Generalized spherical principal component analysis

Generalized Spherical Principal Component Analysis

Sarah Leyder[1*], Jakob Raymaekers[1,2] and Tim Verdonck[1,3]

[1*]Department of Mathematics, University of Antwerp, Middelheimlaan 1, Antwerpen, 2020, Belgium.

[2]QE Econometrics, Quantitative Economics, School of Business and Economics, Maastricht University, Tongersestraat 53 , Maastricht, 6211, LM, The Netherlands.

[3*]IDLab, University of Antwerp - Imec, Middelheimlaan 1, Antwerpen, 2020, Belgium.

*Corresponding author(s). E-mail(s): sarah.leyder@uantwerpen.be;
Contributing authors: jakob.raymaekers@uantwerpen.be; tim.verdonck@uantwerpen.be;

Abstract

Outliers contaminating data sets are a challenge to statistical estimators. Even a small fraction of outlying observations can heavily influence most classical statistical methods. In this paper we propose generalized spherical principal component analysis, a new robust version of principal component analysis that is based on the generalized spatial sign covariance matrix. Theoretical properties of the proposed method including influence functions, breakdown values and asymptotic efficiencies are derived. These theoretical results are complemented with an extensive simulation study and two real-data examples. We illustrate that generalized spherical principal component analysis can combine great robustness with solid efficiency properties, in addition to a low computational cost.

Keywords: Principal component analysis, Robustness, Influence functions, Efficiency, Breakdown value

1 Introduction

A well-known and frequently used technique to analyze the structure of data sets is principal component analysis (PCA). The objective of this technique is usually to construct a new, smaller set of uncorrelated variables using linear combinations of the original variables. These new variables are obtained by preserving as much as possible of the variation present in the original data. Equivalently, they are obtained by projecting the original data on the PCA loading vectors of the data set, i.e. the directions in which the data has the greatest variability. PCA is a key building block in statistical data analysis and is widely used as a first step in clustering, discriminant analysis and regression.

In classical PCA (CPCA), the principal components can be calculated through a spectral decomposition of the covariance matrix. However, it is well-known that this matrix is very sensitive to outliers and potentially heavily influenced by anomalous observations. As a result, the directions of greatest variability are easily attracted towards these outliers, distorting the output of PCA. In order to avoid this, robust PCA methods have been developed which are resistant to such outlying observations. Many different approaches to robust principal component analysis exist, and we provide a brief overview here.

One approach is to use a spectral decomposition of a robust estimate of the covariance matrix. For affine equivariant covariance matrices, this approach was studied by Campbell

(1980) and Boente (1987) who used M-estimators, which unfortunately cannot withstand many outliers. Croux and Haesbroeck (2000) revisited this approach, suggesting instead to use high-breakdown estimators of location and scatter such as S-estimators (Davies, 1987; Rousseeuw and Yohai, 1984) or the MCD (Rousseeuw, 1984). In particular, they derive general expressions for the influence functions and efficiencies of the resulting eigenvector and eigenvalue estimates. A drawback of the approach is that many highly robust covariance estimators are computationally demanding. Additionally, some of them can only be computed when the number of samples is (substantially) larger than the dimension.

A second approach works incrementally by starting from the principal component corresponding with the largest projected robust variance. Each new principal component is then estimated as a maximizer of the projected robust variance conditional on being orthogonal to the already estimated components. This approach is detailed in Li and Chen (1985); Croux and Ruiz-Gazen (2005); Croux et al. (2007). The approach works well when a relatively small number of principal components is required. However, it can also be computationally demanding as the number of projections needed should increase rapidly with the dimension of the data to guarantee a stable performance. A combination of the two approaches mentioned above was used by Hubert et al. (2005) to develop the ROBPCA algorithm. ROBPCA often outperforms the projection-pursuit based methods as well as the covariance-based methods. It remains fairly slow to compute, especially on larger data sets.

To mitigate the computational burden of the previously mentioned approaches to robust principal component analysis, one elegant and popular approach for robust PCA is *spherical principal component analysis* (SPCA), introduced independently by Marden (1999) and Locantore et al. (1999). Spherical PCA starts by projecting the centered data onto a unit sphere before performing classical principal component analysis on this transformed data set. It was studied by, among others, Visuri et al. (2001); Taskinen et al. (2012); Croux et al. (2002). This procedure is equivalent to performing PCA on the *spatial sign covariance matrix* (SSCM)

$$S_{\text{SSCM}}(\boldsymbol{X}) = \mathbb{E}\left[\frac{(\boldsymbol{X}-\boldsymbol{\mu})(\boldsymbol{X}-\boldsymbol{\mu})^\intercal}{\|\boldsymbol{X}-\boldsymbol{\mu}\|^2}\right],$$

with $\boldsymbol{X}$ a p-variate random variable, $\boldsymbol{\mu}$ the location of the distribution of $\boldsymbol{X}$ and $\|.\|$ the Euclidean norm. Under mild assumptions on the underlying distribution, the SSCM is a Fisher consistent estimator of the eigenvectors and it preserves the order of the eigenvalues. The SSCM was studied in detail in Magyar and Tyler (2014); Dürre et al. (2014, 2016); Boente et al. (2019).

In Raymaekers and Rousseeuw (2019) a generalisation to the SSCM was introduced, namely the *generalized spatial sign covariance matrix* (GSSCM). They identified SSCM as a part of a larger class of orthogonally equivariant scatter estimates, namely the generalized spatial sign covariance matrices. Whereas in the SSCM all centered data vectors $\boldsymbol{x}_i$ are given the weight $1/\|\boldsymbol{x}_i\|$, the generalized SSCM assigns different weights, depending on the distribution of the random variable:

$$S_{g_{\boldsymbol{X}}}(\boldsymbol{X}) = \mathbb{E}[g_{\boldsymbol{X}}(\boldsymbol{X}-\boldsymbol{\mu})g_{\boldsymbol{X}}(\boldsymbol{X}-\boldsymbol{\mu})^\intercal] \quad (1)$$
$$\text{with} \quad g_{\boldsymbol{X}}(\boldsymbol{t}) = \boldsymbol{t}\ \xi_{\boldsymbol{X}}(\|\boldsymbol{t}\|),$$

where $\xi_{\boldsymbol{X}} : \mathbb{R}^+ \to \mathbb{R}^+$ is the *radial function*. This radial function plays the role of a weight function which assigns weights based on the Euclidean norm of the observation. In principle, any function can be used here as long as the resulting $S_{g_{\boldsymbol{X}}}(\boldsymbol{X})$ exists and can be computed on finite samples. In Section 2 we will elaborate on the choice of radial function. By using the Euclidean norm, the GSSCM becomes an orthogonally equivariant scatter estimator. In Raymaekers and Rousseeuw (2019), it is shown that the GSSCM inherits the consistency properties of the SSCM in that it is a Fisher consistent estimator of the eigenvectors for elliptical distributions and preserves the ranks of the eigenvalues under the same assumptions.

A new robust method for principal component analysis emerges when we combine the idea of spherical PCA with the GSSCM. Instead of computing the principal components from the SSCM, we can compute them from the generalized SSCM. We refer to this method as *generalized spherical principal component analysis* (GSPCA).

In this paper, we introduce and investigate the GSPCA method from a robustness perspective. The rest of the paper is organized as follows. Section 2 formally introduces GSPCA. Section 3 covers the theoretical properties of the new method, including breakdown values, influence functions and asymptotic variances. Section 4 presents an extensive simulation study comparing GSPCA with the state-of-the-art competitors. An illustration of GSPCA on two real datasets is presented in Section 5. Finally, Section 6 concludes.

2 Generalized spherical principal component analysis

The GSPCA method computes the principal component directions as the eigenvectors of the GSSCM given in Equation (1). For a finite, p-variate data set $\boldsymbol{X}$ containing $\{\boldsymbol{x}_1,...,\boldsymbol{x}_n\}$ the sample GSSCM becomes

$$\begin{aligned} &S_{g_{\boldsymbol{X}}}(\boldsymbol{X}) \\ =&\frac{1}{n}\sum_{i=1}^{n}\xi^2_{\boldsymbol{X}}(\|\boldsymbol{x}_i-T(\boldsymbol{X})\|)\ (\boldsymbol{x}_i-T(\boldsymbol{X}))(\boldsymbol{x}_i-T(\boldsymbol{X}))^\intercal \\ =&\frac{1}{n}\ g_{\boldsymbol{X}}(\boldsymbol{X})^\intercal g_{\boldsymbol{X}}(\boldsymbol{X}), \end{aligned}$$

where T is a (orthogonally equivariant) location estimator for the center of the data matrix $\boldsymbol{X} \in \mathbb{R}^{n\times p}$. In Raymaekers and Rousseeuw (2019), it is suggested to use the *k-step least trimmed squares (LTS) estimator*. This estimator starts from the spatial median, but adds additional iterative steps to improve robustness against outliers. The k-step LTS estimator has a breakdown value of $\lfloor (n+1)/2 \rfloor /n$ for any fixed value of k.

In this paper we will consider the following five radial functions suggested in Raymaekers and Rousseeuw (2019), apart from the evident radial functions $\xi(r) = 1$ and $\xi(r) = 1/r$, respectively corresponding to the classical covariance matrix and the SSCM.

1. Winsorizing (Winsor):

$$\xi_{\boldsymbol{X}}(r) = \begin{cases} 1 & \text{if } r \le Q_2 \\ Q_2/r & \text{if } Q_2 < r \end{cases} \tag{2}$$

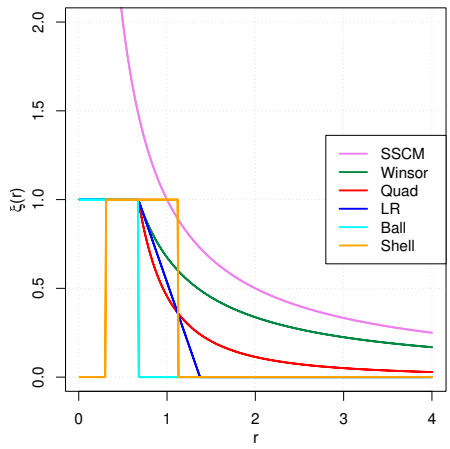


Fig. 1: The different radial functions introduced in Equations (2)-(6).

2. Quadratic Winsor (Quad):

$$\xi_{\boldsymbol{X}}(r) = \begin{cases} 1 & \text{if } r \le Q_2 \\ Q_2^2/r^2 & \text{if } Q_2 < r \end{cases} \tag{3}$$

3. Ball:

$$\xi_{\boldsymbol{X}}(r) = \begin{cases} 1 & \text{if } r \le Q_2 \\ 0 & \text{if } Q_2 < r \end{cases} \tag{4}$$

4. Shell

$$\xi_{\boldsymbol{X}}(r) = \begin{cases} 0 & \text{if } r < Q_1 \\ 1 & \text{if } Q_1 \le r \le Q_3 \\ 0 & \text{if } Q_3 < r \end{cases} \tag{5}$$

5. Linearly Redescending (LR):

$$\xi_{\boldsymbol{X}}(r) = \begin{cases} 1 & \text{if } r \le Q_2 \\ \frac{Q_3^*-r}{Q_3^*-Q_2} & \text{if } Q_2 < r \le Q_3^* \\ 0 & \text{if } Q_3^* < r \end{cases} \tag{6}$$

The cutoffs Q_1,Q_2,Q_3 and Q_3^* are robust estimates of the first, second and third quartile of the distribution of the Euclidean distances. We will come back to these in the next section. Figure 1 illustrates the above radial functions. It is clear that they all go to zero as their argument increases, which is what makes GSPCA more robust than the classical covariance matrix.

For a given data set $\boldsymbol{X}$ we can now use the spectral decomposition of the sample GSSCM to obtain the GSPCA loading vectors:

$$S_{g_{\boldsymbol{X}}}(\boldsymbol{X}) = \widehat{\boldsymbol{V}}_g \widehat{\boldsymbol{\Lambda}}_g \widehat{\boldsymbol{V}}_g^{\intercal}.$$

Here the matrix $\widehat{\boldsymbol{\Lambda}}_g$ is the diagonal matrix containing the eigenvalues $\widehat{\lambda}_{g,i}$ of $S_{g_{\boldsymbol{X}}}(\boldsymbol{X})$ in descending order ($\widehat{\lambda}_{g,1} > \cdots > \widehat{\lambda}_{g,p}$) and the columns of the matrix $\widehat{\boldsymbol{V}}_g$ consist of the corresponding eigenvectors $\widehat{\boldsymbol{v}}_{g,i}$. The GSPCA loading vectors then coincide with these eigenvectors and the principal components correspond to the data projected onto these vectors.

Alternatively we can also state that the principal components are the uncorrelated vectors maximizing the variance of the transformed data. In particular, the i-th loading vector $\widehat{\boldsymbol{v}}_{g,i}$ can be defined as

$$\widehat{\boldsymbol{v}}_{g,i} = \underset{\boldsymbol{a}\in A}{\operatorname{argmax}}\{\operatorname{var}(g_{\boldsymbol{X}}(\boldsymbol{X})\boldsymbol{a})\},$$
$$\text{where } A = \{\boldsymbol{a} \in \mathbb{R}^p \mid \|\boldsymbol{a}\|_2 = 1 \text{ and } \boldsymbol{a}^{\intercal}\widehat{\boldsymbol{v}}_{g,j} = 0 \text{ for all } j = 1,\ldots,i-1\}.$$

It should be clear from the definition that GSPCA generally has a fairly low computational cost. The only uncertain factor is the computation and evaluation of the radial functions g. In the cases considered above however, it requires the computation of the Euclidean norms of the observations followed by robust estimates of the quantiles of these norms. The first can be done in $\mathcal{O}(np)$ time, and the latter in $\mathcal{O}(n)$. Therefore, GSPCA has a computational complexity given by $\mathcal{O}(n+np+np^2+p^3) = \mathcal{O}(np^2+p^3)$, the same as classical PCA.

Additionally, like CPCA, GSPCA has the property that it can be computed through the singular value decomposition (SVD) as well as the spectral composition of the GSSCM. In order to use the SVD for GSPCA, we first need to transform the observations using the function $g_{\boldsymbol{X}}$ of Equation (1) after which we can apply SVD to the transformed data $g_{\boldsymbol{X}}(\boldsymbol{X})$. More precisely, we first compute the singular value decomposition

$$g_{\boldsymbol{X}}(\boldsymbol{X}) = \widehat{\boldsymbol{U}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{V}}^{\intercal},$$

where the orthogonal matrices $\widehat{\boldsymbol{U}}$ and $\widehat{\boldsymbol{V}}$ respectively contain the left and right singular vectors of $g_{\boldsymbol{X}}(\boldsymbol{X})$ and the diagonal matrix $\widehat{\boldsymbol{D}}$ contains the corresponding singular values $\widehat{\delta}_{g,i}$. Next it can easily be shown that the columns $\widehat{\boldsymbol{v}}_{g,i}$ of $\widehat{\boldsymbol{V}}$ correspond to the GSPCA loading vectors and the values $\widehat{\delta}^2_{g,i}/n$ correspond to the GSPCA eigenvalues:

$$\begin{aligned} S_{g_{\boldsymbol{X}}}(\boldsymbol{X}) &= \frac{1}{n}\, g_{\boldsymbol{X}}(\boldsymbol{X})^{\intercal} g_{\boldsymbol{X}}(\boldsymbol{X}) = \frac{1}{n}\, \widehat{\boldsymbol{V}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{U}}^{\intercal}\widehat{\boldsymbol{U}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{V}}^{\intercal} \\ &= \frac{1}{n}\, \widehat{\boldsymbol{V}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{D}}\widehat{\boldsymbol{V}}^{\intercal} \\ &= \widehat{\boldsymbol{V}}\operatorname{diag}(\widehat{\delta}^2_{g,1}/n,\ldots,\widehat{\delta}^2_{g,p}/n)\widehat{\boldsymbol{V}}^{\intercal} = \widehat{\boldsymbol{V}}_g \widehat{\boldsymbol{\Lambda}}_g \widehat{\boldsymbol{V}}_g^{\intercal}. \end{aligned}$$

The SVD procedure has the advantage that the computational cost of GSPCA can be lowered if the required number of principal components is known in advance. In this scenario we can use truncated singular value decomposition on the GSSCM to calculate only the first k singular vectors, which correspond with the first k eigenvectors of the matrix $S_{g_{\boldsymbol{X}}}(\boldsymbol{X})$. This not only avoids the calculation of the remaining principal components, but also does not require the explicit construction of the $p\times p$ covariance estimate $S_{g_{\boldsymbol{X}}}(\boldsymbol{X})$.

3 Theoretical properties

In this section we study the robustness and efficiency of GSPCA. To evaluate robustness, we derive the breakdown value and influence function of GSPCA. We additionally study efficiencies and potential corrections for the eigenvalues.

We emphasize that the results of this section assume the dimension p to be fixed. A word of caution is in place when we consider GSPCA on (ultra-) high-dimensional data with $\frac{p}{n} \not\to 0$. It is known that classical PCA and related methods can behave very unexpectedly in such settings (Bickel et al., 2018; Johnstone and Paul, 2018; Pires and Branco, 2019) and GSPCA is no different. In fact, to a certain extent the behavior of GSPCA resembles that of classical PCA as the dimension grows for the following reason. For many distributions, we have $\|\boldsymbol{X}\|_2^2 = \mathcal{O}_{\Pr}(p)$, causing most radial functions in the GSSCM to give roughly equal weight to all data points as $p \to \infty$, which is exactly what the classical covariance matrix does.

The most commonly discussed issue of PCA in high-dimensions is that without further assumptions, the parameters can no longer be estimated consistently. This means that classical estimators of eigenvectors, eigenvalues, and functions of these quantities tend to be inconsistent (Baik et al., 2005; Baik and Silverstein, 2006; Paul, 2007; Johnstone et al., 2009; Cai et al., 2015).

A second issue in high-dimensional PCA is the attainability of equivariance properties. Whereas in classical statistics, orthogonal equivariance is a natural property of PCA estimators, in high-dimensional statistics this may no longer be relevant as it cannot be reconciled with sparsity assumptions.

When one is concerned about consistent estimation, GSPCA should thus primarily be considered when a regime with $\frac{p}{n} = o(1)$ is the most natural one. In case another regime is more natural, one should proceed with caution. One approach is to make sparsity assumptions on the estimands to guarantee the possibility of consistent estimation. GSPCA can easily be made sparse by multiplying the centered observations with their radial functions before applying any sparsity-inducing PCA method. Another approach is to use it as an exploratory tool or a method for detecting certain types of anomalies. As we will see in our simulation studies and real-data examples, there are still situations with $p \geq n$ where GSPCA can provide insight even though the asymptotic properties may be subpar.

3.1 Breakdown value

The breakdown value is a global measure for robustness defined by Hampel based on an idea of Hodges (Hampel et al., 1986). For a given estimator, it is defined as the smallest fraction of observations in the data set that needs to be changed to carry the estimate arbitrarily far. More specifically, for a scatter estimator S, it is defined as the minimal amount of contamination required to make the largest eigenvalue λ_1 arbitrarily large (explosion) or the smallest eigenvalue λ_p arbitrarily close to zero (implosion):

$$\varepsilon(S,\boldsymbol{X}) = \min\left\{\frac{m}{n} : \sup_{\boldsymbol{X}^*_m} \max[\lambda_1(S(\boldsymbol{X}^*_m)),\lambda_p^{-1}(S(\boldsymbol{X}^*_m))] = \infty\right\},$$

with $\boldsymbol{X}^*_m$ the data set $\boldsymbol{X}$ where m observations have been replaced.

While in many applications, implosion of the covariance matrix is undesirable, in the context of PCA, it is the explosion breakdown that is more relevant. After all, PCA is often most useful when there are indeed directions with (nearly) zero variance which allows for effective dimension reduction. In such a case, the smallest eigenvalue(s) would be (close to) zero and would thus have "breakdown" if we were to use the definition above. In Raymaekers and Rousseeuw (2019) it is shown that the breakdown value of the GSSCM is $\lfloor(n-p+1)/2\rfloor/n$. This seems unsatisfactory, since it is in contrast with the SSCM which has a higher breakdown value, namely $\lfloor(n+1)/2\rfloor/n$. Fortunately, the lower breakdown value of the GSSCM is due to the implosion breakdown value of the "hard redescending" radial functions such as Ball, Shell and LR, in combination with the estimation of the cutoffs Q_1,Q_2,Q_3 and Q_3^* used in these radial functions.

If we are no longer concerned with the implosion breakdown value, we can take a slightly different approach. Denote $d_i := \|\boldsymbol{x}_i - T(\boldsymbol{X})\|$. We estimate the cutoffs using the estimators

$$\begin{aligned}
Q_1 &= \left[\text{med}_i\left(d_i^{2/3}\right) - \text{MAD}_i\left(d_i^{2/3}\right)\right]^{3/2},\\
Q_2 &= \text{med}_i(d_i),\\
Q_3 &= \left[\text{med}_i\left(d_i^{2/3}\right) + \text{MAD}_i\left(d_i^{2/3}\right)\right]^{3/2},\\
Q_3^* &= \left[\text{med}_i\left(d_i^{2/3}\right) + 1.4826\cdot\text{MAD}_i\left(d_i^{2/3}\right)\right]^{3/2}.
\end{aligned}$$

In these definitions med and MAD are the median and the median absolute deviation. Unlike in Raymaekers and Rousseeuw (2019), these estimators for the cutoffs no longer depend on the order statistic $h = \lfloor(n+p+1)/2\rfloor$, which was required to prove the implosion breakdown value of the GSSCM. The dependence on the h-th order statistic which depends on p is undesirable in the context of PCA.

The theorem below states that for our adapted GSSCM, with cutoffs based on the median and MAD, we obtain an explosion breakdown value independent of p, equal to that of the SSCM. The proof can be found in the Supplementary material.

Theorem 1 (Explosion breakdown value) *Given* $\boldsymbol{X} = \{\boldsymbol{x}_1,...,\boldsymbol{x}_n\}$ *an* $n \times p$ *dimensional data set and a*

location estimator $T(\boldsymbol{X})$ with a breakdown value of at least $\lfloor(n+1)/2\rfloor/n$. Suppose that

1. *The radial function takes values in* [0,1].
2. $\forall$ *data sets* $\boldsymbol{X}: \#\{i: \xi(d_i)=1\} \geq \lfloor(n+1)/2\rfloor$.
3. $\forall \boldsymbol{t}: \ \|g(\boldsymbol{t})\| = \|\boldsymbol{t}\|\xi(\|\boldsymbol{t}\|) \leq med_i(d_i)+1.4826\cdot MAD_i(d_i)$.

Then the explosion breakdown value ε of the GSSCM is $\lfloor(n+1)/2\rfloor/n$.

Note that the conditions in the theorem are all satisfied for the radial functions in Equations (2) – (6). In addition, the k-step LTS estimator has a breakdown value of $\lfloor(n+1)/2\rfloor/n$. Hence, we conclude that GSPCA based on the GSSCM is robust up to $\frac{\lfloor(n+1)/2\rfloor}{n} \approx 50\%$ contamination.

3.2 Influence functions of the loading vectors

We now consider the influence functions relevant for GSPCA. The influence function for a statistical functional $\mathcal{T}$ at distribution F is defined as (see Hampel et al. (1986)):

$$\begin{aligned} \mathrm{IF}(x,\mathcal{T},F) &= \lim_{\varepsilon\to 0}\frac{\mathcal{T}(F_{\varepsilon,x})-\mathcal{T}(F)}{\varepsilon} \\ &= \frac{\partial}{\partial\varepsilon}\mathcal{T}(F_{\varepsilon,x})\Big|_{\varepsilon=0}, \end{aligned} \tag{7}$$

with $F_{\varepsilon,x} = (1-\varepsilon)F+\varepsilon\Delta_x$ the distribution contaminated by x where Δ_x is the distribution putting all its mass in x.

In contrast to the breakdown value, the influence function is a local measure of robustness instead of a global one. Informally, it can be interpreted as a measure for the effect that an infinitesimal small amount of contamination has on the functional. Therefore it is a complementary measure and describes a different aspect of our method.

In the context of PCA, the relevant functionals $\mathcal{T}$ in Equation (7) are the eigenvector and eigenvalue functionals, which we denote by $\mathcal{V}_{g,j}$ and $\mathcal{L}_{g,j}$ for $j = 1,...,p$. In this section specifically, we will study the eigenvector functionals, for which the relevant influence function is given by

$$\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,j},F) = \lim_{\varepsilon\to 0}\frac{\mathcal{V}_{g,j}(F_{\varepsilon,\boldsymbol{x}})-\mathcal{V}_{g,j}(F)}{\varepsilon}.$$

In what follows, we assume that F is a distribution with the center at the origin and covariance matrix $\boldsymbol{\Sigma}$ which has distinct eigenvalues. Furthermore, we denote the spectral decomposition of $\boldsymbol{\Sigma}$ by $\boldsymbol{\Sigma} = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\intercal$ where $\boldsymbol{V}$ is a matrix with the eigenvectors $\boldsymbol{v}_i$ in its columns and $\boldsymbol{\Lambda}$ is a diagonal matrix with the eigenvalues $\lambda_1 > ... > \lambda_p$ on its diagonal.

Note that the functionals $\mathcal{V}_{g,j}$ inherit the Fisher consistency properties of the GSSCM in case of elliptically symmetric distributions, see Raymaekers and Rousseeuw (2019). We say that a distribution $F_{\boldsymbol{X}}$ has an elliptically symmetric density if the density can be written as

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = det(\boldsymbol{\Sigma})^{-1/2}\cdot h((\boldsymbol{x}-\boldsymbol{\mu})^\intercal\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})),$$

where h is a positive decreasing function acting on the whitened observations. For such distributions, we thus have that $\mathcal{V}_{g,j}(F) = \boldsymbol{v}_j$.

We now state the expression for the influence functions corresponding to the loading vectors of GSPCA, for which the proof can be found in Section A.2 of the Supplementary material:

Theorem 2 (Influence functions of the eigenvectors of the GSSCM) *For the influence functions of the eigenvectors resulting from GSPCA we have the following analytical expression:*

$$\begin{aligned} &IF(\boldsymbol{x},\mathcal{V}_{g,j},F) = \\ &\sum_{k=1,k\neq j}^{p}\frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big[(\boldsymbol{v}_{g,k}^\intercal g(\boldsymbol{x}))(\boldsymbol{v}_{g,j}^\intercal g(\boldsymbol{x})) \\ &\quad +\boldsymbol{v}_{g,k}^\intercal\int\big(dge(\boldsymbol{X})g(\boldsymbol{X})^\intercal \\ &\quad + g(\boldsymbol{X})dge(\boldsymbol{X})^\intercal\big)dF(\boldsymbol{X})\ \boldsymbol{v}_{g,j}\Big]\boldsymbol{v}_{g,k}, \end{aligned} \tag{8}$$

where $dge(\boldsymbol{X}) = \frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})\Big|_{\varepsilon=0}$. Here it is assumed that F has location $\boldsymbol{\mu}=0$ and that the GSSCM and its influence function exist.

While the above expression looks rather involved, we can simplify it further without (much) loss of generality. First, note that we can assume that F has a diagonal covariance matrix since the GSSCM is orthogonally equivariant. In that case, the eigenvectors of the covariance matrix are the standard unit vectors, i.e. $\boldsymbol{v}_j = \boldsymbol{e}_j$. Additionally, we can assume that F is elliptically symmetric. This assumption is very natural since it is required for the Fisher consistency of

the eigenvector functionals $\mathcal{V}_{g,j}$ of the GSSCM, yielding $\mathcal{V}_{g,j}(F) = \boldsymbol{v}_j$. We then obtain the result below, the proof of which is in Section A.3 of the Supplementary material.

Corollary 1 *Assume that F is a centered elliptically symmetric distribution, that it has a density function and a diagonal covariance matrix $\boldsymbol{\Sigma}$, then the following holds*

$$IF(\boldsymbol{x},\mathcal{V}_{g,j},F) = \sum_{k=1,k\neq j}^{p} \frac{1}{\lambda_{g,j}-\lambda_{g,k}} \big(g(\boldsymbol{x})_k g(\boldsymbol{x})_j\big)\boldsymbol{v}_k,$$

where $g(\boldsymbol{x})_k$ is the k^{th} element of the vector $g(\boldsymbol{x})$.

We thus see that, under reasonable assumptions, the influence function of the eigenvectors of GSPCA reduces to a fairly simple expression. In particular, if we plug in the identity function for g, we obtain the influence function of the loadings of classical PCA. In particular, we obtain $\sum_{k=1,k\neq j}^{p} \frac{\boldsymbol{x}_k\boldsymbol{x}_j}{\lambda_j-\lambda_k}\boldsymbol{v}_k$, corresponding to the result obtained in Croux and Haesbroeck (2000).

Having obtained analytical forms for the influence functions of the loading vectors resulting from GSPCA, we will now visualize them to compare the different radial functions in Equations (2) – (6) and illustrate their robustness. We consider a bivariate normal distribution, $F = \mathcal{N}_2(\mathbf{0}_2,\boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \text{diag}(1,0.5)$, and plot the norm of the influence function of the largest eigenvector for the different radial functions resulting in the 3D-plots shown in Figure 2.

First, we observe that the eigenvectors of the classical covariance matrix are the only ones with an unbounded influence function. Second, all GSSCM radial functions are redescending to zero, except for Winsor, whose influence function looks like a smoothed version of the influence function of the SSCM, suggesting that Winsor will attain higher efficiency than the SSCM. Further, we see that the norms of the influence functions of LR, Ball and Shell look quite similar, all three demonstrate four large spikes whereafter each influence function descends to zero. This is due to their cut-offs at the second or third quantile. Quad only descends to zero in the limit, making it more robust than Winsor, but not quite as robust as the redescending radial functions. Overall these results suggest high robustness for the LR, Ball and Shell radial functions.

To summarize influence function in a single number, we consider its supremum, the gross-error sensitivity (GES). For a functional $\mathcal{T}$ at distribution F, it is defined as:

$$\gamma^*(\mathcal{T},F) = \sup_{\boldsymbol{x}} |\text{IF}(\boldsymbol{x},\mathcal{T},F)|.$$

The gross-error sensitivity measures the maximal influence an infinitesimally small amount of contamination can have on a functional $\mathcal{T}$. Therefore it should be finite for robust estimators and preferably small.

For the loading vectors of GSPCA, we now compute the gross-error sensitivities per radial function. We again consider the bivariate normal distribution $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,0.5))$ and study the first loading vector. We compute the gross-error sensitivity as the supremum of the norm of the corresponding influence function. This results in Table 1.

Once again, we observe that CPCA is not robust,

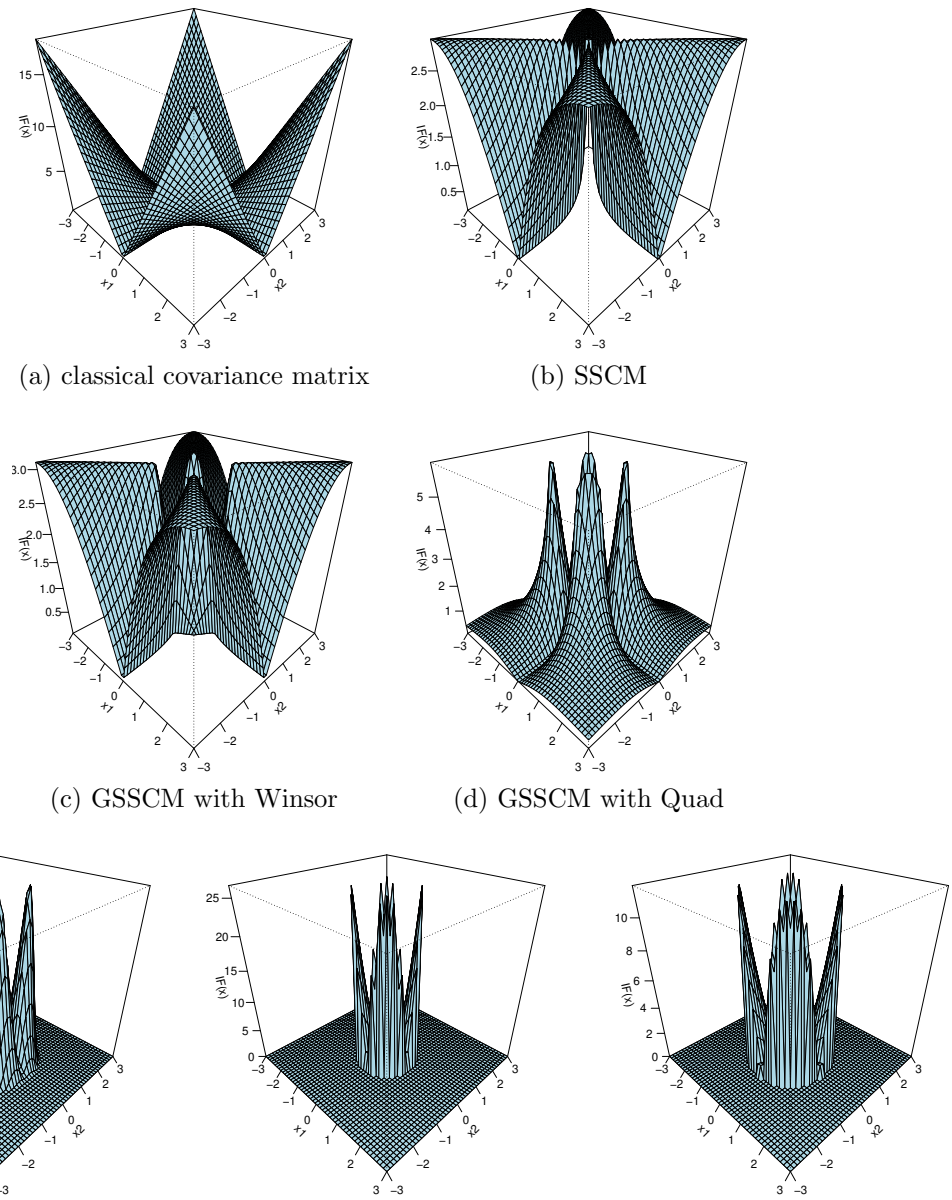
Table 1: Gross-error sensitivities of the largest eigenvector at $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,0.5))$.

	classical	SSCM	Winsor	Quad
γ^*	∞	2.914213	3.100523	6.569927
	LR	Ball	Shell	
	10.74909	29.27686	12.42133	

its GES is unbounded. In contrast, GSPCA has bounded GES for each radial function, implying robustness. The smallest values are attained by SSCM and Winsor. While this may suggest these are the superior methods, the plots of the influence functions do show that the Quad, LR, Ball and Shell methods will have a smaller influence function for most values of the contamination $\boldsymbol{x}$.

3.3 Asymptotic variances and asymptotic relative efficiencies of the loading vectors

We will now use the results on the influence functions to study the asymptotic variances and efficiencies of GSPCA. This will allow us to compare precision among the different radial functions. Following Hampel et al. (1986), we obtain that for well-behaved functionals $\mathcal{T}$, the corresponding estimator $T_n = \mathcal{T}(F_n)$ is asymptotically normal:



(a) classical covariance matrix (b) SSCM

(c) GSSCM with Winsor (d) GSSCM with Quad

(e) GSSCM with LR (f) GSSCM with Ball (g) GSSCM with Shell

Fig. 2: Norm of the influence function of the largest eigenvector at $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,0.5))$ for different radial functions.

$\sqrt{n}(T_n - \mathcal{T}(F)) \xrightarrow{D} \mathcal{N}(0,\text{ASV}(\mathcal{T},F))$ where

$$\text{ASV}(\mathcal{T},F) = \mathbb{E}_F[\text{IF}(\boldsymbol{x},\mathcal{T},F)\text{IF}(\boldsymbol{x},\mathcal{T},F)^\intercal].$$

In order to simplify the exposition, we consider the case of a bivariate normal distribution $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,\gamma))$. Since we are interested in eigenvectors for PCA, we first calculate the asymptotic variance (ASV) of the second element of the largest eigenvector, given that it is important that we estimate this second element close to zero to get a good estimation of the first eigenvector. This is similar to the approach taken in Croux et al.

(2010). We denote the asymptotic variance of the second element of the largest eigenvector simply as ASV_g per radial function.

In Section 3.2 we found analytical expressions for the influence functions of the eigenvectors. Hence, we can use them to calculate the asymptotic variances.

For the second component of the influence function of the first eigenvector, which we denote by $\mathcal{V}_{g,(1,2)}$, we obtain the following expression:

$$\text{IF}(\boldsymbol{x},\mathcal{V}_{g,(1,2)},F) = \frac{1}{\lambda_{g,1}-\lambda_{g,2}} g(\boldsymbol{x})_1 g(\boldsymbol{x})_2.$$

Using this, we obtain the asymptotic variance:

$$\begin{aligned} &\text{ASV}_g(F) \\ =&\frac{1}{(\lambda_{g,1}-\lambda_{g,2})^2}\int (g(\boldsymbol{X})_1 g(\boldsymbol{X})_2)^2 \, dF(\boldsymbol{X}). \end{aligned} \tag{9}$$

When we assume that F is distributed as $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,\gamma))$, we can evaluate the asymptotic variance of Equation (9) for different values of γ. The results of this computation are presented in Figure 3.

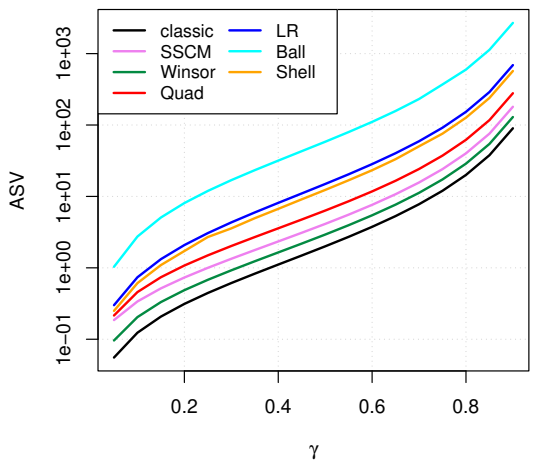


Fig. 3: Asymptotic variance of the second component of the largest eigenvector for different values of γ at bivariate normal distribution $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,\gamma))$.

From the plots in Figure 3, it is clear that Ball performs the poorest of all radial functions by a large margin. LR and Shell perform somewhat average. The lowest ASV is evidently obtained by classical PCA. However, Winsor's ASV is quite close to it, followed by SSCM. Quad also has a low asymptotic variance.

We now use these asymptotic variances to calculate the asymptotic relative efficiency in comparison to the classical method as follows:

$$\text{Eff}_g(F) = \frac{\text{ASV}_{\text{class}}(F)}{\text{ASV}_g(F)},$$

where g refers to the chosen radial function. The closer this value is to one, the more efficient is our method based on the corresponding radial function g. This yields Figure 4. It is immediately clear

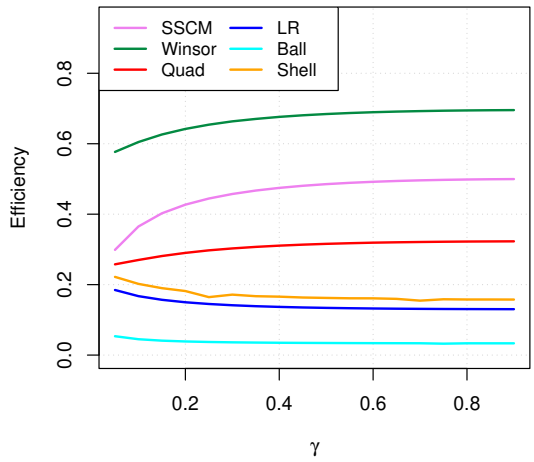


Fig. 4: Asymptotic relative efficiencies of the second component of the largest eigenvector for different values of γ at bivariate normal distribution $\mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,\gamma))$.

that the highest efficiency is obtained by Winsor, it seems to converge to 0.7 for higher values of γ. SSCM is inferior, but still achieves values around 0.5. This is interesting, since it suggests that we can improve on the efficiency of the popular SSCM without sacrificing robustness (in terms of having a bounded influence function). Next we observe Quad, attaining values around 0.3. Shell, LR and Ball have low efficiencies at bivariate

normal distributions. Overall Winsor clearly outperforms the other radial functions in terms of relative efficiency.

The high variance of Ball translates into the lowest efficiency. This does not come as a surprise, as Ball bears similarities to the affine equivariant robust covariance estimators which use only half of the data such as the minimum covariance determinant and minimum volume ellipsoid estimators (Rousseeuw, 1984). These are known to have fairly low efficiencies on Gaussian data but strong robustness properties (Davies, 1992; Butler et al., 1993; Croux and Haesbroeck, 1999; Cator and Lopuhaä, 2012). Ball is also similar to the BACON algorithm (Billor et al., 2000), which also has strong robustness but weak efficiency properties.

The efficiencies obtained in the discussion above may be off-putting at first sight. However, one should keep in mind that these are calculated on multivariate Gaussian data. If more heavy-tailed distributions are considered, the picture can look very different. To illustrate this, we have replicated the analysis for the multivariate Student's t_3-distribution in Section A.6. For this heavy-tailed distribution, the performance of the GSPCA variants has remained stable, but the asymptotic variance of classical PCA exploded. As a result, the asymptotic relative efficiencies of GSPCA all become larger than 1. In Section 4 we present additional evidence that GSPCA does in fact outperform classical PCA when the distributions have heavy tails.

3.4 Influence functions of the eigenvalues

In addition to the eigenvectors, we also study the influence functions of the eigenvalues obtained by GSPCA. Denoting these functionals by $\mathcal{L}_{g,j}$ for $j = 1,...,p$, we could be interested in

$$\mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,j},F) = \lim_{\varepsilon\to 0}\frac{\mathcal{L}_{g,j}(F_{\varepsilon,\boldsymbol{x}})-\mathcal{L}_{g,j}(F)}{\varepsilon}.$$

The problem with using the expression above however, is that the eigenvalue functionals are not Fisher consistent for elliptical distributions unlike their eigenvector counterparts. Otherwise stated, in general we have $\mathcal{L}_{g,j}(F) \neq \lambda_j$, as well as $\mathcal{L}_{g_1,j}(F) \neq \mathcal{L}_{g_2,j}(F)$ for two different functions g_1 and g_2. As a result, we cannot directly compare these influence functions for different functions g which makes them less interesting. Instead, we can correct the estimated eigenvalues to make them Fisher consistent. We then compare the influence functions of these corrected estimators for different radial functions.

3.4.1 Modification of GSPCA for Fisher consistent eigenvalues

To adapt our method for Fisher consistency, we will use the approach suggested in Croux and Ruiz-Gazen (2005). The idea proposed there is that we can correct the eigenvalues by first projecting the data onto the eigenvectors before calculating the eigenvalues. More specifically we proceed as follows:

1. First, we calculate the eigenvectors $\widehat{\boldsymbol{v}}_{g,j}$ through GSPCA as done before. We know that the corresponding functionals are Fisher consistent provided that we have an elliptical distribution.
2. Second, we project the data onto these eigenvectors after which we can use a robust, equivariant scale estimator s to calculate the eigenvalues:

   $$\hat{\lambda}^{(s)}_{g,j} = s^2(\widehat{\boldsymbol{v}}^{\intercal}_{g,j}\boldsymbol{X}).$$

   Provided that the functional $\mathfrak{s}$ corresponding with the estimator s is Fisher consistent, this procedure yields Fisher consistent eigenvalue functionals $\mathcal{L}^{(\mathfrak{s})}_{g,j}$ at elliptically symmetric distributions (see, e.g., Croux and Ruiz-Gazen (2005)).
3. Lastly, should we be interested in the covariance matrix, we can combine the two steps above to obtain a new estimator with functional $\mathcal{S}_{\mathrm{comb}}$ for the covariance matrix:

   $$\mathcal{S}_{\mathrm{comb}}(F) = \sum_{k=1}^{p}\mathcal{L}^{(\mathfrak{s})}_{g,k}(F)\mathcal{V}_{g,k}(F)\mathcal{V}_{g,k}(F)^{\intercal}.$$

   For elliptically symmetric distributions, this covariance matrix is Fisher consistent for $\boldsymbol{\Sigma}$.

In summary, we can correct our method to obtain Fisher consistency for the eigenvalues and the resulting covariance estimate at elliptically symmetric distributions. Next, we will take a look at

the corresponding influence functions and asymptotic variances to evaluate the robustness and efficiency of the adapted GSPCA.

3.4.2 Influence functions of the new eigenvalues

In order to obtain the influence function for the corrected eigenvalue functionals, we need a general property of projections of elliptically symmetric distributions. Denote the distribution of the random vector projected on $\boldsymbol{a}$ as $H^{\boldsymbol{a}}$, i.e., if $\boldsymbol{X} \sim H$ then $\boldsymbol{a}^\intercal \boldsymbol{X} \sim H^{\boldsymbol{a}}$. We will make use of following lemma from Croux and Ruiz-Gazen (2005) on the distribution of this projection $H^{\boldsymbol{a}}$.

Lemma 1 (Projected distributions) *Assume H is an elliptically symmetric distribution with location parameter $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. Then there exists a univariate symmetric distribution F_0 such that*

$$H^{\boldsymbol{a}}(z) = F_0\left(\frac{z-\boldsymbol{\mu}^\intercal \boldsymbol{a}}{\sqrt{\boldsymbol{a}^\intercal \boldsymbol{\Sigma} \boldsymbol{a}}}\right).$$

Lemma 1 ensures that all projections of a random vector with an elliptically symmetric distribution follow the same symmetric distribution F_0 after proper scaling and centering.

We can now obtain an expression for the influence functions of the corrected eigenvalue functionals. The proof can be found in Section A.4 of the Supplementary material.

Theorem 3 (Influence functions of the new eigenvalues) *For the modified eigenvalue functional $\mathcal{L}_{g,k}^{(\mathfrak{s})}$ with $\mathfrak{s}$ an equivariant scale functional, we have following expression for the influence function at elliptically symmetric distributions*

$$IF\left(\boldsymbol{x}, \mathcal{L}_{g,k}^{(\mathfrak{s})}, H\right) = 2\lambda_k \; IF\left(\frac{\boldsymbol{x}^\intercal \boldsymbol{v}_k}{\sqrt{\lambda_k}}, \mathfrak{s}, F_0\right). \tag{10}$$

This is the same expression as the one obtained by Croux and Ruiz-Gazen (2005) in the context of robust PCA based on projection-pursuit. In particular, note that the expression does not depend on the influence function of any of the eigenvectors.

From the expression of the influence function of the eigenvalues, we readily obtain the asymptotic variance of the new eigenvalues at elliptically symmetric distributions:

$$\begin{aligned}
\mathrm{ASV}(\mathcal{L}_{g,k}^{(\mathfrak{s})}, H) &= \mathbb{E}[\mathrm{IF}(\boldsymbol{X}, \mathcal{L}_{g,k}^{(\mathfrak{s})}, H)^2] \\
&= \mathbb{E}\left[4\lambda_k^2 \; \mathrm{IF}\left(\frac{\boldsymbol{v}_k^\intercal \boldsymbol{X}}{\sqrt{\lambda_k}}, \mathfrak{s}, F_0\right)^2\right] \\
&= 4\lambda_k^2 \; \mathrm{ASV}(\mathfrak{s}, F_0).
\end{aligned}$$

Note that the expression for the influence function of the eigenvalues is proportional to the influence function of the scale functional used. This means that known optimality properties of equivariant scale functionals carry over to the estimation of these eigenvalues. In particular, we can leverage the theory on M-estimation to find optimal M-estimators of scale for estimating the eigenvalues. This is stated in Corollary 2, which follows from Theorem 3 and Hampel et al. (1986) (Section 2.5e), and leads us to the use of the median absolute deviation (MAD) defined by

$$\mathrm{MAD}(X) = c \cdot \mathrm{med}|X - \mathrm{med}(X)|.$$

Here $c = 1/\Phi^{-1}\left(\frac{3}{4}\right) \approx 1.4826$ is a consistency factor at normal distributions, hence we obtain Fisher consistency for the estimated eigenvalues.

Corollary 2 *If $F_0 = \mathcal{N}(0,1)$, the median absolute deviation is the most B-robust M-estimator of the eigenvalues, i.e. it has the lowest gross error sensitivity among all (well-behaved) Fisher consistent M-estimators of scale. Similarly, Huber's M-estimator is the optimal B-robust M-estimator of the eigenvalues as it has the highest efficiency for a given bound on the gross error sensitivity.*

We visualize the influence function of the new eigenvalue for the multivariate normal distribution below. Suppose $H \sim \mathcal{N}_2(\mathbf{0}_2, \mathrm{diag}(1,\gamma))$ and thus $F_0 \sim \mathcal{N}(0,1) = \Phi$. The influence function of the MAD is given by (Hampel et al. (1986), p. 107):

$$\mathrm{IF}(x, \mathrm{MAD}, \Phi) = \frac{\mathrm{sign}\left(|x| - \Phi^{-1}\left(\frac{3}{4}\right)\right)}{4\Phi^{-1}\left(\frac{3}{4}\right)\phi\left(\Phi^{-1}\left(\frac{3}{4}\right)\right)}.$$

The influence function of the largest eigenvalue $\lambda_1 = 1$ then becomes 2 $\mathrm{IF}(\boldsymbol{x}_1, \mathrm{MAD}, \Phi)$, which is shown in Figure 5.

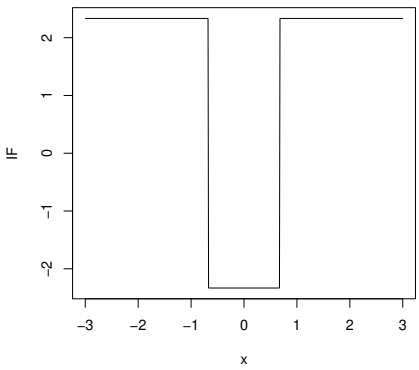


Fig. 5: Influence function of the largest (new) eigenvalue for $x \in [-3,3]$.

3.4.3 Influence function of the new covariance estimator

Finally, as an additional result, we compute the influence function of the new estimate for the covariance matrix S_{comb} (see Section A.5 of the Supplementary material for the proof).

Theorem 4 (Influence function of the new covariance matrix $\boldsymbol{S}_{\text{comb}}$) *Let H be an elliptically symmetric distribution with a diagonal covariance matrix and $\mathfrak{s}$ an equivariant scale functional. Then*

$$IF(\boldsymbol{x},\mathcal{S}_{comb},H) = 2\sum_{k=1}^{p}\lambda_k \; IF\left(\frac{\boldsymbol{v}_k^\intercal \boldsymbol{x}}{\sqrt{\lambda_k}},\mathfrak{s},F_0\right)\boldsymbol{v}_k\boldsymbol{v}_k^\intercal$$
$$+\sum_{k=1}^{p}\lambda_k \sum_{j=1,j\neq k}^{p}\frac{1}{\lambda_{g,k}-\lambda_{g,j}}(g(\boldsymbol{x})_j g(\boldsymbol{x})_k)(\boldsymbol{v}_j\boldsymbol{v}_k^\intercal+\boldsymbol{v}_k\boldsymbol{v}_j^\intercal).$$

Next we compare the influence functions of the new covariance functional using MAD as scale estimator s with the common GSSCM. For this we plot the influence functions for $H \sim \mathcal{N}_2(\mathbf{0}_2,\text{diag}(1,\gamma))$ and focus on the off-diagonal element. For that, we obtain:

$$\text{IF}(\boldsymbol{x},\mathcal{S}_{\text{comb}},H)_{1,2} = \frac{1-\gamma}{\lambda_{g,1}-\lambda_{g,2}}g(\boldsymbol{x})_1 g(\boldsymbol{x})_2.$$

This corresponds to a rescaling of the influence function of the common GSSCM. For $\gamma = 0.5$ and contamination in the direction of $(x,x)^\intercal$ this yields Figure 6.

As we can see from the comparison of the raw GSSCM and our proposed correction, the relative performances among different radial functions

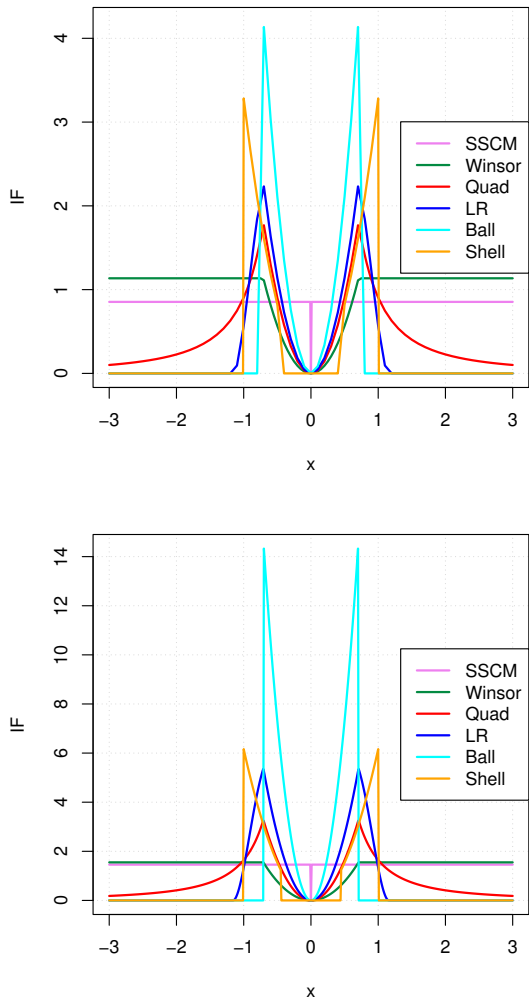


Fig. 6: Influence function of the uncorrected GSSCM $\mathcal{S}$ (top) and the corrected GSSCM $\mathcal{S}_{\text{comb}}$ (bottom) for contamination in the direction of $(x,x)^\intercal$: the off-diagonal element.

remain relatively stable. However, the influence function of Winsor is now similar to that of SSCM except in the center where it is lower. This again speaks in favor of Winsor instead of the popular SSCM. Contamination in the direction of $(x,0)^\intercal$ results in a zero influence function for each radial function since $g((x,0)^\intercal)_2 = 0$.

4 Simulation study

In this section we conduct a simulation study on synthetically sampled data to investigate the performance of generalized spherical PCA using the five radial functions defined in Equations (2) – (6). We compare them with classical PCA, spherical PCA based on the SSCM and ROBPCA.

The ROBPCA method was published by Hubert et al. (2005). It is based on the projection pursuit and the minimum covariance determinant estimator. For our simulation, we will largely base ourselves on the methodology of the study conducted in the prior mentioned ROBPCA paper (Hubert et al., 2005).

4.1 Setting

In the simulation, we will look at uncontaminated data ($\varepsilon = 0$) and data with different levels of contamination ($\varepsilon = 0.1$, $\varepsilon = 0.2$) obtained from the following p-variate Gaussian distribution and the p-variate Student's t-distribution with five degrees of freedom:

$$(1-\varepsilon)\ \mathcal{N}_p(\mathbf{0}_p,\mathbf{\Sigma})+\varepsilon\ \mathcal{N}_p(\tilde{\boldsymbol{\mu}},\tilde{\mathbf{\Sigma}}),$$
$$(1-\varepsilon)\ t_5(\mathbf{0}_p,\mathbf{\Sigma})+\varepsilon\ t_5(\tilde{\boldsymbol{\mu}},\tilde{\mathbf{\Sigma}}).$$

From these distributions, we will repeatedly generate 500 data samples of size n. As the GSSCM scatter estimator is orthogonally equivariant, we only need to look at diagonal covariance matrices $\mathbf{\Sigma}$. More in specific, we consider the following two situations, just as in the ROBPCA paper (Hubert et al., 2005):

1. Low-dimensional data:
   $n = 100$, $p = 4$, $\mathbf{\Sigma} = \text{diag}(8,4,2,1)$, $\tilde{\boldsymbol{\mu}} = f_1 \cdot \boldsymbol{e}_4 = (0,0,0,f_1)^\intercal$, $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma}/f_2$.
2. High-dimensional data:
   $n = 50$, $p = 100$, $\mathbf{\Sigma} = \text{diag}(17,13.5,8,3,1,0.095,...,0.002,0.001)$,
   $\tilde{\boldsymbol{\mu}} = f_1 \cdot \boldsymbol{e}_6$, $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma}/f_2$.

In the low-dimensional case we compute $k = 3$ principal components, whereas in the high-dimensional case we compute $k = 5$ components, as we want our principal component analysis to explain at least 90% of the total variance. Parameter $f_1 \in \{6,8,10,...,20\}$ determines the location shift for the contaminated data in the direction of the $k+1^{\text{th}}$ principal component, $\boldsymbol{e}_4$ or $\boldsymbol{e}_6$. This location of the contamination is typically the hardest to deal with for (robust) scatter and PCA estimators (see Hubert et al. (2014) and Lemma 2 of Louvet et al. (2023)). Simulations on alternative locations for the contamination confirmed this behavior. Parameter $f_2 \in \{1,15\}$ specifies the concentration of the contaminated data. For the k-step LTS estimator we use 5 successive C-steps.

4.2 Maxsub measure

To evaluate performance in our simulation, we will compute the maxsub measure, which calculates the maximal angle between the estimated PCA subspace and the space spanned by $\{\boldsymbol{e}_1,\boldsymbol{e}_2,...,\boldsymbol{e}_k\}$. This can be computed as follows (Hubert et al., 2005):

$$\text{maxsub} = \arccos(\sqrt{\lambda_k})/(\pi/2),$$

where λ_k represents the smallest eigenvalue of $\boldsymbol{I}_{p,k}^\intercal \boldsymbol{V}_{p,k} \boldsymbol{V}_{p,k}^\intercal \boldsymbol{I}_{p,k}$ with $\boldsymbol{I}_{p,k} = [\boldsymbol{e}_1\ \boldsymbol{e}_2\ \cdots\ \boldsymbol{e}_k]$ and $\boldsymbol{V}_{p,k} = [\boldsymbol{v}_1\ \boldsymbol{v}_2\ \cdots\ \boldsymbol{v}_k]$. We divide by $\pi/2$ to standardize the value of the angle. The ideal value of the maxsub measure is zero, the worst value is one.

4.2.1 Uncontaminated data

In a first step, we simulate uncontaminated data ($\varepsilon = 0$). As discussed, we sample from the multivariate normal distribution and the Student's t-distribution. We however do not limit ourselves here to five degrees of freedom for the t-distribution, we also consider t_3, t_2 and t_1. The mean of the maxsub measure for 500 samples is shown in Table 2.

When there is no contamination, we see that for the normal and for the t_5-distribution the best values are attained by classical PCA and GSPCA with the Winsor radial function. This result is in line with the computed efficiencies in Section 3.3, where Winsor attained the highest efficiency of all radial functions for Gaussian data. ROBPCA, SPCA and Quad perform almost as good. The worst results are those of GSPCA with the Ball function, making this method least efficient at no contamination. Shell also performs poorly, especially in the high dimensional case. The low efficiency at no contamination for Ball and Shell is due to the fact that a lot of data points are given a zero weight, see Equation (4) and (5).

For the more heavy-tailed t_1-, t_2- and t_3-distributions, the results are different. As the

Table 2: Maxsub measure at uncontaminated data.

	n	p	CPCA	SSCM	Winsor	Quad	LR	Ball	Shell	ROBPCA
Normal	100	4	**.100**	.137	.113	.132	.163	.257	.149	.138
	50	100	**.216**	.275	.227	.243	.272	.332	.311	.245
t_5	100	4	.134	.131	**.113**	.137	.163	.243	.170	.138
	50	100	.303	.274	**.254**	.277	.313	.395	.378	.281
t_3	100	4	.185	.122	**.113**	.135	.159	.222	.174	.140
	50	100	.419	.272	**.267**	.294	.332	.416	.407	.305
t_2	100	4	.249	.129	**.119**	.142	.162	.235	.186	.156
	50	100	.554	**.280**	.285	.318	.358	.455	.445	.354
t_1	100	4	.403	.120	**.117**	.155	.173	.236	.204	.183
(Cauchy)	50	100	.710	**.277**	.304	.366	.403	.494	.497	.495

degrees of freedom in the t-distribution decrease, making it more heavy tailed, the performance of CPCA deteriorates. GSPCA with Winsor and SPCA on the other hand, are not affected by the heavy tails. Hence in this scenario, the use of Winsor or the SSCM is preferred over the use of CPCA, even if there is no contamination.

4.2.2 Contaminated data

In a second step, we look at the mean of the maxsub measure for 500 samples of data with different levels of contamination ($\varepsilon = 0.1$ and $\varepsilon = 0.2$). The results for even higher levels of contamination are qualitatively similar and discussed in Section A.7 of the Supplementary material. Figures 7 to 10 present the results for multivariate normal and t_5-data in the low- and high-dimensional case.

Low-dimensional, normal data:
For the low-dimensional, multivariate normal data (Figure 7) we can clearly observe that CPCA, SPCA and Winsor fail. Their maxsub measure is close to one, implying that outliers influenced the estimated PCA subspace to the extent that one of the estimated principal components is orthogonal to span$\{\boldsymbol{e}_1,...,\boldsymbol{e}_k\}$.

The poor result for these radial functions can be explained by their influence functions in Section 3.2, where we saw that the influence functions of the eigenvectors of the classical covariance matrix, the SSCM and the GSSCM of Winsor did not redescend to zero in contrast with the others. The maxsub measure of Quad also attains some high values, especially for $\varepsilon = 0.2$ and with f_1 rather small. When f_1 is larger, indicating that the outliers are shifted far enough from the center of the distribution, Quad performs better. In contrast, we attain very good maxsub measure results for LR, Shell and Ball, whose values are as good as the values for ROBPCA. These three radial functions have redescending influence functions.

High-dimensional, normal data:
In the high-dimensional normal case (Figure 8), Quad performs worse and fails, just as CPCA, SPCA and Winsor. LR, Shell and Ball are able to distinguish the outliers when $f_1 \geq 10$, in which case their results are comparable to ROBPCA. However, for $f_1 < 10$, only ROBPCA achieves low values.

Low-dimensional, t_5-data:
For t_5-data in the low-dimensional case (Figure 9), the results are similar to low-dimensional normal data. The only difference is that for $\varepsilon = 0.2$ and $f_1 = 6$, all methods perform poorly, except for Ball who still attains a low value for the maxsub measure. So in this situation, when the outliers are very close to the regular observations, Ball is the only one who can distinguish them.

High-dimensional, t_5-data.
Lastly, for the high-dimensional multivariate t_5-data (Figure 10), we observe that once again CPCA, SPCA, Winsor and Quad fail. However, Ball and ROBPCA perform well once f_1 is greater than 10. LR and Shell attain higher values, and they only achieve good results when f_1 exceeds 12 or 14.
In summary, we can conclude that for the multivariate normal data GSPCA with radial functions LR, Shell and Ball performs comparable to ROBPCA. For the t_5-data, Ball achieves the same results as ROBPCA. LR and Shell perform somewhat worse, but can still be considered as good alternatives. The fact that GSPCA with the Ball radial function achieves such great results is somewhat surprising, since from its definition in Equation (4) it follows that half of the data points get weight zero and therefore do not contribute anymore. However, this cutoff makes Ball very

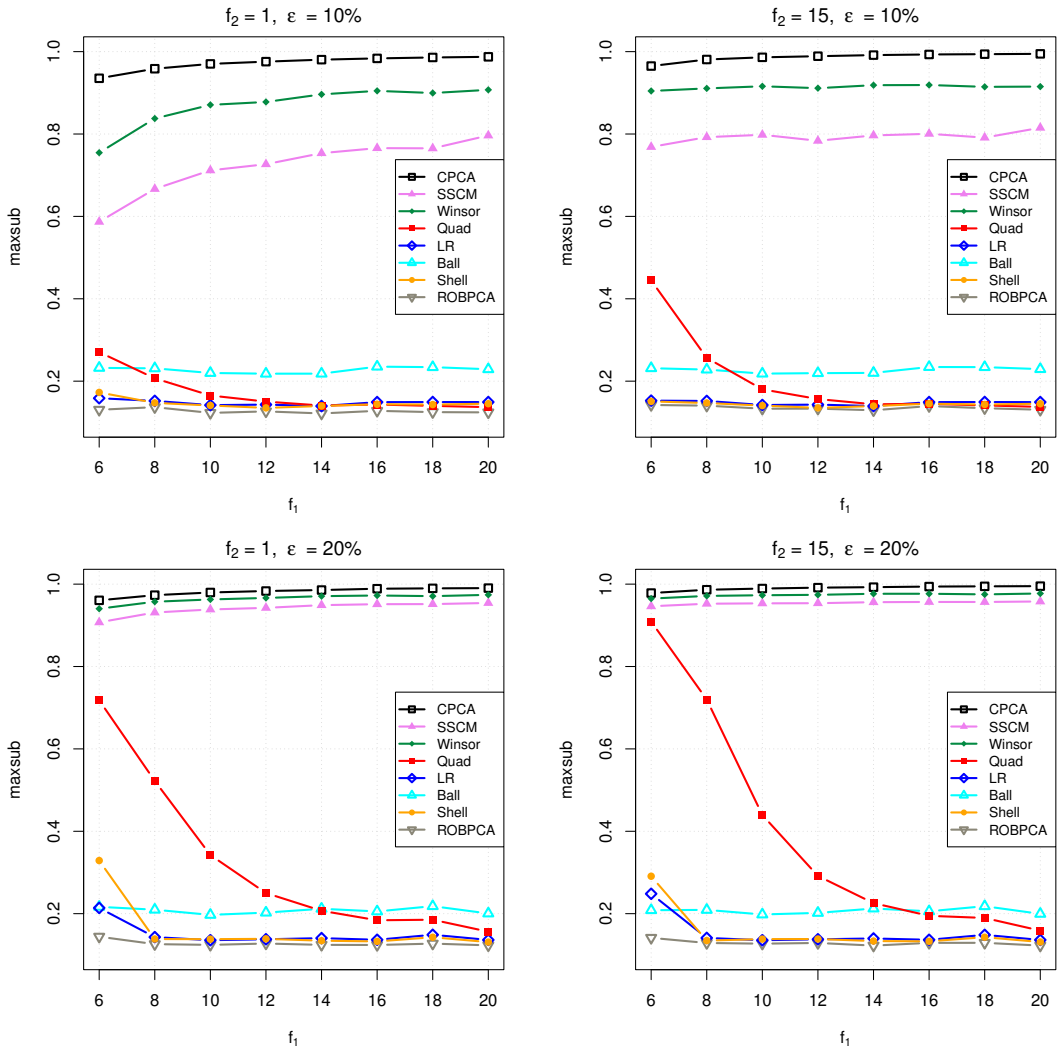


Fig. 7: Maxsub measure for low-dimensional, multivariate normal data.

robust, as can be seen by the quickly redescending influence functions of its eigenvectors in Section 3.2.

4.3 Computation time

One of the biggest advantages of our new PCA method is its computational speed. It has the same computational complexity as classical PCA, while ROBPCA is a significantly slower method. To illustrate this, we will compare the computation time of CPCA and ROBPCA with GSPCA using the radial functions Ball and LR.

We will use the same setting as previously and sample multivariate normal data, from the low- and high-dimensional case, with the mean vector and covariance matrix as specified in Section 4.1. We set $\varepsilon = 0$ (no contamination) and measure the total computational time for 100 runs. The computations were done on a 2.50 GHz core i5 processor (7^{th} gen) and results are shown in Table 3 and 4.

From both tables it is clear that GSPCA outperforms ROBPCA when we consider the computation time. The time required for GSPCA is comparable to that of classical PCA for both the low- and high-dimensional situation. However, the time needed for ROBPCA increases significantly

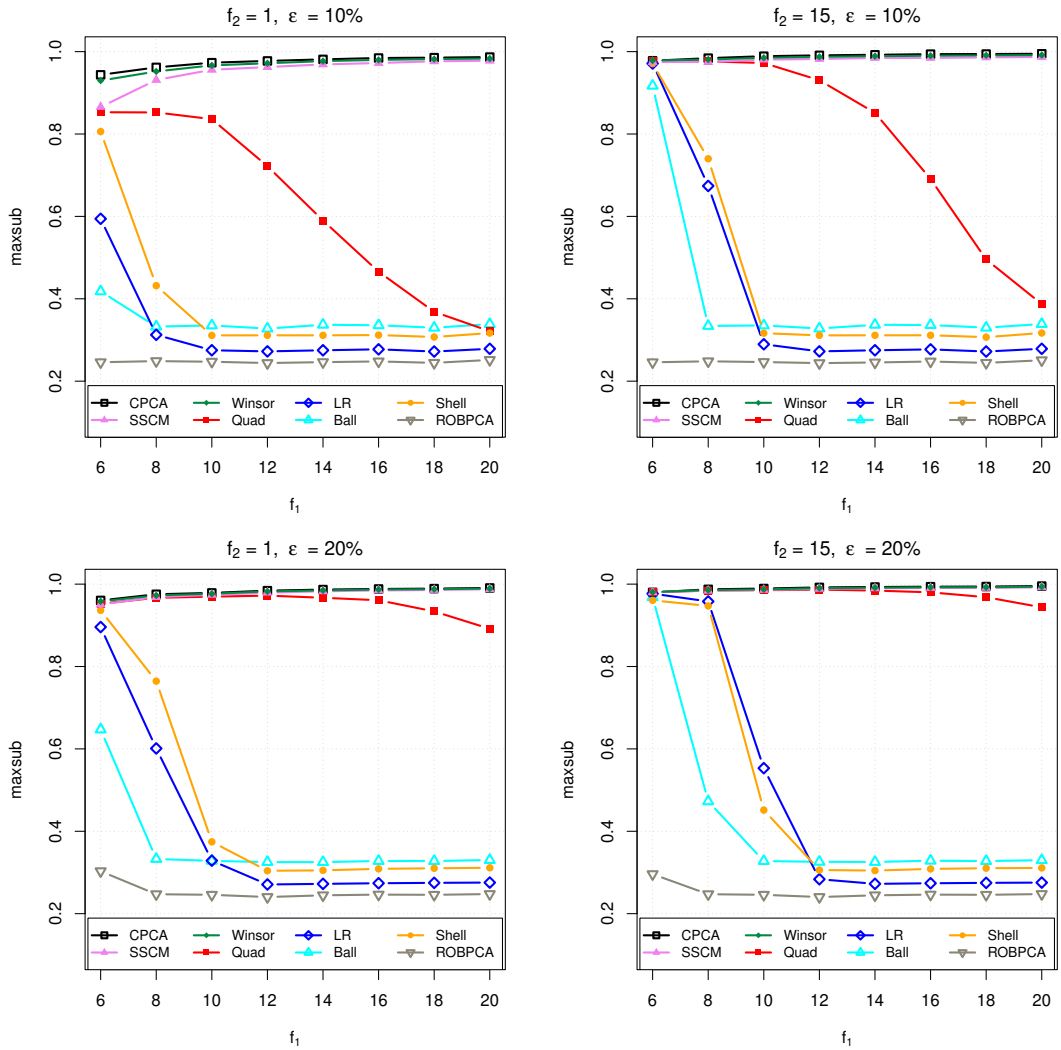


Fig. 8: Maxsub measure for high-dimensional, multivariate normal data.

when the sample size n becomes larger. We conclude that a lot of time can be saved by using GSPCA instead of ROBPCA.

5 Data Examples

In this section we illustrate the performance of GSPCA on two real data sets: the Top Gear car data set and a surveillance video of a beach, previously studied in a different context in Rousseeuw et al. (2018).

5.1 Top Gear data

The first data set is the Top Gear car data set from the R package RobustHD, which has been studied frequently in the context of robustness and PCA. We consider the numeric variables Price, Displacement, BHP, Torque, Acceleration, Top-Speed, MPG, Weight, Length, Width and Height and remove incomplete observations, retaining 245 observations. Next we scale the data and perform CPCA and GSPCA combined with the LR radial function, retaining 3 principal components to explain at least 85% of the total variance. To illustrate the robustness of GSPCA, we use the

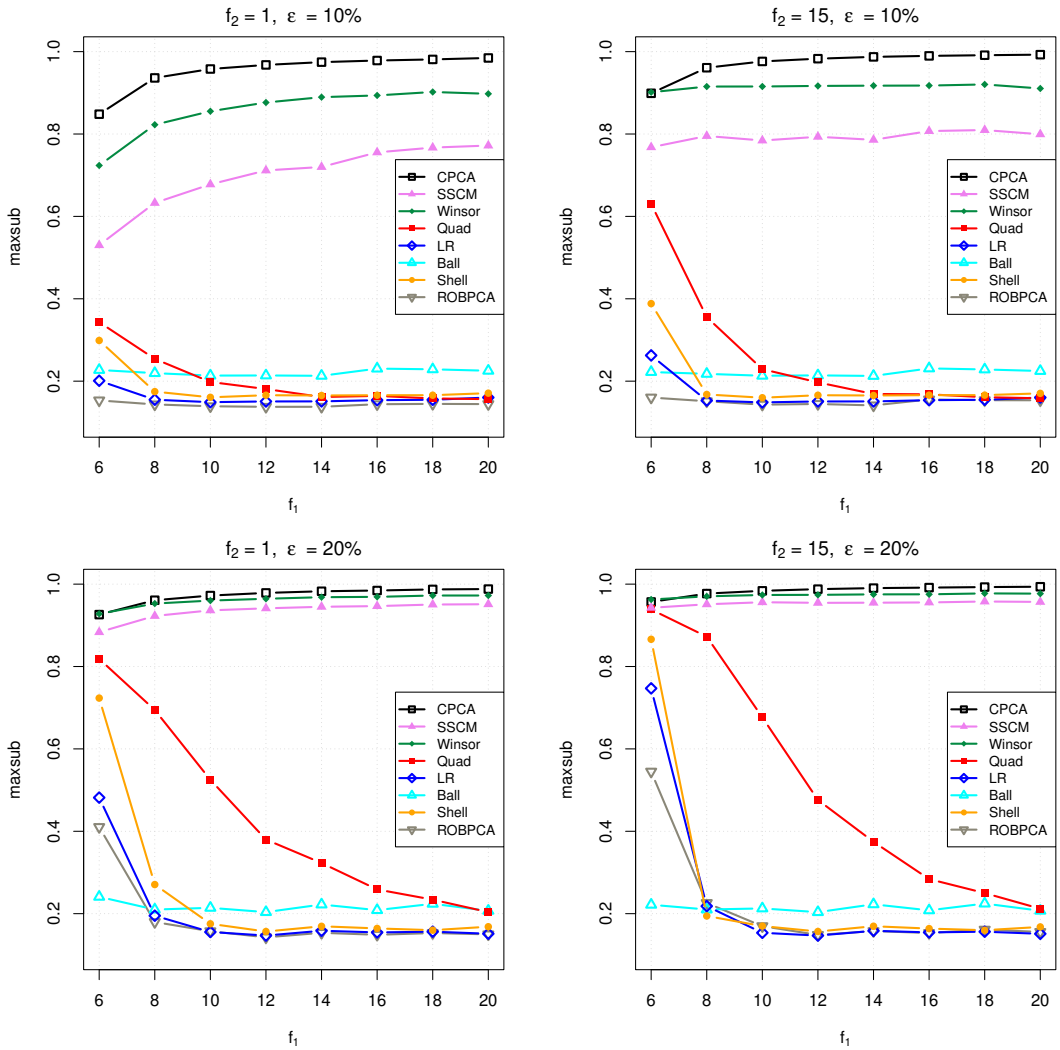


Fig. 9: Maxsub measure for low-dimensional, multivariate t_5-data.

Table 3: Total computational time for 100 runs (in seconds) for low-dimensional, multivariate normal data ($p = 4$, $k = 3$) for various values of n.

n	50	100	150	200	250	300	350	400	450	500
CPCA	0.38	0.25	0.36	0.32	0.33	0.35	0.34	0.37	0.47	0.41
GSPCA LR	0.25	0.22	0.5	0.47	0.46	0.6	0.55	0.58	0.61	0.73
GSPCA Ball	0.23	0.23	0.38	0.37	0.41	0.45	0.52	0.58	0.62	0.71
ROBPCA	1.66	4.74	13.18	28.61	54.52	92.37	146.24			

diagnostic plot defined by Hubert et al. (2005) to classify the flagged outliers, resulting in Figure 11.

We observe that GSPCA identifies many bad leverage points, outlying in score distance and orthogonal distance, while CPCA only flags 2 of them as bad leverage points. This suggests that CPCA was heavily influenced by the bad leverage points identified by GSPCA. We also ran ROBPCA on the data, and its results were very similar to those of GSPCA.

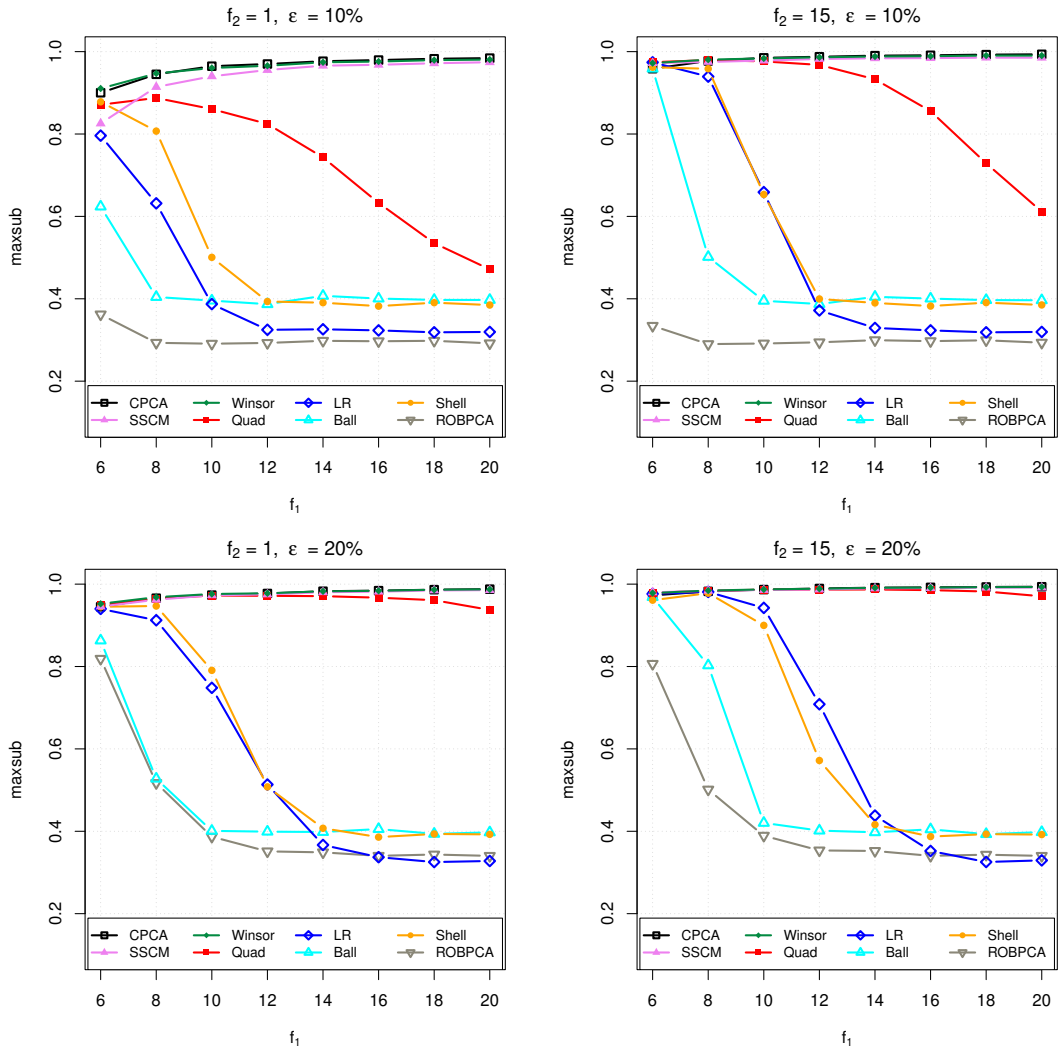


Fig. 10: Maxsub measure for high-dimensional, multivariate t_5-data.

Table 4: Total computational time for 100 runs (in seconds) for high-dimensional, multivariate normal data ($p = 100$, $k = 5$) for various values of n.

n	50	100	150	200	250	300	350	400	450	500
CPCA	0.89	1.34	1.88	2.28	2.86	2.93	3.58	3.88	4.18	4.6
GSPCA LR	1.11	1.38	1.65	2	2.3	2.56	3.41	3.66	3.76	4.56
GSPCA Ball	1.14	1.4	1.58	2	2.19	2.67	3.11	3.45	4.11	4.35
ROBPCA	2.78	9.34	23.78	50.18	95.04	154.86	240.79			

5.2 Video data

The second data set is comprised of a surveillance video of a beach, consisting of 633 frames of 160×128 pixels in the RGB color model. The data was previously studied in the context of outlier detection for functional data analysis (Rousseeuw et al., 2018) and originates from Li et al. (2004). In the video we see a beach scenery, where in frame 483 a man comes into view, in frame 489 he disappears behind a tree and in frames 493 to 633 he

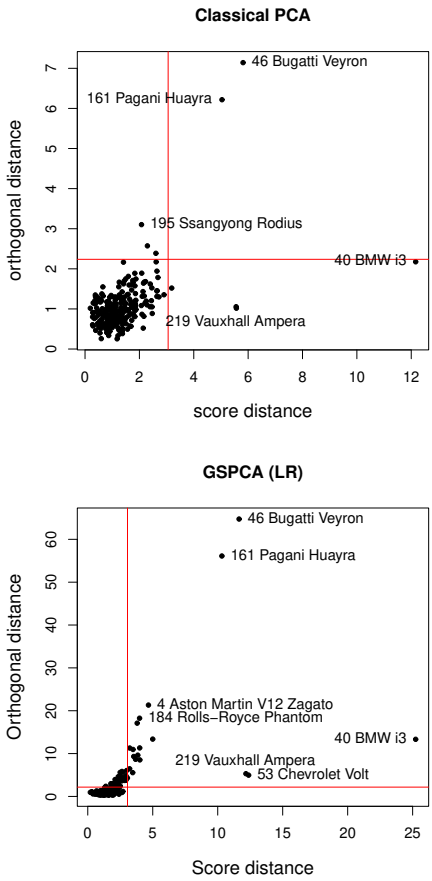


Fig. 11: Diagnostic plots for CPCA and GSPCA on the Top Gear data.

remains in view, see for example Figure 12. Hence from a robustness perspective, we consider frames 489 to 633 as outliers, while frames 1 to 488 are considered normal observations.

To test GSPCA, we run our new method using the LR radial function and the classical method on the high dimensional, video data. As the video is color coded in the RGB model, we have three data sets (**X1**, **X2** and **X3**) of 633 observations with 20480 variables. We select the number of principal components to explain at least 85% of the variance. The resulting diagnostic plots of the principal component analyses are shown in Figure 13 for the three colors.

Figure 13 shows that GSPCA distinguishes the normal observations (no man, black points) from the outliers (man in view, colored points). The method flags all outliers as orthogonal outliers or bad leverage points. It also separates the frames where the man is behind the tree (green). For CPCA, the distinction between the outliers and the normal observations is less clear. Many of the frames containing the man are flagged as good leverage points, indicating that CPCA was strongly influenced by these points. CPCA also doesn't separate the frames where the man is behind the tree.

It is also worth noting that GSPCA distinguishes one point with a large score distance of 20. This observation corresponds to the first frame which is a bit lighter compared to the others, probably due to starting the video recording. CPCA also separates this point, but as a point with a large orthogonal distance.

To illustrate the effect of the outliers on the analysis, we calculate the predicted values $\hat{\mathbf{X}}_{(k)}$ of the principal component analysis

$$\hat{\mathbf{X}}_{(k)} = (\boldsymbol{X}_{n,p} - \mathbf{1}_n T(\boldsymbol{X})^\intercal)\boldsymbol{V}_{p,k}\boldsymbol{V}_{p,k}^\intercal + \mathbf{1}_n T(\boldsymbol{X})^\intercal,$$

with k the number of principal components to explain 85% of the variance, and T the location estimator. From these, we compute the residuals from CPCA and GSPCA to detect outlying pixels and to study the difference between the original frames and the predictions:

$$\boldsymbol{r} = \boldsymbol{X} - \hat{\mathbf{X}}_{(k)}.$$

We also standardize the residuals per frame using the mean and standard deviation for CPCA, and the median and MAD for GSPCA. In Figure 14 we show them for frames 108 (no man), 487 (man left), 491 (man behind tree) and 564 (man right). Here all scaled residuals below outlier cutoff 2.5 are given the same color.

From the residual plot we observe that CPCA only has few mild residuals, consisting mostly of noise around the man. In GSPCA, on the contrary, the entire man has very high residuals and is clearly detected by the analysis. The waves of the sea also have significant residuals as they move throughout the video.

6 Conclusion

We proposed and studied *generalized spherical principal component analysis* as a new robust version of PCA. GSPCA calculates the loading

Fig. 12: Video frames 108, 487, 491 and 564.

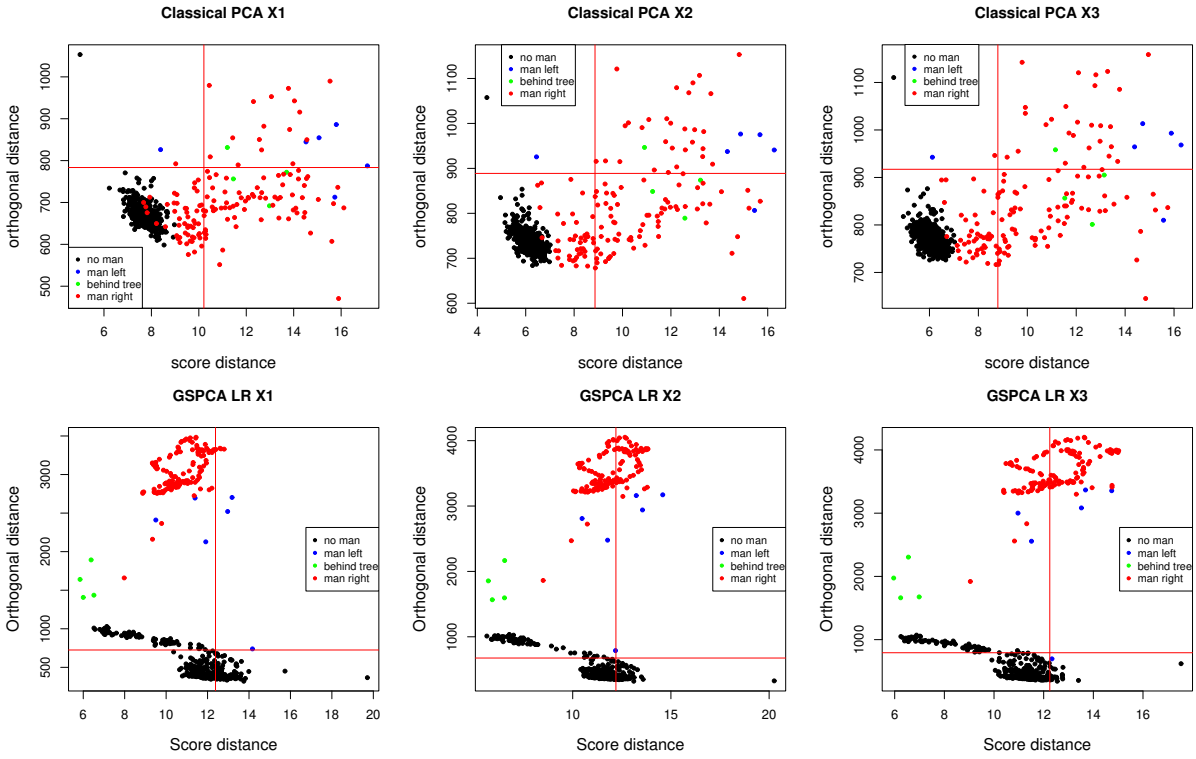


Fig. 13: Diagnostic plots for CPCA and GSPCA on the video data in RGB color.

vectors on the generalized spatial sign covariance matrix (GSSCM) instead of on the classical covariance matrix. The GSSCM can be used with different radial functions, and we compared five in this work: Winsor, Quad, LR, Ball and Shell.

We studied the robustness properties of GSPCA by deriving influence functions and breakdown values. It turns out that GSPCA has a breakdown value of $\lfloor (n+1)/2 \rfloor / n$, meaning that GSPCA can resist up to 50% contamination in a data set. We demonstrated that all radial functions except Winsor, had bounded influence functions which redescended to zero, implying the robustness of GSPCA. In addition to robustness properties, we studied the efficiency of GSPCA on Gaussian data. It turns out that Winsor is the most efficient version of GSPCA, followd by Quad, LR, Shell and Ball. A simulation study complements the theoretical results and confirms the robustness properties of GSPCA. Additionally, it showed that the efficiency properties of GSPCA greatly improve when the data is heavy-tailed.

GSPCA was further illustrated on two real-data examples. The first data set was the well-studied, moderately sized Top Gear data set. The second data set consisted of high-dimensional

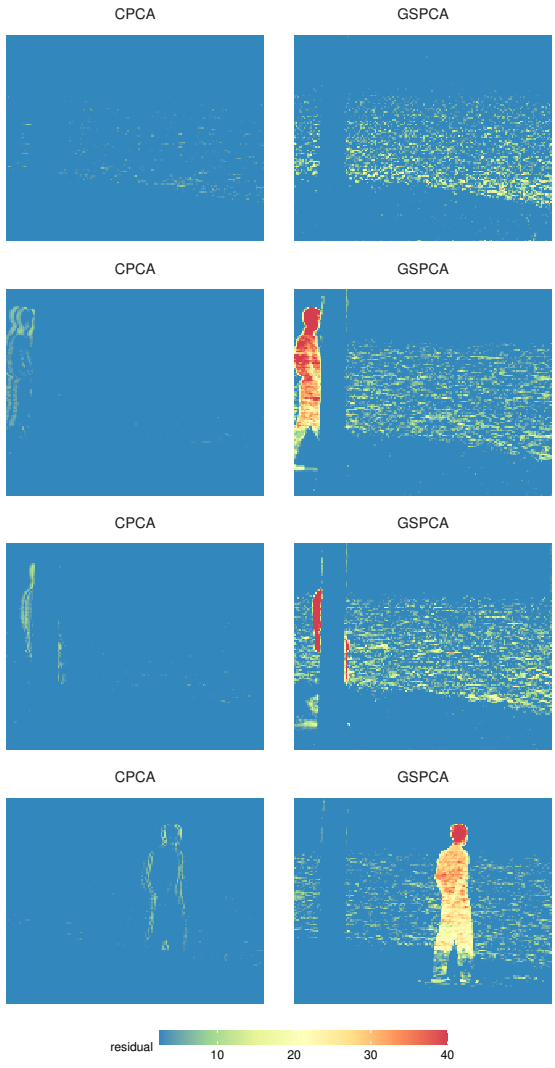


Fig. 14: Scaled residuals of data set **X2** for frames 108, 487, 491 and 564.

video data. For both examples, GSPCA demonstrated favorable robustness properties, whereas CPCA was clearly influenced by outliers.

In summary, we conclude that GSPCA is a valuable alternative for robust PCA. It is however important to combine GSPCA with the appropriate radial function. As a general advice without much information on the specific use case the LR radial function strikes a good balance between robustness and reasonable efficiency. In case efficiency is key, Winsor is advised whereas Ball would only be advised when very large amounts of contamination could be present in the data.

Further research directions could include using GSPCA to test the dimension of the subspace (Nordhausen et al., 2022), using GSPCA in principal component regression, or the development of higher-order GSPCA in the style of PARAFAC (Bro, 1997).

Acknowledgments. First author is supported by Fonds Wetenschappelijk onderzoek - Vlaanderen (FWO) as a PhD fellow Fundamental Research (PhD fellowship 11K5523N).

Appendix A Supplementary material

A.1 Proof of Theorem 1

We follow the train of thought of the proof in Raymaekers and Rousseeuw (2019):

Proof Part 1: $\varepsilon \geq \lfloor (n+1)/2 \rfloor / n$:

Given an $m < \lfloor (n+1)/2 \rfloor$, change m of the observations in $\boldsymbol{X}$ obtaining the contaminated data set $\mathbf{X}_m^*$. As location estimates we have $T(\boldsymbol{X})$ and $T(\mathbf{X}_m^*)$ and we define $c_1 = \max_i \|\boldsymbol{x}_i - T(\boldsymbol{X})\| < \infty$. Since m/n is smaller than the breakdown value of the location estimator, we have that there exist a constant c_2 such that $\|T(\boldsymbol{X}) - T(\mathbf{X}_m^*)\| \leq c_2 < \infty$. Hence by the triangle inequality, we get $d_i^* := \|\boldsymbol{x}_i - T(\mathbf{X}_m^*)\| \leq c_1 + c_2 < \infty$. Therefore we get $\text{med}_i(d_i^*) \leq c_1 + c_2$ and hence $\text{med}_i(d_i^*) + 1.4826 \cdot \text{MAD}_i(d_i^*) \leq 2.4826 \cdot \text{med}_i(d_i^*) \leq 2.4826 \cdot (c_1 + c_2)$. Using condition 3, this yields $\|g(t)\| \leq 2.4826 \cdot (c_1 + c_2)$. Now we compute:

$$
\begin{aligned}
\lambda_{\max} &= \sup_{\|\boldsymbol{u}\|=1} \boldsymbol{u}^\intercal S_g(\mathbf{X}_m^*)\boldsymbol{u} \\
&= \sup_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{u}^\intercal g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*)) g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))^\intercal \boldsymbol{u} \\
&= \sup_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} \left[\boldsymbol{u}^\intercal g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))\right]^2 \\
&\leq \sup_{\|\boldsymbol{u}\|=1} \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{u}\|^2 \left\|g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))\right\|^2 \\
&\leq \left(2.4826 \cdot (c_1 + c_2)\right)^2 < \infty.
\end{aligned}
$$

Hence we have shown that the largest eigenvalue of $S_g(\mathbf{X}_m^*)$ is bounded.

Part 2: $\varepsilon \leq \lfloor (n+1)/2 \rfloor / n$:

Given an $m > \lfloor (n+1)/2 \rfloor$, replace the last m observations from $\boldsymbol{X}$ yielding the contaminated data set $\mathbf{X}_m^* = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-m}, \boldsymbol{x}_{n-m+1}^*, \ldots, \boldsymbol{x}_n^*\}$. WLOG, by using location equivariance, we can assume that the mean of $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-m}$ is zero. For the other data points,

$j \in \{n-m+1,...,n\}$, we put $\boldsymbol{x}_j^* = \lambda \boldsymbol{a}_j$, with $\boldsymbol{a}_j$ such that $\min_{i\in\{n-m+1,...,n\}} \|\boldsymbol{a}_j - \boldsymbol{a}_i\| \geq 1$ and such that for all $\lambda > 1$: $\min_{i\in\{1,...,n-m\}} \|\lambda \boldsymbol{a}_j - \boldsymbol{x}_i\| \geq \lambda$. We can do this by placing the $\boldsymbol{a}_j$'s outside of the convex hull of $\boldsymbol{X}$, far enough apart from each other.

Further, we consider an increasing sequence $(\lambda_k)_k > 1$. Then for every λ_k the set $\{\boldsymbol{x}_{n-m+1}^*,...,\boldsymbol{x}_n^*\}$ must contain at least one point such that $\xi(\|\boldsymbol{x}_i^* - T(\mathbf{X}_m^*)\|) = 1$ by condition 2, say $\boldsymbol{x}_b^*$. The set $\mathbf{X}_m^*$ contains other points with weight 1, take one arbitrarily and call it $\boldsymbol{x}_c^*$. By the previous paragraph, we then have $\|\boldsymbol{x}_b^* - \boldsymbol{x}_c^*\| \geq \lambda$ and hence $\|\boldsymbol{x}_b^* - T(\mathbf{X}_m^*)\| + \|\boldsymbol{x}_c^* - T(\mathbf{X}_m^*)\| \geq \lambda$ and further $\|\boldsymbol{x}_b^* - T(\mathbf{X}_m^*)\|^2 + \|\boldsymbol{x}_c^* - T(\mathbf{X}_m^*)\|^2 \geq \lambda^2/2$. We can then compute:

$$
\begin{aligned}
&\textstyle\sum_{j=1}^p \lambda_j(S_g(\mathbf{X}_m^*)) = \operatorname{trace}(S_g(\mathbf{X}_m^*)) \\
&= \frac{1}{n}\sum_{i=1}^n \operatorname{trace}\big[g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))^\intercal\big] \\
&= \frac{1}{n}\sum_{i=1}^n \big\|g(\boldsymbol{x}_i^* - T(\mathbf{X}_m^*))\big\|^2 \\
&\geq \frac{1}{n}\Big(\big\|\boldsymbol{x}_b^* - T(\mathbf{X}_m^*)\big\|^2 + \big\|\boldsymbol{x}_c* - T(\mathbf{X}_m^*)\big\|^2\Big) \\
&\geq \lambda^2/(2n).
\end{aligned}
$$

For an unbounded increasing sequence of λ's this becomes arbitrarily large and hence $\varepsilon \leq \lfloor (n+1)/2 \rfloor / n$. □

A.2 Proof of Theorem 2

To proof the theorem, we use the following expression for the influence function of the GSSCM found in the paper of Raymaekers and Rousseeuw (2019):

$$
\begin{aligned}
\mathrm{IF}(\boldsymbol{x},\mathcal{S}_g,F) &= g(\boldsymbol{x})g(\boldsymbol{x})^\intercal - \mathcal{S}_g(F) \\
&\quad + \frac{\partial}{\partial\varepsilon}\int g_\varepsilon(\boldsymbol{X})g_\varepsilon(\boldsymbol{X})^\intercal dF(\boldsymbol{X})\Big|_{\varepsilon=0}. \qquad \text{(A1)}
\end{aligned}
$$

Here the last term can be expanded as:

$$
\begin{aligned}
&\frac{\partial}{\partial\varepsilon}\int g_\varepsilon(\boldsymbol{X})g_\varepsilon(\boldsymbol{X})^\intercal dF(\boldsymbol{X})\Big|_{\varepsilon=0} = \\
&\int\Big(\Big\{\frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})\Big|_{\varepsilon=0}\Big\}g(\boldsymbol{X})^\intercal + g(\boldsymbol{X})\Big\{\frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})^\intercal\Big|_{\varepsilon=0}\Big\}\Big)dF(\boldsymbol{X})
\end{aligned}
$$

To ease the notational burden, we write $dge(\boldsymbol{x})$ for the derivative to ε of $g_\varepsilon(\boldsymbol{x})$ in $\varepsilon = 0$, for which we have:

$$
\begin{aligned}
dge(\boldsymbol{x}) &= \frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})\Big|_{\varepsilon=0} \\
&= \boldsymbol{X}\frac{\partial}{\partial\varepsilon}\xi_\varepsilon(\|\boldsymbol{X}\|)\Big|_{\varepsilon=0}. \qquad \text{(A2)}
\end{aligned}
$$

Explicit expressions for Equation (A2) per radial function can be found in appendix A.3 of the GSSCM paper (Raymaekers and Rousseeuw, 2019).

Additionally, we make use of following lemma published in Croux and Haesbroeck (2000) to compute the influence functions of the eigenvectors of the GSSCM:

Lemma 2 (Croux & Haesbroeck 2000) *Given $\mathcal{S} : F \to SPD(p)$ a statistical functional such that $IF(\boldsymbol{x},\mathcal{S},F)$ exists. Let $\boldsymbol{v}_{g,1},...,\boldsymbol{v}_{g,p}$ and $\lambda_{g,1},...,\lambda_{g,p}$ be the eigenvectors and eigenvalues of $\mathcal{S}(F)$. Then the influence function of $\boldsymbol{v}_{g,j}$ is given by:*

$$
IF(\boldsymbol{x},\mathcal{V}_{g,j},F) = \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big(\boldsymbol{v}_{g,k}^\intercal IF(\boldsymbol{x},\mathcal{S},F)\boldsymbol{v}_{g,j}\Big)\boldsymbol{v}_{g,k}.
$$

The proof for this lemma follows from Lemma 2.1 in Sibson (1979). We are now ready to proof the theorem.

Proof Making use of Lemma 2 and Equation (A1) for the influence function of the GSSCM, we note:

$$
\begin{aligned}
&\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,j},F) \\
&= \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big(\boldsymbol{v}_{g,k}^\intercal \mathrm{IF}(\boldsymbol{x},\mathcal{S}_g,F)\boldsymbol{v}_{g,j}\Big)\boldsymbol{v}_{g,k} \\
&= \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big(\boldsymbol{v}_{g,k}^\intercal g(\boldsymbol{x})g(\boldsymbol{x})^\intercal \boldsymbol{v}_{g,j} - \boldsymbol{v}_{g,k}^\intercal \mathcal{S}_g(F)\boldsymbol{v}_{g,j} \\
&\qquad + \boldsymbol{v}_{g,k}^\intercal \frac{\partial}{\partial\varepsilon}\int g_\varepsilon(\boldsymbol{X})g_\varepsilon(\boldsymbol{X})^\intercal dF(\boldsymbol{X})\Big|_{\varepsilon=0}\boldsymbol{v}_{g,j}\Big)\boldsymbol{v}_{g,k} \\
&= \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big((\boldsymbol{v}_{g,k}^\intercal g(\boldsymbol{x}))(\boldsymbol{v}_{g,j}^\intercal g(\boldsymbol{x})) - \boldsymbol{v}_{g,k}^\intercal \lambda_{g,j}\boldsymbol{v}_{g,j} \\
&\qquad + \boldsymbol{v}_{g,k}^\intercal \frac{\partial}{\partial\varepsilon}\int g_\varepsilon(\boldsymbol{X})g_\varepsilon(\boldsymbol{X})^\intercal dF(\boldsymbol{X})\Big|_{\varepsilon=0}\boldsymbol{v}_{g,j}\Big)\boldsymbol{v}_{g,k} \\
&= \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Bigg((\boldsymbol{v}_{g,k}^\intercal g(\boldsymbol{x}))(\boldsymbol{v}_{g,j}^\intercal g(\boldsymbol{x})) - 0 \\
&\qquad + \boldsymbol{v}_{g,k}^\intercal \int\Big(\Big\{\frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})\Big|_{\varepsilon=0}\Big\}g(\boldsymbol{X})^\intercal \\
&\qquad + g(\boldsymbol{X})\Big\{\frac{\partial}{\partial\varepsilon}g_\varepsilon(\boldsymbol{X})^\intercal\Big|_{\varepsilon=0}\Big\}\Big)dF(\boldsymbol{X})\ \boldsymbol{v}_{g,j}\Bigg)\boldsymbol{v}_{g,k} \\
&= \sum_{k=1,k\neq j}^p \frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big((\boldsymbol{v}_{g,k}^\intercal g(\boldsymbol{x}))(\boldsymbol{v}_{g,j}^\intercal g(\boldsymbol{x}))
\end{aligned}
$$

$$+\boldsymbol{v}_{g,k}^{\intercal}\int\big(dge(\boldsymbol{X})g(\boldsymbol{X})^{\intercal} +g(\boldsymbol{X})dge(\boldsymbol{X})^{\intercal})dF(\boldsymbol{X})\ \boldsymbol{v}_{g,j}\Big)\boldsymbol{v}_{g,k}.$$

□

A.3 Proof of Corollary 1

Proof When we assume that F is a centered elliptically symmetric distribution with diagonal covariance matrix, we have that $\boldsymbol{v}_{g,j} = \boldsymbol{v}_j = \boldsymbol{e}_j$. Hence Equation (8) simplifies to:

$$\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,j},F) = \sum_{k=1,k\neq j}^{p}\frac{1}{\lambda_{g,j}-\lambda_{g,k}}\Big(g(\boldsymbol{x})_k g(\boldsymbol{x})_j + \int\{dge(\boldsymbol{X})_k\ g(\boldsymbol{X})_j + g(\boldsymbol{X})_k\ dge(\boldsymbol{X})_j\}dF(\boldsymbol{X})\Big)\boldsymbol{v}_k. \tag{A3}$$

Next we proof that the above integral in the second term equals zero. For this, we first know that $g(\boldsymbol{x}) = \boldsymbol{x}\cdot\xi(\|\boldsymbol{x}\|)$, hence $g(\boldsymbol{x})_i = \boldsymbol{x}_i\cdot\xi(\|\boldsymbol{x}\|)$. Second, if we study appendix A.3 of the GSSCM paper (Raymaekers and Rousseeuw, 2019), one can find that for every suggested radial function $dge(\boldsymbol{x})$ is of the form $dge(\boldsymbol{x}) = \text{scalar}\cdot\boldsymbol{x}\cdot f(\|\boldsymbol{x}\|)$. Therefore, we have that $dge(\boldsymbol{x})_i = \text{scalar}\cdot\boldsymbol{x}_i\cdot f(\|\boldsymbol{x}\|)$. Last, the density function of a centered elliptically symmetric distribution has the following form: $f_{\boldsymbol{X}}(\boldsymbol{x}) = \det(\boldsymbol{\Sigma})^{-1/2}\cdot h(\boldsymbol{x}^{\intercal}\boldsymbol{\Sigma}^{-1}\boldsymbol{x})$ with $\boldsymbol{\Sigma}$ diagonally assumed here, hence it is an even function in each of its variables $\boldsymbol{x}_i$. Putting all this together, the integral in the second term of Equation (A3) is of following form

$$\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\boldsymbol{x}_k\boldsymbol{x}_j\cdot f(\|\boldsymbol{x}\|)\xi(\|\boldsymbol{x}\|)\cdot\omega(\boldsymbol{x})\cdot d\boldsymbol{x}_1 d\boldsymbol{x}_2 ... d\boldsymbol{x}_p$$

for $k \neq j$,

where $\omega(\boldsymbol{x})$ is an even function in each variable. Then one has that this integral is equal to zero because of symmetry reasons. Hence the second term in Equation (A3) becomes zero for centered elliptically symmetric distributions. □

A.4 Proof of Theorem 3

Using Lemma 1, we can find an expression for the influence functions of our new eigenvalues:

Proof

$$\mathcal{L}_{g,k}^{(\mathfrak{s})}(H) = s^2(\boldsymbol{v}_{g,k}^{\intercal}\boldsymbol{X}) = \mathfrak{s}^2(H^{\boldsymbol{v}_{g,k}})$$

$$\Longrightarrow \mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,k}^{(\mathfrak{s})},H) = \frac{d}{d\varepsilon}\Big(\mathfrak{s}^2\Big(H_{\varepsilon,\boldsymbol{x}}^{\boldsymbol{v}_{g,k}(H_{\varepsilon,\boldsymbol{x}})}\Big)\Big)\Big|_{\varepsilon=0}$$

$$= \frac{d}{d\varepsilon}\Big(\mathfrak{s}^2(H_{\varepsilon,\boldsymbol{x}}^{\boldsymbol{v}_k})\Big)\Big|_{\varepsilon=0} + \left(\frac{d}{d\boldsymbol{a}}\Big(\mathfrak{s}^2(H^{\boldsymbol{a}})\Big)\Big|_{\boldsymbol{a}=\boldsymbol{v}_{g,k}}\right)^{\intercal}\cdot\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H).$$

We can use that $\mathfrak{s}^2(H^{\boldsymbol{a}}) = \boldsymbol{a}^{\intercal}\boldsymbol{\Sigma}\boldsymbol{a}$, as $\mathfrak{s}$ is equivariant, which yields:

$$\frac{d}{d\boldsymbol{a}}\Big(\mathfrak{s}^2(H^{\boldsymbol{a}})\Big)\Big|_{\boldsymbol{a}=\boldsymbol{v}_{g,k}} = \left(\frac{d}{d\boldsymbol{a}}(\boldsymbol{a}^{\intercal}\boldsymbol{\Sigma}\boldsymbol{a})\Big|_{\boldsymbol{a}=\boldsymbol{v}_{g,k}}\right) = 2\boldsymbol{\Sigma}\boldsymbol{v}_{g,k} = 2\boldsymbol{\Sigma}\boldsymbol{v}_k = 2\lambda_k\boldsymbol{v}_k.$$

We can then continue our calculation:

$$\mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,k}^{(\mathfrak{s})},H) = \frac{d}{d\varepsilon}\Big(\mathfrak{s}^2(H_{\varepsilon,\boldsymbol{x}}^{\boldsymbol{v}_k})\Big)\Big|_{\varepsilon=0} + 2\lambda_k\boldsymbol{v}_k^{\intercal}\cdot\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H) = \mathrm{IF}(\boldsymbol{v}_k^{\intercal}\boldsymbol{x},\mathfrak{s}^2,H^{\boldsymbol{v}_k}) + 2\lambda_k\boldsymbol{v}_k^{\intercal}\cdot\mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H).$$

When we calculated the influence function of the eigenvector $\mathcal{V}_{g,k}$ of the GSSCM, we saw that this function had no component in the direction of $\boldsymbol{v}_k$. Therefore we can remove the second term. Hence we obtain:

$$\mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,k}^{(\mathfrak{s})},H) = \mathrm{IF}(\boldsymbol{v}_k^{\intercal}\boldsymbol{x},\mathfrak{s}^2,H^{\boldsymbol{v}_k}) = \boldsymbol{v}_k^{\intercal}\boldsymbol{\Sigma}\boldsymbol{v}_k\ \mathrm{IF}\left(\frac{\boldsymbol{v}_k^{\intercal}\boldsymbol{x}}{\sqrt{\boldsymbol{v}_k^{\intercal}\boldsymbol{\Sigma}\boldsymbol{v}_k}},\mathfrak{s}^2,F_0\right) = 2\lambda_k\ \mathrm{IF}\left(\frac{\boldsymbol{v}_k^{\intercal}\boldsymbol{x}}{\sqrt{\lambda_k}},\mathfrak{s},F_0\right).$$

□

A.5 Proof of Theorem 4

Proof

$$\mathrm{IF}(\boldsymbol{x},\mathcal{S}_{\mathrm{comb}},H) = \frac{d}{d\varepsilon}\left(\sum_{k=1}^{p}\mathcal{L}_{g,k}^{(\mathfrak{s})}(H_{\varepsilon,\boldsymbol{x}})\mathcal{V}_{g,k}(H_{\varepsilon,\boldsymbol{x}})\mathcal{V}_{g,k}(H_{\varepsilon,\boldsymbol{x}})^{\intercal}\right)\Big|_{\varepsilon=0}$$

$$= \sum_{k=1}^{p}\Big[\mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,k}^{(\mathfrak{s})},H)\boldsymbol{v}_k\boldsymbol{v}_k^{\intercal} + \lambda_k\ \mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H)\ \boldsymbol{v}_k^{\intercal} + \lambda_k\boldsymbol{v}_k\ \mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H)^{\intercal}\Big].$$

Assuming that the original covariance matrix is diagonal and that H is elliptically symmetric, we can use Corollary 1 and Equation (10) to obtain:

$$\mathrm{IF}(\boldsymbol{x},\mathcal{S}_{\mathrm{comb}},H) = \sum_{k=1}^{p}\Big[\mathrm{IF}(\boldsymbol{x},\mathcal{L}_{g,k}^{(\mathfrak{s})},H)\boldsymbol{v}_k\boldsymbol{v}_k^{\intercal} + \lambda_k\ \mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H)\ \boldsymbol{v}_k^{\intercal} + \lambda_k\boldsymbol{v}_k\ \mathrm{IF}(\boldsymbol{x},\mathcal{V}_{g,k},H)^{\intercal}\Big]$$

$$= \sum_{k=1}^{p}2\lambda_k\ \mathrm{IF}\left(\frac{\boldsymbol{v}_k^{\intercal}\boldsymbol{x}}{\sqrt{\lambda_k}},\mathfrak{s},F_0\right)\boldsymbol{v}_k\boldsymbol{v}_k^{\intercal} + \sum_{k=1}^{p}\lambda_k\Big[\sum_{j=1,j\neq k}^{p}\frac{1}{\lambda_{g,k}-\lambda_{g,j}}(g(\boldsymbol{x})_j g(\boldsymbol{x})_k)\boldsymbol{v}_j\boldsymbol{v}_k^{\intercal}$$

$$+\sum_{j=1,j\neq k}^{p}\frac{1}{\lambda_{g,k}-\lambda_{g,j}}(g(\boldsymbol{x})_j g(\boldsymbol{x})_k)\boldsymbol{v}_k\boldsymbol{v}_j^\intercal\Big]$$
$$=2\sum_{k=1}^{p}\lambda_k\ \mathrm{IF}\left(\frac{\boldsymbol{v}_k^\intercal\boldsymbol{x}}{\sqrt{\lambda_k}},\mathfrak{s},F_0\right)\boldsymbol{v}_k\boldsymbol{v}_k^\intercal$$
$$+\sum_{k=1}^{p}\lambda_k\sum_{j=1,j\neq k}^{p}\frac{1}{\lambda_{g,k}-\lambda_{g,j}}(g(\boldsymbol{x})_j g(\boldsymbol{x})_k)(\boldsymbol{v}_j\boldsymbol{v}_k^\intercal+\boldsymbol{v}_k\boldsymbol{v}_j^\intercal)$$

□

A.6 Efficiency on heavy-tailed distributions

We repeat the analysis of Section 3.3 on the asymptotic variance and relative efficiency of GSPCA for heavy-tailed distributions. We focus in particular on the multivariate t_3-distribution. For this, we need a variation on Proposition 5 of Raymaekers and Rousseeuw (2019). For X a p-dimensional t-distributed random variable with scale matrix $\boldsymbol{\Sigma}$ and ν degrees of freedom we have:

$$\|\boldsymbol{X}\|^2=\left\|\sqrt{\nu/U}\ \boldsymbol{Y}\right\|^2=\nu/U\ \|\boldsymbol{Y}\|^2,$$

with $U\sim\chi^2_\nu$ and $\boldsymbol{Y}\sim\mathcal{N}(\boldsymbol{0}_p,\boldsymbol{\Sigma})$. The distribution of $\|\boldsymbol{Y}\|^2$ is a convolution of gamma distributions. We thus obtain that the distribution of the Euclidean distance for a multivariate t-distribution is a scaled ratio of a convolution of gamma distributions and a chi-square distribution. Using this result, we can compute the asymptotic variances of GSPCA for the multivariate t-distribution. For a bivariate t_3-distribution with covariance matrix diag(1,γ), we obtain the asymptotic variances and relative efficiencies shown in Figure A1. We can clearly see a very different picture now. While the relative performances of the GSPCA variants have remained stable, the asymptotic variance of classical PCA exploded. As a result, the asymptotic relative efficiencies of GSPCA all become larger than 1. The relative efficiencies are somewhat wobbly due to the tricky numerical integration, but the overall picture is clear.

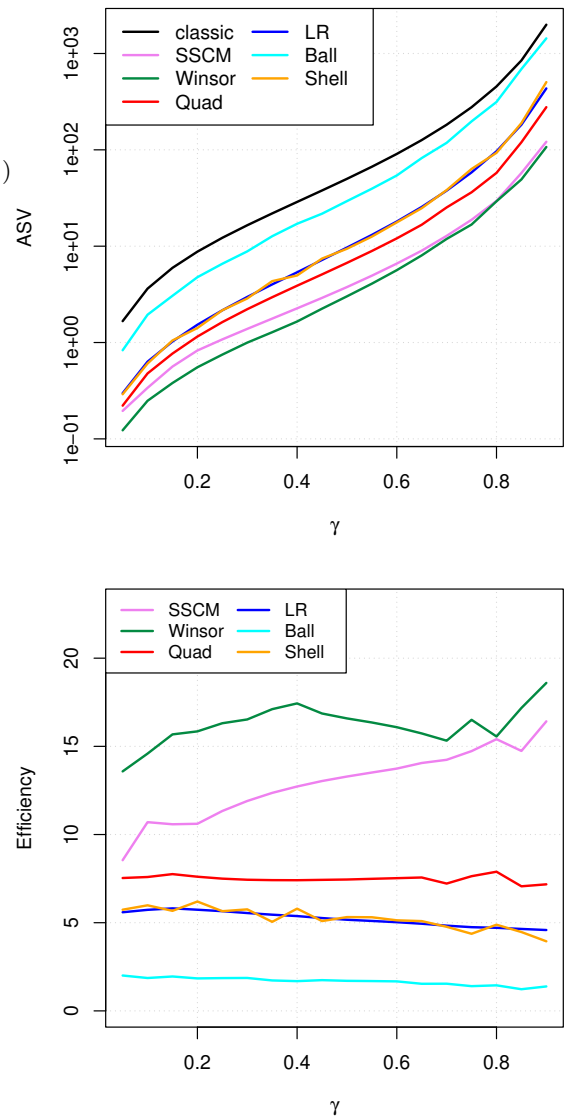


Fig. A1: Asymptotic variances (top) and relative efficiencies (bottom) for a bivariate t_3-distribution with covariance matrix diag(1,γ).

A.7 Simulations for increasing contamination levels

Given that the asymptotic explosion breakdown value of GSPCA is 50%, it is interesting to investigate how GSPCA reacts to contamination levels close to that number. Of course, the higher the contamination level, the more difficult the problem gets in general. In order to illustrate this

empirically, we rerun part of the simulation study with increasing levels of contamination.

Figure A2 below shows the result for low dimensional Gaussian data with $\varepsilon \in \{0.2, 0.25, 0.3, 0.4, 0.45, 0.49\}$. It is clear that all methods slowly deteriorate when we increase the level of contamination ε. ROBPCA with the default robustness parameter $\alpha = 0.75$ fails when $\varepsilon \geq 0.25$. LR, Shell, Ball and ROBPCA with $\alpha = 0.5$ maintain a high level of resistance until contamination reaches 40%, beyond which they start to fail as we approach the asymptotic breakdown value of 50%.

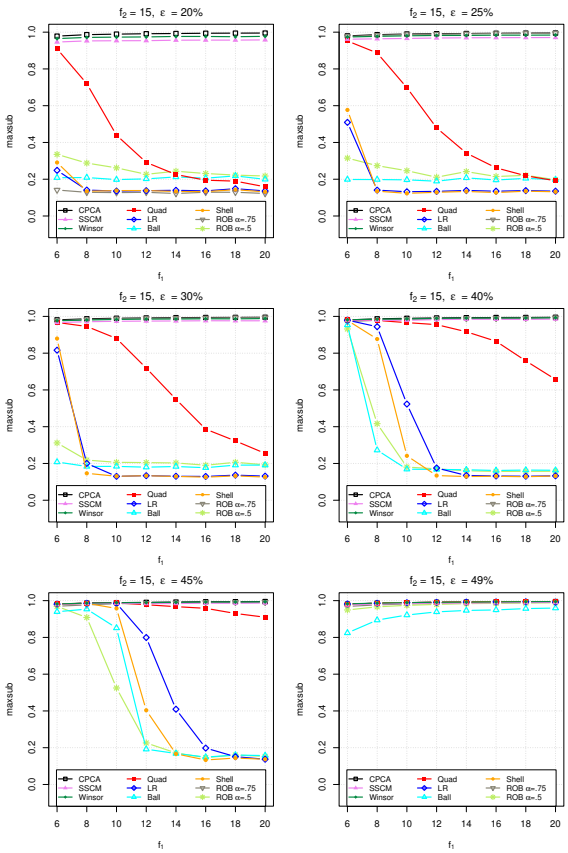


Fig. A2: Low dimensional, multivariate normal data with increasing levels of contamination.

We repeated the same exercise for the high-dimensional Gaussian setting, the results of which are presented in Figure A3. We can see that the relative performances are largely similar. To have any chance at obtaining a reasonable estimate in this setting, the radial function needs to be hard-redescending. As we reach the breakdown point, all methods start to fail. ROBPCA performs best here, which is somewhat expected given its use of the MCD and Stahel-Donoho outlyingness.

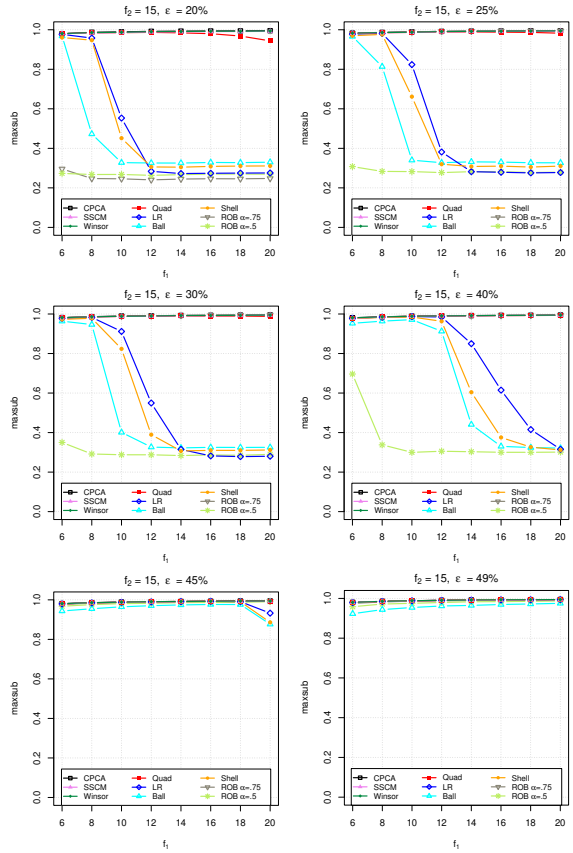


Fig. A3: High dimensional, multivariate normal data with increasing levels of contamination.

References

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6): 1382–1408, 2006.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. 2005.

P. J. Bickel, G. Kur, and B. Nadler. Projection pursuit in high dimensions. *Proceedings*

of the National Academy of Sciences, 115(37): 9151–9156, 2018.

N. Billor, A. S. Hadi, and P. F. Velleman. Bacon: blocked adaptive computationally efficient outlier nominators. *Computational statistics & data analysis*, 34(3):279–298, 2000. doi: https://doi.org/10.1016/S0167-9473(99)00101-2.

G. Boente. Asymptotic theory for robust principal components. *Journal of Multivariate Analysis*, 21(1):67–78, 1987. doi: https://doi.org/10.1016/0047-259X(87)90099-6.

G. Boente, D. Rodriguez, and M. Sued. The spatial sign covariance operator: Asymptotic results and applications. *Journal of Multivariate Analysis*, 170:115–128, 2019. doi: https://doi.org/10.1016/j.jmva.2018.10.002.

R. Bro. Parafac. tutorial and applications. *Chemometrics and Intelligent Laboratory Systems*, 38(2):149–171, 1997. ISSN 0169-7439. doi: https://doi.org/10.1016/S0169-7439(97)00032-4. URL https://www.sciencedirect.com/science/article/pii/S0169743997000324.

R. Butler, P. Davies, and M. Jhun. Asymptotics for the minimum covariance determinant estimator. *The Annals of Statistics*, pages 1385–1400, 1993. doi: https://doi.org/10.1214/aos/1176349264.

T. T. Cai, T. Liang, and H. H. Zhou. Law of log determinant of sample covariance matrix and optimal estimation of differential entropy for high-dimensional gaussian distributions. *Journal of Multivariate Analysis*, 137:161–172, 2015. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2015.02.003. URL https://www.sciencedirect.com/science/article/pii/S0047259X1500038X.

N. A. Campbell. Robust procedures in multivariate analysis i: Robust covariance estimation. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 29(3):231–237, 1980. doi: https://doi.org/10.2307/2346896.

E. A. Cator and H. P. Lopuhaä. Central limit theorem and influence function for the mcd estimators at general multivariate distributions. *Bernoulli*, 18(2):520–551, 2012. doi: https://www.jstor.org/stable/23238609.

C. Croux and G. Haesbroeck. Influence function and efficiency of the minimum covariance determinant scatter matrix estimator. *Journal of Multivariate Analysis*, 71(2):161–190, 1999. doi: https://doi.org/10.1006/jmva.1999.1839.

C. Croux and G. Haesbroeck. Principal component analysis based on robust estimators of the covariance or correlation matrix: Influence functions and efficiencies. *Biometrika*, 87(3): 603–618, 09 2000. doi: https://doi.org/10.1093/biomet/87.3.603.

C. Croux and A. Ruiz-Gazen. High breakdown estimators for principal components: the projection-pursuit approach revisited. *Journal of Multivariate Analysis*, 95:206–226, 2005. doi: https://doi.org/10.1016/j.jmva.2004.08.002.

C. Croux, E. Ollila, and H. Oja. Sign and rank covariance matrices: statistical properties and application to principal components analysis. In *Statistical data analysis based on the L1-norm and related methods*, pages 257–269. Springer, 2002. doi: 10.1007/978-3-0348-8201-9_22.

C. Croux, P. Filzmoser, and M. R. Oliveira. Algorithms for projection–pursuit robust principal component analysis. *Chemometrics and Intelligent Laboratory Systems*, 87(2):218–225, 2007. doi: https://doi.org/10.1016/j.chemolab.2007.01.004.

C. Croux, C. Dehon, and A. Yadine. The k-step spatial sign covariance matrix. *Advances in Data Analysis and Classification*, 4:137–150, 2010. doi: https://doi.org/10.1007/s11634-010-0062-7.

L. Davies. The asymptotics of rousseeuw's minimum volume ellipsoid estimator. *The Annals of Statistics*, 20(4):1828–1843, 1992. doi: https://doi.org/10.1214/aos/1176348891.

P. L. Davies. Asymptotic behaviour of s-estimates of multivariate location parameters and dispersion matrices. *The Annals of Statistics*, pages 1269–1292, 1987. doi: https://www.jstor.org/stable/2241828.

A. Dürre, D. Vogel, and D. E. Tyler. The spatial sign covariance matrix with unknown location. *Journal of Multivariate Analysis*, 130:107–117, 2014. doi: https://doi.org/10.1016/j.jmva.2014.05.004.

A. Dürre, D. E. Tyler, and D. Vogel. On the eigenvalues of the spatial sign covariance matrix in more than two dimensions. *Statistics & Probability Letters*, 111:80–85, 2016. doi: https://doi.org/10.1016/j.spl.2016.01.009.

F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. John

Wiley & Sons, 1986.

M. Hubert, P. Rousseeuw, and K. Vanden Branden. Robpca: A new approach to robust principal component analysis. *Technometrics*, 47(1):64–79, 02 2005. doi: 10.1198/004017004000000563.

M. Hubert, P. Rousseeuw, and K. Vakili. Shape bias of robust covariance estimators: an empirical study. *Statistical Papers*, 55:15–28, 2014.

I. M. Johnstone and D. Paul. Pca in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292, 2018.

I. M. Johnstone, A. Y. Lu, B. Nadler, D. M. Witten, T. Hastie, R. Tibshirani, and J. O. Ramsay. On consistency and sparsity for principal components analysis in high dimensions [with comments]. *Journal of the American Statistical Association*, 104(486):682–703, 2009. ISSN 01621459. URL http://www.jstor.org/stable/40592215.

G. Li and Z. Chen. Projection-pursuit approach to robust dispersion matrices and principal components: primary theory and monte carlo. *Journal of the American Statistical Association*, 80 (391):759–766, 1985. doi: https://doi.org/10.2307/2288497.

L. Li, W. Huang, I. Y.-H. Gu, and Q. Tian. Statistical modeling of complex backgrounds for foreground object detection. *IEEE Transactions on Image Processing*, 13(11):1459–1472, 2004. doi: 10.1109/TIP.2004.836169.

N. Locantore, J. Marron, D. Simpson, N. Tripoli, J. Zhang, and K. Cohen. Robust principal component analysis for functional data. *Test*, 8(1): 1–73, 06 1999. doi: https://doi.org/10.1007/BF02595862.

G. Louvet, J. Raymaekers, G. Van Bever, and I. Wilms. The influence function of graphical lasso estimators. *Econometrics and Statistics*, 2023. ISSN 2452-3062. doi: https://doi.org/10.1016/j.ecosta.2023.03.004. URL https://www.sciencedirect.com/science/article/pii/S2452306223000266.

A. F. Magyar and D. E. Tyler. The asymptotic inadmissibility of the spatial sign covariance matrix for elliptically symmetric distributions. *Biometrika*, 101(3):673–688, 2014. doi: https://doi.org/10.1093/biomet/asu020.

J. I. Marden. Some robust estimates of principal components. *Statistics & probability letters*, 43(4):349–359, 1999. doi: https://doi.org/10.1016/S0167-7152(98)00272-7.

K. Nordhausen, H. Oja, and D. E. Tyler. Asymptotic and bootstrap tests for subspace dimension. *Journal of Multivariate Analysis*, 188:104830, 2022. ISSN 0047-259X. doi: https://doi.org/10.1016/j.jmva.2021.104830. URL https://www.sciencedirect.com/science/article/pii/S0047259X21001081. 50th Anniversary Jubilee Edition.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

A. Pires and J. Branco. High dimensionality: The latest challenge to data analysis. *arXiv preprint arXiv:1902.04679*, 2019.

J. Raymaekers and P. Rousseeuw. A generalized spatial sign covariance matrix. *Journal of Multivariate Analysis*, 171:94–111, 2019. doi: https://doi.org/10.1016/j.jmva.2018.11.010.

P. Rousseeuw and V. Yohai. Robust regression by means of s-estimators. In *Robust and nonlinear time series analysis*, pages 256–272. Springer, 1984. doi: https://doi.org/10.1007/978-1-4615-7821-5_15.

P. J. Rousseeuw. Least median of squares regression. *Journal of the American statistical association*, 79(388):871–880, 1984. doi: {10.1080/01621459.1984.10477105}.

P. J. Rousseeuw, J. Raymaekers, and M. Hubert. A measure of directional outlyingness with applications to image data and video. *Journal of Computational and Graphical Statistics*, 27 (2):345–359, 2018. doi: 10.1080/10618600.2017.1366912.

R. Sibson. Studies in the robustness of multidimensional scaling: Perturbational analysis of classical scaling. *Journal of the royal statistical society series b-methodological*, 41(2):217–229, 1979.

S. Taskinen, I. Koch, and H. Oja. Robustifying principal component analysis with spatial sign vectors. *Statistics & Probability Letters*, 82(4): 765–774, 2012. doi: https://doi.org/10.1016/j.spl.2012.01.001.

S. Visuri, H. Oja, and V. Koivunen. Subspace-based direction-of-arrival estimation using nonparametric statistics. *IEEE Transactions on Signal Processing*, 49(9):2060–2073, 2001. doi: 10.1109/78.942634.