Robust functional principal component analysis

Juan Lucas Bali and Graciela Boente

Abstract When dealing with multivariate data, like classical PCA, robust PCA searches for directions with maximal dispersion of the data projected on it. Instead of using the variance as a measure of dispersion, a robust scale estimator s_n may be used in the maximization problem. In this paper, we review some of the proposed approaches to robust functional PCA including one which adapts the projection pursuit approach to the functional data setting.

1 Introduction

Functional data analysis provides modern analytical tools for data that are recoded as images or as a continuous phenomenon over a period of time. Because of the intrinsic nature of these data, they can be viewed as realizations of random functions often assumed to be in $L^2(\mathscr{I})$, with $\mathscr{I}$ a real interval or a finite dimensional Euclidean set.

Principal Components Analysis (PCA) is a standard technique used in the context of multivariate analysis as a dimension–reduction technique. The goal is to search for directions with maximal dispersion of the data projected on it. The classical estimators are obtained taking as dispersion the sample variance leading to estimators which are sensitive to atypical observations. To overcome this problem, Li and Chen (1985) proposed a procedure based on the principles of projection-pursuit to define the estimator of the first direction as

$$\widehat{\mathbf{a}} = \text{argmax}_{\mathbf{a}:\|\mathbf{a}\|=1} s_n(\mathbf{a}^{\mathrm{T}}\mathbf{x}_1, \cdots, \mathbf{a}^{\mathrm{T}}\mathbf{x}_n)$$

Juan Lucas Bali · Graciela Boente
Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires and CONICET, Argentina,
e-mail: lbali@dm.uba.ar e-mail: gboente@dm.uba.ar

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are i.i.d. $\mathbf{x}_i \in \mathbb{R}^p$ and s_n is a robust scale estimator. The subsequent eigenvectors are then obtained by imposing orthogonality conditions. When dealing with high dimensional data, the projection pursuit approach is preferable to the plug–in approach that estimates the principal components as the eigenvectors of a robust estimator of the covariance matrix. Effectively, as pointed out by Tyler (2010), when the dimension is larger than the sample size, the only affine equivariant multivariate location statistic is the sample mean vector and any affine equivariant scatter matrix must be proportional to the sample covariance matrix, with the proportionality constant not being dependent on the data. Hence, in that case, any affine equivariant scatter estimator looses its robustness, so that most commonly used robust scatter estimators should be avoided for high dimensional data and projection methods become useful. Croux and Ruiz-Gazen (2005) derived the influence functions of the projection–pursuit principal components, while their asymptotic distribution was studied in Cui *et al.* (2003). A maximization algorithm for obtaining $\widehat{\mathbf{a}}$ was proposed in Croux and Ruiz-Gazen (1996) and adapted for high dimensional data in Croux *et al.* (2007).

When dealing with functional data, an approach to functional principal component analysis (FPCA) is to consider the eigenvalues and eigenfunctions of the sample covariance operator. In a very general setting, Dauxois *et al.* (1982) studied their asymptotic properties. However, this approach may produce rough principal components and in some situations, smooth ones may be preferable. One argument in favour of smoothed principal components is that smoothing might reveal more interpretable and interesting features of the modes of variation for functional data. To provide smooth estimators, Boente and Fraiman (2000) considered a kernel approach by regularizing the trajectories. A different approach was proposed by Rice and Silverman (1991) and studied by Pezzulli and Silverman (1993). It consists on imposing an additive roughness penalty to the sample variance. On the other hand, Silverman (1996) considered estimators based on penalizing the norm rather than the sample variance. More recent work on estimation of the principal components and the covariance function includes Hall and Hosseini–Nasab (2006), Hall *et al.* (2006) and Yao and Lee (2006).

Not much work has been done in the area of robust functional data analysis. Of course, when $X \in L^2(\mathscr{I})$, it is always possible to reduce the functional problem to a multivariate one by evaluating the observations on a common output grid or by using the coefficients of a basis expansion, as in Locantore *et al.* (1999). However, as mentioned by Gervini (2008) discretizing the problem has several disadvantages which include the choice of the robust scatter estimators when the size of the grid is larger than the number of trajectories, as discussed above, the selection of the grid and the reconstruction of the functional estimators from the values over the grid. Besides, the theoretical properties of these procedures are not studied yet and they may produce an avoidable smoothing bias see, for instance, Zhang and Chen (2007). For this reason a fully functional approach to the problem is preferable. To avoid unnecessary smoothing steps, Gervini (2008) considered a functional version of the estimators defined in Locantore *et al.* (1999) and derived their consistency and influence function. Also, Gervini (2009) developed robust functional principal component estimators for sparsely and irregularly observed functional data and used it for outlier detection. Recently, Sawant *et al.* (2011) consider a robust approach of principal components based on a robust eigen–analysis of the coefficients of the observed data on some known basis. On the other hand, Hyndman and Ullah (2007) give an application of a robust projection-pursuit approach, applied to smoothed trajectories. Recently, Bali *et al.* (2011) considered robust estimators of the functional principal directions using a projection–pursuit approach that may include a penalization in the scale or in the norm and derived their consistency and qualitative robustness.

In this paper, we review some notions related with robust estimation for functional data. The paper is organized as follows, Section 2 states some preliminary concepts and notation that will be helpful along the paper. Section 3 states the principal component problem, Section 4 review the robust proposals previously studied while a real data example is given in Section 5.

2 Preliminaries and notation

Let us consider independent identically distributed random elements $X_1,\dots,X_n$ in a separable Hilbert space $\mathscr{H}$ (often $L^2(\mathscr{I})$) with inner product $\langle\cdot,\cdot\rangle$ and norm $\|u\| = \langle u,u\rangle^{1/2}$ and assume that $\mathbb{E}\|X_1\|^2 < \infty$. Denote by $\mu \in \mathscr{H}$ the mean of $X \sim X_1$, $\mu = \mathbb{E}(X)$ and by $\Gamma_X : \mathscr{H} \to \mathscr{H}$ the covariance operator of X. Let $\otimes$ stand for the tensor product on $\mathscr{H}$, e.g., for $u,v \in \mathscr{H}$, the operator $u \otimes v : \mathscr{H} \to \mathscr{H}$ is defined as $(u\otimes v)w = \langle v,w\rangle u$. With this notation, the covariance operator Γ_X can be written as $\Gamma_X = \mathbb{E}\{(X-\mu)\otimes(X-\mu)\}$, which is just the functional version of the variance–covariance matrix in the classical multivariate analysis. The operator Γ_X is linear, self-adjoint and continuous. Moreover, it is a Hilbert–Schmidt operator having a countable number of eigenvalues, all of them being real.

Let $\mathscr{F}$ denote the Hilbert space of Hilbert–Schmidt operators with inner product defined by $\langle \mathsf{H}_1,\mathsf{H}_2\rangle_{\mathscr{F}} = \text{trace}(\mathsf{H}_1\mathsf{H}_2) = \sum_{\ell=1}^{\infty}\langle \mathsf{H}_1 u_\ell, \mathsf{H}_2 u_\ell\rangle$ and norm $\|\mathsf{H}\|_{\mathscr{F}} = \langle \mathsf{H},\mathsf{H}\rangle_{\mathscr{F}}^{1/2} = \{\sum_{\ell=1}^{\infty}\|\mathsf{H}u_\ell\|^2\}^{1/2}$, where $\{u_\ell : \ell \geq 1\}$ is any orthonormal basis of $\mathscr{H}$, while H_1, H_2 and H are Hilbert-Schmidt operators, i.e., such that $\|\mathsf{H}\|_{\mathscr{F}} < \infty$. Choosing an orthonormal basis $\{\phi_\ell : \ell \geq 1\}$ of eigenfunctions of Γ_X related to the eigenvalues $\{\lambda_\ell : \ell \geq 1\}$ such that $\lambda_\ell \geq \lambda_{\ell+1}$, we get $\|\Gamma_X\|_{\mathscr{F}}^2 = \sum_{\ell=1}^{\infty}\lambda_\ell^2$.

The Karhunen-Loève expansion for the process leads to $X = \mu + \sum_{\ell=1}^{\infty}\lambda_\ell^{1/2} f_\ell\,\phi_\ell$, where the random variables $\{f_\ell : \ell \geq 1\}$ are the standardized coordinates of $X-\mu$ on the basis $\{\phi_\ell : \ell \geq 1\}$, that is, $\lambda_m^{1/2} f_m = \langle X-\mu, \phi_m\rangle$. Note that $\mathbb{E}(f_m) = 0$, while $\mathbb{E}(f_m^2) = 1$ if $\lambda_m \neq 0$, $\mathbb{E}(f_m\ f_s) = 0$ for $m \neq s$, since $\text{COV}(\langle u,X-\mu\rangle,\langle v,X-\mu\rangle) = \langle u,\Gamma_X v\rangle$. This expansion shows the importance of an accurate estimation of the principal components as a way to predict the observations and examine their atypicity.

3 The problem

As in multivariate analysis, there are two major approaches to develop robust estimators of the functional principal components. The first aims at developing robust estimates of the covariance operator, which will then generate robust FPCA procedures. The second approach aims directly at robust estimates of the principal direction bypassing a robust estimate of the covariance operator. They are based, respectively, on the following properties of the principal components

- **Property 1.** The principal component correspond to the eigenfunction of Γ_X related to the largest eigenvalues.
- **Property 2.** The first principal component maximizes $\text{var}(\langle\alpha, X\rangle)$ over $\mathscr{S} = \{\alpha : \|\alpha\| = 1\}$. The subsequent are obtained imposing orthogonality constraints to the first ones.

Let $X_1, \cdots, X_n$, $1 \le i \le n$, be independent observations from $X \in \mathscr{H}$, $X \sim P$ with mean μ and covariance operator Γ_X. An natural way to estimate the covariance operators Γ_X is to consider the empirical covariance operator given by $\widehat{\Gamma}_X = \sum_{j=1}^n \left(X_j - \overline{X}\right) \otimes \left(X_j - \overline{X}\right)/n$, where $\overline{X} = \sum_{j=1}^n X_j/n$. Dauxois *et al.* (1982) proved that $\sqrt{n}\left(\widehat{\Gamma}_X - \Gamma_X\right)$ converges in distribution to a zero mean gaussian random element U of $\mathscr{F}$. Besides, they derived the asymptotic behaviour of the eigenfunctions of the empirical covariance operator, leading to a complete study on the behaviour of the classical unsmoothed estimators of the principal components. As mentioned in the Introduction, smooth estimators of the covariance operators were studied in Boente and Fraiman (2000) where also the asymptotic behaviour of its eigenfunctions was obtained. This approach to principal components follows the lines established by **Property 1**.

As is well known, FPCA is a data analytical tool to describe the major modes of variation of the process as a way to understand it and also to predict each curve. Once we have estimators $\widehat{\phi}_\ell$ for the ℓ-th principal component, $1 \le \ell \le m$, one can predict each observation through $\widehat{X}_i = \overline{X} + \sum_{\ell=1}^m \widehat{\xi}_{i\ell}\, \widehat{\phi}_\ell$, where $\widehat{\xi}_{i\ell}$ are the scores of X_i in the basis of principal components, i.e., $\widehat{\xi}_{i\ell} = \langle X_i - \overline{X}, \widehat{\phi}_\ell\rangle$. In this sense, FPCA offers an effective way for dimension reduction.

However, FPCA based on the sample covariance operator is not robust. Hence, if one suspects that outliers may be present in the sample, robust estimators should be preferred. We recall that robust statistics seeks for reliable procedures when a small amount of atypical observations arise in the sample. In most cases, the estimators are functionals over the set of probability measures evaluated at the empirical probability measure and in this case, robustness is related to continuity of the functional with respect to the Prohorov distance.

In a functional setting influential observations may occur in several different ways. As mentioned by Locantore *et al.* (1999) they may correspond to atypical trajectories entirely outlying, that is, with extreme values for the L^2 norm, also to isolated points within otherwise typical trajectories (corresponding to a single extreme measurement) or they can be related to an extreme on some principal components, being the latter the more difficult to detect. In the functional case, these type of observations may have a significantly impact on the empirical covariance operator even if they may not be outlying in the sense of being faraway of their center. Detection of such observations is not easy and has been recently investigated by Sun and Genton (2011).

As an example for each type of influential observations, Figure 1 shows $n = 100$ trajectories generated using a finite Karhunen–Loève expansion, $X_i = Z_{i1}\phi_1 + Z_{i2}\phi_2 + Z_{i3}\phi_3$ where $\phi_1(x) = \sin(4\pi x)$, $\phi_2(x) = \cos(7\pi x)$ and $\phi_3(x) = \cos(15\pi x)$. The uncontaminated trajectories correspond to $Z_{ij} \sim N(0, \sigma_j^2)$ with $\sigma_1 = 4$, $\sigma_2 = 2$ and $\sigma_3 = 1$, Z_{ij} independent for all $1 \leq i \leq n$ and $1 \leq j \leq 3$. The atypical observations are plotted in thick lines and they correspond in each case to

a) add randomly to 10% the trajectories a factor of 12,
b) replace $X_2(t)$ by $X_2(t) + 25$ when $-0.4 < t < -0.36$
c) generate the random variables $Z_{i,j}$ as $Z_{i1} \sim N(0, \sigma_1^2)$,

$$\begin{pmatrix} Z_{i2} \\ Z_{i3} \end{pmatrix} \sim (1-\varepsilon)\, N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \mathrm{diag}\left(\sigma_2^2, \sigma_3^2\right)\right) + \varepsilon\, N\left(\begin{pmatrix} 4 \\ 4 \end{pmatrix}, \mathrm{diag}\left(0.01, 0.01\right)\right)$$

where $\varepsilon = 0.1$, leading in this case to 10 atypical observations, labelled 5, 7, 17, 32, 33, 40, 47, 69, 88 and 95.

It is clear that the influential observations can be clearly distinguished from the plots in cases a) and b) while they are more difficult to identify in c). The boxplot of the scores $s_{i,j} = \langle X_i - \mu, \phi_j \rangle$, for $1 \leq j \leq 3$ are provided in d), where the outliers in the boxplot correspond to the atypical observations. It is worth noting that the interdistance procedure described in Gervini (2010) only detects observation 33 as outlier and identify four of the uncontaminated trajectories, labelled 64, 71, 84 and 39, as atypical. However, in practice the practitioner cannot construct the scores $s_{i,j}$ and only scores from estimators of the principal directions can be used. For that reason, it is important to provide reliable estimators of the principal directions less sensitive to influential observations.

4 Robust proposals for FPCA

Recalling **Property 1** of the principal components, an approach to robust functional principal components is to consider the spectral value decomposition of a robust covariance or scatter operator. The spherical principal components, which were proposed by Locantore *et al.* (1999) and further developed by Gervini (2008), apply this approach using the spatial covariance operator defined as $\mathsf{V} = \mathbb{E}\left(Y \otimes Y\right)$, where $Y = (X-\eta)/\|X-\eta\|$ with η being the spatial median, defined in Gervini (2008), that is $\eta = \mathrm{argmin}_{\alpha \in \mathscr{H}} \mathbb{E}\left(\|X-\alpha\| - \|X\|\right)$. The estimators of the principal directions are then the eigenfunctions of the sample version of V, that is, $\widehat{\mathsf{V}} = \sum_{i=1}^n Y_i \otimes Y_i/n$, where $Y_i = (X_i - \widehat{\eta})/\|X_i - \widehat{\eta}\|$ and $\widehat{\eta} = \mathrm{argmin}_{\alpha \in \mathscr{H}} \sum_{i=1}^n \left(\|X_i - \alpha\| - \|X_i\|\right)/n$.

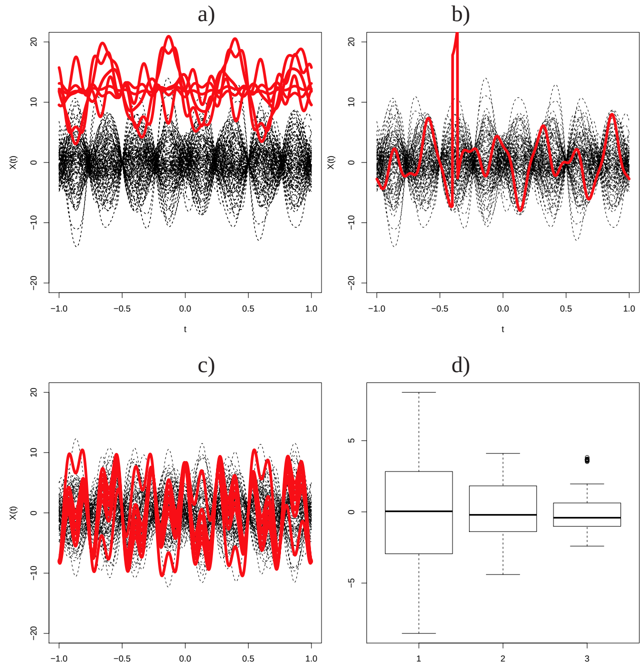

Fig. 1 Different influential trajectories with a) large values on the L^2 norm b) a extreme value over a small interval and c) extreme score on a principal component. Boxplot of the scores of the generated data c) over ϕ_j, $1 \le j \le 3$.

Gervini (2008) studied the properties of the eigenfunctions of $\widehat{\mathsf{V}}$ for functional data concentrated on an unknown finite–dimensional space. It is easy to see, that if $X = \mu + \sum_{\ell=1}^{\infty} \lambda_\ell^{1/2} f_\ell \phi_\ell$ and f_ℓ have a symmetric distribution which ensures that $\eta = \mu$, then, the functional spherical principal components estimate the true directions since V has the same eigenfuntions as γ. Indeed, $\mathsf{V} = \sum_{\ell \ge 1} \widetilde{\lambda}_\ell \phi_\ell \otimes \phi_\ell$ where $\widetilde{\lambda}_\ell = \lambda_\ell \mathbb{E}\left(f_\ell^2 (\sum_{s\ge 1} \lambda_s f_s^2)^{-1}\right)$.

From a different point of view, taking into account **Property 2**, Bali *et al.* (2011) considered a projection–pursuit approach combined with penalization to obtain robust estimators of the principal directions which provide robust alternatives to the estimators defined by Rice and Silverman (1991) and Silverman (1996).

To define these estimators, denote as $P[\alpha]$ for the distribution of $\langle \alpha, X \rangle$ when $X \sim P$. Given $\sigma_{\rm R}(F)$ a robust univariate scale functional, define $\sigma : \mathscr{H} \to \mathbb{R}$ as the map $\sigma(\alpha) = \sigma_{\rm R}(P[\alpha])$. Let $s_n^2 : \mathscr{H} \to \mathbb{R}$ be the empirical version of σ^2, that is, $s_n^2(\alpha) = \sigma_{\rm R}^2(P_n[\alpha])$, where $\sigma_{\rm R}(P_n[\alpha]\rangle)$ stands for the functional $\sigma_{\rm R}$ computed at the empirical distribution of $\langle \alpha, X_1 \rangle, \dots, \langle \alpha, X_n \rangle$.

Moreover, let us consider $\mathscr{H}_{\rm S}$, the subset of "smooth elements"of $\mathscr{H}$ and $D : \mathscr{H}_{\rm S} \to \mathscr{H}$ a linear operator, referred as the "differentiator". Using D, they define the symmetric positive semidefinite bilinear form $\lceil \cdot, \cdot \rceil : \mathscr{H}_{\rm S} \times \mathscr{H}_{\rm S} \to \mathbb{R}$, where $\lceil \alpha, \beta \rceil =$

$\langle D\alpha, D\beta\rangle$. The "penalization operator" is then defined as $\Psi : \mathscr{H}_{\mathrm{S}} \to \mathbb{R}$, $\Psi(\alpha) = \lceil \alpha, \alpha\rceil$, and the penalized inner product as $\langle \alpha,\beta\rangle_\tau = \langle \alpha,\beta\rangle + \tau\lceil\alpha,\beta\rceil$. Therefore, $\|\alpha\|_\tau^2 = \|\alpha\|^2 + \tau\Psi(\alpha)$. Besides, let $\{\delta_i\}_{i\geq 1}$ be a basis of $\mathscr{H}$ and denote $\mathscr{H}_{p_n}$ the linear space spanned by $\delta_1,\ldots,\delta_{p_n}$ and $\mathscr{S}_{p_n} = \{\alpha \in \mathscr{H}_{p_n} : \|\alpha\| = 1\}$.

The robust projection pursuit estimators are then defined as

$$\begin{cases} \widehat{\phi}_1 = {}_{\alpha\in\mathscr{H}_{p_n},\|\alpha\|_\tau=1}\left\{s_n^2(\alpha) - \rho\Psi(\alpha)\right\} \\ \widehat{\phi}_m = \operatorname{argmax}_{\alpha\in\widehat{\mathscr{B}}_{m,\tau}}\left\{s_n^2(\alpha) - \rho\Psi(\alpha)\right\} \quad 2\leq m, \end{cases} \tag{1}$$

where $\widehat{\mathscr{B}}_{m,\tau} = \{\alpha \in \mathscr{H}_{p_n} : \|\alpha\|_\tau = 1, \langle\alpha,\widehat{\phi}_j\rangle_\tau = 0\ ,\ \forall\, 1\leq j\leq m-1\}$. In the above definition, we understand that the products $\rho\Psi(\alpha)$ or $\tau\Psi(\alpha)$ are defined as 0 when $\rho = 0$ or $\tau = 0$ respectively, even when $\alpha \notin \mathscr{H}_{\mathrm{S}}$ for which case $\Psi(\alpha) = \infty$ and when $p_n = \infty$, $\mathscr{H}_{p_n} = \mathscr{H}$.

With this definition and by taking $p_n = \infty$, the robust raw estimators are obtained when $\rho = \tau = 0$, while the robust estimators penalizing the norm and scale correspond to $\rho = 0$ and $\tau = 0$, respectively. On the other hand, the basis expansion approach correspond a finite choice for p_n and $\tau = \rho = 0$.

Bali *et al.* (2011) derived the qualitative robustness of these estimators and show that they turn out to be consistent to the functional principal component directions defined as

$$\begin{cases} \phi_{\mathrm{R},1}(P) = \operatorname{argmax}_{\|\alpha\|=1}\sigma(\alpha) \\ \phi_{\mathrm{R},m}(P) = \operatorname{argmax}_{\|\alpha\|=1,\alpha\in\mathscr{B}_m}\sigma(\alpha), \quad 2\leq m\ , \end{cases}$$

where $\mathscr{B}_m = \{\alpha\in\mathscr{H} : \langle\alpha,\phi_{\mathrm{R},j}(P)\rangle = 0,\ 1\leq j\leq m-1\}$. To provide an explanation of what the directions $\phi_{\mathrm{R},m}(P)$ represent, assume that there exists here exists a constant $c > 0$ and a self–adjoint, positive semidefinite and compact operator Γ_0, such that for any $\alpha\in\mathscr{H}$, $\sigma^2(\alpha) = c\langle\alpha,\Gamma_0\alpha\rangle$. Moreover, denote by $\lambda_1 \geq \lambda_2 \geq \ldots$ are the eigenvalues of Γ_0 and by ϕ_j the eigenfunction of Γ_0 associated to λ_j. Assume that for some $q\geq 2$, and for all $1\leq j\leq q$, $\lambda_1 > \lambda_2 > \ldots > \lambda_q > \lambda_{q+1}$, then $\phi_{\mathrm{R},j}(P) = \phi_j$. Conditions that guarantee that $\sigma^2(\alpha) = c\langle\alpha,\Gamma_0\alpha\rangle$ when a robust scale is used are discussed in Bali *et al.* (2011) where also the results of an extensive simulation study showing the advantages of using robust procedures are reported.

As an example, we compute the robust projection–pursuit estimators for the generated data in Figure 1 c). The robust estimators correspond to an $M-$scale with score function the Tukey's function $\chi_c(y) = \min\left(3\,(y/c)^2 - 3\,(y/c)^4 + (y/c)^6, 1\right)$ with tuning constant $c = 1.56$ and breakdown point $1/2$. We have also computed the classical estimators which correspond to select σ_{R} as the standard deviation (SD). Figure 2 report the results corresponding to the raw estimators of each principal component. The solid line correspond to the true direction while the line with triangles to the estimators. From these plots we observe the sensitivity of the classical procedure to the influential observations introduced.

As detection rule, Figure 3 gives parallel boxplots of the scores $\widehat{s}_{i,j} = \langle X_i - \widehat{\mu}, \widehat{\phi}_j\rangle$ when $\widehat{\phi}_j$ are the classical and robust estimators. For the classical estimators, $\widehat{\mu} = \overline{X}$, while for the robust ones $\widehat{\mu} = \operatorname{argmin}_{\theta\in\mathscr{H}}\sum_{i=1}^n\left(\|X_i - \theta\| - \|X_i\|\right)/n$. Due to a masking effect, the boxplots of the scores over the classical estimators do not

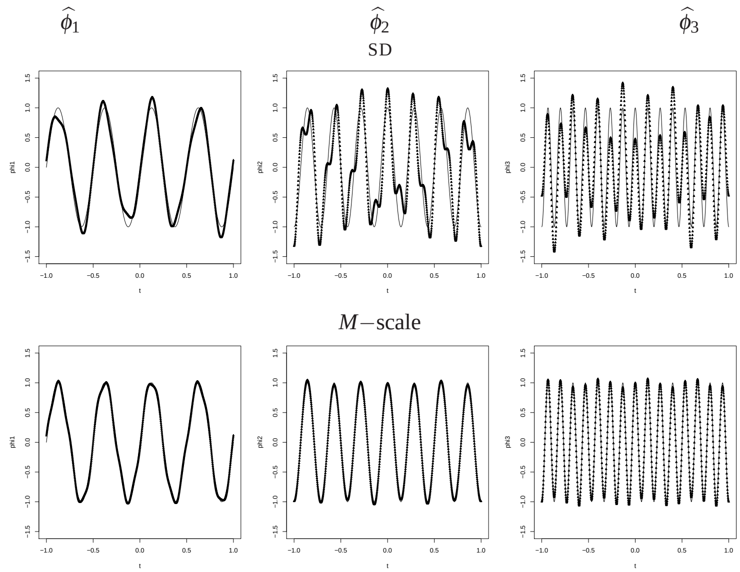

Fig. 2 Estimators of the principal directions for the generated data c). The solid line correspond to the true direction while the line with triangles to the estimators.

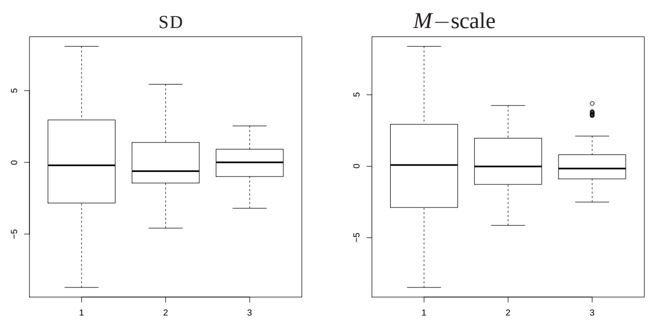

Fig. 3 Boxplots of the estimated scores $\langle X_i, \widehat{\phi}_j \rangle$ for the generated data c).

reveal any outlier. On the other hand, when using the robust projection–pursuit estimators the largest values of $\widehat{s}_{i,3}$ correspond the atypical observations generated. It is worth noting that the same conclusions are obtained if the plots of $\|\widehat{r}_i\| = \|X_i - \widehat{\mu} - \sum_{j=1}^{3} \widehat{s}_{i,j} \widehat{\phi}_j\|$ are considered (see Figure 4). The residual plot corresponding to the $M-$scale show clearly that the residual norm of the atypical observations are out of bound. On the other hand, when considering the eigenfunctions of the sample covariance operator, the observations with the largest residuals correspond to those labelled 19, 39, 40, 71 and 94, that is, only one of the atypical observations appears with a large residual, so leading to the wrong conclusions. Hence, robust procedures should be preferred.

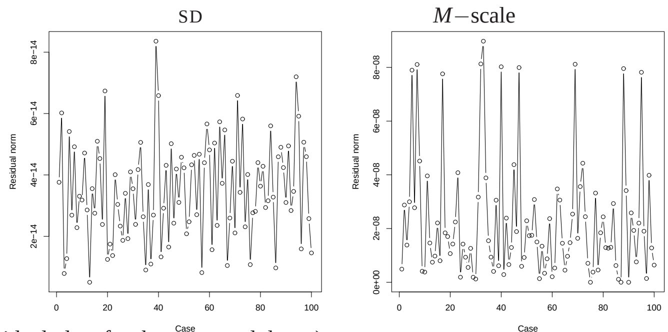

Fig. 4 Residual plots for the generated data c).

5 Lip data example

The following example was considered in Gervini (2008) to show the effect of outliers on the functional principal components. A subject was asked to say the word *bob* 32 times and the position of lower lip was recorded at each time point. Lip movement data was originally analyzed by Malfait and Ramsay (2003). In Figure 5, the plotted curves correspond to the 32 trajectories of the lower lip versus time. Three of these curves (plotted with thick lines on Figure 5 (a)) seem to be out of line, with delayed second peaks. To determine whether or not these curves are within the normal range of variability, it is necessary to estimate accurately the principal components.

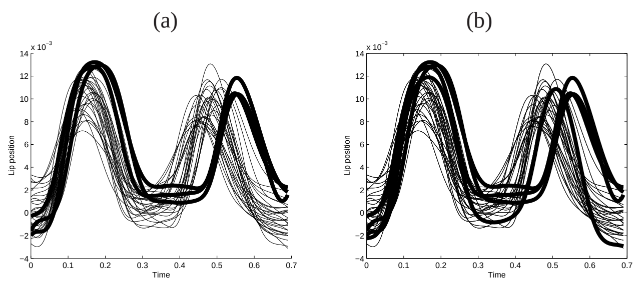

Fig. 5 Lip-movement data. Smoothed lower-lip trajectories of an individual pronouncing *bob* 32 times. (a) The trajectories 24,25 and 27 are indicated with thick lines (b) The trajectories 14, 24,25 and 27 are indicated with thick lines.

As in Gervini (2008), we have estimated 5 principal directions using the robust projection–pursuit estimators defined in (1) related to the $M-$scale with Tukey's score function. The robust and classical principal components are given in Figure 6 where the classical and robust raw estimators are plotted with a solid line and with a broken line, respectively. We refer to Gervini (2008) to understand the type of variability explained by these components. Besides, as described therein a positive component score will be associated with curves that show a large first peak and a delayed second peak, as those observed in the three atypical curves.

Figure 7 presents the parallel boxplots of the scores $\widehat{s}_{i,j} = \langle X_i - \widehat{\mu}, \widehat{\phi}_j \rangle$ when $\widehat{\phi}_j$ are the robust estimators together with the plot of the norm of the residuals $\|\widehat{r}_i^{(q)}\| = \|X_i - \widehat{\mu} - \sum_{j=1}^q \widehat{s}_{i,j} \widehat{\phi}_j\|$ where $\widehat{\mu} = \text{argmin}_{\theta \in \mathscr{H}} \sum_{i=1}^n (\|X_i - \theta\| - \|X_i\|)/n$. We only present the plots for the robust fit since we have already shown that when considering the classical one a masking effect may appear.

The residual plot corresponding to the $M-$scale shows clearly that the residual norm of the atypical observations are out of bound. Figure 7 also present the boxplots of $\|\widehat{r}_i\|$. Due to the skewness of the distribution of the norm, we have considered the adjusted boxplots (see Hubert and Vandervieren, 2008) instead of the usual ones. The two outliers appearing in the boxplot of the robust residuals $\|\widehat{r}_i^{(1)}\|$ and $\|\widehat{r}_i^{(5)}\|$ correspond to the observations 24 and 25. It is worth noticing that the trajectory labelled 14 also corresponds to the large negative outlier appearing in the scores $\widehat{s}_{i,4}$ while the observations 24, 25 and 27 appear as outliers with large negative scores $\widehat{s}_{i,2}$. Trajectory 14 is almost completely explained by the first four principal component, since the minimum and maximum of $\widehat{r}_i^{(4)}$ equal -1.084×10^{-18} and 9.758×10^{-19}, respectively. Figure 7 (d) gives the residual curves $\widehat{r}_i^{(4)}$ which do not suggest that a finite four–dimensional Karhunen–Loève representation suffices to explain the behaviour of the data while observation 14 may be explained by $\widehat{\phi}_1, \dots, \widehat{\phi}_4$ with the largest absolute scores on the first and fourth estimated component. Figure 5(b) indicates with thick lines the observations 14, 24, 25 and 27. From this plot, the curve related to observation 14 has a large first peak, a very smooth second peak while its fourth peak is clearly smaller and occurring before the majority of the data.

Acknowledgements This research was partially supported by Grants 276 from the Universidad de Buenos Aires, PIP 216 from CONICETand PICT 821 from ANPCYT at Buenos Aires, Argentina.

References

1. Bali, J. L., Boente, G., Tyler, D. E. and Wang, J. L. (2011). Robust Functional Principal Components: a projection-pursuit approach. *Annals of Statistics*, **39**, 2852-2882.
2. Boente, G. and Fraiman, R. (2000). Kernel-based functional principal components. *Statistics and Probabability Letters*, **48**, 335-345.
3. Croux, C., Filzmoser, P. and Oliveira M.R. (2007). Algorithms for Projection-Pursuit Robust Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **87**, 218-225.
4. Croux, C. and Ruiz–Gazen, A. (1996). A fast algorithm for robust principal components based on projection pursuit. In *Compstat: Proceedings in Computational Statistics*, ed. A. Prat, Heidelberg: Physica–Verlag, 211-217.

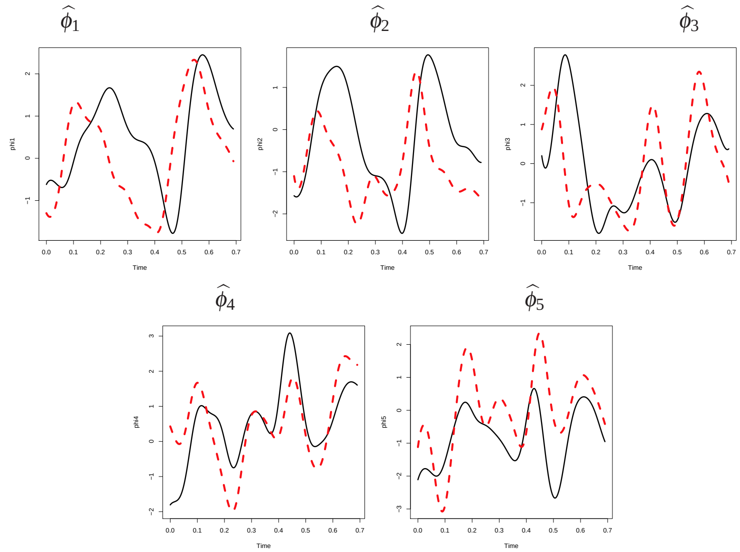

Fig. 6 Estimators of the principal directions for the lip movement data. The solid line correspond to the classical direction while the broken line to the robust ones.

5. Croux, C. and Ruiz–Gazen, A. (2005). High Breakdown Estimators for Principal Components: the Projection–Pursuit Approach Revisited. *Journal of Multivariate Analysis*, **95**, 206-226.
6. Cui, H., He, X. and Ng, K. W. (2003). Asymptotic Distribution of Principal Components Based on Robust Dispersions. *Biometrika*, **90**, 953-966.
7. Dauxois, J., Pousse, A. and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, **12**, 136-154.
8. Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.
9. Gervini, D. (2009) Detecting and handling outlying trajectories in irregularly sampled functional datasets. *Annals of Applied Statistics*, **3**, 1758-1775.
10. Gervini, D. (2010). tlier detection and trimmed estimation for general functional data. To appear in *Statistica Sinica*. doi:10.5705/ss.2010.282
11. Hall, P. and Hosseini–Nasab, M. (2006). On properties of functional principal components analysis. *Journal of the Royal Statistical Society, Series B*, **68**, 10926.
12. Hall, P., Müller, H.-G. and Wang, J.-L. (2006). Properties of principal component methods for functional and longitudinal data analysis. *Annals of Statistics*, **34**, 1493517.
13. Hubert, M. and Vandervieren, E. (2008). An adjusted boxplot for skewed distributions. *Computational Statistics and Data Analysis*, **52**, 5186-5201.
14. Hyndman, R. J. and S. Ullah (2007). Robust forecasting of mortality and fertility rates: A functional data approach. *Computational Statistics and Data Analysis*, **51**, 4942-4956.
15. Li, G. and Chen, Z. (1985). Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo. *Journal of the American Statistical Association*, **80**, 759-766.
16. Locantore, N., Marron, J. S., Simpson, D. G., Tripoli, N., Zhang, J. T. and Cohen, K. L. (1999). Robust principal components for functional data (with Discussion). *Test*, **8**, 1-73.
17. Pezzulli, S. D. and Silverman, B. W. (1993). Some properties of smoothed principal components analysis for functional data. *Computational Statistics*, **8**, 1-16.

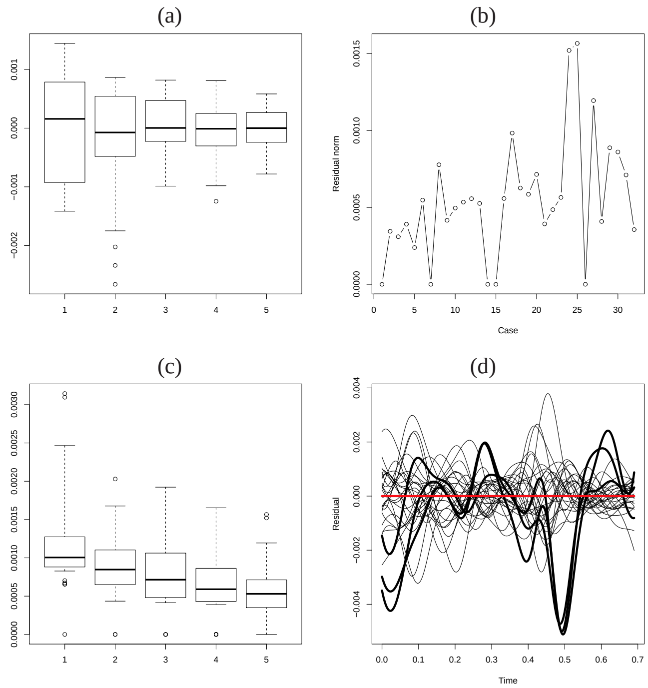

Fig. 7 Lip movement data (a) Boxplot of the scores $\widehat{s}_{i,j} = \langle X_i - \widehat{\mu}, \widehat{\phi}_j \rangle$, (b) Residual plots $\|\widehat{r}_i^{(5)}\|$, (c) Adjusted boxplots of $\|\widehat{r}_i^{(q)}\|$, $1 \leq q \leq 5$ and (d) Residuals plot $\widehat{r}_i^{(4)}$ based on a robust fit. The thick lines correspond to the observations 24, 25 and 27 while the red thick horizontal line to the trajectory 14.

18. Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. New York: Springer-Verlag.
19. Rice, J. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, **53**, 233-243.
20. Sawant, P., Billor, N. and Shin, H. (2011) Functional outlier detection with robust functional principal component analysis. *Computational Statistics*, **27**, 83-102.
21. Silverman, B. W. (1996). Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, **24**, 1-24.
22. Sun, Y. and Genton, M. G. (2011). Functional boxplots. *Journal of Computational and Graphical Statistics*, **20**, 316-334.
23. Tyler, D. (2010) A note on multivariate location and scatter statistics for sparse data sets. *Statistics and Probability Letters*, **80**, 1409-1413
24. Yao, F. and Lee, T. C. M. (2006). Penalized spline models for functional principal component analysis. *Journal of the Royal Statistical Society, Series B*, **68**, 325.
25. Zhang, J.-T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, **35**, 1052-1079.