A Comment on "Robust Principal Component Analysis for Functional Data"

Christophe Croux
Université Libre de Bruxelles *

1. Introduction

First of all I would like to congratulate the authors with their paper, which presents a very nice analysis of a functional data set from ophthalmology using robust principal components. In this comment, I will focus on the newly proposed method for robust PCA.

Several robustifications for PCA have been proposed in the past. The most simple idea is to compute eigenvalues and eigenvectors of a robust estimator of the covariance or correlation matrix of the data. Many simulation studies, starting with (Devlin et al. 1981), have been carried out to find out which robust estimator should be used, and recently some more theoretical results were obtained by Croux and Haesbroeck (1999). As was pointed out by the authors, these methods require that the number of variables d is higher than the number of observations n, making them less useful for functional data analysis.

Another approach to robustify PCA, based on projection pursuit (PP), has been considered by Li and Chen (1985). As one knows, a classical principal component is determined by the direction for which the projections of the data onto that direction have maximal standard deviation, under the constraint of orthogonality with all previously determined components. Instead of maximizing the standard deviation, one uses now a robust dispersion measure as "PP-index", resulting in a robust PCA. Since the principal components are computed sequentially, this approach can be used even in the high dimensional case $n < d$.

*ECARES, Université Libre de Bruxelles, CP-139, Av. F.D. Roosevelt 50, B-1050 Brussels, Belgium, ccroux@ulb.ac.be.

The method proposed in this paper has both a projection aspect and an eigenanalysis aspect. A important virtue of this method is it simplicity and ease of implementation. In contrast with many other highly robust multivariate statistical procedures, the required computation time is extremely limited.

2. Some Statistical Properties

For a sample $X = \{X_1, \ldots, X_n\} \subset I\!R^d$, the proposed robust PCA is carried out by computing the eigenvectors $v_1(X), \ldots, v_k(X)$ of the matrix

$$\tilde{\Sigma}_n(X) = \sum_{i=1}^{n} \frac{(X_i - \hat{\theta}_n)(X_i - \hat{\theta}_n)^t}{\|X_i - \hat{\theta}_n\| \, \|X_i - \hat{\theta}_n\|}, \tag{1}$$

with $k =$rank$(\tilde{\Sigma}_n(X))$ and $\hat{\theta}_n$ the L_1 location estimator. The "robust eigenvectors" $v_1(X), \ldots, v_k(X)$ are the vectors of interest since the data will be projected on them.

Equivariance Properties: Although $\tilde{\Sigma}_n$ is no *affine equivariant* covariance matrix estimator, it is orthogonal equivariant which suffices in the context of PCA. Indeed, denote $\alpha\Gamma X + b = \{\alpha\Gamma x_1 + b, \ldots, \alpha\Gamma x_n + b\}$ where Γ is an orthogonal matrix, b a vector in $I\!R^d$ and α a scalar, then the usual equivariance property holds

$$v_j(\alpha\Gamma X + b) = \Gamma v_j(X), \tag{2}$$

for $j = 1, \ldots, k$.

By first prescaling the data, for example by dividing them by the coordinatewise MAD, an equivalent of a correlation based PCA is obtained. This procedure is called elliptical PCA by the authors, and one has the additional equivariance property $v_j(DX) = v_j(X)$ for any diagonal matrix D.

Influence Function: The authors claim that outliers have bounded influence on their procedure. This can be made formal. To keep things simple, suppose that we are in the bivariate normal case, and due to (2) suppose w.l.o.g.

$$X_1, \ldots, X_n \overset{iid}{\sim} F = N(0, \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}), \qquad 0 < \gamma < 1.$$

The functional corresponding to the $\tilde{\Sigma}_n(X)$ is given by

$$\tilde{\Sigma}(G) = \int \frac{(y - T(G))(y - T(G))^t}{\|y - T(G)\|^2} dG(y)$$

for an arbitrary distribution G. Denote then $v_1(G), \ldots, v_d(G)$ the eigenvectors of $\tilde{\Sigma}(G)$. It is not difficult to show that $\tilde{\Sigma}(F) = N\left(0, \text{diag}(1/(\sqrt{\gamma}+1), 1/(\sqrt{\gamma^{-1}}+1)\right)$, implying Fisher consistency for the eigenvectors at bivariate normal distributions. Like in Critchley (1985), one can prove quite easily that the influence function for v_1 is given by

$$IF((x_1, x_2), v_1, F) = \frac{1-\gamma}{(1-\sqrt{\gamma})^2} \frac{x_1 x_2}{\|x\|^2} v_2(F) \tag{3}$$

and analogously for the second eigenvector. ¿From (3) boundedness of the influence function follows immediately.

Efficiency: Since emphasis in the paper was on the use of the proposed method as a tool for exploratory data analysis, efficiency considerations are less important but nevertheless interesting. Take once again the simple case where the data come from the bivariate normal distribution F. Assuming that the functional $\tilde{\Sigma}$ is sufficiently regular, the asymptotic variance of v_1 equals $\text{ASV}(v_1, F) = \frac{\sqrt{\gamma}}{2(1-\sqrt{\gamma})^2} v_2(F) v_2(F)^t$, which needs to be compared with the asymptotic variance of the classical estimator of the first eigenvector $\frac{\gamma}{(\gamma-1)^2} v_2(F) v_2(F)^t$. In Figure 1 the associated efficiency (defined as the ratio of the traces of the asymptotic variance matrices) is pictured as a function of γ.

The efficiency of the method depends thus on γ and never exceeds 50%: the more spherical the distribution, the higher the efficiency of the method. This is in contracts with most other methods for robust PCA, where the efficiencies are independent of γ. The same problem will arise for the elliptical version of the method.

3. Some suggestions

(i) It is not so obvious to interpret the eigenvalues of $\tilde{\Sigma}_n$. As a measure of dispersion of the data in the direction of $v_j(X)$, one could compute

$$\hat{\lambda}_j = S_n(v_j(X)^t X_1, \ldots, v_j(X)^t X_n),$$

for $j = 1, \ldots, k$, with S_n a robust univariate scale estimator like the MAD. Moreover, unlike the eigenvalues of $\tilde{\Sigma}_n$, the $\hat{\lambda}_j$ will be consistent estimators for the eigenvalues of covariance matrices of normal distributions.

(ii) A generalization of (1) is given by

$$\tilde{\Sigma}_n(X) = \sum_{i=1}^{n} w_i \frac{(X_i - \hat{\theta}_n)(X_i - \hat{\theta}_n)^t}{\|X_i - \hat{\theta}_n\| \; \|X_i - \hat{\theta}_n\|}$$

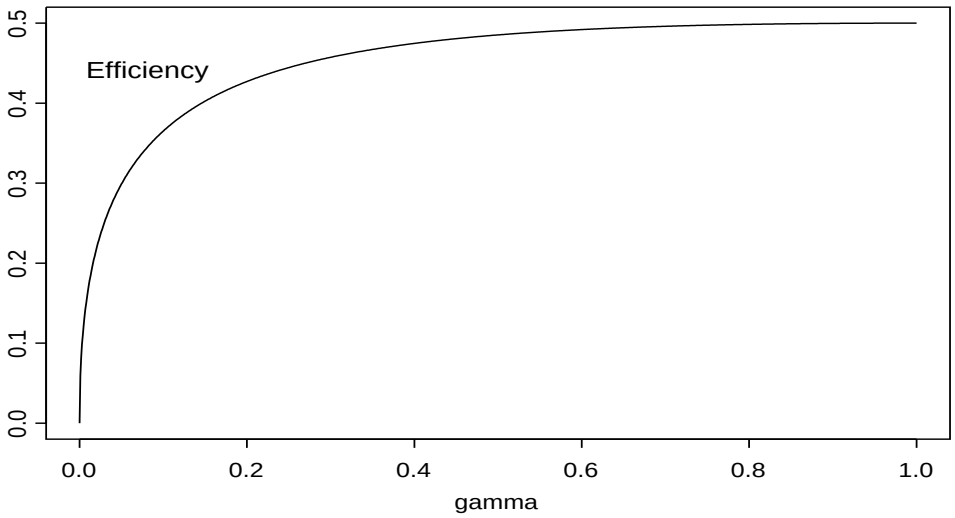

Figure 1: Efficiency of the proposed estimator for the first eigenvector of a bivariate normal distribution as a function of γ, where γ equals the second divided by the first population eigenvalue.

where the assigned weights w_i depends only on the rank of $\|X_i - \hat{\theta}_n\|$ and $\sum_{i=1}^{n} w_i = 1$. The location counterpart of the above estimator has been studied by Hössjer and Croux (1995). By choosing the weights properly, higher efficiencies can be obtained while not loosing too much robustness.

(iii) The choice of the starting value for the algorithm computing the L_1 estimator is not crucial, but the coordinatewise median might yield faster convergence in noisy data sets than the sample mean. Using the Newton steps of Bedall and Zimmerman (1979), the computation time of the L_1 estimator could be even further reduced.

4. References

Bedall, F.K., and Zimmermann, H (1979), "[Algorithm AS 143] The Mediancentre," *Applied Statistics*, 28, 325-328.

Critchley, F. (1985), "Influence in Principal Component Analysis", *Biometrika*, 72, 627–636.

Croux, C., and Haesbroeck, G. (199x), "Principal Component Analysis based on Robust Estimators of the Covariance or Correlation Matrix: Influence Functions and Efficiencies,"

Preprint, University of Brussels (ULB), http://www.sig.egss.ulg.ac.be/Haesbroeck/.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981), "Robust Estimation of Dispersion Matrices and Principal Components", *Journal of the American Statistical Association*, 76, 354–362.

Hössjer, O., and Croux, C. (1995), "Generalizing Univariate Signed Rank Statistics for Testing and Estimating a Multivariate Location Parameter," *Nonparametric Statistics,* 4, 293–308.

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.