lfloor \beta n \rfloor + 2$ in this case, at least one of the outliers is not cut off and then $\|\hat{\mu}^{(m)}\| \to \infty$ as $m \to \infty$. For the other case, $\lfloor \beta n \rfloor + 2 > \lceil \alpha n \rceil$, define the outliers $\tilde{X}_i^{(m)} = mX_0$ for $i = 1, \ldots, k$; then $\tilde{r}_i^{(m)} = 0$ for $i = 1, \ldots, k$ (because $k = \lceil \alpha n \rceil$ in this case), but $r_i^{(m)} > 0$ for the non-outliers (except in the trivial situation where all the sample points are identical). The number of non-outliers is $n - k = n - \lceil \alpha n \rceil > n - \lfloor \beta n \rfloor - 2$, so $n - k \geq n - \lfloor \beta n \rfloor - 1 \geq \lfloor \beta n \rfloor - 1$ (because $\beta \leq .50$). Therefore the $n - \lceil (1-\beta) n \rceil + 1 = \lfloor \beta n \rfloor + 1$ observations that are cut off include at most two outliers, and since $\lceil \alpha n \rceil \geq 3$ by hypothesis, there is at least one outlier that is not cut off, so $\|\hat{\mu}^{(m)}\| \to \infty$ as $m \to \infty$.

For $\varepsilon_n^*(\widehat{\mathfrak{C}})$ the proof is similar, because $\|\widehat{\mathfrak{C}}\| \leq \sum_{i=1}^n w(X_i)\|X_i\|^2 / \sum_{i=1}^n w(X_i) + \|\hat{\mu}\|$ and the preceding proof is also valid for $\sum_{i=1}^n w(X_i)\|X_i\|^2 / \sum_{i=1}^n w(X_i)$.

Proof of Proposition 3

By definition, $\alpha \leq P\{\|X - v\| \leq r_P(v)\}$ for any $v \in \mathscr{H}$. Since $\|X - v\| \geq |\|X\| - \|v\||$, it follows that $\alpha \leq P\{\|v\| - r_P(v) \leq \|X\|\}$. But if $\alpha > 0$, there is a finite $K_{\alpha,P}$ such that $P\{\|X\| > K_{\alpha,P}\} < \alpha/2$, say. Therefore $\|v\| - r_P(v) \leq K_{\alpha,P}$ for any $v \in \mathscr{H}$. Then, if $\beta > 0$, $w_P(v) > 0$ implies $r_P(v) \leq G_P^{-1}(1-\beta)$, which is finite, so

$$\begin{aligned} E_P\{w_P(X)\|X\|^k\} &\leq E_P[I\{\|X\| \leq K_{\alpha,P} + G_P^{-1}(1-\beta)\}\|X\|^k] \\ &\leq \{K_{\alpha,P} + G_P^{-1}(1-\beta)\}^k < \infty \end{aligned}$$

for any $k > 0$.

Proof of Proposition 4

Suppose $r_P(v) > r_P(w)$. By definition, $r_P(v) = \min\{\delta : P(B_\delta(v)) \geq \alpha\}$, so $P(B_{r_P(w)}(v)) < \alpha$. But $P(B_{r_P(w)}(v)) \geq P(B_{r_P(w)}(w))$ by hypothesis, and $P(B_{r_P(w)}(w)) \geq \alpha$ by definition of $r_P(w)$, a contradiction. Then it must be $r_P(v) \leq r_P(w)$.

Proof of the equivariance of μ_P and $\mathfrak{C}_P$

Let $X \sim P$ and $\tilde{X} = a\mathfrak{U}X + b \sim \tilde{P}$. Then

$$\begin{aligned} F_{\tilde{P}}(t; v) &= P(\|a\mathfrak{U}X + b - v\| \leq t) \\ &= P(\|X - \mathfrak{U}^*(v - b)/a\| \leq t/|a|) \\ &= F_P(t/|a|; \mathfrak{U}^*(v - b)/a), \end{aligned}$$

so $r_{\tilde{P}}(v) = |a| r_P(\mathfrak{U}^*(v - b)/a)$ for all $v \in \mathscr{H}$. Then

$$\begin{aligned} G_{\tilde{P}}(t) &= \tilde{P}\{r_{\tilde{P}}(\tilde{X}) \leq t\} \\ &= P\{r_{\tilde{P}}(a\mathfrak{U}X + b) \leq t\} \\ &= P\{|a| r_P(X) \leq t\} \\ &= G_P(t/|a|), \end{aligned}$$

which implies that

$$\begin{aligned} w_{\tilde{P}}(v) &= g[G_{\tilde{P}}\{r_{\tilde{P}}(v)\}] \\ &= g[G_P\{r_P(\mathfrak{U}^*(v - b)/a)\}] \\ &= w_P(\mathfrak{U}^*(v - b)/a) \end{aligned}$$

for all $v \in \mathscr{H}$. Therefore $E_{\tilde{P}}\{w_{\tilde{P}}(\tilde{X})\} = E_P\{w_{\tilde{P}}(a\mathfrak{U}X + b)\} = E_P\{w_P(X)\}$ and

$$E_{\tilde{P}}\{w_{\tilde{P}}(\tilde{X})\tilde{X}\} = E_P\{w_{\tilde{P}}(a\mathfrak{U}X + b)(a\mathfrak{U}X + b)\} = E_P\{w_P(X)(a\mathfrak{U}X + b)\}.$$

It follows that $\mu_{\tilde{P}} = a\mathfrak{U}\mu_P + b$. In a similar way we obtain that $\mathfrak{C}_{\tilde{P}}(f, g) = a^2 \mathfrak{C}_P(\mathfrak{U}^* f, \mathfrak{U}^* g)$ for all f and g.

Proof of Proposition 5

It follows immediately from the equivariance of μ_P. Let $X \sim P$, $X - \mu_0 \sim P_1$ and $-X + \mu_0 \sim P_2$. Then $\mu_{P_1} = \mu_P - \mu_0$ and $\mu_{P_2} = -\mu_P + \mu_0$. But $P_1 = P_2$ by hypothesis, so $\mu_P - \mu_0 = -\mu_P + \mu_0$, which implies that $\mu_P = \mu_0$.

Proof of Proposition 6

This result is also a direct consequence of the equivariance of the estimators. Since $\mathfrak{C}(\cdot)$ is location invariant we will assume, without loss of generality, that $\mu_0 = 0$. Then, since the Z_ks have a symmetric distribution about 0, by Proposition 4 we have $\mu_P = 0$. For any f and g in $\mathscr{H}$ we have

$$\begin{aligned}
\mathfrak{C}_P(f,g) &= \frac{E_P\{w_P(X)\langle X,f\rangle\langle X,g\rangle\}}{E_P\{w_P(X)\}} \\
&= \frac{1}{E_P\{w_P(X)\}}\sum_{j\in\mathscr{I}}\sum_{k\in\mathscr{I}} E_P\{w_P(X)\lambda_{0j}^{1/2}Z_j\lambda_{0k}^{1/2}Z_k\}\langle\phi_{0j},f\rangle\langle\phi_{0k},g\rangle \\
&= \frac{1}{E_P\{w_P(X)\}}\sum_{j\in\mathscr{I}}\sum_{k\in\mathscr{I}} E_P\{w_P(X)\langle\phi_{0j},X\rangle\langle\phi_{0k},X\rangle\}\langle\phi_{0j},f\rangle\langle\phi_{0k},g\rangle.
\end{aligned}$$

We will show that $E_P\{w_P(X)\langle\phi_{0j},X\rangle\langle\phi_{0k},X\rangle\} = 0$ for $j \neq k$. (To simplify notation, we will use the tensor-product operator: for a and b in $\mathscr{H}$, $a \otimes b$ denotes the operator $\mathfrak{T}$ defined as $\mathfrak{T}(f,g) = \langle a,f\rangle\langle b,g\rangle$.) Since $\{\phi_{0k}\}$ is an orthonormal system in $\mathscr{H}$, if it is not complete we extend it to an orthonormal basis of $\mathscr{H}$, $\{\check{\phi}_{0k}\}$, and correspondingly we extend the sequences $\{\lambda_{0k}\}$ and $\{Z_k\}$ by adding zeros, so that we can write $X = \sum_k \check{\lambda}_{0k}^{1/2}\check{Z}_k\check{\phi}_{0k}$ with probability one. Now, any operator of the form $\mathfrak{S} = \sum_k a_k(\check{\phi}_{0k}\otimes\check{\phi}_{0k})$ with $a_k = \pm 1$ is unitary and self-adjoint, so we know by the proof of equivariance above that if $\tilde{P}$ denotes the probability distribution of $\mathfrak{S}X$, then $w_{\tilde{P}}(v) = w_P(\mathfrak{S}v)$ for all $v \in \mathscr{H}$. But under the current assumptions the $\check{Z}_k$s are independent and symmetrically distributed around 0 (including the artificially added zeros), so $\mathfrak{S}X$ and X have the same distribution; that is, $\tilde{P} = P$ and then $w_P(v) = w_P(\mathfrak{S}v)$ for all $v \in \mathscr{H}$. Therefore

$$\begin{aligned}
E_P\{w_P(X)\langle\check{\phi}_{0j},X\rangle\langle\check{\phi}_{0k},X\rangle\} &= E_P\{w_P(\mathfrak{S}X)\langle\check{\phi}_{0j},\mathfrak{S}X\rangle\langle\check{\phi}_{0k},\mathfrak{S}X\rangle\} \\
&= E_P\{w_P(X)\langle\mathfrak{S}\check{\phi}_{0j},X\rangle\langle\mathfrak{S}\check{\phi}_{0k},X\rangle\}.
\end{aligned}$$

If, in particular, we take $\mathfrak{S}_j$ to be the sign-change operator for the jth coordinate ($a_j = -1$ and $a_k = 1$ for any $k \neq j$), we have $\mathfrak{S}_j\check{\phi}_{0j} = -\check{\phi}_{0j}$ and $\mathfrak{S}_j\check{\phi}_{0k} = \check{\phi}_{0k}$ for any $k \neq j$, so

$$E_P\{w_P(X)\langle \mathfrak{S}_j \check{\phi}_{0j}, X\rangle \langle \mathfrak{S}_j \check{\phi}_{0k}, X\rangle\} = -E_P\{w_P(X)\langle \check{\phi}_{0j}, X\rangle \langle \check{\phi}_{0k}, X\rangle\}.$$

This implies that $E_P\{w_P(X)\langle \phi_{0j}, X\rangle \langle \phi_{0k}, X\rangle\} = 0$ for $j \neq k$, as we wanted to prove. Then, in tensor-product notation,

$$\mathfrak{C}_P = \sum_{k \in \mathscr{I}} \tilde{\lambda}_{0k}(\phi_{0k} \otimes \phi_{0k})$$

with

$$\tilde{\lambda}_{0k} = \frac{E_P\{w_P(X)|\langle \phi_{0k}, X\rangle|^2\}}{E_P\{w_P(X)\}},$$

as claimed.

Now, since $F_P(t; v) = P(\|X - v\| \leq t) = P\{\sum_{k \in \mathscr{I}} (\check{\lambda}_{0k}^{1/2} \check{Z}_k - \theta_k)^2 \leq t\}$, with $\theta_k = \langle \check{\phi}_{0k}, v\rangle$, it is clear that $r_P(v)$ depends on P only through $\{\lambda_{0k}^{1/2} Z_k\}$ and does not depend on μ_0 or on the ϕ_{0k}s. It is also clear that $r_P(v)$ depends on v only through its Fourier coefficients $\{\theta_k\}$, so the distribution of $r_P(X)$, $G_P(t)$, depends on P only through $\{\lambda_{0k}^{1/2} Z_k\}$ (because these are the Fourier coefficients of X). Consequently, the weight function $w_P(v) = g[G_P\{r_P(v)\}]$ depends on P only through $\{\lambda_{0k}^{1/2} Z_k\}$, and depends on v only through $\{\theta_k\}$, implying once again that the distribution of $w_P(X)$ depends only on $\{\lambda_{0k}^{1/2} Z_k\}$. Since $\langle \phi_{0k}, X\rangle = \lambda_{0k}^{1/2} Z_k$, it is clear then that $\tilde{\lambda}_{0k}$ depends on P only through $\{\lambda_{0k}^{1/2} Z_k\}$ and does not depend on μ_0 or the ϕ_{0k}s.

Finally, suppose that the Z_ks are identically distributed in addition to being independent, and that $\lambda_{0j} = \lambda_{0k}$. Then the distribution of X remains unchanged if we switch $\lambda_{0j}^{1/2} Z_j$ with $\lambda_{0k}^{1/2} Z_k$. More formally, define the "switch operator" $\mathfrak{S}_{jk} = (\phi_{0j} \otimes \phi_{0k} + \phi_{0k} \otimes \phi_{0j}) + \{\mathfrak{I} - (\phi_{0j} \otimes \phi_{0k} + \phi_{0k} \otimes \phi_{0j})\}$, where $\mathfrak{I}$ is the identity operator. Then, if $X = \sum_{k \in \mathscr{I}} \lambda_{0k}^{1/2} Z_k \phi_{0k}$, we have $\mathfrak{S}_{jk} X = \lambda_{0j}^{1/2} Z_j \phi_{0k} + \lambda_{0k}^{1/2} Z_k \phi_{0j} + \sum_{l \neq j,k} \lambda_{0l}^{1/2} Z_l \phi_{0l}$, so X and $\mathfrak{S}_{jk} X$ have the same distribution. Moreover, since $\mathfrak{S}_{jk}$

is unitary and self-adjoint, $w_P(v) = w_P(\mathfrak{S}v)$ for all $v \in \mathscr{H}$, as before, so

$$\begin{aligned} E_P\{w_P(X)|\langle\phi_{0j}, X\rangle|^2\} &= E_P\{w_P(\mathfrak{S}_{jk}X)|\langle\phi_{0j}, \mathfrak{S}_{jk}X\rangle|^2\} \\ &= E_P\{w_P(X)|\langle\mathfrak{S}_{jk}\phi_{0j}, X\rangle|^2\} \\ &= E_P\{w_P(X)|\langle\phi_{0k}, X\rangle|^2\}. \end{aligned}$$

Then $\tilde{\lambda}_{0j} = \tilde{\lambda}_{0k}$, as claimed.

2 Third Example: Human Gait Curves

The data for this example consists of angle trajectories of the hip during the gait cycle for 39 children; it was collected at the Motion Analysis Laboratory of San Diego's Children's Hospital (Olshen et al. 1989, Ramsay and Silverman 2005). The curves begin and end at the points where the heel strikes the ground, but have been rescaled so that $t \in [0, 1]$ for all subjects. The raw data, consisting of 20 observations per individual, was smoothed using B-splines; the smoothed curves are shown in Figure 1.

One of the curves highlighted in Figure 1 is clearly atypical, showing a much lower hip angle at the beginning of the cycle than the other curves. But there are other sources of variability in the data that may produce outliers that are not so easy to visualize. For example: the minimum hip angle for each curve normally occurs near the middle of the gait cycle, but there is a lot of variability in the timing; the second curve highlighted in Figure 1 seems to be atypical in this respect, showing an unusually delayed occurrence of the minimum hip angle. It is hard to tell, however, to what extent these curves will have a negative influence on the estimators.

We computed the trimmed mean and the trimmed principal components with $\alpha = .50$ and $\beta = .10$. This value of β was suggested by a boxplot of the radii, which shows only two extreme values (corresponding to the curves highlighted in Figure 1). Boxplots of the radii for smaller values of α consistently show these two curves among the two or three observations with largest radii.

We computed only the first 10 components, since $\hat{\lambda}_{10}/\hat{\lambda}_1 \approx .001$. The proportion

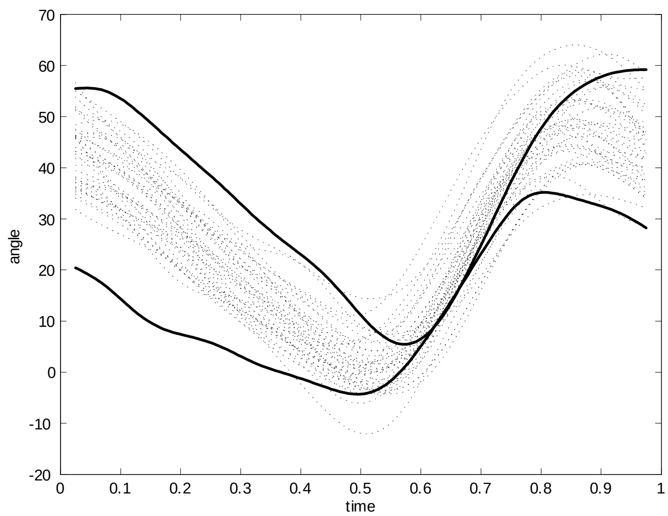

Figure 1: Human Gait Example. Trajectories of hip angles for 39 subjects (dotted line), with two atypical trajectories highligted (solid line).

of the variability explained by the first three sample principal components is .71, .13 and .09, while the respective proportions for the trimmed principal components are .66, .13 and .12, respectively. The estimators of the mean and the effects of the principal components on the mean are shown in Figure 2.

We see that the sample mean and the trimmed mean are very similar, and so are the first sample principal component and the first trimmed principal component. But the striking difference between the estimators occurs for the second and third components. The second sample principal component (Figure 2(c)) is associated with the steepness of the curves: positive component scores correspond to "flat" curves and negative component scores correspond to curves with pronounced minimum and maximum hip angles. The third sample principal component (Figure 2(e)) is associated with time variability: curves with positive component scores show early minimum hip angles, while curves with negative component scores show delayed minimum hip angles. In contrast, the *second* trimmed principal component (Figure 2(d)) is the one associated with time variability, while the third one (Figure 2(f)) is associated with steepness. In fact, Figures 2(c) and 2(f) are

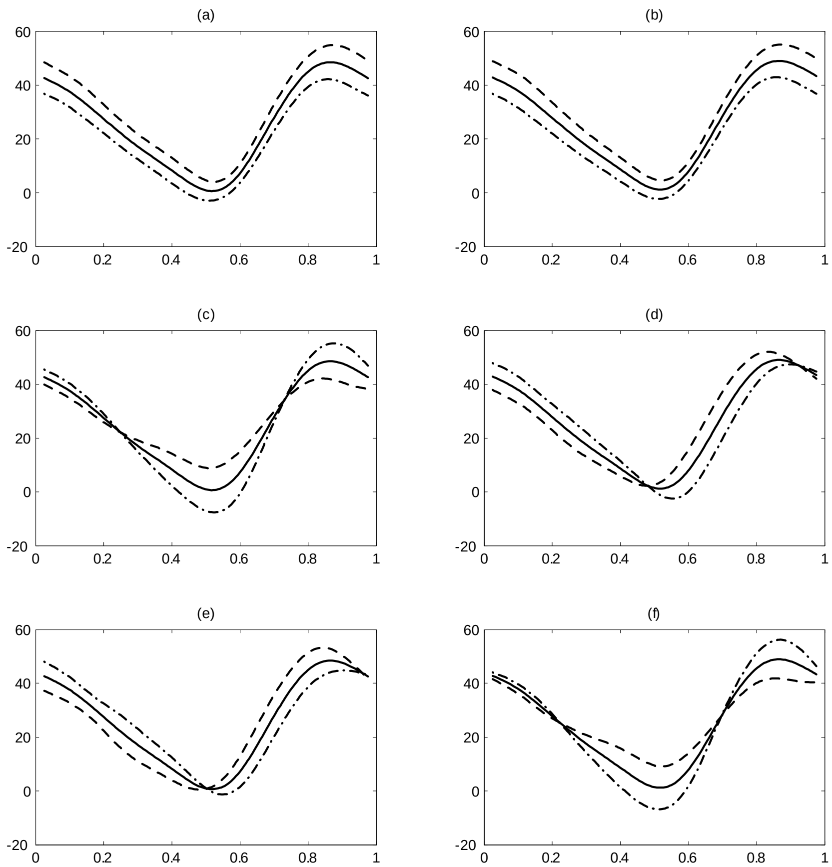

Figure 2: Human Gait Example. Estimators of the mean (solid line), and the mean plus (dashed line) and minus (dash-dot line) a constant times the principal components are shown, for (a) first sample principal component, (b) first trimmed principal component, (c) second sample principal component, (d) second trimmed principal component, (e) third sample principal component, and (f) third trimmed principal component.

virtually identical, and so are Figures 2(e) and 2(d). This is a typical situation of *component reversal*, caused by only two outliers.

This example shows that outlying curves often have a more deleterious effect on the higher-order components than on the mean and the first principal component. It also explains why some outliers are hard to detect visually: higher-order components tend to be associated with subtle forms of variability that are hard to assess by visual inspection.

References

Olshen, R., Biden, E., Wyatt, M., and Sutherland, D. (1989), "Gait Analysis and the Bootstrap," *The Annals of Statistics*, 17, 1419–1440.

Ramsay, J. O., and Silverman, B. W. (2005), *Functional Data Analysis* (2nd ed.), Springer Series in Statistics, New York: Springer.