Sliced inverse median difference regression

Stephen Babos[1] · Andreas Artemiou[1] 　

Abstract
In this paper we propose a sufficient dimension reduction algorithm based on the difference of inverse medians. The classic methodology based on inverse means in each slice was recently extended, by using inverse medians, to robustify existing methodology at the presence of outliers. Our effort is focused on using differences between inverse medians in pairs of slices. We demonstrate that our method outperforms existing methods at the presence of outliers. We also propose a second algorithm which is not affected by the ordering of slices when the response variable is categorical with no underlying ordering of its values.

1 Introduction

Sufficient Dimension Reduction (SDR) is a class of dimension reduction techniques used in regression to address the high dimensionality of a predictor vector $X \in \mathbb{R}^p$ when a response variable Y (assumed univariate without loss of generality) is regressed on X. In other words, in SDR we are trying to estimate a $p \times d$ $(d < p)$ matrix $\boldsymbol{\beta}$ such that

$$Y \perp\!\!\!\perp X | \boldsymbol{\beta}^\mathsf{T} X \tag{1}$$

The space spanned by the columns of $\boldsymbol{\beta}$ is called a Dimension Reduction Subspace (DRS). There are many different $\boldsymbol{\beta}$'s that satisfy (1) and the main objective is to estimate the one with the minimum dimension d. The minimum DRS is known as the Central Dimension Reduction Space (CDRS) or simply the Central Space (CS) and is denoted with $\mathcal{S}_{Y|X}$. There are some mild conditions of existence of the CS

✉ Andreas Artemiou
    ArtemiouA@cardiff.ac.uk

[1]   School of Mathematics, Cardiff University, Senghennydd Road, Cardiff CF24 4AG, Wales, UK

which we assume that they hold in this paper (see Yin et al. 2008). A number of methods have been proposed in the SDR literature. The most well known class of methods that has been developed and is being used most frequently is probably the class of methods based on inverse moments—see Li (1991), Cook and Weisberg (1991), Li and Wang (2007) among others. A comprehensive review of this methodology can be found in Li (2018).

There are two main drawbacks in the inverse-moment based methodology like Sliced Inverse Regression (SIR) which was introduced by Li (1991). First is the dependence on moments which suffer on the presence of outliers. To address this, Gather et al. (2001), Dong et al. (2015) and Christou (2018) suggested using inverse medians instead of means. Second is the dependence on the number of slices the range of Y is discretized into, especially in methods like Sliced Average Variance Estimation (SAVE—Cook and Weisberg 1991) where the second inverse moment is used as well. Zhu et al. (2010) suggested a cumulative slicing approach (Cumulative Mean Estimation—CUME) to avoid using the number of slices as a tuning parameters. Artemiou and Tian (2015) identified that the solution given by the cumulative approach will suffer in accuracy in cases where the response is categorical and in computational time in cases where we have massive datasets. They proposed Sliced Inverse Mean Difference (SIMD) approach which uses the difference between slice means and suggested two different algorithms for their method. One is the "left vs right" (LVR) which is more appropriate to be used when the response variable is continuous and the other is the "one vs another" (OVA) which is more appropriate when the response is categorical. The LVR approach was demonstrated to be theoretically equivalent to CUME, but computationally faster, and unlike CUME the OVA approach allowed SIMD to be used for categorical responses.

Although SIMD has certain advantages over similar methods, it suffers in accuracy at the presence of outliers. In this paper, we take a look in addressing this by proposing the use of Sliced Inverse Median Difference (SIMeD) which uses the difference between slice medians. As with SIMD, we propose the use of two different algorithms as they were proposed in Li et al. (2011). The "left vs right" (LVR) which gives computational speed and works better when the response is continuous and the "one vs another" (OVA) which gives better accuracy when the response is categorical. Although there are different definitions for the multivariate median, in this work we focus on the use of L1 median (spatial median) which was used in Dong et al. (2015). The main reason is that theoretical results require the uniqueness of the median and the L1 median is the only one which is shown to be unique for $p > 2$ (Hettmansperger and McKean 1998).

The paper is structured as follows. We discuss previous methodology in Sect. 2 and then in Sect. 3 we discuss the newly proposed method (SIMeD). In Sect. 4 we demonstrate the performance with numerical studies and we close with a small discussion section.

2 Review of previous methodology

In this section we provide a review of the methodology previously proposed in the SDR framework which mostly relates to our proposal. We will discuss Sliced Inverse Regression (SIR), Cumulative Mean Estimation (CUME), Sliced Inverse Mean Difference Regression (SIMD) and Sliced Inverse Median (SIME). Let X be the p dimensional predictor vector, Y be the response variable and $Z = \Sigma^{-1/2}(X - E(X))$ the standardized predictors, where $\Sigma = \text{var}(X)$.

2.1 Sliced inverse regression (SIR)

SIR was introduced by Li (1991) and is considered the first method introduced in the SDR framework. As with all the methods we discuss in this section, SIR depends on the linearity assumption (or the linear conditional mean assumption) which is satisfied when the predictors have an elliptical distribution. The author proposed to standardize the predictors and then use the inverse mean $E(Z|Y)$. By performing an eigenvalue decomposition of the variance-covariance matrix $\text{var}(E(Z|Y))$ one can find the directions which span the CS, $\mathcal{S}_{Y|Z}$. One then can use the invariance principle (see Cook 1998) to find the direction that span $\mathcal{S}_{Y|X}$.

One of the most important aspect of SIR is the use of inverse conditional means. The conditioning on Y is achieved by discretizing Y, that is, by slicing into a number of intervals (denoted with $I_1, \ldots, I_H$ where H the number of intervals). Therefore, the inverse mean $E(Z|Y)$ is in practice replaced with $E(Z|Y \in I_i)$ for $i = 1, \ldots, H$ and thus, it is the eigenvalue decomposition of $\text{var}(E(Z|Y \in I_i))$ which is used to find the directions which span the CS, $\mathcal{S}_{Y|Z}$.

2.2 Cumulative mean estimation (CUME)

Although, it was shown that SIR is robust to the number of slices, Zhu et al. (2010) proposed the use of Cumulative Mean Estimation (CUME) which removes the need of tuning for the number of slices. In essence they proposed the use of the eigenvalue decomposition of the matrix $\text{var}(E(Z|Y \leq y))$ to find the directions which span the CS, $\mathcal{S}_{Y|Z}$. In practice this means that one increases the value of y and recalculates the mean each time the value of y is large enough for a new observation to be included in the range $\{Y \leq y\}$.

2.3 Sliced inverse mean difference (SIMD)

Sliced Inverse Mean Difference (SIMD) was introduced by Artemiou and Tian (2015) to address issues of the aforementioned methodology. First, they identified that in case of massive datasets, the cumulative approach by CUME will be computationally very costly. Second, they identified that in the case of a response variable that has no underlying ordering, CUME wouldn't be able to work. Therefore, they proposed the use of the mean difference between subsets of the data

(not necessarily slices) using two different algorithms which were influenced by the work of Li et al. (2011).

The first one, called "left vs right" (LVR) used the eigenvalue decomposition of $\mathrm{var}(E(\mathbf{Z}|Y > q_i) - E(\mathbf{Z}|Y \leq q_i))$ where $i = 1, \ldots H - 1$ and q_i is the cutoff point between slice i and slice $i + 1$. This algorithm is shown to be theoretically equivalent to CUME but computationally more efficient due to the use of fewer comparisons. As with SIR it was shown to be robust to the number of slices. The second algorithm, called "one vs another" (OVA) uses the eigenvalue decomposition of $\mathrm{var}(E(\mathbf{Z}|Y \in I_j) - E(\mathbf{Z}|Y \in I_i))$ where $i = 1, \ldots H - 1$ and $j = i, \ldots, H$ (essentially one takes the difference between the means of two slices). This method was shown to be equivalent to SIR, and it was also more appropriate than CUME to be used in cases where Y was categorical. In cases where Y is categorical different orderings of the values of Y can lead to different results in CUME and the LVR approach and therefore comparing the difference between all possible pair of slices made more sense as the result in these case were not affected by the ordering of the values of Y.

2.4 Sliced inverse median (SIME)

Finally, the work by Dong et al. (2015) showed that to robustify SIR one can use the inverse L1 median instead of means. They preferred L1 median due to its uniqueness in cases where $p \geq 2$. In their work they defined the inverse L1 median as:

Definition 1 The inverse L1 median is given by $\tilde{m} = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| | Y)$, where $\|\cdot\|$ is the Euclidean norm.

Similarly, we denote with $\tilde{m}_{\mathbf{Z}} = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{Z} - \boldsymbol{\mu}\| | Y)$. Therefore they performed an eigenvalue decomposition of the matrix: $\mathrm{var}(\tilde{m}_{\mathbf{Z}})$ to find the the vectors which span the CS $\mathcal{S}_{Y|\mathbf{Z}}$. The authors demonstrated the robustness in the presence of outliers of this algorithms as well as the good overall performance when there are no outliers. More recently, similar ideas were proposed by Christou (2018) but instead of using the L1 median, they suggest using the Oja and the Tukey median.

3 Sliced inverse median difference (SIMeD)

In this section the main idea of this paper is presented and an estimation algorithm and a method for determining the dimension of the CS is given.

3.1 Using inverse medians in sufficient dimension reduction

Here we present the foundations of our proposal. Let (X, Y) be a random pair where $X \in \mathbb{R}^p$ denotes the predictors and the response Y is assumed univariate without loss of generality. We also assume that the range of Y is discretized in H slices, which we denote with $J_1, \ldots, J_H$. Also we define the dividing points between two slices with

$q_1, \ldots q_{H-1}$, that is q_k separates slice J_k from slice J_{k+1}. Finally, let $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ the standardized predictors, where $\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{X})$ and $\boldsymbol{\mu} = E(\mathbf{X})$.

Using the definition of the inverse L1 median one can define now the difference between medians. We use A_1 and A_2 to be two disjoint sets of Ω, the support of Y. Then we define, $\tilde{Y}$ to be the discretized version of Y,

$$\tilde{Y} = I(Y \in A_1) - I(Y \in A_2)$$

where $I(\cdot)$ denotes the indicator function. Therefore, one can define the difference between two medians to be:

$$\tilde{m}_d = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, \tilde{Y} = 1) - \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, \tilde{Y} = -1).$$

Depending on the way one chooses to define the sets A_1 and A_2, then the definition of $\tilde{m}_d$ may take different forms. In this article, we will use two different definitions [as it was the case for the difference between means in the proposal of SIMD by Artemiou and Tian (2015)]. The first is called "left vs right" (LVR) and being more appropriate in cases where the response is continuous or at least a discrete variable with a logical numerical ordering of the values it can take. The second is called "one vs another" and is more appropriate when Y is categorical. With the LVR approach we select multiple cutoff points q and for each cutoff points we split the data to those above (A_1) and those below (A_2) the cutoff point. For the OVA approach we split the data points in a number of slices (like in previous slicing methods) and we select all possible pairs of slices, (one slice of the pair is A_1 and the other is A_2).

The above two methods lead to the following definitions.

Definition 2 The "left vs right" sliced inverse median difference (SIMeD$_{\mathrm{LVR}}$) is defined as:

$$\tilde{m}_i^{\mathrm{LVR}} = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, Y > q_i) - \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, Y \le q_i), \qquad (2)$$

for $i = 1, \ldots, H - 1$. Similarly the "one vs another" sliced inverse median difference (SIMeD$_{\mathrm{OVA}}$) is defined as:

$$\tilde{m}_{i,j}^{\mathrm{OVA}} = \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, Y \in J_i) - \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} E(\|\mathbf{X} - \boldsymbol{\mu}\| \,|\, Y \in J_j), \qquad (3)$$

where $i = 2, \ldots, H$, $j = 1, \ldots H - 1$ and $j < i$.

To simplify the notation we will avoid using the superscripts LVR and OVA from the mathematical notation of the median and the subscripts from the name in the rest of the paper in cases where it is clear which approach is being used.

From the definitions, one can see that each time the LVR algorithm is applied for each cutoff point one uses all the points, while each time the OVA algorithm is applied only the points on the two selected slices are used. The fact that the LVR algorithm uses all the points for different cutoff points, allows it to be more accurate than SIME which uses only the points within a slice. For example, if one compares SIME with H

slices where the $H - 1$ cutoff points between the slices are used as the cutoff points of the LVR algorithm of SIMeD, then the SIMeD LVR algorithm performs more accurately and is more robust to the number of slices H. The LVR though will not perform in cases where the response variable is categorical, with no sense of ordering between the values, while SIME will still perform perfectly fine. The main problem with categorical responses in the LVR algorithm of SIMeD is the fact that it is constructed assuming the responses can be ordered and different orderings of the categories will give different answers. In SIME, as well as in earlier methods like SIR, this is not a problem since we use the points of each slide independently to find the inverse medians. Having this in mind, we propose the use of the OVA algorithm which finds the differences between the medians of all possible pairs of slices. Therefore in the case of a categorical response, the OVA algorithm will not be affected by different orderings of the categorical response. The following theorem gives the theoretical justification and we will demonstrate the advantages of each algorithm in the numerical experiments section later. Note that when using the medians we need the stronger assumption of ellipticity of the predictors instead of the more classic Linear Conditional Mean assumption in the SDR literature.

Theorem 1 *If X are elliptically distributed then $\tilde{m}_i^{\text{LVR}} \in \mathcal{S}_{Y|X}$ and $\tilde{m}_{i,j}^{\text{OVA}} \in \mathcal{S}_{Y|X}$*

The proof of this theorem is pretty straight forward. Using the fact that Dong et al. (2015) showed that the L1 median $\tilde{m} \in \mathcal{S}_{Y|X}$ then $\tilde{m}_i^{\text{LVR}}$ and $\tilde{m}_{i,j}^{\text{OVA}}$ are linear combinations of vectors in $\mathcal{S}_{Y|X}$ and therefore they belong in it.

3.2 Estimation algorithm

The estimation algorithm in the LVR case is as follows:

1. Standardize data to find $\mathbf{Z} = \hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X} - \hat{\boldsymbol{\mu}})$ using robust estimators for location and scale. (we give details below)
2. Divide the range of Y into H slices using cutoff points $q_1, \ldots, q_{H-1}$.
3. For each q_i, $i = 1, \ldots, H - 1$ calculate the median difference using the sample version of formula (2), which we denote with $\hat{\tilde{m}}_i$.
4. Do an eigenvalue decomposition of the matrix $\hat{V}$ to find the largest d eigenvalues and the corresponding eigenvectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_d$, where $\hat{V} = \sum_{i=1}^{H-1} \hat{\tilde{m}}_i \hat{\tilde{m}}_i^{\mathsf{T}}$.
5. Find $\boldsymbol{\beta}_k = \hat{\boldsymbol{\Sigma}}^{1/2} \boldsymbol{v}_k$, for $k = 1, \ldots, d$ to estimate the vectors that span $\mathcal{S}_{Y|X}$.

Similarly, the estimation algorithm in the OVA case is as follows:

1. Standardize data to find $\mathbf{Z} = \hat{\boldsymbol{\Sigma}}^{-1/2}(\mathbf{X} - \hat{\boldsymbol{\mu}})$ using robust estimators for location and scale. (we give details below)
2. Divide the range of Y into H slices and find all possible pairs, that is $\binom{H}{2}$ pairs.
3. For each pair (i, j), $i = 1, \ldots, H$ and $j = 1, \ldots, i$ calculate the median difference using the sample version of formula (3), which we denote with $\hat{\tilde{m}}_{i,j}$.

4. Do an eigenvalue decomposition of the matrix $\hat{V}$ to find the largest d eigenvalues and the corresponding eigenvectors $v_1, \ldots, v_d$, where $\hat{V} = \sum_{i=1}^{\binom{H}{2}} \hat{\hat{m}}_{i,j} \hat{\hat{m}}_{i,j}^{\mathsf{T}}$.

5. Find $\boldsymbol{\beta}_k = \hat{\boldsymbol{\Sigma}}^{1/2} v_k$, for $k = 1, \ldots, d$ to estimate the vectors that span $\mathcal{S}_{Y|X}$.

We suggest the use of a robust estimator from $\boldsymbol{\Sigma} = \mathrm{var}(X)$ for both algorithms to provide better results. Since we are using a robust estimator for the mean, it makes sense to use a robust estimator for $\boldsymbol{\Sigma}$. If we don't, then outliers will have an effect on the estimator of $\boldsymbol{\Sigma}$ and therefore our results will be affected as well. Therefore, we use the minimum covariance determinant (MCD) estimator for $\boldsymbol{\Sigma}$, which is defined as the covariance matrix which has the minimum determinant among all possible covariance matrices using a set of k points where $n/2 < k < n$. It is implemented in function *covMcd* in package *robustbase* in R. (see Maechler et al. 2018)

3.3 Order determination tests

There are a number of ideas that were introduced in the literature for determining the dimension of the CS. Sequential tests were proposed for several methods in the literature [see for example Li (1991) for SIR, Shao et al. (2007) for SAVE, Artemiou and Tian (2015) for SIMD]. Bura and Yang (2011) proposed a unified approach. The second method was a BIC-type criterion which was proposed by Zhu et al. (2006) and also used for other methods like CUME in Zhu et al. (2010) and Li et al. (2011). Recently, a method called the ladle plot was proposed by Luo and Li (2016). In this paper, we use the BIC type criterion as the asymptotic distribution, which is needed for sequential tests, has very complicated variance structure when the estimators are based on medians. Also the ladle plot requires bootstrapping which is computationally more expensive than the other two approaches.

In the BIC-type criterion one tries to maximize:

$$G_n(k) = \sum_{i=1}^{k} \lambda_i - \lambda_1 c_1(n) c_2(k)$$

where $k \in \{1, \ldots, p\}$, λ_i the ith eigenvalue of candidate matrix $\hat{V}$ (in the estimation algorithm), $\lambda_1 \geq \cdots \geq \lambda_p$, $c_1(n)$ is a function of n which converge in probability to 0 as $n \to \infty$ and $c_2(k)$ is a nondecreasing sequence of numbers that depends on k.

The asymptotic properties and more specifically the asymptotic consistency of the method were studied in Zhu et al. (2006). Their result assumes that as long as we have a consistent estimator of the candidate matrix (matrix $\hat{V}$ in our estimation algorithm in Sect. 3.2) then the BIC criterion can estimate consistently the dimension d of the central subspace. Christou (2018) has shown the consistency of similar criterion when using the Tukey and Oja medians in each slice. We can use similar arguments in this case to prove the consistency of our criterion. The only step that will be different is the use of consistency of the estimator of the L1 median which was proved in Brown (1983).

4 Numerical experiments

In this section we discuss numerical experiments where we demonstrate the improvement in performance when Sliced Inverse Median Regression (SIMeD) is used compared to previous methodology. We start with simulated datasets and we then use 3 real datasets.

4.1 Simulated datasets

We run three different simulation experiments using the following models:

$$\text{Model I:}\quad Y = X_1 + X_2 + \sigma\varepsilon$$
$$\text{Model II:}\quad Y = X_1/[0.5 + (X_2 + 1)^2] + \sigma\varepsilon,$$
$$\text{Model III:}\quad Y = 1 + 0.6X_1 - 0.4X_2 + 0.8X_3 + \sigma\varepsilon,$$
$$\text{Model IV:}\quad Y = (1 + (\sigma/2)\varepsilon))X_1,$$
$$\text{Model V:}\quad Y = X_1(X_1 + X_2 + 1) + \sigma\varepsilon,$$

where $X = (X_1, \ldots, X_p)$ ($p = 10, 20$ or 30) are either simulated form $N_p(\boldsymbol{0}, \boldsymbol{I})$, or $X \sim$ Multivariate Cauchy, or even a mixture of two normals where a big proportion is drawn from $N_p(\boldsymbol{0}, \boldsymbol{I})$ and a small proportion is drawn from $N_p(\boldsymbol{0}, k\boldsymbol{I})$ where $k \in \mathbb{R}$. Furthermore, $\varepsilon \sim N(0, 1)$ and $n = 100$, $H = 5, 10, 20$. We run 100 simulations and we report the average trace correlation among all simulations. Trace correlation is calculated as follows:

$$r = \frac{\text{trace}\{\boldsymbol{P_B}\boldsymbol{P_{\hat{B}}}\}}{d}, \tag{4}$$

where $\boldsymbol{P_B}$ is the projection matrix on the true subspace and $\boldsymbol{P_{\hat{B}}}$ the projection matrix on the estimated subspace. It takes values between 0 and 1 and the closer to 1 the better the estimation. We compare our newly proposed algorithm with SIR (Li 1991), SIME (Dong et al. 2015) and SIMD (Artemiou and Tian 2015). Here we have to be careful as SIR and SIME use individual slices and SIMD and SIMeD use comparison between slices. To make sure we have a fair setting for a comparison, the number of slices are chosen so that each slice contains approximately the same number of points. The cutoff points between the slices in SIR and SIME are used as the cutoff points of the LVR algorithm for SIMD and SIMeD, or in other words, those cutoff points are the $(kn/H)^{\text{th}}$ percentiles, for $k = 1, \ldots, H - 1$. Also, note that, unless otherwise noted, for SIMD and SIMeD we run the LVR approach.

In the first experiment we compare the performance of the four algorithms on the five models with fixed number of slices $H = 10$ and different predictor dimension p when the predictors are distributed from one of the three following scenarios:

- a standard multivariate normal,
- from a standard multivariate Cauchy where the components are pairwise independent (hence there are outliers)

Table 1 Comparison of results for SIR, SIME, SIMD, SIMeD for different values of p for the 5 models

Model	p	Distribution	Method			
			SIR	SIME	SIMD	SIMeD
I	10	a	0.99 (0.004)	0.97 (0.016)	0.99 (0.004)	0.98 (0.011)
		b	0.61 (0.274)	0.83 (0.139)	0.62 (0.250)	0.84 (0.137)
		c	0.73 (0.159)	0.94 (0.051)	0.74 (0.149)	0.96 (0.031)
	20	a	0.98 (0.007)	0.94 (0.041)	0.98 (0.006)	0.95 (0.027)
		b	0.49 (0.279)	0.62 (0.196)	0.53 (0.247)	0.69 (0.179)
		c	0.80 (0.124)	0.87 (0.080)	0.80 (0.123)	0.92 (0.044)
	30	a	0.97 (0.012)	0.90 (0.045)	0.97 (0.010)	0.93 (0.033)
		b	0.39 (0.274)	0.39 (0.238)	0.42 (0.250)	0.52 (0.202)
		c	0.82 (0.094)	0.74 (0.136)	0.83 (0.088)	0.85 (0.072)
II	10	a	0.82 (0.096)	0.81 (0.081)	0.85 (0.065)	0.84 (0.063)
		b	0.39 (0.208)	0.50 (0.131)	0.39 (0.170)	0.49 (0.108)
		c	0.53 (0.129)	0.77 (0.114)	0.58 (0.120)	0.82 (0.065)
	20	a	0.64 (0.112)	0.61 (0.109)	0.72 (0.074)	0.70 (0.076)
		b	0.26 (0.178)	0.31 (0.154)	0.28 (0.162)	0.34 (0.141)
		c	0.49 (0.122)	0.61 (0.095)	0.55 (0.118)	0.69 (0.072)
	30	a	0.55 (0.089)	0.51 (0.088)	0.62 (0.075)	0.60 (0.074)
		b	0.16 (0.161)	0.18 (0.142)	0.18 (0.172)	0.23 (0.148)
		c	0.43 (0.094)	0.49 (0.088)	0.51 (0.082)	0.59 (0.072)
III	10	a	0.99 (0.006)	0.91 (0.020)	0.99 (0.005)	0.98 (0.013)
		b	0.58 (0.231)	0.81 (0.121)	0.58 (0.225)	0.83 (0.118)
		c	0.72 (0.147)	0.94 (0.036)	0.72 (0.143)	0.96 (0.022)
	20	a	0.98 (0.010)	0.94 (0.028)	0.98 (0.009)	0.95 (0.023)
		b	0.53 (0.263)	0.62 (0.191)	0.54 (0.243)	0.68 (0.174)
		c	0.81 (0.121)	0.86 (0.072)	0.81 (0.113)	0.91 (0.038)
	30	a	0.96 (0.018)	0.88 (0.059)	0.96 (0.014)	0.91 (0.044)
		b	0.46 (0.278)	0.46 (0.231)	0.48 (0.252)	0.55 (0.193)
		c	0.81 (0.119)	0.75 (0.129)	0.82 (0.112)	0.85 (0.075)
IV	10	a	0.99 (0.004)	0.97 (0.017)	0.99 (0.003)	0.98 (0.013)
		b	0.57 (0.399)	0.83 (0.207)	0.61 (0.386)	0.86 (0.159)
		c	0.75 (0.123)	0.94 (0.035)	0.76 (0.120)	0.96 (0.022)
	20	a	0.99 (0.007)	0.95 (0.026)	0.99 (0.007)	0.96 (0.020)
		b	0.48 (0.398)	0.62 (0.290)	0.52 (0.388)	0.67 (0.262)
		c	0.80 (0.131)	0.85 (0.110)	0.81 (0.121)	0.91 (0.062)
	30	a	0.97 (0.015)	0.90 (0.048)	0.98 (0.011)	0.93 (0.034)
		b	0.37 (0.384)	0.36 (0.308)	0.43 (0.368)	0.51 (0.290)
		c	0.83 (0.101)	0.72 (0.149)	0.84 (0.100)	0.85 (0.070)
V	10	a	0.63 (0.145)	0.67 (0.130)	0.73 (0.109)	0.76 (0.097)
		b	0.17 (0.155)	0.37 (0.172)	0.27 (0.187)	0.44 (0.168)
		c	0.41 (0.148)	0.60 (0.152)	0.45 (0.147)	0.67 (0.139)

Table 1 continued

Model	p	Distribution	Method			
			SIR	SIME	SIMD	SIMeD
	20	a	0.42 (0.154)	0.43 (0.148)	0.55 (0.132)	0.55 (0.127)
		b	0.12 (0.132)	0.20 (0.150)	0.16 (0.146)	0.24 (0.164)
		c	0.33 (0.135)	0.38 (0.141)	0.40 (0.137)	0.47 (0.127)
	30	a	0.30 (0.114)	0.28 (0.118)	0.41 (0.120)	0.40 (0.115)
		b	0.10 (0.135)	0.09 (0.093)	0.14 (0.147)	0.13 (0.110)
		c	0.26 (0.123)	0.25 (0.116)	0.33 (0.119)	0.32 (0.111)
Time in seconds			11.55	89.01	16.53	215.87

The *distriution* column indicates whether X were drawn from multivariate standard Normal (a), from a multivariate standard Cauchy (b) or from a mixture of multivariate Normal distributions where 95% of the points are from standard normal and 5% from $N(0, 10I)$ (c). The number of slices $H = 10$. The last line gives information of the time in seconds needed to run all the simulations on the table for each method

- from a mixture of multivariate normal distributions (where we can control the number of outliers).

As we can see in Table 1, our proposed algorithms, SIMeD, outperforms by a large margin SIMD in both scenarios which produce outliers while the two methods perform similarly when there are no outliers. Also, in all the settings SIMeD is more robust than SIME which is the algorithm that uses the median in each slice as was proposed in Dong et al. (2015). A similar advantage was demonstrated between SIMD and SIR in Artemiou and Tian (2015). Finally, Artemiou and Tian (2015) observed that the OVA approach of SIMD is equivalent to the SIR algorithm. Similarly here, we observe that the OVA approach of the SIMeD method is equivalent to SIME therefore we do not report results.

In the second experiment we compare the four methods on different number of slices in an effort to check the robustness of the algorithms against the number of slices. The results for the number of slices $H = 5, 10, 20$ are presented in Table 2 where we can see that all algorithms are robust when there are no outliers. When we have outliers in the sample the algorithms using the inverse medians (SIME and SIMeD) are more robust to the number of slices. Especially our newly proposed algorithms for models I and II shows no real difference in performance for different number of slices whether there are outliers or not.

We further study the effect of the variance and the proportion of outliers in the case when we have mixture of multivariate normals and the results are shown in Tables 3 and 4 where we can see that our proposed methodology SIMeD is not affected at all by any of the two tuning parameters. SIME is as well not affected but SIMeD outperforms it on all settings we have explored.

Finally, we run the four experiments to verify the performance of our BIC criterion as was described in section 3.3. We set $c_1(n) = (1/2)n^{-3/5}$ and $c_2(n) = k(k + 1)/2$ and we run it for two different sample sizes $n = 200$ and $n = 400$. We choose to run the test on model I which has dimension $d = 1$ and model II which

Table 2 Comparison of results for SIR, SIME, SIMD, SIMeD for different values of *H* for the 5 models

Model	H	Distribution	Method			
			SIR	SIME	SIMD	SIMeD
I	5	a	0.99 (0.008)	0.97 (0.017)	0.99 (0.009)	0.97 (0.016)
		b	0.56 (0.290)	0.80 (0.140)	0.58 (0.274)	0.81 (0.138)
		c	0.70 (0.156)	0.96 (0.026)	0.71 (0.156)	0.96 (0.022)
	10	a	0.99 (0.005)	0.97 (0.017)	0.99 (0.005)	0.98 (0.014)
		b	0.63 (0.280)	0.79 (0.140)	0.64 (0.264)	0.81 (0.126)
		c	0.73 (0.159)	0.94 (0.051)	0.74 (0.149)	0.96 (0.031)
	20	a	0.99 (0.003)	0.96 (0.028)	0.99 (0.003)	0.98 (0.015)
		b	0.66 (0.240)	0.74 (0.184)	0.68 (0.207)	0.81 (0.153)
		c	0.79 (0.121)	0.73 (0.236)	0.79 (0.110)	0.94 (0.200)
II	5	a	0.83 (0.077)	0.82 (0.082)	0.84 (0.070)	0.83 (0.072)
		b	0.36 (0.224)	0.46 (0.161)	0.37 (0.223)	0.48 (0.165)
		c	0.55 (0.134)	0.80 (0.093)	0.58 (0.122)	0.83 (0.075)
	10	a	0.81 (0.101)	0.78 (0.107)	0.85 (0.066)	0.83 (0.072)
		b	0.35 (0.230)	0.47 (0.157)	0.37 (0.186)	0.48 (0.133)
		c	0.53 (0.129)	0.77 (0.114)	0.58 (0.120)	0.82 (0.065)
	20	a	0.76 (0.121)	0.72 (0.123)	0.84 (0.071)	0.82 (0.077)
		b	0.36 (0.228)	0.47 (0.168)	0.37 (0.091)	0.48 (0.119)
		c	0.49 (0.159)	0.67 (0.135)	0.57 (0.130)	0.80 (0.075)
III	5	a	0.98 (0.008)	0.97 (0.018)	0.98 (0.008)	0.97 (0.017)
		b	0.56 (0.234)	0.79 (0.150)	0.56 (0.228)	0.79 (0.147)
		c	0.71 (0.133)	0.96 (0.023)	0.71 (0.142)	0.96 (0.020)
	10	a	0.99 (0.006)	0.97 (0.020)	0.99 (0.005)	0.97 (0.016)
		b	0.64 (0.215)	0.84 (0.116)	0.63 (0.209)	0.85 (0.103)
		c	0.72 (0.147)	0.94 (0.036)	0.72 (0.143)	0.96 (0.022)
	20	a	0.99 (0.004)	0.96 (0.036)	0.99 (0.004)	0.98 (0.015)
		b	0.68 (0.238)	0.67 (0.211)	0.67 (0.228)	0.76 (0.187)
		c	0.77 (0.140)	0.73 (0.251)	0.77 (0.123)	0.83 (0.205)
IV	5	a	0.99 (0.007)	0.97 (0.019)	0.99 (0.007)	0.97 (0.018)
		b	0.55 (0.390)	0.78 (0.228)	0.58 (0.372)	0.79 (0.222)
		c	0.71 (0.144)	0.96 (0.022)	0.71 (0.142)	0.97 (0.019)
	10	a	0.99 (0.003)	0.97 (0.018)	0.99 (0.003)	0.98 (0.013)
		b	0.58 (0.371)	0.83 (0.211)	0.60 (0.358)	0.85 (0.184)
		c	0.75 (0.123)	0.94 (0.035)	0.76 (0.120)	0.96 (0.022)
	20	a	0.99 (0.003)	0.96 (0.019)	0.99 (0.003)	0.98 (0.011)
		b	0.58 (0.408)	0.75 (0.274)	0.63 (0.371)	0.85 (0.190)
		c	0.78 (0.140)	0.76 (0.235)	0.78 (0.123)	0.86 (0.192)
V	5	a	0.61 (0.149)	0.65 (0.148)	0.66 (0.135)	0.69 (0.130)
		b	0.19 (0.155)	0.29 (0.178)	0.23 (0.163)	0.32 (0.188)
		c	0.41 (0.144)	0.63 (0.144)	0.42 (0.146)	0.65 (0.137)

Table 2 continued

Model	H	Distribution	Method			
			SIR	SIME	SIMD	SIMeD
	10	a	0.63 (0.145)	0.67 (0.130)	0.73 (0.109)	0.76 (0.097)
		b	0.17 (0.155)	0.37 (0.172)	0.27 (0.187)	0.44 (0.168)
		c	0.41 (0.148)	0.60 (0.152)	0.45 (0.147)	0.67 (0.139)
	20	a	0.57 (0.171)	0.57 (0.161)	0.74 (0.119)	0.77 (0.096)
		b	0.25 (0.195)	0.50 (0.204)	0.41 (0.215)	0.59 (0.176)
		c	0.36 (0.156)	0.47 (0.152)	0.42 (0.146)	0.56 (0.144)
Time in seconds			14.12	97.51	21.57	185.09

The *distribution* column indicates whether X were drawn from multivariate standard Normal (a), from a multivariate standard Cauchy (b) or from a mixture of multivariate Normal distributions where 95% of the points are from standard normal and 5% from $N(0, 10I)$ (c). The dimension of the predictor $p = 10$. The last line gives information of the time in seconds needed to run all the simulations on the table for each method

Table 3 Comparison of results for SIR, SIME, SIMD, SIMeD on the effect of the variance of the Normal distribution the outliers are drawn from for the 5 models when the data are drawn from a mixture of Normal distributions

Model	SV	Method			
		SIR	SIME	SIMD	SIMeD
I	5	0.90 (0.081)	0.96 (0.025)	0.91 (0.072)	0.97 (0.015)
	10	0.74 (0.128)	0.94 (0.041)	0.75 (0.123)	0.96 (0.022)
	20	0.62 (0.179)	0.94 (0.040)	0.62 (0.174)	0.96 (0.023)
II	5	0.64 (0.137)	0.76 (0.110)	0.68 (0.120)	0.82 (0.080)
	10	0.54 (0.153)	0.77 (0.118)	0.58 (0.130)	0.83 (0.079)
	20	0.45 (0.148)	0.77 (0.120)	0.49 (0.135)	0.83 (0.064)
III	5	0.89 (0.083)	0.95 (0.029)	0.89 (0.077)	0.96 (0.019)
	10	0.73 (0.150)	0.94 (0.042)	0.74 (0.140)	0.96 (0.023)
	20	0.60 (0.192)	0.94 (0.035)	0.62 (0.177)	0.96 (0.021)
IV	5	0.90 (0.085)	0.95 (0.031)	0.90 (0.080)	0.97 (0.019)
	10	0.74 (0.123)	0.95 (0.032)	0.74 (0.119)	0.97 (0.017)
	20	0.57 (0.188)	0.95 (0.032)	0.58 (0.181)	0.97 (0.018)
V	5	0.42 (0.165)	0.58 (0.166)	0.47 (0.165)	0.67 (0.132)
	10	0.41 (0.148)	0.60 (0.152)	0.45 (0.147)	0.67 (0.139)
	20	0.38 (0.147)	0.62 (0.142)	0.40 (0.157)	0.69 (0.131)

The column "sv" indicates that 5% of the points are drawn from multivariate Normal with mean 0 and covariance matrix equal to "sv" times the identity matrix. All other points are drawn from a standard multivariate Normal distribution

has dimension $d = 2$. We can see from the results in Table 5 that our BIC criterion works well for SIMeD in all settings and it outperforms the same criterion for SIME especially when the predictors are drawn from a multivariate standard Cauchy distribution.

Table 4 Comparison of results for SIR, SIME, SIMD, SIMeD on the effect of the proportion of outliers for the 5 models when the data are drawn from a mixture of Normal distributions

Model	k	Method			
		SIR	SIME	SIMD	SIMeD
I	1	0.91 (0.108)	0.97 (0.020)	0.91 (0.105)	0.97 (0.015)
	3	0.80 (0.148)	0.96 (0.028)	0.81 (0.144)	0.97 (0.018)
	5	0.73 (0.159)	0.94 (0.051)	0.74 (0.149)	0.96 (0.031)
II	1	0.76 (0.104)	0.79 (0.110)	0.80 (0.084)	0.84 (0.066)
	3	0.61 (0.150)	0.77 (0.112)	0.65 (0.132)	0.83 (0.068)
	5	0.53 (0.129)	0.77 (0.114)	0.58 (0.120)	0.82 (0.065)
III	1	0.92 (0.108)	0.97 (0.022)	0.92 (0.103)	0.97 (0.016)
	3	0.82 (0.139)	0.95 (0.024)	0.83 (0.132)	0.97 (0.019)
	5	0.72 (0.147)	0.94 (0.036)	0.72 (0.143)	0.96 (0.022)
IV	1	0.94 (0.082)	0.97 (0.017)	0.94 (0.080)	0.98 (0.012)
	3	0.83 (0.134)	0.96 (0.023)	0.84 (0.128)	0.97 (0.015)
	5	0.75 (0.123)	0.94 (0.035)	0.76 (0.120)	0.96 (0.022)
V	1	0.58 (0.158)	0.64 (0.154)	0.64 (0.137)	0.73 (0.113)
	3	0.41 (0.148)	0.60 (0.152)	0.45 (0.147)	0.67 (0.139)
	5	0.40 (0.150)	0.60 (0.147)	0.43 (0.150)	0.69 (0.127)

The column k gives the proportion of points drawn from multivariate Normal with mean 0 and covariance matrix equal to 10 times the identity matrix. All other points are drawn from a standard multivariate Normal distribution

4.2 Real datasets

We run experiments using three different real data sets. With the first two, we do demonstrate the effectiveness of the SIMeD method in the presence of outliers in a dataset with few predictors and in one with high number of predictors. With the last real data experiment we demonstrate the usefulness of the OVA algorithm when we have discrete responses with no logical ordering.

4.2.1 Concrete Strength dataset

We use the concrete strength dataset (see Yeh 1998) which has 8 predictors (Cement, Blast Furnace Slag, Fly Ash, Water, Superplasticizer, Coarse Aggregate, Fine Aggregate, Age) and 1 response variable (Concrete Strength). The dataset has 1030 observations and we use 10 slices. We know from the description of the dataset that age is very significant along with some ingredients with most important being the water and the superplasticizer. We use the 4 methods (SIR, SIMD, SIME, SIMeD) to get an initial estimate of the first direction. Table 6 shows the first direction and we can see that SIR and SIMD which are based on means rank water second and age third (with cement close fourth) while age is ranked first and second

Table 5 Proportion of correct estimation of the BIC criterion for models I through V when $p = 10$ and $H = 5$

		$n = 200$		$n = 400$	
Model	Distribution	SIME	SIMed	SIME	SIMed
I	a	0.82	0.94	0.93	1.00
	b	0.18	0.74	0.27	0.90
	c	0.08	0.80	0.34	0.96
II	a	0.43	1.00	0.66	1.00
	b	0.25	0.66	0.34	0.66
	c	0.13	1.00	0.38	1.00
III	a	0.71	0.91	0.87	1.00
	b	0.31	0.69	0.36	0.91
	c	0.03	0.74	0.25	0.95
IV	a	0.16	0.94	0.22	1.00
	b	0.09	0.70	0.17	0.86
	c	0.08	0.80	0.29	0.97
V	a	0.39	0.93	0.59	0.97
	b	0.03	0.53	0.01	0.80
	c	0.01	0.90	0.01	0.97

The *distribution* column indicates whether we draw our X's from a multivariate standard Normal (a), from a multivariate standard Cauchy (b) or from a mixture of multivariate Normal distributions where 95% of the points are from standard normal and 5% from $N(0, 10I)$ (c)

Table 6 Concrete strength: first direction for the 4 methods

Method	Cement	Blast	Fly Ash	Water	Superplasticizer	Coarse Agg.	Fine Agg.	Age
SIR	-0.225	-0.169	-0.132	0.507	-0.764	-0.001	0.012	-0.251
SIMD	-0.251	-0.193	-0.150	0.599	-0.671	0.000	0.002	-0.263
SIME	0.128	0.096	-0.038	-0.325	0.650	-0.019	-0.011	0.666
SIMeD	-0.115	-0.091	0.033	0.272	-0.787	0.002	-0.004	-0.532

For SIMD and SIMeD the LVR approach is used

in SIME and SIMeD, respectively, and water is ranked third (cement is fourth but further away than the other 3).

To demonstrate the ability of SIMeD to behave well under outliers, we run 100 experiments where we randomly select 30 observations and we multiply them by 10. Then we calculate the distance of the original coefficient vector we have in Table 6 to the new vector of coefficients (calculated with outliers) for each one of the methods. The boxplot in Fig. 1 of the 100 trace correlation distances defined in (4) for each method, show that SIMeD clearly outperforms the methods based on means and it also slightly outperforms the SIME method.
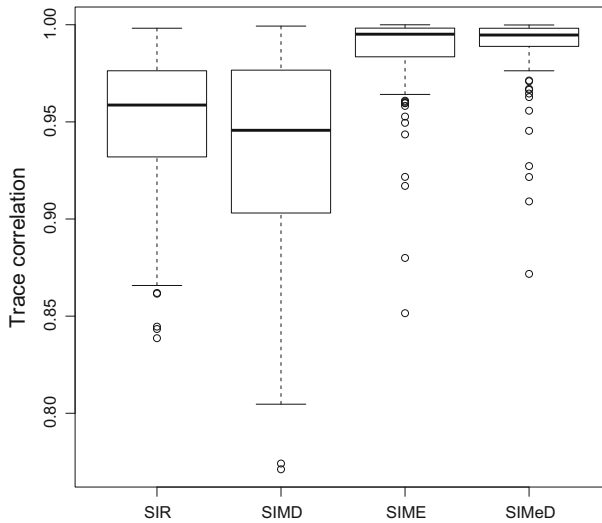
Fig. 1 Performance of 100 repetitions of the experiment on the introduction of outliers in concrete strength dataset

Table 7 Mean and median distances between the original coefficient vector (using the true data) to the new coefficient vector (where 100 points are outliers)

	SIR	SIMD	SIME	SIMeD
Mean	0.46 (0.285)	0.60 (0.246)	0.42 (0.361)	0.62 (0.241)
Median	0.49 (0.270)	0.65 (0.271)	0.29 (0.404)	0.50 (0.261)

In parenthesis we have the standard deviation for the means and the median absolute deviation of the medians

4.2.2 Superconductivity Dataset

The superconductivity data set (see Hamidieh 2018) was obtained from UC Irvine machine learning repository (see Dua and Karra-Taniskidou 2017) and we use it to demonstrate the effectiveness in the methodology in handling outliers in high dimensional cases. The dataset tries to predict the critical temperature of a superconductor using 81 selected features. There are 21263 observations. We calculate the original predictor vector. We randomly select 100 points and we multiply them by 10 to create outliers and we repeat the experiment 100 times. We calculate the distance of the original coefficient vector (using the true data) to the new coefficient vector (where 100 points are outliers). Table 7 gives the mean distances of those 100 experiments for all 4 methods, as well as the median value. As we can see the two methods that are based on differences between slices, i.e. SIMD and SIMeD, outperform the other two. SIMeD has better mean distance and
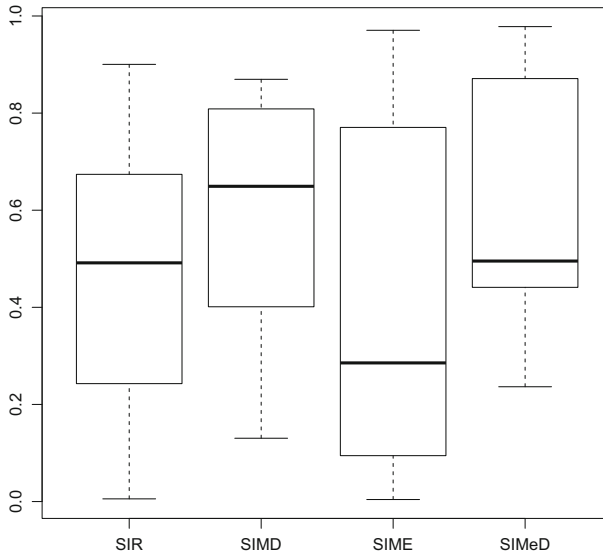
Fig. 2 Performance of 100 repetitions of the experiment on the introduction of outliers in the superconductivity dataset

Table 8 Iris data first direction for the 4 methods

Method	Sepal length	Sepal width	Petal length	Petal width
SIR	− 0.209	− 0.386	0.554	0.707
SIMD	− 0.182	− 0.145	0.405	0.884
SIME	− 0.138	− 0.069	0.077	0.985
SIMeD	− 0.140	0.031	0.111	0.983

For SIMD and SIMeD the LVR approach is used

SIMD has better median distance. In Fig. 2 we can see the range of values for the 4 methods. We first note that SIR and SIME can be influenced a lot (distances closed to 0) or not at all (distances close to 0.9). At the same time SIMD has a range between (0.1 and 0.9) and SIMeD (0.25 to 1). As expected the range of distances for all methods is much bigger in this dataset, due to the very high-dimensional nature of the dataset but our proposed method (SIMeD) performs better than SIR and SIME and is competitive with SIMD.

4.2.3 Iris dataset

We run the the algorithms on the Iris data where we have four predictors sepal length, sepal width, petal length and petal width. There are 150 observations, 50 from each of three species (setosa, versicolor, virginica). Table 8 shows the first direction when dimension reduction methods are applied. One can see that both

Table 9 Iris data first direction for the LVR and the OVA algorithm of the SIMeD method under 2 different coding schemes for the response

Method	Sepal length	Sepal width	Petal length	Petal width
LVR (scheme 1)	− 0.140	0.031	0.111	0.983
LVR (scheme 2)	− 0.230	− 0.592	0.672	− 0.381
OVA (scheme 1)	− 0.143	− 0.066	0.089	0.984
OVA (scheme 2)	− 0.143	− 0.066	0.089	0.984

SIME and SIMeD give a larger coefficient to petal width which is well known to be the one mostly separating the three species.

To emphasize the usefulness of the OVA algorithm when the response is categorical we show the first direction when the three species are first coded as 1 for setosa, 2 for versicolor and 3 for virginica and then when we keep setosa as 1 but we code virginica as 2 and versicolor as 3. We see in Table 9 that the OVA approach does not change estimate while the LVR changes a lot and gives a complete different direction.

5 Discussion

In this work we propose a new algorithm for robust dimension reduction using inverse median differences between slices and which we call SIMeD. We demonstrate the advantages of this algorithm over using a similar algorithm based on inverse means (SIMD) which was proposed by Artemiou and Tian (2015) at the presence of outliers and we also demonstrate that in any case it performs better than just using sliced inverse median regression (SIME) proposed by Dong et al. (2015). Moreover, we propose the OVA algorithm of SIMeD and demonstrate its advantage in the case of categorical responses. Although in this paper we choose the L1 median due to its uniqueness for $p \geq 2$, we expect similar results to hold when other multivariate medians are used, for example Tukey and Oja median which were discussed in the SDR framework by Christou (2018).

References

Artemiou A, Tian L (2015) Using slice inverse mean difference for sufficient dimension reduction. Stat Probab Lett 106:184–190

Brown BM (1983) Statistical uses of the spatial mean. J R Stat Soc B 45:25–30

Bura E, Yang J (2011) Dimension estimation in sufficient dimension reduction: a unifying approach. J Multivar Anal 102:130–142

Christou E (2018) Robust dimension reduction using sliced inverse median regression. Stat Pap. https://doi.org/10.1007/s00362-018-1007-z

Cook RD (1998) Regression graphics: ideas for studying regressions through graphics. Wiley, New York

Cook RD, Weisberg S (1991) Discussion of "sliced inverse regression for dimension reduction". J Am Stat Assoc 86:316–342

Dong Y, Yu Z, Zhu L (2015) Robust inverse regression for dimension reduction. J Multivar Anal 134:71–81

Dua D, Karra-Taniskidou E (2017) UCI machine learning repository. http://archive.ics.uci.edu/ml. University of California, School of Information and Computer Science, Irvine, CA

Gather U, Hilker T, Becker C (2001) A robustified version of sliced inverse regression. In: Fernholz LT, Morgenthaler S, Stahel W (eds) Statistics in genetics and in the environmental sciences. Birkhäuser: Basel, pp 147–157

Hamidieh K (2018) A data-driven statistical model for predicting the critical temperature of a superconductor. Comput Mater Sci 154:346–354

Hettmansperger TP, McKean JW (1998) Robust nonparametric statistical methods. Arnold/Wiley, London, New York

Li B (2018) Sufficient dimension reduction. Methods and applications with R. Chapman and Hall/CRC, New York

Li B, Artemiou A, Li L (2011) Principal support vector machine for linear and nonlinear sufficient dimension reduction. Ann Stat 39:3182–3210

Li B, Wang S (2007) On directional regression for dimension reduction. J Am Stat Assoc 102:997–1008

Li K-C (1991) Sliced inverse regression for dimension reduction (with discussion). J Am Stat Assoc 86:316–342

Luo W, Li B (2016) Combining eigenvalues and variation of eigenvectors for order determination. Biometrika 103:875–887

Maechler M, Rousseeuw P, Croux C, Todorov V, Ruckstuhl A, Salibian-Barrera M, Verbeke T, Koller M, Conceicao ELT, di Palma MA (2018) Robustbase: Basic Robust Statistics R package version 0.93–3. http://CRAN.R-project.org/package=robustbase

Shao Y, Cook RD, Weisberg S (2007) Marginal tests with sliced average variance estimation. Biometrika 94:285–296

Yeh I-C (1998) Modeling of strength of high performance concrete using artificial neural networks. Cem Concr Res 28(12):1797–1808

Yin X, Li B, Cook RD (2008) Successive direction extraction for estimating the central subspace in a multiple-index regression. J Multivar Anal 99:1733–1757

Zhu LP, Zhu LX, Feng ZH (2010) Dimension reduction in regression through cumulative slicing estimation. J Am Stat Assoc 105:1455–1466

Zhu LX, Miao B, Peng H (2006) On sliced inverse regression with large dimensional covariates. J Am Stat Assoc 101:630–643