

The influence function of the Stahel-Donoho covariance estimator of smallest outlyingness

M. Debruyne ^{a,*}, M. Hubert ^b,

^a*Department of Mathematics and Computer Science, Universiteit Antwerpen,
Campus Middelheim, Middelheimlaan 1G, B-2020 Antwerpen, Belgium*

^b*Department of Mathematics - LStat, K.U.Leuven, Celestijnenlaan 200B, B-3001
Leuven, Belgium*

Abstract

In traditional multivariate location and scatter estimation based on the Stahel-Donoho outlyingness, a weight function is applied, usually calibrated with respect to the multivariate Gaussian distribution. Other robust methods compute the covariance matrix of a fixed size subset of the data (e.g. the MCD estimator). In this paper we study a combination of both ideas. Location and scatter are estimated using a fixed size subset of the data containing the points with smallest Stahel-Donoho outlyingness. Local robustness and asymptotic relative efficiency are investigated.

Key words: Covariance estimation, robustness, influence function, efficiency.

1 Introduction

Traditional estimators of multivariate location and scatter are the sample mean and sample covariance matrix. They are optimal if the data follow a normal distribution, but they are also very sensitive to outliers. Many robust alternative have been proposed in the literature. The Stahel-Donoho ([Stahel, 1981](#); [Donoho, 1982](#)) estimator was one of the first introduced high-breakdown and affine equivariant estimators of multivariate location and scatter. It is based on the outlyingness of observations, which is obtained by projecting the observation on univariate directions. The original Stahel-Donoho estimator computes a weighted mean and covariance matrix, with weights inverse proportional to the outlyingness. In this paper we denote this approach as the

* Corresponding author. Tel: +32-(0)32-653887.

Email address: michiel.debruyne@ua.ac.be

‘weighted outlyingness’ Stahel-Donoho estimator (SD_{wo}). Theoretical properties of this estimator were obtained by [Gervini \(2002\)](#) and [Zuo et al. \(2004\)](#). A different approach is the Minimum Covariance Determinant (MCD) estimator ([Rousseeuw, 1984](#)). It is defined as the mean and covariance matrix of the subset of $(1 - \alpha)n$ observations (with $0 < \alpha < 1/2$) for which the determinant of the covariance matrix is smallest. A fast algorithm to find this subset was presented in [Rousseeuw and Van Driessen \(1999\)](#). Theoretical properties were investigated in [Croux and Haesbroeck \(1999\)](#). Consider a combination of both ideas: the mean and covariance matrix of the $(1 - \alpha)n$ observations with smallest Stahel-Donoho outlyingness, denoted as SD_{so} . This estimator was introduced in [Hubert et al. \(2005\)](#) as the first stage of the robust PCA algorithm ROBPCA. It was further used in robust regression and classification methods ([Hubert and Verboven, 2003](#); [Hubert and Vanden Branden, 2003](#); [Vanden Branden and Hubert, 2005](#)). Good empirical results were reported, e.g. in [Hubert and Engelen \(2004\)](#). In this article local robustness and efficiency of the SD_{so} estimator are studied. Section 2 contains notations and definitions. In Section 3 an expression for the influence function of SD_{so} is derived. In Section 4 gross error sensitivities and asymptotic relative efficiencies of SD_{so} are analyzed for several distributions and are compared to results for the SD_{wo} and MCD estimator. Section 5 contains some results concerning ROBPCA.

2 Definitions

2.1 Location and scatter functionals

Consider X a p -dimensional random variable with distribution F . Denote F^a the univariate distribution of $a^t X$, $a \in \mathbb{R}^p$. Let $m(\cdot)$ and $s(\cdot)$ be univariate estimators of location and scale. Following [Stahel \(1981\)](#) and [Donoho \(1982\)](#) the outlyingness $r(x; F)$ of a point $x \in \mathbb{R}^p$ is defined as

$$r(x; F) = \sup_{a \in \mathbb{R}^p} \left| \frac{a^t x - m(F^a)}{s(F^a)} \right|$$

Consider the region $A(F) = \{x \in \mathbb{R}^p : r^2(x, F) \leq qr_\alpha(F)\}$ with $qr_\alpha(F)$ the smallest value such that $P(r^2(x, F) \leq qr_\alpha(F)) = 1 - \alpha$. Then the location

estimator T_{so} and the scatter estimator V_{so} are defined respectively as:

$$\begin{aligned} T_{so}(F) &= \frac{1}{1-\alpha} \int_{A(F)} x dF(x) \\ V_{so}(F) &= \frac{c_\alpha}{1-\alpha} \int_{A(F)} (x - T_{so}(F))(x - T_{so}(F))^t dF(x) \end{aligned} \quad (1)$$

The factor c_α is chosen in order to ensure Fisher consistency at the specified model.

2.2 Elliptical distributions

To analyze the asymptotic properties of these functionals, we restrict ourselves to the class of p -dimensional elliptical distributions around μ . This means that the corresponding density is of the form

$$f(x) = |\Sigma|^{-1/2} g((x - \mu)^t \Sigma^{-1} (x - \mu))$$

with g a monotone decreasing function, $\mu \in \mathbb{R}^p$ the location parameter and the symmetric positive-definite matrix Σ the scatter parameter. Due to affine equivariance we can set $\mu = 0$ and $\Sigma = I$ (I being the $p \times p$ identity matrix) without loss of generality. The distribution F then becomes spherical, $F^a := F^1$ is identical for every a on the unit sphere in \mathbb{R}^p and satisfies $m(F^a) = 0$ and $s(F^a) = s_0 \in \mathbb{R}$. Denote q_α as the $(1 - \alpha)$ quantile of the distribution of $R^2 := X^t X$. An important elliptical model is the multivariate normal $\mathcal{N}_p(\mu, \Sigma)$ with

$$g(t) = \frac{e^{-t/2}}{(2\pi)^{p/2}}.$$

Then $R^2 \sim \chi_p^2$. In order to study the robustness against deviations from the Gaussian model, we also consider the class of multivariate Student distributions with ν degrees of freedom $\mathcal{T}_{p,\nu}(\mu, \Sigma)$, with density

$$g(t) = \frac{\Gamma(\frac{\nu+p}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{p/2}} \left(\frac{\nu}{\nu+t} \right)^{\frac{\nu+p}{2}}$$

Then $R^2/p \sim \mathcal{F}_{p,\nu}$. To obtain Fisher-consistency the factor c_α in (1) should equal

$$c_\alpha = (1 - \alpha) \left(\frac{\pi^{p/2}}{\Gamma(p/2 + 1)} \int_0^{\sqrt{q_\alpha}} r^{p+1} g(r^2) dr \right)^{-1}.$$

Then $V_{so}(F) = I$ for a spherical distribution F with density function g .

2.3 Influence functions

Let T and V be location and scatter functionals. Denote $F_{\epsilon,z} = (1 - \epsilon)F + \epsilon\Delta_z$ with Δ_z a Dirac distribution with all probability mass at $z \in \mathbb{R}^p$. Then the influence function ([Hampel et al., 1986](#)) of T at F is

$$IF(z; T, F) = \lim_{\epsilon \downarrow 0} \frac{T(F_{\epsilon,z}) - T(F)}{\epsilon}.$$

The definition of $IF(z; V, F)$ is analogous. For any affine equivariant T and V the influence functions at spherical distributions can be written in the following form ([Croux and Haesbroeck, 2000](#)) for some real functions w_μ, w_1, w_2 :

$$\begin{aligned} IF(z; T, F) &= w_\mu(\|z\|)z \\ IF(z; V, F) &= w_1(\|z\|)zz^t + w_2(\|z\|)I. \end{aligned}$$

Useful summaries of local robustness are the information standardized gross error sensitivities. For the location estimator this equals

$$\gamma_T^*(F) = \sup_{z \in \mathbb{R}^p} \left(IF(z; T, P)^t \mathcal{F}_1(F) IF(z; T, P) \right)^{1/2}$$

with \mathcal{F}_1 the Fisher information matrix of F for location $\mathcal{F}_1 = E(l^2(R^2)XX^t)$ and $l(t) = -2g'(t)/g(t)$. Denote $d_\mu = E(R^2w_\mu(R^2))/p$ and $d_\mu^* = E(R^2l(R^2))/p$. Then

$$\gamma_T^*(F) = \sup_{t \geq 0} |tw_\mu(t^2)|d_\mu^* \frac{1}{2}.$$

Not only robustness, but also the asymptotic relative efficiency with respect to the maximum likelihood estimator can be retrieved from the influence function under appropriate conditions. Then

$$ARE(T, F) = (d_\mu d_\mu^*)^{-1}.$$

For the scatter estimator we restrict ourselves to a shape component. This is a function S such that $S(\lambda\Sigma) = S(\Sigma)$ for all $\lambda > 0$, for example the eigenvectors of Σ . An advantage is that the asymptotic relative efficiency of $S(V) = p V/\text{tr}(V)$ is determined by a single scalar ([Gervini, 2002](#)):

$$\begin{aligned} ARE(S, F) &= (d_1 d_1^*)^{-1} \\ \gamma_S^*(F) &= \sup_{t \geq 0} |t^2 w_1(t^2)| \left(\frac{1}{2} \left(1 - \frac{1}{p} \right) d_1^* \right)^{1/2} \end{aligned}$$

with $d_1 = E(R^4 w_1(R^2))/(p(p+2))$ and $d_1^* = E(R^4 l(R^2))/(p(p+2))$. For the $\mathcal{N}_p(0, I)$ distribution, $d_1^* = d_\mu^* = 1$. For $\mathcal{T}_{p,\nu}(\mu, \Sigma)$, $d_1^* = d_\mu^* = (p+\nu)/(p+\nu+2)$. To summarize: just like in [Gervini \(2002\)](#) we will use γ_T^* , γ_S^* , $ARE(T)$ and $ARE(S)$ to analyze the estimators.

3 The influence function of SD_{so}

For the univariate estimators m and s we restrict ourselves to the class of M-estimators. From now on we assume that the pair (m, s) solves

$$\begin{aligned} E\psi\left(\frac{Z-m}{s}\right) &= 0 \\ E\rho\left(\frac{Z-m}{s}\right) &= K, \end{aligned} \quad (2)$$

with both ψ and ρ non-decreasing and bounded in $[0, \infty)$ and both $\psi(|u|)$ and $\rho'(|u|)$ strictly positive in a neighborhood of zero (except at zero). Moreover, ψ is odd, twice continuously differentiable and $\psi'(u)u$ and $\psi''(u)u^2$ are bounded. The function ρ is even, $\rho(0) < K < \rho(\infty)$ and ρ is twice continuously differentiable with $\rho'(u)u$ and $\rho''(u)u^2$ bounded.

Theorem 1 *With the notation and assumptions above, the influence functions at a spherical distribution F are given by*

$$\begin{aligned} IF(z; T_{so}, F) &= w_\mu(\|z\|)z \\ IF(z; V_{so}, F) &= w_1(\|z\|)zz^t + w_2(\|z\|)I \end{aligned}$$

where

$$\begin{aligned} w_\mu(\|z\|) &= \frac{1}{1-\alpha} \mathbf{1}(\|z\|^2 \leq q_\alpha) + \frac{g(q_\alpha)}{1-\alpha} \frac{d_0(\|z\|)}{\|z\|} \\ w_1(\|z\|) &= \frac{c_\alpha}{1-\alpha} \mathbf{1}(\|z\|^2 \leq q_\alpha) + \frac{c_\alpha}{1-\alpha} g(q_\alpha) \frac{d_1(\|z\|)}{\|z\|^2} \\ w_2(\|z\|) &= -1 + \frac{c_\alpha q_\alpha}{p} \left(1 - \frac{\mathbf{1}(\|z\|^2 \leq q_\alpha)}{1-\alpha}\right) - \frac{c_\alpha}{1-\alpha} g(q_\alpha) \left(\frac{q_\alpha d_2(\|z\|)}{p} - \frac{d_3(\|z\|)}{p-1}\right) \end{aligned}$$

with

$$\begin{aligned} d_0(t) &= \frac{4\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \frac{\sqrt{q_\alpha}}{s_0} \int_0^{\sqrt{q_\alpha}} \frac{\psi(x_1 t/s_0)}{E_F(\psi'(X_1/s_0))} x_1 (q_\alpha - x_1^2)^{\frac{p-3}{2}} dx_1 \\ d_1(t) &= \frac{4\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \frac{q_\alpha}{s_0^2} \int_0^{\sqrt{q_\alpha}} \frac{\rho(x_1 t/s_0) - K}{E_F(\rho'(X_1/s_0)X_1/s_0)} \frac{px_1^2 - q_\alpha}{p-1} (q_\alpha - x_1^2)^{\frac{p-3}{2}} dx_1 \\ d_2(t) &= \frac{4\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \frac{q_\alpha}{s_0^2} \int_0^{\sqrt{q_\alpha}} \frac{\rho(x_1 t/s_0) - K}{E_F(\rho'(X_1/s_0)X_1/s_0)} (q_\alpha - x_1^2)^{\frac{p-3}{2}} dx_1 \\ d_3(t) &= \frac{4\pi^{\frac{p-1}{2}}}{\Gamma(\frac{p-1}{2})} \frac{q_\alpha}{s_0^2} \int_0^{\sqrt{q_\alpha}} \frac{\rho(x_1 t/s_0) - K}{E_F(\rho'(X_1/s_0)X_1/s_0)} (q_\alpha - x_1^2)^{\frac{p-1}{2}} dx_1 \end{aligned}$$

For the proof of this theorem, we refer to the appendix. Note that Theorem 3 in [Gervini \(2002\)](#) provides analogous results for the SD_{so} estimator,

whereas [Croux and Haesbroeck \(1999\)](#) derived the influence function of the MCD estimator.

4 Gross error sensitivities and asymptotic relative efficiencies

As outlined in [Section 2.3](#) the expressions for w_μ and w_1 in [Theorem 1](#) can be used to find gross error sensitivities and asymptotic relative efficiencies of the location and shape estimators. We restrict ourselves to the shape component, since the results for the location estimator are very similar. In [\(2\)](#) we chose $\psi(u) = u/(1 + u^2)$, $\rho(u) = u^2/(1 + u^2)$ and $K = 1/2$. Then the resulting univariate M-estimators m and s attain the maximal asymptotic breakdown point of 50%. For the asymptotic efficiencies, we assume asymptotic normality and Fréchet differentiability. [Figure 1](#) gives an overview of the results. Part [\(a\)](#) depicts the ARE at Gaussian distributions in several dimensions, as a function of α . The ARE obviously decreases as α increases. Moreover it increases as the dimension increases. In very high dimensions, one can imagine that data generated from a spherical distribution lie on a sphere. Then every point is just as outlying as any other, and thus the region of smallest outlyingness is like a random subset of size $1 - \alpha$. One can indeed observe that the ARE converges to $1 - \alpha$ as p converges to infinity, both for MCD (dashed) as SD_{so} (solid). However, for smaller p the difference between both estimators is surprisingly high, especially for higher α . [Figure 1\(b\)](#) shows the gross error sensitivities at a Gaussian distribution as a function of α . For SD_{so} the GES is uniformly smaller than for MCD, with differences again being larger in smaller dimensions and with larger α . At Student distributions these conclusions stay the same no matter the degrees of freedom: SD_{so} outperforms MCD with respect to ARE ([Figure 1\(c\)](#)) and GES (part [\(d\)](#)).

For SD_{wo} different weight functions have been proposed in the literature, e.g Huber-type by [Maronna and Yohai \(1995\)](#) and exponential by [Zuo et al. \(2004\)](#). [Gervini \(2002\)](#) recommends a Gaussian weight function

$$w(r) = \frac{\phi(r/c)}{\phi(1)} \tag{3}$$

with cut-off $c = \sqrt{\chi_{p,1-\delta}^2}$ and $\delta = 0.1$. A comparison with SD_{so} and MCD is a bit hampered due to the presence of the parameters δ and α both with a different meaning. Note for instance that the breakdown value of SD_{wo} always attains the maximal value 0.5 irrespective of δ . To achieve this for SD_{so} as well, $\alpha = 0.5$ should be chosen, but this comes at a large cost in efficiency. However, in practice a reweighting step is often applied to improve efficiency. Denote (T_0, V_0) initial robust estimators of multivariate location and scatter.

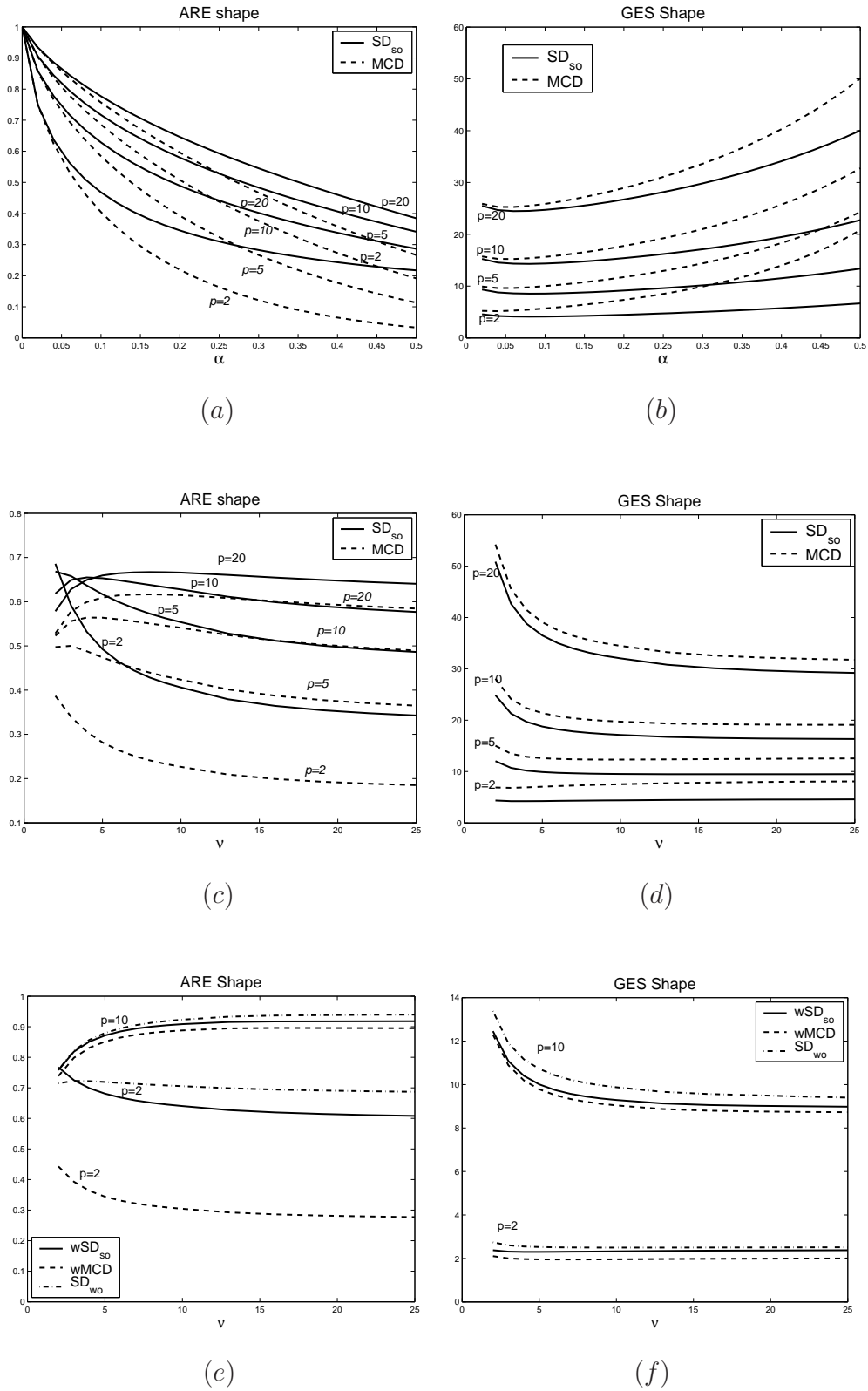


Fig. 1. Left: asymptotic relative efficiencies. Right: gross error sensitivities. Top: at $\mathcal{N}(0, I)$ as a function of α . Middle: at $\mathcal{T}_\nu(0, I)$ as a function of ν for $\alpha = 0.25$. Under: at $\mathcal{T}_\nu(0, I)$ as a function of ν for $\alpha = 0.5$ and one additional reweighting step.

Then one step reweighted estimators are defined as

$$T_1 = \frac{E(w(X)X)}{E(w(X))} \quad \text{and} \quad V_1 = c_1 \frac{E(w(X)(X - T_1)(X - T_1)^t)}{E(w(X))}$$

for some consistency factor c_1 and where the weights are computed from the initial estimators by $w(X) = w((X - T_0)^t \Sigma^{-1} (X - T_0))$ with $w : [0, \infty) \rightarrow \mathbb{R}$ a weight function. Expressions for the influence functions of T_1 and V_1 in terms of the influence functions of T_0 and V_0 are obtained in [Lopuhaä \(1999\)](#). Let us now compare the one-step reweighted MCD, one-step reweighted SD_{so} (both with $\alpha = 0.5$) and SD_{wo} all using the same weight function from (3). This is done in Figure 1(e) for the ARE's. The SD_{so} (solid) estimator keeps its bonus in efficiency with respect to the MCD estimator (dashed) after a reweighting step as well, especially in lower dimensions. In higher dimensions the difference decreases. For a larger number of degrees of freedom the SD_{wo} estimator is more efficient than the one step reweighted SD_{so} . For small ν both estimators perform almost equal.

Two additional remarks should be kept in mind in practice. First note that the maximum bias is not considered here. We already mentioned that the breakdown value of SD_{wo} does not depend on the cutoff c in (3). However, the maximum bias does. Choosing a small δ in (3) will of course increase the maximum bias. The MCD estimator with small α on the other hand is known to possess a small maximum bias. Second, for finite n the asymptotic results might not always be appropriate. The χ^2 distribution for example can be a pretty poor choice for the cut-off in (3). Therefore, an unweighted trimmed estimator such as MCD or SD_{so} can still be very useful, especially in high-dimensional applications such as PCA, when the assumption of large n is inappropriate.

5 Influence function of ROBPCA

The ROBPCA algorithm ([Hubert et al., 2005](#)) performs robust principal component analysis. The first step consists of determining the covariance estimator V_{so} . Then all data points are projected onto the subspace spanned by the first k eigenvectors of V_{so} . In this subspace, the MCD covariance estimator is applied. The eigenvalues and eigenvectors of this matrix provide the robust principal components. Let us introduce the following notation. Let F be a p -dimensional elliptical distribution with zero mean and covariance matrix $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d)$ with $\lambda_i \neq \lambda_j$ for every $i \neq j$. Let F_k be the k -dimensional distribution of the projection of F onto the first k eigenvectors of $V_{so}(F)$. These k eigenvectors are collected in the $k \times p$ matrix $P_k(F)$. Denote $V_{\text{robPCA}}(F)$ the MCD estimator of scatter of F_k . Its eigenvalues (in decreas-

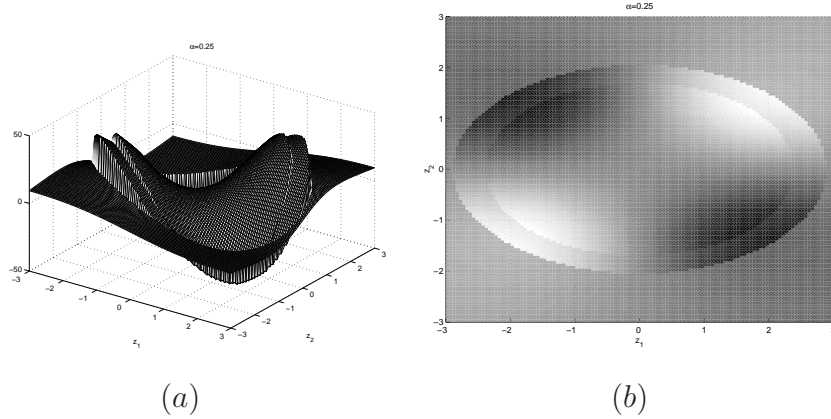


Fig. 2. $IF([z_1, z_2, 0]; v_{r,1}, F)_2$ at a three dimensional Gaussian distribution, with $\alpha = 0.25$. (a) Three dimensional view and (b) two dimensional view from above with grey scale indicating the height of the influence function.

ing order) and its corresponding eigenvectors are denoted as $(\lambda_{r,j}(F), v_{r,j}(F))$. With this notation, the following theorem holds.

Theorem 2 For $j = 1, \dots, k$, the influence functions of $\lambda_{r,j}(F)$ and $v_{r,j}(F)$ at the distribution F are given by

$$\begin{aligned}
 IF(z; \lambda_{r,j}, F) &= IF(P_k(F)z; MCD, F_k)_{jj}. \\
 IF(z; v_{r,j}, F)_i &= IF(P_k(F)z; v_{MCD,j}, F_k)_i - \frac{w_1(\|\Sigma^{-\frac{1}{2}}z\|)z_i z_j (\lambda_i + \lambda_j)}{(\lambda_j - \lambda_i)^2}, \quad i \neq j \\
 &= 0, \quad \text{if } i = j.
 \end{aligned}$$

Figure 2 shows an example with $d = 3$ and $k = 2$. The influence function of the second component of $v_{r,1}$ is plotted for all points in the plane $z_3 = 0$, for a Gaussian distribution with covariance matrix $\Sigma = \text{diag}(2, 1, 0.5)$. Part (a) of Figure 2 is the three dimensional view, whereas part (b) is the same plot seen from above, with a greyscale indicating the value of the influence function (large negative=black, around zero=grey, large positive=white). One clearly distinguishes two ellipses, which reflect the two estimators in play: the three-dimensional V_{so} estimator of which the first two eigenvectors provide a dimension reduction to a two-dimensional subspace, and then the two-dimensional MCD estimator applied in this subspace. Moreover the influence function is clearly bounded, with large outliers outside the ellipses having minimal effect.

Results about influence functions of other robust PCA methods can be found in [Croux and Haesbroeck \(2000\)](#) and [Cui et al. \(2003\)](#).

6 Conclusion

The Stahel-Donoho estimator of smallest outlyingness was defined. The influence function was obtained and local robustness and asymptotic relative efficiencies were analyzed. With the same α , SD_{so} consistently outperforms the MCD estimator. With $\alpha = 0.5$ and a one step reweighting this bonus remains and the values for SD_{so} approach, but do not beat, those for the classical SD_{wo} estimator.

The SD_{so} estimator is used as the first step of a robust PCA algorithm by [Hubert et al. \(2005\)](#). Influence functions for the corresponding eigenvalues and eigenvectors were calculated. More details about further extensions, e.g. towards PLS-regression ([Hubert and Vanden Branden, 2003](#)), can be found in [Debruyne \(2007\)](#).

7 Appendix

Proof of Theorem 1

Consider the contaminated distribution $F_{\epsilon,z} = (1 - \epsilon)F + \epsilon\Delta_z$. Then

$$\begin{aligned} V_{so}(F_{\epsilon,z}) &= c_\alpha \left(\frac{\int_{A(F_{\epsilon,z})} xx^t dF_{\epsilon,z}(x)}{1 - \alpha} - T_{so}(F_{\epsilon,z})T_{so}(F_{\epsilon,z})^t \right) \\ &= c_\alpha \left(\frac{1 - \epsilon}{1 - \alpha} \int_{A(F_{\epsilon,z})} xx^t dF(x) + \frac{\epsilon}{1 - \alpha} \mathbf{1}(z \in A(F_{\epsilon,z}))zz^t - T_{so}(F_{\epsilon,z})T_{so}(F_{\epsilon,z})^t \right). \end{aligned}$$

Deriving the latter expression to ϵ and evaluate at $\epsilon = 0$ yields

$$IF(z; V_{so}, F) = -I + \frac{c_\alpha}{1 - \alpha} \frac{\partial}{\partial \epsilon} \int_{A(F_{\epsilon,z})} xx^t dF(x) \Big|_{\epsilon=0} + \frac{c_\alpha}{1 - \alpha} \mathbf{1}(\|z\|^2 \leq q_\alpha)zz^t. \quad (4)$$

To further evaluate the second term, we assume z lying on the x_1 -axis. Then

$$\begin{aligned} &\frac{c_\alpha}{1 - \alpha} \frac{\partial}{\partial \epsilon} \int_{A(F_{\epsilon,z})} xx^t dF(x) \Big|_{\epsilon=0} \\ &= \frac{c_\alpha}{1 - \alpha} \frac{\partial}{\partial \epsilon} \int_{x_1(\epsilon)}^{x_2(\epsilon)} dx_1 \int_{C(\epsilon, x_1)} dx_2 \dots dx_p xx^t g(\|x\|^2) \Big|_{\epsilon=0} \end{aligned} \quad (5)$$

for certain $x_1(\epsilon)$ and $x_2(\epsilon)$. Because of symmetry, $C(F_{\epsilon,z}, x_1)$ is a $(p - 1)$ -dimensional ball. Denote its radius with $r = \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}$ and transform to polar coordinates as follows: $(x_2 \dots x_p)^t \longrightarrow re(\theta)$. Denote the Jacobian of this

transformation as $J(\theta, r)$. Then the right-hand size of (5) can be written as

$$\frac{c_\alpha}{1-\alpha} \int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \frac{\partial}{\partial \epsilon} \int_0^{\sqrt{q_\alpha(F_{\epsilon,z}, x_1)}} dr \times \int_{\Theta} d\theta J(\theta, r) \left(x_1 r e(\theta)^t \right)^t \left(x_1 r e(\theta)^t \right) g(x_1^2 + r^2)|_{\epsilon=0}$$

which by Leibniz' rule equals

$$\frac{c_\alpha}{1-\alpha} \int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \frac{\partial}{\partial \epsilon} \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}|_{\epsilon=0} \int_{\Theta} d\theta J(\theta, \sqrt{q_\alpha(F, x_1)}) \times \left(x_1 \sqrt{q_\alpha(F, x_1)} e(\theta)^t \right)^t \left(x_1 \sqrt{q_\alpha(F, x_1)} e(\theta)^t \right) g(x_1^2 + q_\alpha(F, x_1)). \quad (6)$$

We thus need to evaluate $\frac{\partial}{\partial \epsilon} \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}|_{\epsilon=0}$. By definition of $A(F_{\epsilon,z})$ we have $\int_{A(F_{\epsilon,z})} dF_{\epsilon,z}(x) = 1 - \alpha$. Deriving both sides of this equality we find after some calculations that

$$- \int_{A(F)} dF(x) + \mathbf{1}(z \in A(F)) + \int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \frac{\partial}{\partial \epsilon} \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}|_{\epsilon=0} \int_{\Theta} d\theta J(\theta, \sqrt{q_\alpha(F, x_1)}) g(q_\alpha) = 0. \quad (7)$$

Now take a point $x(\epsilon)$ on the boundary of $A(F_{\epsilon,z})$ with first coordinate x_1 . Then

$$q_\alpha(F_{\epsilon,z}) = r^2(x(\epsilon), F_{\epsilon,z}).$$

This yields

$$\begin{aligned} \frac{\partial}{\partial \epsilon} q_\alpha(F_{\epsilon,z})|_{\epsilon=0} &= \frac{\partial}{\partial \epsilon} r^2(x(0), F_{\epsilon,z})|_{\epsilon=0} + \frac{\partial}{\partial \epsilon} r^2(x(\epsilon), F_{0,z})|_{\epsilon=0} \\ &= \frac{\partial}{\partial \epsilon} r^2(x(0), F_{\epsilon,z})|_{\epsilon=0} + \frac{\partial}{\partial \epsilon} q_\alpha(F_{\epsilon,z}, x_1)|_{\epsilon=0} \end{aligned} \quad (8)$$

and so we find

$$\frac{\partial}{\partial \epsilon} \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}|_{\epsilon=0} = \frac{\frac{\partial}{\partial \epsilon} q_\alpha(F_{\epsilon,z})|_{\epsilon=0} - \frac{\partial}{\partial \epsilon} r^2(x(0), F_{\epsilon,z})|_{\epsilon=0}}{2\sqrt{q_\alpha(F, x_1)}}.$$

Substituting this in (7), we find an expression for $\frac{\partial}{\partial \epsilon} q_\alpha(F_{\epsilon,z})|_{\epsilon=0}$. Using equality (8), we find after some calculations that

$$\begin{aligned} \frac{\partial}{\partial \epsilon} q_\alpha(F_{\epsilon,z}, x_1)|_{\epsilon=0} &= \frac{1 - \alpha - \mathbf{1}(\|z\|^2 \leq q_\alpha)}{\frac{\int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \sqrt{q_\alpha(F, x_1)}^{p-3}}{2} \int_{\Theta} d\theta J_\theta(\theta) g(q_\alpha)} \\ &+ \frac{\int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \frac{\partial}{\partial \epsilon} r^2(x(0), F_{\epsilon,z})|_{\epsilon=0} \sqrt{q_\alpha(F, x_1)}^{p-3}}{\int_{-\sqrt{q_\alpha}}^{\sqrt{q_\alpha}} dx_1 \sqrt{q_\alpha(F, x_1)}^{p-3}} - \frac{\partial}{\partial \epsilon} r^2(x(0), F_{\epsilon,z})|_{\epsilon=0} \end{aligned} \quad (9)$$

from which $\frac{\partial}{\partial \epsilon} \sqrt{q_\alpha(F_{\epsilon,z}, x_1)}|_{\epsilon=0}$ can be computed. Plugging in this expression in (6) and simplifying the integrals (thereby using Theorem 2 in [Gervini \(2002\)](#)), finally gives us following expression for the second term in (4):

$$\begin{aligned} & \frac{c_\alpha}{1-\alpha} \frac{q_\alpha}{p} \left(1 - \alpha - \mathbf{1}(\|z\|^2 \leq q_\alpha)\right) I + \frac{c_\alpha}{1-\alpha} g(q_\alpha) \left(\frac{q_\alpha d_1}{2p}\right) I \\ & - \frac{c_\alpha}{1-\alpha} g(q_\alpha) \frac{d_2}{2(p-1)} I - \frac{c_\alpha}{1-\alpha} g(q_\alpha) \frac{d_3}{2z_1^2} \begin{pmatrix} z_1^2 & 0_{1,p-1} \\ 0_{p-1,1} & 0_{p-1,p-1} \end{pmatrix} \end{aligned}$$

The affine equivariance of V_{so} leads to the result in Theorem 1 for general z . For the location part, a completely similar proof can be given.

Proof of Theorem 2

Let us derive $\frac{\partial}{\partial \epsilon} V_{\text{robPCA}}(F_{\epsilon,z})|_{\epsilon=0}$. So first we project the contaminated distribution $F_{\epsilon,z}$ onto a k -dimensional subspace. The distribution of this projection is then given by

$$F_{\epsilon,z}^{\text{proj}} = (1 - \epsilon)F_\epsilon + \epsilon \Delta_{P_k(F_{\epsilon,z})z}$$

where F_ϵ is a k -dimensional elliptical distribution with covariance matrix $P_k(F_{\epsilon,z})\Sigma P_k(F_{\epsilon,z})^t$. Now $V_{\text{robPCA}}(F_{\epsilon,z})$ is nothing but the MCD estimator of covariance applied to $F_{\epsilon,z}^{\text{proj}}$ and thus one can write (with $V_r = V_{\text{robPCA}}$):

$$\begin{aligned} V_r(F_{\epsilon,z}) &= c_\alpha \left\{ \frac{1}{1-\alpha} \int_{B(F_{\epsilon,z})} xx^t dF_{\epsilon,z}(x) - T_r(\epsilon)T_r(\epsilon)^t \right\} \\ &= c_\alpha \left\{ \frac{1-\epsilon}{1-\alpha} \int_{B(F_{\epsilon,z})} xx^t dF_\epsilon(x) + \frac{\epsilon}{1-\alpha} \mathbf{1}(P_k(F_{\epsilon,z})z \in B(F_{\epsilon,z})) P_k(F_{\epsilon,z})zz^t P_k(F_{\epsilon,z})^t \right. \\ & \quad \left. - T_r(\epsilon)T_r(\epsilon)^t \right\} \end{aligned}$$

with $T_r(\epsilon)$ the MCD estimator of location of $F_{\epsilon,z}^{\text{proj}}$ and $B(F_{\epsilon,z}) = \{x \in \mathbb{R}^d : (x - T_r(\epsilon))^t V_r(F_{\epsilon,z})^{-1} (x - T_r(\epsilon)) \leq q_\alpha(\epsilon)\}$ for a certain positive number $q_\alpha(\epsilon)$. From the above expression, one can proceed similarly to the proof of Theorem 1 in [Croux and Haesbroeck \(1999\)](#). After some calculations one then finds

$$IF(z; V_r, F) = IF(P_k(F)z; MCD, F) + \frac{\partial}{\partial \epsilon} (P_k(F_{\epsilon,z})\Sigma P_k(F_{\epsilon,z}))|_{\epsilon=0} \quad (10)$$

Denote $\lambda_{r,j}(F)$ and $v_{r,j}(F)$ the j th eigenvalue and eigenvector. Then

$$\begin{aligned} IF(z; \lambda_{r,j}, F) &= IF(z; V_r, F)_{jj} \\ IF(z; v_{r,j}, F)_i &= \frac{IF(z; V_r, F)_{ij}}{\lambda_j - \lambda_i}, \end{aligned}$$

see for example ([Croux and Haesbroeck, 2000](#)). Combining (10) with Theorem 1 to calculate $\frac{\partial}{\partial \epsilon} (P_k(F_{\epsilon,z}))$ yields the result in Theorem 2.

References

- Croux, C. and Haesbroeck, G., 1999. Influence function and efficiency of the Minimum Covariance Determinant scatter matrix estimator. *J. Mult. An.*, 71, 161–190.
- Croux, C. and Haesbroeck, G., 2000. Principal components analysis based on robust estimators of the covariance or correlation matrix: influence functions and efficiencies. *Biometrika*, 87, 603–618.
- Cui, H., He, X. and Ng, K.W., 2003. Asymptotic distributions of principal components based on robust dispersions. *Biometrika*, 90, 953–966.
- Debruyne, M., 2007. Robustness of censored depth quantiles, PCA and kernel based regression, with new tools for model selection. PhD thesis, K.U.Leuven, Leuven, available at <http://hdl.handle.net/1979/933>
- Donoho, D.L., 1982. Breakdown properties of multivariate location estimators. Qualifying paper, Harvard University, Boston.
- Gervini, D., 2002. The influence function of the Stahel–Donoho estimator of multivariate location and scatter. *Stat. Prob. Letters*. 60, 425–435.
- Hampel, F.R., Ronchetti E.M., Rousseeuw, P.J. and Stahel, W.A., 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Hubert, M. and Engelen, S., 2004. Robust PCA and classification in bioinformatics. *Bioinformatics*, 20, 1728–1736.
- Hubert, M., Rousseeuw, P.J. and Vanden Branden, K., 2005. ROBPCA: a new approach to robust principal components analysis. *Technometrics*, 47:64–79.
- Hubert, M. and Vanden Branden, K., 2003. Robust methods for Partial Least Squares Regression. *J. Chemometr.*, 17, 537–549.
- Hubert, M. and Verboven, S., 2003. A robust PCR method for high-dimensional regressors. *J. Chemometr.*, 17, 438–452.
- Lopuhaä, H.P., 1999. Asymptotics of reweighted estimators of multivariate location and scatter. *Ann. Stat.*, 27, 1638–1665.
- Maronna, R.A. and Yohai, V.J., 1995. The behavior of the Stahel-Donoho robust multivariate estimator. *J. Am. Stat. Assoc.*, 90, 330–341.
- Rousseeuw, P.J., 1984. Least median of squares regression. *J. Am. Stat. Assoc.*, 79, 871–880.
- Rousseeuw, P.J. and Van Driessen, K., 1999. A fast algorithm for the Minimum Covariance Determinant estimator. *Technometrics*, 41, 212–223.
- Stahel, W.A., 1981. *Robuste Schätzungen: infinitesimale Optimalität und Schätzungen von Kovarianzmatrizen*. PhD thesis, ETH Zürich.
- Vanden Branden, K. and Hubert, M., 2005. Robust classification in high dimensions based on the SIMCA method. *Chem. Intell. Lab. Syst.*, 79, 10–21.
- Zuo, Y., Cui, H., He, X., 2004. On the Stahel-Donoho Estimator and Depth-weighted Means of Multivariate Data. *Ann. Stat.* 33, 167–188.