Robust Sliced Inverse Regression Procedures

Ursula Gather, Torsten Hilker, and Claudia Becker

Abstract

Sliced Inverse Regression (SIR) is a promising technique for the purpose of dimension reduction. Several properties of this relatively new method have been examined already, but little attention has been paid to robustness aspects. We show that SIR is very sensitive towards outliers in the data. Therefore a generalized estimation procedure which allows for robustness properties, especially for a high breakdown point, is proposed.

AMS 1991 Subject Classifications: Primary 62F35; secondary 62H12.
Key words: Dimension reduction; Outliers; High breakdown procedures.

1 Introduction

Many statistical methods suffer from the so–called "curse of dimensionality". As an example, consider classical nonparametric regression techniques which are used if a flexible modelling of a completely unknown functional relationship is the goal of the data analysis. It is well known that such methods (mainly nearest neighbor techniques) work successfully if the explanatory variable is low–dimensional. The theoretical concepts can be transferred to the situation of a high–dimensional explanatory variable, but it is impossible to use the resulting methods practically. The reason for this is that with samples of realistic size the high–dimensional euclidean space cannot be filled densely enough with observations. Consequently, locally, there are not sufficiently many observations to keep both the bias, and the variance of the classical nonparametric regression estimates small.
One possibility to overcome this problem is the construction of methods which are especially designed for the situation of a high–dimensional explanatory variable such as Projection Pursuit Regression (Friedman and Stuetzle, 1981) and MARS (Friedman, 1991).
Another way is offered by techniques for reducing the dimension of the regressor space first before applying usual nonparametric regression techniques. Led by this idea, Li (1991)

considers the following model for dimension reduction

$$y = g(\boldsymbol{\beta}_1'\boldsymbol{x}, \ldots, \boldsymbol{\beta}_K'\boldsymbol{x}, \varepsilon), \tag{1.1}$$

where y denotes a real-valued response variable, $\boldsymbol{x}$ is the d-dimensional random vector of explanatory variables, and ε is a random error. The link function g is assumed to be completely unknown. The unknown directions $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ are called effective dimension reduction (edr) directions. They span the edr space $\mathcal{B}$ and have to be estimated from the data. For this purpose Li (1991) suggests Sliced Inverse Regression (SIR), which uses the fact that under certain conditions concerning the distribution of $\boldsymbol{x}$, the centered inverse regression curve $E(\boldsymbol{x}|y) - E(\boldsymbol{x})$ falls into a subspace spanned by $\boldsymbol{\Sigma}\boldsymbol{\beta}_k$, $k = 1, \ldots, K$, where $\boldsymbol{\Sigma}$ denotes the covariance matrix of $\boldsymbol{x}$, cf. Li (1991, Theorem 3.1). Estimating the edr directions based upon obervations $(\boldsymbol{x}_i, y_i)$, $i = 1, \ldots, n$, of $(\boldsymbol{x}, y)$ is then carried out in five steps:

1. Standardization: $\boldsymbol{z}_i = \widehat{\boldsymbol{\Sigma}}^{-1/2}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})$, $i = 1, \ldots, n$, where $\widehat{\boldsymbol{\Sigma}}$ and $\overline{\boldsymbol{x}}$ denote the sample covariance matrix and the sample mean of $\boldsymbol{x}_i$, $i = 1, \ldots, n$, respectively.

2. Slicing: Divide the range of y into H slices $I_1, \ldots, I_H$. Let n_h be the number of observations of y which fall into I_h, $h = 1, \ldots, H$, and let $p_h = n_h/n$.

3. Sample mean-vector within each slice: $\widehat{\boldsymbol{m}}_h = \sum_{y_i \in I_h} \boldsymbol{z}_i / n_h$.

4. Principal Component Analysis (PCA): Let $\widehat{\boldsymbol{V}} = \sum_{h=1}^{H} p_h \widehat{\boldsymbol{m}}_h \widehat{\boldsymbol{m}}_h'$ and compute the eigenvalues $\widehat{\lambda}_1 \geq \cdots \geq \widehat{\lambda}_d$ and corresponding eigenvectors $\widehat{\boldsymbol{\eta}}_1, \ldots, \widehat{\boldsymbol{\eta}}_d$ of $\widehat{\boldsymbol{V}}$ with $\widehat{\boldsymbol{\eta}}_i'\widehat{\boldsymbol{\eta}}_j = \delta_{ij}$ where δ_{ij} denotes the Kronecker symbol.

5. Retransformation: The edr directions are estimated by $\widehat{\boldsymbol{\beta}}_k = \widehat{\boldsymbol{\Sigma}}^{-1/2}\widehat{\boldsymbol{\eta}}_k$, $k = 1, \ldots, K$.

The matrix $\boldsymbol{\Sigma}$ and $\mu = E(\boldsymbol{x})$ can be used instead of $\widehat{\boldsymbol{\Sigma}}$ and $\bar{\boldsymbol{x}}$ if they are known. Because of the partition of the data into slices and the estimation of the inverse regression curve, this estimation scheme is called Sliced Inverse Regression (SIR) (Li, 1991). The slices are very often determined by choosing the number H of slices beforehand and by partitioning the sample size n into slice sizes $n_1, \ldots, n_H$ with $\sum_{h=1}^{H} n_h = n$. Let $y_{(1)} \leq \cdots \leq y_{(n)}$ be the ordered observations of y. The n_1 smallest observations of y are used to form the first slice I_1, the next n_2 smallest observations are used for the second slice I_2, and so on. For our considerations we will adopt this procedure. Usually, slices with equal sizes are chosen.

Remark 1.1 *The matrix* $\widehat{\boldsymbol{V}}$ *is given by* $\widehat{\boldsymbol{V}} = \widehat{\boldsymbol{\Sigma}}^{-1/2} \sum_{h=1}^{H} \widehat{p}_h (\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{x}})' \widehat{\boldsymbol{\Sigma}}^{-1/2}$ *with* $\overline{\boldsymbol{x}}_h = \sum_{i:y_i \in I_h} \boldsymbol{x}_i / n_h$, $h = 1, \ldots, H$. *Therefore, the estimated edr directions are also eigenvectors of*

$$\widetilde{\boldsymbol{V}} = \widehat{\boldsymbol{\Sigma}}^{-1} \sum_{h=1}^{H} \widehat{p}_h (\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{x}})(\overline{\boldsymbol{x}}_h - \overline{\boldsymbol{x}})'. \tag{1.2}$$

More details and comments on the estimation scheme can be found in Li (1991). SIR has been discussed in several articles with emphasis on its asymptotic properties (Li, 1991; Hsing and Carroll, 1992; Kötter, 1996; Zhu and Fang, 1996) and on the estimation of the dimension K of the edr space (Li, 1991; Schott, 1994; Ferré, 1997). Since design conditions concerning the distribution of $\boldsymbol{x}$ are essential for the application of SIR, cf. Li (1991, Condition 3.1), this aspect is another main topic in the discussion, cf. Cook and Weisberg (1991), Härdle and Tsybakov (1991), Hall and Li (1993), Cook and Nachtsheim (1994). Carroll and Li (1992) use SIR in a general nonlinear regression model with measurement error in the predictor variables. Aragon and Saracco (1997) consider SIR in the situation of small sample sizes. Bura (1997) uses a multivariate linear model for the inverse regression curve. Two nonparametric methods for testing the hypothesis $H_0 : K = 1$ versus the alternative $H_1 : K > 1$ are proposed by Sheather and McKean (1997). SIR is implemented in the package XploRe (Härdle, Klinke and Turlach, 1995). There are also more sophisticated methods for estimating the edr direction which are based on second moments (Cook and Weisberg, 1991) and on properties of so called principal Hessian directions (Li, 1992).

One aspect of SIR which has not been treated yet is the sensitivity of this method with respect to outliers in the data. Although Li (1991, p. 319) mentions that "[...] it would help the analysis if closer examination of the distribution of $\boldsymbol{x}$ can be made so that outliers can be removed [...]", and also that for the standardization of $\boldsymbol{x}$ robust estimators "may be preferable" (p. 320), an analysis of the robustness issue is still outstanding. The estimation scheme SIR uses estimators which are known to be extremely nonrobust (sample mean and covariance, classical PCA). Hence, it can be that some bad points in the data completely ruin the results of SIR. To some extent, this contradicts Li's opinion, that the whole problem only concerns the design points "that are under our control", and that "even in the observational study, we may screen out bad design points" (p. 320). We agree in that the whole problem only concerns the design points (see below), but the model assumptions are such that we cannot really control the design points since we are dealing with a correlation model, where $\boldsymbol{x}$ is random, too. Moreover, we are in the situation of high–dimensional regressor points, such that we cannot just screen out bad observations but have to solve the problem of identifying outliers in high–dimensional data (see Rocke, 1996, Becker and Gather, 1997). Therefore, a sensitivity analysis of SIR in outlier situations seems necessary, and we discuss this in Section 2.

The third section contains our proposal for generalized procedures which allow the estimation of edr directions even in the case when there are some heavily corrupted points in the data. The proposed procedures are called Generalized SIR (GSIR). The main feature is that the concept of SIR is maintained but that instead of the special estimators we allow for other alternatives, too. We investigate conditions under which such GSIR procedures are affine equivariant. As a worst case criterion for the robustness of a GSIR procedure we transfer the concept of the finite-sample breakdown point to the situation investigated here. In the fourth section, we discuss properties of a particular robust GSIR procedure called DAME which we get by using some special robust estimators in the different steps of the estimation scheme.

A simulation study which is described in the last section indicates that DAME is successful in handling situations with some severe outliers in the data, while not loosing much efficiency when there are no outliers.

2 Sensitivity of SIR to outliers

Consider the situation that in a sample $(\boldsymbol{x}_1, y_1), \ldots, (\boldsymbol{x}_n, y_n)$ of $(\boldsymbol{x}, y)$ of size n, one observation is replaced by an arbitrarily chosen point. Without loss of generality we assume that

the observations are ordered such that $y_1, \dots, y_{n_1}$ and the corresponding observations of $\boldsymbol{x}$ belong to the first slice, $y_{n_1+1}, \dots, y_{n_1+n_2}$ to the second slice and so on. Now we replace $\boldsymbol{x}_1$ by $\tilde{\boldsymbol{x}}_{1,t} = t\boldsymbol{\beta} + \boldsymbol{u}$ with $t \in \mathbb{R}$, $t > 0$, $\boldsymbol{u}, \boldsymbol{\beta} \in \mathbb{R}^d$, $\boldsymbol{\beta} \neq \mathbf{o}$. Thus, we get a corrupted sample for each value of t and if we let t tend to infinity, the interpretation is that one observation of $\boldsymbol{x}$ is pulled away into the direction of $\boldsymbol{\beta}$. The corresponding observation y_1 of y is not altered. In general we can observe that altering observations of y only has an indirect effect on the results of SIR since those observations are only used to construct the slices (cf. Li, 1991).
How does the above contamination change the estimation of the edr directions? The following theorem gives an answer. In order to measure the success of the estimation, we look at the canonical correlations between $\boldsymbol{\beta}_1'\boldsymbol{x}, \dots, \boldsymbol{\beta}_K'\boldsymbol{x}$ on the one hand and $\widehat{\boldsymbol{\beta}}_{1,t}'\boldsymbol{x}, \dots, \widehat{\boldsymbol{\beta}}_{K,t}'\boldsymbol{x}$ on the other hand, where $\widehat{\boldsymbol{\beta}}_{1,t}, \dots, \widehat{\boldsymbol{\beta}}_{K,t}$ are estimated with the corresponding corrupted sample for each $t > 0$. For the concept of canonical correlations see Kshirsagar (1972). Here, they are denoted by $\varrho_{1,t} \geq \cdots \geq \varrho_{K,t}$ where the index t expresses the dependence on the corrupted sample. Moreover, let $\boldsymbol{B} = [\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_K]$ and let $\widehat{\boldsymbol{B}}_t = [\widehat{\boldsymbol{\beta}}_{1,t}, \dots, \widehat{\boldsymbol{\beta}}_{K,t}]$ be the matrix consisting of the estimated edr directions as column vectors.

Theorem 2.1 *Let $\boldsymbol{\beta} = \boldsymbol{\Sigma}\boldsymbol{\beta}_1$, $\overline{\boldsymbol{x}}_1^* = \sum_{i=2}^{n_1} \boldsymbol{x}_i/(n_1 - 1)$ and $\overline{\boldsymbol{x}}^* = \sum_{i=2}^{n} \boldsymbol{x}_i/(n-1)$. If there is no $c \in \mathbb{R}$ with $\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^* = c\boldsymbol{\beta}$ and if there exist $\varepsilon > 0$ and an unbounded and strictly increasing sequence $\{t_m\}_{m \in \mathbb{N}}$ in $\mathbb{R}$ with $\det(\widehat{\boldsymbol{B}}_{t_m}' \widehat{\boldsymbol{B}}_{t_m}) \geq \varepsilon$ for all $m \in \mathbb{N}$, then $\lim_{m\to\infty} \varrho_{K,t_m}^2 = 0$.*

The Theorem tells us that under some mild conditions the smallest canonical correlation between the true and the estimated edr directions tends to zero if there is only one bad point in the sample which is pulled away into the direction $\boldsymbol{\Sigma}\boldsymbol{\beta}_1$. Consequently, one 'large' outlier can make it impossible to estimate all K edr directions reliably. The proof of this Theorem is given in the Appendix.
A result similar to Theorem 2.1 can be derived when $\boldsymbol{\Sigma}$ is known and is not estimated from the data:

Theorem 2.2 *For known covariance matrix $\boldsymbol{\Sigma}$,*

$$\lim_{t\to\infty} \widehat{\boldsymbol{\beta}}_{1,t} = \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}{\sqrt{\boldsymbol{\beta}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\beta}}} \quad \text{and} \quad \lim_{t\to\infty} \widehat{\lambda}_{1,t} = \infty, \tag{2.1}$$

and if $\boldsymbol{\beta}'\boldsymbol{\beta}_k = 0$, $k = 1, \dots, K$, then it holds that $\lim_{t\to\infty} \varrho_{K,t}^2 = 0$.

The Theorem shows that the first estimated edr direction is extremely influenced by only one bad point which is pulled away into the direction of $\boldsymbol{\beta}$.

In the same sense as above, this demonstrates the sensitivity of SIR to one 'large' outlier. The remarkable difference to the case of unknown $\boldsymbol{\Sigma}$ lies in the choice of $\boldsymbol{\beta}$. Contrary to the situation in Theorem 2.1, $\boldsymbol{\beta}$ is orthogonal to the true edr directions in Theorem 2.2. This difference is due to the fact that in the situation of unknown $\boldsymbol{\Sigma}$, the estimate of $\boldsymbol{\Sigma}$ (the sample covariance matrix) breaks down when one observation is pulled away to infinity. Especially, from (5.3) it follows that the appropriately oriented eigenvector of $\widehat{\boldsymbol{\Sigma}}_t$ corresponding to the largest eigenvalue tends to $\boldsymbol{\beta}$, the eigenvalue itself tends to infinity. As a consequence, $\boldsymbol{\beta}$ is an eigenvector of $\lim_{t\to\infty} \widehat{\boldsymbol{\Sigma}}_t^{-1}$ corresponding to the smallest possible eigenvalue 0. It follows that with increasing t all estimated edr directions corresponding to positive eigenvalues become orthogonal to $\boldsymbol{\beta}$.

3 Generalized SIR

It is obvious from the preceding section, that the sensitivity of SIR to outliers is due to the fact that nonrobust estimators are used. Therefore, we propose to maintain the clever estimation scheme of SIR but to replace all nonrobust estimators by alternatives with desirable theoretical properties but keeping the method feasible. Thus, we use a location estimator T_1 and a positive definite estimator C_1 of scatter in the standardization step of the estimation procedure. Moreover, a location estimator T_2 and a positive semidefinite estimator C_2 of scatter are used in the steps 3. and 4. respectively. Here, C_2 is applied to a set M containing the estimated locations T_2(slice h) from step 3., each with suitably chosen multiplicity w_h,$h = 1, \ldots, H$. We call this procedure Generalized SIR (GSIR). Given a sample $(\boldsymbol{x}, y)_n$ of n observations of $(\boldsymbol{x}, y)$ we denote the covariance estimator computed in the fourth step of GSIR by $GSIR((\boldsymbol{x}, y)_n)$.

3.1 Affine Equivariance of GSIR

The property of affine equivariance is of course a desirable property of multivariate statistical procedures. For GSIR this is defined as follows.

Definition 3.1 *A GSIR procedure is affine equivariant if for each nonsingular matrix* $\boldsymbol{A} \in \mathbb{R}^{d\times d}$*, for each* $\boldsymbol{v} \in \mathbb{R}^d$*, for each* $a \in \mathbb{R}$*,* $a \neq 0$*, for each* $b \in \mathbb{R}$ *and for each sample* $(\boldsymbol{x}, y)_n$*: If* $\widehat{\boldsymbol{\eta}}_i$ *is an eigenvector of* $GSIR((\boldsymbol{x}, y)_n)$ *with corresponding eigenvalue* $\widehat{\lambda}_i$*, then*

$$C_{1,n}^{1/2}(\boldsymbol{A}\boldsymbol{x}_n + \boldsymbol{v})(\boldsymbol{A}')^{-1}C_{1,n}^{-1/2}(\boldsymbol{x}_n)\widehat{\boldsymbol{\eta}}_i \tag{3.1}$$

is an eigenvector of $GSIR((\boldsymbol{A}\boldsymbol{x}+\boldsymbol{v}, ay+b)_n)$ *with corresponding eigenvalue* $\widehat{\lambda}_i$*,* $i = 1, \ldots, d$*.*

It is not necessary that the estimators T_1, C_1, T_2, C_2 are all affine equivariant in order to get a GSIR procedure with this property.

Lemma 3.2 *A GSIR procedure is affine equivariant, if* C_1 *is affine equivariant, if* T_2 *and* C_2 *are orthogonal equivariant and if*

$$n_h = n_{H-h+1} \text{ and } w_h = w_{H-h+1}, \quad h = 1, \ldots, \lfloor H/2 \rfloor, \tag{3.2}$$

where $\lfloor x \rfloor$ *denotes the largest integer smaller than or equal to* $x \in \mathbb{R}$*.*

Proof: First note that because of (3.2) the slices remain unaltered regardless of the values of a and b.

Now let $(\boldsymbol{z})_n$ and $(\tilde{\boldsymbol{z}})_n$ be the standardized samples belonging to $(\boldsymbol{x})_n$ and $(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{v})_n$, respectively. Since C_1 is affine equivariant we find $(\tilde{\boldsymbol{z}})_n = \boldsymbol{P}(\boldsymbol{z})_n + \boldsymbol{u}$, where $\boldsymbol{P}'\boldsymbol{P} = \boldsymbol{I}_d$. Consequently, only orthogonal equivariance of T_2 and C_2 is required to achieve the affine eqivariance of the GSIR procedure. $\square$

One easily sees that the choice of T_1 does not influence the results of the estimation scheme such that, apart from numerical considerations, T_1 can be chosen arbitrarily.

3.2 The finite-sample breakdown point

In order to compare different GSIR procedures with respect to their robustness properties we need some appropriate criteria. One worst-case performance measure which is used in many different situations like estimation of (multivariate) location and scatter or regression is the finite-sample breakdown point, cf. Donoho and Huber (1983). We transfer this criterion to the situation of a GSIR procedure, where we want to capture the breakdown behavior with respect to outliers (but not with respect to small changes in the data).

As the estimates of the edr directions are mainly determined by the covariance estimate $GSIR((\boldsymbol{x}, y)_n)$, it is near at hand to consider the possible explosion of the largest and the implosion of the smallest eigenvalue of this matrix in order to define the breakdown of GSIR. In this way one would use the usual definition of breakdown of a covariance estimator (cf. Lopuhaä and Rousseeuw, 1991). But this definition is not suitable here. Just note the fact that in high–dimensional situations one may want to choose a number H of slices which is smaller than the dimension d of the explanatory variable. This could be crucial since the dimension K of the edr space is assumed to be small ($K \ll d$). But then the covariance estimate $GSIR((\boldsymbol{x}, y)_n)$ is based upon a maximum of H different points in a d-dimensional space such that the smallest eigenvalue of this matrix should become zero in this case. Consequently, the implosion of the smallest eigenvalue of $GSIR((\boldsymbol{x}, y)_n)$ is not a useful indicator of breakdown here.

Also, the problem cannot be solved by looking only at the largest eigenvalue. Consider the situation of Section 2 and note that

$$\widehat{\boldsymbol{V}} = SIR((\boldsymbol{x}, y)_n) = \boldsymbol{I}_d - \sum_{h=1}^{H} \widehat{\boldsymbol{\Sigma}}^{-1/2} \left(\frac{1}{n} \sum_{i: y_i \in I_h} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_h)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_h)' \right) \widehat{\boldsymbol{\Sigma}}^{-1/2}$$

for each sample $(\boldsymbol{x}, y)_n$. It follows that the eigenvalues of $SIR((\boldsymbol{x}, y)_n)$ are bounded in $[0, 1]$ which is also true for every corrupted sample. On the other hand, Theorems 2.1 and 2.2 show the extreme sensitivity of SIR to outliers. Consequently, it is not possible to characterize the breakdown of GSIR by just looking at the largest eigenvalue of $GSIR((\boldsymbol{x}, y)_n)$. We therefore suggest the following definition.

Definition 3.3 *A corrupted sample* $(\boldsymbol{x}, y)_{n,m}$ *is constructed from* $(\boldsymbol{x}, y)_n$ *by replacing* m *observations of* $(\boldsymbol{x}, y)_n$ *by arbitrary points* $(\tilde{\boldsymbol{x}}_i, \tilde{y}_i) \in \mathbb{R}^{d+1}$, $i = 1, \ldots, m$. *Let* $\varepsilon_1^\star := \varepsilon_1^\star((\boldsymbol{x})_n, C_1)$

be the finite-sample breakdown point of C_1. *Define* $\varepsilon^\star_{2,+} := \varepsilon^\star_{2,+}((\boldsymbol{x}, y)_n, C_1, T_2, C_2, H, \vec{n}, \vec{w})$

$$:= \min_{1\leq m\leq n} \left\{ \frac{m}{n} : \sup_{(\boldsymbol{x},y)_{n,m}} |\lambda_1 \left(GSIR((\boldsymbol{x}, y)_{n,m})\right) - \lambda_1 \left(GSIR((\boldsymbol{x}, y)_n)\right)| = \infty \right\}$$

and $\varepsilon^\star_{2,-} := \varepsilon^\star_{2,-}((\boldsymbol{x}, y)_n, C_1, T_2, C_2, H, \vec{n}, \vec{w})$

$$:= \min_{1\leq m\leq n} \left\{ \frac{m}{n} : \sup_{(\boldsymbol{x},y)_{n,m}} \left| \frac{1}{\lambda_1 \left(GSIR((\boldsymbol{x}, y)_{n,m})\right)} - \frac{1}{\lambda_1 \left(GSIR((\boldsymbol{x}, y)_n)\right)} \right| = \infty \right\}.$$

The finite-sample breakdown point of GSIR is defined as

$$\varepsilon^\star((\boldsymbol{x}, y)_n, C_1, T_2, C_2, H, \vec{n}, \vec{w}) := \min \left\{ \varepsilon^\star_1, \varepsilon^\star_{2,+}, \varepsilon^\star_{2,-} \right\}.$$

We give some comments on this definition:

Remark 3.4 *1. The estimator* C_1 *is used for standardization of the data in the first step of GSIR and for retransformation of the estimated directions in the last step. If for a sequence* $((\boldsymbol{x})_{n,m;k})_{k\in\mathbb{N}}$ *of corrupted samples the largest eigenvalue of* $C_1((\boldsymbol{x})_{n,m;k})$ *tends to infinity or if the smallest eigenvalue tends to zero, then for every vector* $\boldsymbol{u} \in \mathbb{R}^d$, $\|\boldsymbol{u}\| = 1$, *the transformed vector* $C_1^{-1/2}((\boldsymbol{x})_{n,m;k})\boldsymbol{u}$ *must be contained in a true subspace of* $\mathbb{R}^d$. *This means that in case of a breakdown of* C_1 *the associated GSIR procedure can only estimate directions from a true subspace of* $\mathbb{R}^d$ *as edr directions. This is a crucial point if the edr directions are not contained in this subspace which is something we do not know. Theorem 2.1 can be seen as a special case of this fact, showing that the estimated edr directions for a sequence of corrupted samples are finally contained in the orthogonal complement of the space spanned by the direction of corruption.*

2. As a second indicator of breakdown of a GSIR procedure the largest eigenvalue of the covariance estimate used in the fourth step is considered. If there is no breakdown of C_1 *then this eigenvalue can yield a breakdown of a GSIR procedure, for example by becoming arbitrarily large for corrupted samples (consider again Theorem 2.2 for SIR with known covariance matrix* $\boldsymbol{\Sigma}$*).*

3. Moreover, we speak of a breakdown of a GSIR procedure if the largest eigenvalue can be brought arbitrarily close to zero by corruption of a sample. In such a situation all eigenvalues *tend to zero which implies that the covariance matrices estimated in step 4 of GSIR tend to the zero matrix. In this situation, a sensible estimation of the edr directions is no longer possible.*

4. A stronger and rather natural definition of breakdown of GSIR could consider not only the eigenvalues mentioned above but the estimated edr directions themselves. One might then speak of a breakdown of GSIR only in situations for which the smallest canonical correlation between $\boldsymbol{\beta}'_1\boldsymbol{x}, \ldots, \boldsymbol{\beta}'_K\boldsymbol{x}$ *and* $\widehat{\boldsymbol{\beta}}'_1\boldsymbol{x}, \ldots, \widehat{\boldsymbol{\beta}}'_K\boldsymbol{x}$ *can become zero by the corruption of a given sample. But this nice looking condition would be almost impossible to verify. For example, consider the situation that the* Kth *and the* $(K+1)th$ *eigenvalue of* $GSIR((\boldsymbol{x}, y)_n)$ *are almost equal, the corresponding eigenvectors* $\widehat{\boldsymbol{\eta}}_K = \boldsymbol{u}, \widehat{\boldsymbol{\eta}}_{K+1} = \boldsymbol{v}$ *being orthogonal (*$\boldsymbol{\Sigma} = \boldsymbol{I}_d$ *known). Then it is possible that small changes of only a few data points alter the results in such a way that vectors very close to* $\boldsymbol{v}$ *and* $\boldsymbol{u}$ *are eigenvectors to the* Kth *and to the* $(K+1)th$ *largest eigenvalue respectively. Hence, small fluctuations in the data could have this dramatic effect on the results, being very difficult to analyse. Moreover, we concentrate here on gross errors which means that we want to provide GSIR procedures which are able to produce still good estimates of the edr directions in situations where some data points are outliers w.r.t. the assumed distribution of* $\boldsymbol{x}$ *(cf. Davies and Gather, 1993).*

4 A specific GSIR procedure: DAME

In this section, we investigate the robustness properties of a special affine equivariant GSIR procedure. Our choice of robust estimators in the different steps of the estimation scheme is as follows:

For the standardization step one takes an affine equivariant covariance estimator C_1 which fulfills two conditions: From a theoretical point of view, the estimator should possess a high finite-sample breakdown point, and from a practical point of view, it must be able to work successfully with high–dimensional data. With respect to these restrictions we use an S-estimator of scatter as defined in Lopuhaä (1989) where the translated biweight function as defined by Rocke (1993) is used as ρ-function in order to achieve an improved performance in high–dimensional situations, e.g. in comparison with the ordinary biweight function. The translated biweight function possesses two parameters which are chosen in such a way that the S-estimator achieves the maximal possible finite-sample breakdown point and an asymptotic rejection probability of .01 (Rocke, 1993).

The choice of robust estimators T_2 and C_2 in the next steps is facilitated on the one hand by the fact that orthogonal equivariant estimators can be used. But on the other hand one has to keep in mind that the observations are partitioned into the different slices. If the number H of slices is large, we might have only very few observations in some slices. But also, we have few slices if each slice has to contain a large number of observations. Hence, we have to deal with the problem that either the location estimator T_2 has to be applied to only a few observations, or that the covariance estimator C_2 is based on a few different points in a high–dimensional space.

For T_2, we choose the L_1-median which combines the following advantages which are important here: The L_1-median is an orthogonal equivariant location estimator with maximal possible finite-sample breakdown point in this class of estimators (cf. Lopuhaä and Rousseeuw, 1991) and there exist algorithms to compute this estimator, e.g. Bedall and Zimmermann (1979). Moreover, in this step of the estimation scheme it is sufficient to get a coarse estimate of the inverse regression curve such that it seems justifiable to use the L_1-median. As an alternative the orthomedian proposed by Grübel (1996) can be applied. Other generalizations of the univariate median to the multivariate case are discussed by Small (1990), in work on data depth (Liu, 1990) and on the bagplot (Rousseeuw and Ruts, 1997).

Based upon the projection pursuit principle, Li and Chen (1985) propose a robust estimation method for principal components. For a sample $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in $\mathbb{R}^d$ the first principal component is defined as a direction $\boldsymbol{a}_1 \in S(d) = \{\boldsymbol{a} \in \mathbb{R}^d : \boldsymbol{a}'\boldsymbol{a} = 1\}$ for which a robust univariate scale estimator applied to the projected sample $\boldsymbol{a}'\boldsymbol{x}_1, \ldots, \boldsymbol{a}'\boldsymbol{x}_n$ becomes maximal. The corresponding first eigenvalue is computed by applying the scale estimator to the projected sample $\boldsymbol{a}_1'\boldsymbol{x}_1, \ldots, \boldsymbol{a}_1'\boldsymbol{x}_n$. In order to find the second principal component only directions orthogonal to $\boldsymbol{a}_1$ are considered for the maximization problem. Continuing in this way, one finally gets all principal components and a corresponding covariance estimator. We use this estimation procedure here for C_2 for several reasons: First of all it reflects the intention of SIR. In order to get estimates of the edr directions in model 1.1 one is interested in finding the directions in which the estimated inverse regression curve shows the greatest variability. This is exactly what the described projection pursuit estimator does. Furthermore, the procedure can be applied even when there is only a small number of observations. This is no problem, because the method is based upon a univariate scale estimator which has to be applied to projections of the data. As described above, this is an important feature here, since the number of slices can be small compared to the dimension d.

Finally, a univariate scale estimator has to be chosen. We use a modification of an estimator proposed by Rousseeuw and Croux (1993) as an alternative to the MAD. This modification showed up best results also in our simulations. For a univariate sample $u_1, \ldots, u_n$ this estimator is defined by

$$RCQ_\alpha(u_1, \ldots, u_n) := \{|u_i - u_j| : 1 \leq i < j \leq n\}_{(k)}$$

with $k = (\lfloor n\alpha \rfloor + 1)\lfloor n\alpha \rfloor / 2$, $\alpha \in [0.5, 1)$, which means that the k-th order statistic of the absolute differences $|u_i - u_j|$, $1 \leq i < j \leq n$, is determined.

Now we are able to state some results concerning the finite-sample breakdown point of this special GSIR procedure which we call DAME (**D**imension **A**djustment **Me**thod). We concentrate here on the practically relevant case that slices of equal sizes are used, i.e. $n_1 = \cdots = n_H =: n_S$. Therefore, we use the weights $w_1 = \cdots = w_H = 1$. The following conditions concerning the sample $(\boldsymbol{x}, y)_n$ are needed:

(R1) For each partition of the observations of $(\boldsymbol{x})_n$ into the slices and for each standardization of these observations according to $\boldsymbol{x} \mapsto \boldsymbol{A}\boldsymbol{x}$ with $\boldsymbol{A} \in \mathbb{R}^{d \times d}$, $\boldsymbol{A}$ positive definite, $\lambda_d(\boldsymbol{A}) \geq \lambda_0$ for a fixed $\lambda_0 > 0$, there exists $d_0 > 0$ with $\|\widehat{\boldsymbol{m}}_{h_1} - \widehat{\boldsymbol{m}}_{h_2}\| \geq d_0$ for all $h_1, h_2 \in \{1, \ldots, H\}$ with $h_1 \neq h_2$, where $\widehat{\boldsymbol{m}}_h$ denotes T_2 applied to the standardized observations in the h-th slice.

(R2) There exists $d_0 > 0$, such that for each partition of the observations of $(\boldsymbol{x})_n$ into the slices it holds that $\|\widehat{\boldsymbol{m}}_{h_1} - \widehat{\boldsymbol{m}}_{h_2}\| \geq d_0$ for all $h_1, h_2 \in \{1, \ldots, H\}$ with $h_1 \neq h_2$.

This condition is similar to the assumption that a sample is in general position which is often claimed in multivariate robust framework.

Theorem 4.1 *Let $(\boldsymbol{x}, y)_n$ be a sample with $(\boldsymbol{x})_n$ in general position and with Σ unknown.*
1. Let $\lfloor (n_S+1)/2 \rfloor \cdot \lceil H(1-\alpha) \rceil < \lfloor (n-d+1)/2 \rfloor$, then

$$\varepsilon^{\star}_{2,+}((\boldsymbol{x}, y)_n, DAME) = \frac{\lceil H(1-\alpha) \rceil}{H n_S} \left\lfloor \frac{n_S+1}{2} \right\rfloor .$$

2. Let $(\lfloor H\alpha \rfloor + 1)\lfloor (n_S+1)/2 \rfloor < \lfloor (n-d+1)/2 \rfloor$. For $n_S \geq 3$ and under Condition (R1):

$$\frac{\lfloor H\alpha \rfloor}{H n_S} \leq \varepsilon^{\star}_{2,-}((\boldsymbol{x}, y)_n, DAME) \leq \frac{\lfloor H\alpha \rfloor + 1}{H n_S} \left\lfloor \frac{n_S+1}{2} \right\rfloor .$$

3. Let $\lfloor H\alpha \rfloor < \lfloor (n-d+1)/2 \rfloor$. For $n_S = 2$ and under Condition (R2) we have

$$\varepsilon^{\star}_{2,-}((\boldsymbol{x}, y)_n, DAME) = \frac{\lfloor H\alpha \rfloor}{2H}.$$

The proof is given in the Appendix.

Theorem 4.1 shows that DAME possesses good breakdown point properties compared with SIR. The parameter α allows for a certain flexibility in adjusting the procedure to different outlier situations. This is due to the fact that resistance against explosion of the largest and against implosion of the smallest eigenvalue seem to be somewhat contradicting aims. This becomes clear if one compares the explosion breakdown point in 1. and the lower bound for the implosion breakdown point in 2. with respect to the choice of α. The simulations in the next section show that and how DAME works and how it compares with SIR in some typical outlier situations.

Of course, besides looking at breakdown properties, one should also consider other robustness properties. For example, the bias of GSIR under outlier situations should be taken into account as well, because it is well known that a bounded bias under outliers and efficiency are contradictory aims. Another problem lies in determining the dimension K of the edr space. In this paper, we assume that K is known, but in practical situations K has to be estimated in a possibly robust way, too, taking into account the possibility of outliers in the data. Some research is done in this direction currently.

5 Simulations

In order to investigate the performance of DAME, we have conducted a simulation study. Some of the results are shown here.

An important aspect concerning the application of robust procedures is the availability of algorithms to compute the estimates. For the above DAME procedure, algorithms for the various steps of the procedure have to be chosen.

For the computation of the S-estimator in the first step of DAME we use the hybrid algorithm as described by Rocke and Woodruff (1996). This algorithm is based upon partitioning the data into cells. Within each cell MCD-estimates are computed and are used as starting points for an iterative computation of the desired estimates. Rocke and Woodruff (1996) demonstrate that partitioning the data leads to improved estimates especially in high–dimensional situations.

In order to compute the L_1-median in the third step of DAME, the well-known algorithm of Bedall and Zimmermann (1979) is applied.

Finally, for the projection pursuit estimator in the fourth step of DAME an algorithm proposed by Croux and Ruiz-Gazen (1996) is used. This algorithm limits the search for the k-th principal component to some specific directions which are chosen depending on the data points and the $k-1$ principal components which have already been found. Of course, the original optimization problem is not solved exactly in this way, but the algorithm is easy to implement and it seems that the results are sufficiently precise. In our study we use $RCQ_{0.5}$ as well as the MAD as univariate scale estimators.

We restrict to two models here:

$$y = x_1 + 0.1 \cdot \varepsilon \tag{5.1}$$

$$y = \frac{x_1}{0.5 + \sqrt{1.5 + x_2}} + 0.1 \cdot \varepsilon \tag{5.2}$$

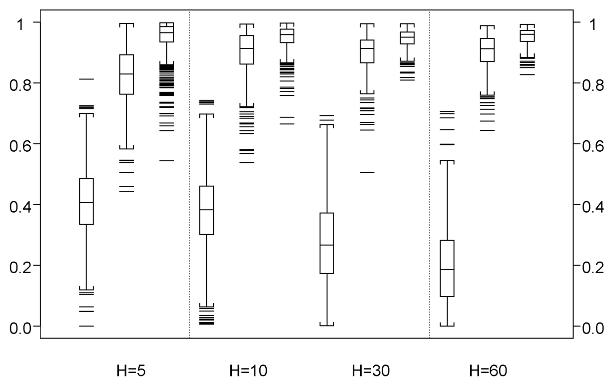
The first model corresponds to the situation $K = 1$ with $\beta_1 = (1, 0, \ldots, 0)^T$, in the second model we have $K = 2$ and $\beta_1 = (1, 0, \ldots, 0)^T$ and $\beta_2 = (0, 1, 0, \ldots, 0)^T$.

We generate $n = 300$ data points. The dimension d equals 10; $x_1, \ldots, x_{10}$ and ε are i.i.d. random variables from a standard normal distribution. We contaminate the samples by adding $2\sqrt{\chi^2_{d;.999}} \cdot \beta_1$ to 30 randomly chosen data points, which means that 10% of the observations in our samples are contaminated with large outliers (also see Rocke, 1993). For the numbers of slices we use $H = 5, 10, 30, 60$.

The results are reported, using as performance criterion the absolute value $|corr(\boldsymbol{\beta}_1^T \boldsymbol{x}, \widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x})|$ of the correlation between $\boldsymbol{\beta}_1^T \boldsymbol{x}$ and the estimated variable $\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x}$ for the first model. For the two component model we use the canonical correlations between $\boldsymbol{\beta}_1^T \boldsymbol{x}, \boldsymbol{\beta}_2^T \boldsymbol{x}$ and $\widehat{\boldsymbol{\beta}}_1^T \boldsymbol{x}, \widehat{\boldsymbol{\beta}}_2^T \boldsymbol{x}$. This means that one looks at the angle between the model edr's and the estimated edr's. The results of 500 replications are reported graphically using boxplots.

Figure 1 shows the results for the first model. For each value of H, three boxplots are

Figure 1: Simulation results for model 1: Boxplots of correlations between true and estimated edr direction for SIR and two versions of DAME

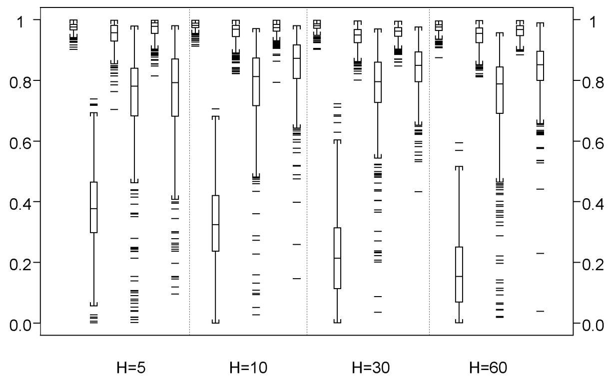


shown. The first one presents the results for SIR, the second and the third boxplot show the results for DAME using the MAD and $RCQ_{0.5}$ as univariate scale esimators in the fourth step, respectively.

First of all, we observe that the results of SIR are rather bad in the simulated outlier

situations. Due to the outliers in the data, SIR is not able to give good estimates of the edr direction $\boldsymbol{\beta}_1$. In contrast to this, DAME is more successful. Especially with the use of $RCQ_{0.5}$ as univariate scale estimator in the fourth step we get good results.
For the second model, the results are very similar which is reflected by figure 2. Here, the

Figure 2: Simulation results for model 2: Boxplots of correlations between true and estimated first and second edr directions, respectively, for SIR and two versions of DAME



first and the second boxplot show the results of SIR, the third and fourth boxplot represent DAME using the MAD and the last two boxplots give the results for DAME using $RCQ_{0.5}$. We see that dimension reduction based upon classical SIR is not successful as only the first canonical correlation achieves high values whereas the second canonical correlation gives low values. The robust DAME procedures again show up better results. Moreover, the use of $RCQ_{0.5}$ instead of the MAD seems to lead to improved results in most cases. Only for $H = 5$ there seems to be a slight difference between the two methods.
Summing up it may be said that the effects of outliers on SIR are dramatic. The results of this method can become totally unreliable if there are a few bad outliers in the data.

Robustified SIR-methods, such as DAME proposed above, can be used to overcome these problems.

Acknowledgements

This work was partially supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 475.

Appendix: Proofs

Proof of Theorem 2.1

We first consider two Lemmas which are needed to prove Theorem 2.1.

The sample covariance matrix of the corrupted sample is given by

$$\widehat{\boldsymbol{\Sigma}}_t = \boldsymbol{S} + \frac{n-1}{n^2} t^2 \boldsymbol{\beta}_t \boldsymbol{\beta}_t', \tag{5.3}$$

with $\boldsymbol{S} = \sum_{i=2}^{n} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}^*)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}^*)'/n$, $\boldsymbol{\beta}_t = \boldsymbol{\beta} + (\boldsymbol{u} - \overline{\boldsymbol{x}}^*)/t$ and $\overline{\boldsymbol{x}}^* = \sum_{i=2}^{n} \boldsymbol{x}_i/(n-1)$. We use the fact that $\widehat{\boldsymbol{\Sigma}}_t$ is positive definite if $\boldsymbol{S}$ is positive definite. Then we have

$$\lim_{t\to\infty} \widehat{\boldsymbol{\Sigma}}_t^{-1} = \boldsymbol{S}^{-1} - \frac{\boldsymbol{S}^{-1}\boldsymbol{\beta}\boldsymbol{\beta}'\boldsymbol{S}^{-1}}{\boldsymbol{\beta}'\boldsymbol{S}^{-1}\boldsymbol{\beta}}, \quad \lim_{t\to\infty} t^2 \cdot \widehat{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\beta}_t = \frac{n^2\boldsymbol{S}^{-1}\boldsymbol{\beta}}{(n-1)\boldsymbol{\beta}'\boldsymbol{S}^{-1}\boldsymbol{\beta}}. \tag{5.4}$$

We now consider the limiting behavior of the matrix $\widetilde{\boldsymbol{V}}_t$ defined as $\widetilde{\boldsymbol{V}}$ in (1.2), but for the case of corrupted samples with t tending to infinity. Using decomposition formulae (Griffiths and Hill, 1985) we get

$$\begin{aligned} \frac{1}{t}\widetilde{\boldsymbol{V}}_t = \frac{1}{t}\boldsymbol{I}_d \quad &- \quad \frac{1}{nt}\widehat{\boldsymbol{\Sigma}}_t^{-1}\left(\boldsymbol{S}_1^* + \sum_{h=2}^{H} \boldsymbol{S}_h\right) \\ &- \quad \frac{1}{n}\left(1 - \frac{1}{n_1}\right)\left(t\widehat{\boldsymbol{\Sigma}}_t^{-1}\boldsymbol{\beta}_t + \widehat{\boldsymbol{\Sigma}}_t^{-1}(\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*)\right)\left(\boldsymbol{\beta}_t + \frac{\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*}{t}\right)' \end{aligned}$$

with $\boldsymbol{S}_1^* = \sum_{i=2}^{n_1} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_1^*)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_1^*)'$, $\boldsymbol{S}_h = \sum_{i:y_i \in I_h} (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_h)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_h)'$.

Using (5.4), yields the following Lemma:

Lemma 5.1 *For positive definite $\boldsymbol{S}$, we have that* $\lim_{t\to\infty} \widetilde{\boldsymbol{V}}_t/t = \widetilde{\boldsymbol{V}}^*$ *with*

$$\widetilde{\boldsymbol{V}}^* = \frac{1}{n}\left(1 - \frac{1}{n_1}\right)\left(\frac{\boldsymbol{\beta}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*)}{\boldsymbol{\beta}'\boldsymbol{S}^{-1}\boldsymbol{\beta}}\boldsymbol{S}^{-1}\boldsymbol{\beta} - \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*)\right) \cdot \boldsymbol{\beta}', \tag{5.5}$$

where $\overline{\boldsymbol{x}}_1^* = \sum_{i=2}^{n_1} \boldsymbol{x}_i/(n_1 - 1)$ *and* $\overline{\boldsymbol{x}}^* = \sum_{i=2}^{n} \boldsymbol{x}_i/(n-1)$.

As the estimated edr directions are eigenvectors of $\widetilde{\boldsymbol{V}}_t$, we can use Lemma 5.1 to find results about the limiting behavior of those directions. We need the following Lemma, the proof of which is omitted here.

Lemma 5.2 *For each $t > 0$ let $\boldsymbol{M}_t \in \mathbb{R}^{d\times d}$ be a matrix with only real eigenvalues, and $\boldsymbol{v}_t$, $\|\boldsymbol{v}_t\| = 1$, an eigenvector of $\boldsymbol{M}_t$ with corresponding eigenvalue λ_t. If $\lim_{t\to\infty} \boldsymbol{M}_t = \boldsymbol{u}_1 \cdot \boldsymbol{u}_2'$, $\boldsymbol{u}_1, \boldsymbol{u}_2 \in \mathbb{R}^d$, $\boldsymbol{u}_1, \boldsymbol{u}_2 \neq \mathbf{o}$, $\boldsymbol{u}_2'\boldsymbol{u}_1 = \mathbf{o}$ then $\lim_{t\to\infty}(\boldsymbol{u}_2'\boldsymbol{v}_t) = 0$.*

Note that the conditions of Lemma 5.2 are fulfilled for $\widetilde{\boldsymbol{V}}_t$ and $\widetilde{\boldsymbol{V}}^*$. The matrix $\widetilde{\boldsymbol{V}}^*$ possesses the structure $\widetilde{\boldsymbol{V}}^* = \boldsymbol{u}_1\boldsymbol{u}_2'$ with $\boldsymbol{u}_1'\boldsymbol{u}_2 = 0$ and $\boldsymbol{u}_2 = \boldsymbol{\beta}$ which is the direction of contamination. We now can prove Theorem 2.1.

Proof of Theorem 2.1. Without loss of generality we assume $\boldsymbol{\beta}_i'\boldsymbol{\Sigma}\boldsymbol{\beta}_j = \delta_{ij}$, $i, j = 1, \dots, K$. Note that $\widetilde{\boldsymbol{V}}^*$ possesses the structure used in Lemma 5.2 with $\boldsymbol{u}_2 = \boldsymbol{\beta}$ and

$$\boldsymbol{u}_1 = \frac{1}{n}\left(1 - \frac{1}{n_1}\right)\left(\frac{\boldsymbol{\beta}'\boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*)}{\boldsymbol{\beta}'\boldsymbol{S}^{-1}\boldsymbol{\beta}}\boldsymbol{S}^{-1}\boldsymbol{\beta} - \boldsymbol{S}^{-1}(\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*)\right)$$

and $\boldsymbol{u}_2'\boldsymbol{u}_1 = 0$. Moreover we have $\boldsymbol{u}_1 \neq \mathbf{o}$, since $\overline{\boldsymbol{x}}^* - \overline{\boldsymbol{x}}_1^*$ is not a scalar multiple of $\boldsymbol{\beta}$.
Using the additional assumptions, it follows that there exist $b_0 > 0$ and $b_1 < \infty$ with $b_0 \leq \|\widehat{\boldsymbol{\beta}}_{i,t_m}\|^2 \leq b_1$, $i = 1, \dots, K$, $m \in \mathbb{N}$.
Analogously we find that there exist an upper bound $l_1 < \infty$ and a lower bound $l_d > 0$ for the eigenvalues of $\widehat{\boldsymbol{B}}_{t_m}'\boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_{t_m}$, $m \in \mathbb{N}$, such that $\widehat{\boldsymbol{B}}_{t_m}'\boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_{t_m}$ is positive definite for each $m \in \mathbb{N}$.
The estimated edr directions $\widehat{\boldsymbol{\beta}}_{i,t}$, $i = 1, \dots, K$, are eigenvectors of $\widetilde{\boldsymbol{V}}_t$ and consequently of $\widetilde{\boldsymbol{V}}_t/t$ for each $t > 0$. Without loss of generality let $\widehat{\boldsymbol{\beta}}_{i,t}$, $i = 1, \dots, K$, be oriented in the sense of Lemma 5.2. Since $\boldsymbol{u}_2 = \boldsymbol{\beta}$ we have

$$\lim_{m\to\infty} \widehat{\boldsymbol{\beta}}_{i,t_m}'\boldsymbol{\Sigma}\boldsymbol{\beta}_1 = \lim_{m\to\infty} \widehat{\boldsymbol{\beta}}_{i,t_m}'\boldsymbol{\beta} = 0, \ i = 1, \dots, K. \tag{5.6}$$

We now get $\varrho^2_{1,t_m}, \ldots, \varrho^2_{K,t_m}$ as eigenvalues of

$$\boldsymbol{A}_{t_m} := \left(\boldsymbol{B}'\boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_{t_m}\right)\left(\widehat{\boldsymbol{B}}'_{t_m}\boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_{t_m}\right)^{-1}\left(\widehat{\boldsymbol{B}}'_{t_m}\boldsymbol{\Sigma}\boldsymbol{B}\right),$$

since $\boldsymbol{B}'\boldsymbol{\Sigma}\boldsymbol{B} = \boldsymbol{I}_K$. With $\boldsymbol{e}_1 = (1, 0, \ldots, 0)'$ we find

$$\boldsymbol{e}_1'\boldsymbol{A}_{t_m}\boldsymbol{e}_1 = \left(\boldsymbol{\beta}_1'\boldsymbol{\Sigma}\widehat{\boldsymbol{\beta}}_{i,t_m}\right)_{i=1,\ldots,K}\left(\widehat{\boldsymbol{B}}'_{t_m}\boldsymbol{\Sigma}\widehat{\boldsymbol{B}}_{t_m}\right)^{-1}\left(\widehat{\boldsymbol{\beta}}'_{i,t_m}\boldsymbol{\Sigma}\boldsymbol{\beta}_1\right)_{i=1,\ldots,K}.$$

Looking at (5.6), we have $\lim_{m\to\infty}\boldsymbol{e}_1'\boldsymbol{A}_{t_m}\boldsymbol{e}_1 = 0$. This completes the proof since $\varrho^2_{K,t_m} \leq \boldsymbol{e}_1'\boldsymbol{A}_{t_m}\boldsymbol{e}_1$ for all $m \in \mathbb{N}$. $\square$

Proof of Theorem 4.1

Let $r := \max\{\|\boldsymbol{x}_1\|, \ldots, \|\boldsymbol{x}_n\|\}$. Without loss of generality let $\overline{\boldsymbol{x}} = \mathbf{o}$. Moreover let $S_1 = \{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n_1}\}$ and $S_h = \{\boldsymbol{x}_{n_1+\cdots+n_{h-1}+1}, \ldots, \boldsymbol{x}_{n_1+\cdots+n_h}\}$, $h = 2, \ldots, H$, such that S_h denotes the set of points of $(\boldsymbol{x})_n$ which fall into the h-th slice.

We corrupt $(\boldsymbol{x})_n$ in the following manner: Let $\tilde{\boldsymbol{u}} \in \mathbb{R}^d$, $\tilde{\boldsymbol{u}}'\tilde{\boldsymbol{u}} = 1$, and replace S_1 by $S_1^{t_1}$ with $t_1 \in \mathbb{R}$ and

$$S_1^{t_1} = \left\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\lfloor n_S/2\rfloor}, t_1\tilde{\boldsymbol{u}}, t_1\tilde{\boldsymbol{u}} - (\boldsymbol{x}_1 - t_1\tilde{\boldsymbol{u}}), \ldots, t_1\tilde{\boldsymbol{u}} - (\boldsymbol{x}_{\lfloor n_S/2\rfloor} - t_1\tilde{\boldsymbol{u}})\right\}$$

if n_S is odd, and by

$$S_1^{t_1} = \left\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{\lfloor n_S/2\rfloor}, t_1\tilde{\boldsymbol{u}} - (\boldsymbol{x}_1 - t_1\tilde{\boldsymbol{u}}), \ldots, t_1\tilde{\boldsymbol{u}} - (\boldsymbol{x}_{\lfloor n_S/2\rfloor} - t_1\tilde{\boldsymbol{u}})\right\}$$

if n_S is even. The slices $S_2, \ldots, S_{\lceil H(1-\alpha)\rceil}$ are corrupted in an analogous manner using $\tilde{\boldsymbol{u}}$ in each case. For any choice $t_1, \ldots, t_{\lceil H(1-\alpha)\rceil} \in \mathbb{R}$ we get a corrupted sample $(\boldsymbol{x})_{n,m}$ with $m = \lceil H(1-\alpha)\rceil\lfloor(n_S+1)/2\rfloor$. Note that C_1 cannot break down in this situation. Consequently, there exists a lower bound $L_d > 0$ for the smallest eigenvalue of $C_1((\boldsymbol{x})_{n,m})$ and for each corrupted sample $(\boldsymbol{x})_{n,m}$.

The special structure of the contaminated sample and the fact that T_2 is orthogonal equivariant imply that for the corrupted slices the application of T_2 yields $t_h\boldsymbol{u}$, $h = 1, \ldots, \lceil H(1-\alpha)\rceil$, with $\boldsymbol{u} = C_1^{-1/2}((\boldsymbol{x})_{n,m})\tilde{\boldsymbol{u}}$. For the non-corrupted slices the application of T_2 yields a location estimate $\widehat{\boldsymbol{m}}_h$ with $\|\widehat{\boldsymbol{m}}_h\| \leq R := rL_d^{-1/2}$, $h = \lceil H(1-\alpha)\rceil + 1, \ldots, H$.

We now choose $T > 0$ and $t_h = (R + 2hT)/\|\boldsymbol{u}\|$, $h = 1, \ldots, \lceil H(1-\alpha)\rceil$.

For the fourth step of the DAME procedure we look at the projection of the location estimates delivered in the third step onto the direction $\boldsymbol{u}/\|\boldsymbol{u}\|$. We get $\boldsymbol{u}'t_h\boldsymbol{u}/\|\boldsymbol{u}\| = R + 2hT$, $h = 1, \ldots, \lceil H(1-\alpha) \rceil$, and $\boldsymbol{u}'\widehat{\boldsymbol{m}}_h/\|\boldsymbol{u}\| \in [-R, R]$, $h = \lceil H(1-\alpha) \rceil + 1, \ldots, H$. Hence, a number of $\lfloor H\alpha \rfloor(\lfloor H\alpha \rfloor - 1)/2 < \lfloor H\alpha \rfloor(\lfloor H\alpha \rfloor + 1)/2$ absolute differences between points of the projected sample are smaller or equal to $2R$, the other differences are all larger than T. Therefore the scale estimator RCQ_α yields a value larger than T, too. Looking at the definition of the projection pursuit estimator C_2, we see that the largest eigenvalue of this estimator exceeds T^2. Therefore, it can be made arbitrarily large by choosing T appropriately. This shows that $\lceil H(1-\alpha) \rceil \lfloor (n_S+1)/2 \rfloor / n$ is an upper bound for the breakdown point $\varepsilon^\star_{2,+}$.

Now consider the situation that without loss of generality the location estimates $\widehat{\boldsymbol{m}}_1, \ldots, \widehat{\boldsymbol{m}}_{\lfloor H\alpha \rfloor + 1}$ remain bounded: $\|\widehat{\boldsymbol{m}}_h\| \leq L_0$, $h = 1, \ldots, \lfloor H\alpha \rfloor + 1$, $L_0 > 0$. Then for each $\boldsymbol{u} \in \mathbb{R}^d$, $\|\boldsymbol{u}\| = 1$, we have

$$|\boldsymbol{u}'\widehat{\boldsymbol{m}}_{h_1} - \boldsymbol{u}'\widehat{\boldsymbol{m}}_{h_2}| \leq 2L_0$$

with $h_1, h_2 \in \{1, \ldots, \lfloor H\alpha \rfloor + 1\}$, $h_1 \neq h_2$. This means that $\lfloor H\alpha \rfloor(\lfloor H\alpha \rfloor + 1)/2$ absolute differences between points of the projected sample are bounded by $2L_0$. With respect to the definitions of RCQ_α and the projection pursuit covariance estimator in the fourth step of DAME, we see that the largest eigenvalue of the covariance estimator remains bounded. It follows that this eigenvalue can explode only if $\lceil H(1-\alpha) \rceil$ of the location estimates in the third step can be made arbitrarily large. Since the L_1-median is used as a location estimator, this means that one has to corrupt a minimum of $\lfloor (n_S+1)/2 \rfloor$ data points in each of the corresponding $\lceil H(1-\alpha) \rceil$ slices, such that $\lceil H(1-\alpha) \rceil \lfloor (n_S+1)/2 \rfloor / n$ is a lower bound for the breakdown point $\varepsilon^\star_{2,+}$, too. This completes the proof of the first part of the theorem.

The remaining parts can be proved by similar considerations. $\square$

References

Aragon, Y., Saracco, J. (1997), "Sliced Inverse Regression (SIR): An Appraisal of Small Sample Alternatives to Slicing," *Computational Statistics*, 12, 109–130.

Becker, C., Gather, U. (1997), "The Masking Breakdown Point of Multivariate Outlier Identification Rules," *submitted.*

Bedall, F. K., Zimmermann, H. (1979), "Algorithm AS 143: The Mediancentre," *Applied Statistics*, 23, 325–328.

Bura, E. (1997), "Dimension Reduction via Parametric Inverse Regression," L_1-*Statistical Procedures and Related Topics*, MS Lecture Notes, Volume 31, 215–228.

Carroll, R. J., Li, K.-C. (1992), "Measurement Error Regression With Unknown Link: Dimension Reduction and Data Visualization," *Journal of the American Statistical Association*, 87, 1040–1050.

Cook, R. D., Nachtsheim, C. J. (1994), "Reweighting to Achieve Elliptically Contoured Covariates in Regression," *Journal of the American Statistical Association*, 89, 592–599.

Cook, R. D., Weisberg, S. (1991), "Comment on "Sliced Inverse Regression for Dimension Reduction" by K.-C. Li," *Journal of the American Statistical Association*, 86, 328–332.

Croux, C., Ruiz–Gazen, A. (1996), "A Fast Algorithm for Robust Principal Components Based on Projection Pursuit," in: A. Prat (ed.), *Compstat, Proceedings in Computational Statistics*, 12th Symposium, Barcelona 1996, Physika–Verlag, Heidelberg, 211–216.

Donoho, D. L., Huber, P. J. (1983), "The Notion of Breakdown Point," in: P. J. Bickel, K. A. Doksum, J. L. Hodges (eds.), *A Festschrift for Erich Lehmann*, Wadsworth, Belmont, CA, 157–184.

Ferré, L. (1997), "Dimension Choice for Sliced Inverse Regression," *Student*, 2, 95–108.

Friedman, J. H. (1991), "Multivariate Adaptive Regression Splines," invited paper with discussion, *The Annals of Statistics*, 19, 1–141.

Friedman, J. H., Stuetzle, W. (1981), "Projection Pursuit Regression," *Journal of the American Statistical Association*, 76, 817–823.

Griffiths, P., Hill, I. H. (eds.) (1985), *Applied Statistics Algorithms*, Royal Statistical Society, Ellis Horwood Limited.

Grübel, R. (1996), "Orthogonalization of Multivariate Location Estimators: The Orthomedian," *The Annals of Statistics*, 24, 1457–1473.

Härdle, W., Tsybakov, A. B. (1991), "Comment on "Sliced Inverse Regression for Dimension Reduction" by K.-C. Li," *Journal of the American Statistical Association*, 86, 333–335.

Härdle, W., Klinke, S., Turlach, B. A. (1995), *XploRe: An Interactive Computing Environment*, Springer, New York.

Hall, P., Li, K.–C. (1993), "On Almost Linearity of Low Dimensional Projections from High Dimensional Data," *The Annals of Statistics*, 21, 867–889.

Hsing, T., Carroll, R. J. (1992), "An Asymptotic Theory for Sliced Inverse Regression," *The Annals of Statistics*, 20, 1040–1061.

Kötter, T. (1996), "An Asymptotic Result for Sliced Inverse Regression," *Computational Statistics*, 11, 113–136.

Kshirsagar, A. M. (1972), *Multivariate Analysis*, Marcel Dekker, New York.

Li, K.–C. (1991), "Sliced Inverse Regression for Dimension Reduction (with discussion)," *Journal of the American Statistical Association*, 86, 316–342.

Li, K.–C. (1992), "On Principal Hessian Directions for Data Visualization and Dimension Reduction: Another Application of Stein's Lemma," *Journal of the American Statistical Association*, 87, 1025–1039.

Li, G., Chen, Z. (1985), "Projection–Pursuit Approach to Robust Dispersion Matrices and Principal Components: Primary Theory and Monte Carlo," *Journal of the American Statistical Association*, 80, 759–766. Correction: *Journal of the American Statistical Association*, 80, 1084.

Liu, R. Y. (1990), "On a Notion of Data Depth Based on Random Simplices," *The Annals of Statistics*, 18, 405–414.

Lopuhaä, H. P. (1989), "On the Relation between S–Estimators and M–Estimators of Multivariate Location and Covariance," *The Annals of Statistics*, 17, 1662–1683.

Lopuha, H. P., Rousseeuw, P. J. (1991), "Breakdown Points of Affine Equivariant Estimators of Multivariate Location and Covariance Matrices," *The Annals of Statistics*, 19, 229–248.

Rocke, D. M. (1993), "Robustness Properties of S–Estimators of Multivariate Location and Shape in High Dimension," *Working Paper*, Graduate School of Management, University of California, Davis.

Rocke, D. M., Woodruff, D. L. (1996), "Identification of Outliers in Multivariate Data," *Journal of the American Statistical Association*, 91, 1047–1061.

Rousseeuw, P. J., Croux, C. (1993), "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, 88, 1273–1283.

Rousseeuw, P. J., Ruts, I. (1997), "The Bagplot: A Bivariate Box-and-Whiskers Plot," *Technical Report*, University of Antwerp, submitted.

Schott, J. R. (1994), "Determining the Dimensionality in Sliced Inverse Regression," *Journal of the American Statistical Association*, 89, 141–148.

Sheather, S. J., McKean, J. W. (1997), "A Comparison of Procedures based on Inverse Regression," L_1*-Statistical Procedures and Related Topics*, IMS Lecture Notes, Volume 31, 271–278.

Small, C. G. (1990), "A Survey of Multidimensional Medians," *International Statistical Review*, 58, 263–277.

Zhu, L.–X., Fang, K.–T. (1996), "Asymptotics for Kernel Estimate of Sliced Inverse Regression," *The Annals of Statistics*, 24, 1053–1068.