Communications for Statistical Applications and Methods
2021, Vol. 28, No. 2, 205–215

https://doi.org/10.29220/CSAM.2021.28.2.205
Print ISSN 2287-7843 / Online ISSN 2383-4757

Iterative projection of sliced inverse regression with fused approach

Hyoseon Han[a], Youyoung Cho[a], Jae Keun Yoo[1,a]

[a]Department of Statistics, Ewha Womans University, Korea

Abstract

Sufficient dimension reduction is useful dimension reduction tool in regression, and sliced inverse regression (Li, 1991) is one of the most popular sufficient dimension reduction methodologies. In spite of its popularity, it is known to be sensitive to the number of slices. To overcome this shortcoming, the so-called fused sliced inverse regression is proposed by Cook and Zhang (2014). Unfortunately, the two existing methods do not have the direction application to large p-small n regression, in which the dimension reduction is desperately needed. In this paper, we newly propose seeded sliced inverse regression and seeded fused sliced inverse regression to overcome this deficit by adopting iterative projection approach (Cook *et al.*, 2007). Numerical studies are presented to study their asymptotic estimation behaviors, and real data analysis confirms their practical usefulness in high-dimensional data analysis.

Keywords: central subspace, fused reduction, inverse regression, iterative projection, large p-small n regression, sufficient dimension reduction

1. Introduction

In modern society, data is collected and stored through various channels such as mobile, table, and IoT, and big data analysis using this data is used in many ways. However, as the amount of data increases, the curse of dimensionality is one of the inevitable problems. As the dimensions increase, the space of data to be explained increases exponentially. This means that more data are needed to explain, but, consequently, the explanatory power is reduced. Therefore, in order to overcome this problem, we intend to reduce the dimension without loss of information.

Sufficient dimension reduction (SDR) provides effective tools for the problem in regression of $Y \in \mathbb{R}^1|\mathbf{X} \in \mathbb{R}^p$. The SDR replaces the p-dimensional predictors $\mathbf{X}$ with its lower-dimensional projection $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ without loss of information on $Y \in \mathbb{R}^1|\mathbf{X} \in \mathbb{R}^p$, where $p \geq 2$ and $\boldsymbol{\eta} \in \mathbb{R}^{p\times d}$ with $d \leq p$. This is equivalently expressed as follows:

$$Y \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \tag{1.1}$$

where $\perp\!\!\!\perp$ stands for statistical independence. Naturally, SDR pursues the estimation of $\boldsymbol{\eta}$ to satisfy (1.1) and to have the minimal dimension. The subspace spanned by the columns of such $\boldsymbol{\eta}$ is called the *central subspace* $\mathcal{S}_{Y|\mathbf{X}}$. Hereafter, $\boldsymbol{\eta}$ and d will represent an orthonormal basis matrix and the structural dimension of $\mathcal{S}_{Y|\mathbf{X}}$. And, the d-dimensional linear projection $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ is called sufficient predictors.

[1] Corresponding author: Department of Statistics, Ewha Womans University, 52, Ewhayeodae-gil, Seodaemun-gu, Seoul 03760, Republic of Korea. E-mail: peter.yoo@ewha.ac.kr

Published 31 March 2021 / journal homepage: http://csam.or.kr

One of the popular SDR methods to estimate $\mathcal{S}_{Y|\mathbf{X}}$ is sliced inverse regression (SIR) (Li, 1991). The method requires so-called *linearity condition* such that $E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})$ is linear in $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$. To implement SIR, a categorization of the response Y, called *slicing* is crucial. Intuitively, the sample performance of SIR essentially depends on the choice of the slices, that is, the number of categories of Y. That is, different choices of the numbers of slices cause non-consistent structural dimension and basis estimation of $\mathcal{S}_{Y|\mathbf{X}}$. However, there is no recommendation for adequate choices of the number of slices. This is always critical to SIR despite its popularity.

Recently, Cook and Zhang (2014) proposed fused sliced inverse regression (FSIR) to overcome this deficit. The FSIR constructs a new kernel matrix by fusing kernel matrices constructed from various choices of numbers of slices. This new kernel matrix balances the different information acquired from non-nested slices and the similar information acquired from nested slices. We will discuss this nestness of slices in later section. This balancing makes FSIR more robust to the number of slices than SIR, and increase the estimation accuracy of $\mathcal{S}_{Y|\mathbf{X}}$.

Unfortunately, FSIR requires the inverse of the sample covariance matrix of predictors, and it is not invertible with $p \geq n$. This indicates that FSIR does not have a direct application to large p-small n regression, which is quite common these days. Therefore, the extension of FISR to large p-small n regression must be done in this big data era. For this, we will take a route of an iterative projection approach suggested by Cook *et al.* (2007). Upto date, there is no iterative projection approach for SIR, so the paper newly provides so-called seeded SIR. Based on this seeded SIR, we finally propose seeded FSIR. This is the primary purpose of the paper.

The organization of the paper is as follows. In Section 2, fused sliced inverse regression and nestness are introduced. Section 3 is devoted to developing seeded sliced inverse regression and seeded fused sliced inverse regression. In Section 4, numerical studies and real data example are presented. We summarize our work in Section 5.

2. Sliced and fused sliced inverse regressions and nestness

2.1. Sliced and fused sliced inverse regressions

As discussed in Introduction, the fused sliced inverse regression (FSIR) (Cook and Zhang, 2014) is based on sliced inverse regression (SIR) (Li, 1991). Letting $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X})$ and $\mathbf{Z} = \boldsymbol{\Sigma}^{-1/2}(\mathbf{X} - E(\mathbf{X}))$ with $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Sigma}^{-1/2} = \boldsymbol{\Sigma}^{-1}$, we have $\mathcal{S}_{Y|\mathbf{X}} = \boldsymbol{\Sigma}^{-1/2}\mathcal{S}_{Y|\mathbf{Z}}$ according to Yoo (2016a).

In Li (1991), it is shown that $E(\mathbf{Z}|Y) \in \mathcal{S}_{\mathbf{Z}|Y}$ under certain condition, which is not an issue here. In other words, a subspace spanned by $E(\mathbf{Z}|Y = y)$ varying the values of y is a subspace of $\mathcal{S}_{\mathbf{Z}|Y}$. That is, once $E(\mathbf{Z}|Y)$ is restored, we can, at least partially, capture $\mathcal{S}_{\mathbf{Z}|Y}$. In SDR context, it is normally assumed that $\mathcal{S}(E(\mathbf{Z}|Y)) = \mathcal{S}_{\mathbf{Z}|Y}$. For more about the two conditions mentioned here, readers are recommended to read Yoo (2016a, 2016b). Hereafter without mentioning clearly, we assume that $\mathcal{S}(E(\mathbf{Z}|Y)) = \mathcal{S}_{Y|\mathbf{Z}}$.

Since there is no parametric assumption so far, it is crucial to estimate $E(\mathbf{Z}|\mathbf{Y})$ non-parametically. Li (1991) suggests a simple procedure, which is a categorization of Y. This categorization of $\mathbf{Y}$ and the category are called slicing and slice, respectively. Once the original $\mathbf{Y}$ is categorized into h groups, $E(\mathbf{Z}|\mathbf{Y} = y)$ can be estimated by the sample mean within each group.

Define $\mathbf{K} = ((n_1/n)\bar{\mathbf{Z}}_1, (n_2/n)\bar{\mathbf{Z}}_2, \ldots, (n_h/n)\bar{\mathbf{Z}}_h)$, where $\bar{\mathbf{Z}}_k = (1/n_k)\sum_{i\in H_k}\mathbf{Z}_i$, H_k is the k^{th} slice, and n_k stands for the sample size of the k^{th} slice. Then, spectral-decompose $\mathbf{M}_{\mathrm{SIR}} = \mathbf{K}\mathbf{K}^{\mathrm{T}}$ such that

$$\mathbf{M}_{\mathrm{SIR}} = \sum_{i=1}^{p} \lambda_i \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^{\mathrm{T}},$$

where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$.

Then, a subspace spanned by the first d-largest eigenvectors pre-multiplied by $\boldsymbol{\Sigma}^{-1/2}$, that is, $\boldsymbol{\Sigma}^{-1/2}(\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_d)$, is the estimate of $\mathcal{S}_{Y|\mathbf{X}}$. The estimation approach of $\mathcal{S}_{Y|\mathbf{X}}$ through $\mathbf{M}_{\text{SIR}}$ is called *sliced inverse regression* (SIR).

As the name of the method indicating, the crucial step of practical implementation of SIR is slicing. That is, its estimation performance clearly depends on the number of slices, but there is no thumb rule for its proper selection.

To overcome this deficiency of SIR, Cook and Zhang (2014) propose a fused approach. Let $\mathbf{K}_g$ be $\mathbf{K}$ constructed with g slices. Then, we construct the following non-decreasing sequence:

$$\mathbf{K}_{(g)} = (\mathbf{K}_2, \mathbf{K}_3, \ldots, \mathbf{K}_g) \quad \text{for } g = 3, 4, \ldots, h.$$

We rule $g = 2$ out in $\mathbf{K}_{(g)}$ because $\mathbf{K}^{(2)} = \mathbf{K}_2$.

Since the columns of each $\mathbf{K}_g$ spans $\mathcal{S}_{Y|\mathbf{Z}}$, those of $\mathbf{K}_{(g)}$ also span $\mathcal{S}_{Y|\mathbf{Z}}$. This directly implies that $\mathbf{K}_{(g)}$ are fully informative to $\mathcal{S}_{Y|\mathbf{Z}}$, so $\mathbf{M}_{\text{FSIR}} = \mathbf{K}_{(g)}\mathbf{K}_{(g)}^{\text{T}}$ becomes another kernel matrix to estimate $\mathcal{S}_{Y|\mathbf{Z}}$. The estimation of $\mathcal{S}_{Y|\mathbf{X}}$ thru $\mathbf{M}_{\text{FSIR}}$ is called *fused sliced inverse regression* (FSIR). Cook and Zhing (2014) shows that FSIR is more robust to the number of slices and has more accurate estimation of $\mathcal{S}_{Y|\mathbf{X}}$ than SIR.

In the implementation of SIR and FSIR, the unknown population quantities are replaced with their usual moment estimators.

2.2. Nestness

This subsection is devoted to explain intuitively why FSIR is, at least, robust to the numbers of slices. In SIR, the slicing is usually done for each slice to have equal number of observations, which induce that the slicing results are nested for multiples of prime numbers. For example, suppose that one categorizes Y into two, four and six groups. The four and six groups completely and exhaustively fall into the two groups. If another one categorizes Y into three, six and nine groups, then the six and nine groups again completely and exhaustively fall into the three group. This property of slicing is called *nestness*. Because of the nestness of the slicing, fusing kernel matrices of SIR resulted from various numbers of slices combine heterogeneous information acquired by the different prime numbers of slices and homogeneous information provided by the same prime numbers of slices. The heterogeneous information relieves sensitiveness to the numbers of slices and the homogeneous information increases the estimation accuracy.

3. Seeded sliced and seeded fused sliced inverse regressions

In both SIR and FSIR, practically, the inversion of the sample covariance matrix of $\mathbf{X}$ is commonly required. However, if the sample size n is less than the number of predictors p, the inversion is not possible. Therefore, in the so-called large p-small n regression, in which dimension reduction of $\mathbf{X}$ is on urgent demand, the application of SIR and FSIR is not possible.

To overcome this deficit, we adopt seeded dimension reduction suggested by Cook *et al.* (2007). Basic idea of the seeded dimension reduction is iterative projection of seed matrix, which is informative to $\boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$, onto $\boldsymbol{\Sigma}$.

Then, the following relation should be noted:

$$\begin{aligned} E(\mathbf{Z}|Y) \in \mathcal{S}_{Y|\mathbf{Z}} &\Leftrightarrow \boldsymbol{\Sigma}^{-\frac{1}{2}} E(\mathbf{Z}|Y) \in \boldsymbol{\Sigma}^{-\frac{1}{2}} \mathcal{S}_{Y|\mathbf{Z}} \\ &\Leftrightarrow \boldsymbol{\Sigma}^{-\frac{1}{2}} E(\mathbf{Z}|Y) \in \mathcal{S}_{Y|\mathbf{X}} \\ &\Leftrightarrow \boldsymbol{\Sigma}^{-1} E(\mathbf{X}|Y) \in \mathcal{S}_{Y|\mathbf{X}} \\ &\Leftrightarrow E(\mathbf{X}|Y) \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}. \end{aligned}$$

By the last equivalence of $E(\mathbf{X}|Y) \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$, $E(\mathbf{X}|Y)$ can be a seed matrix. This directly implies that the following two matrices of $\mathbf{K}_g^x$ and $\mathbf{K}_{(g)}^x$ also can be seed matrices:

$$\mathbf{K}_g^x \in \mathbb{R}^{p\times g} = \left(\bar{\mathbf{X}}_1, \ldots, \bar{\mathbf{X}}_g\right) \quad \text{and} \quad \mathbf{K}_{(g)}^x \in \mathbb{R}^{p\times(0.5g(g+1)-1)} = \left(\mathbf{K}_2^x, \mathbf{K}_3^x, \ldots, \mathbf{K}_g^x\right),$$

where $\bar{\mathbf{X}}_k = (1/n_k)\sum_{i\in H_k} \mathbf{X}_i$.

Next, by projecting $\mathbf{K}_g^x$ and $\mathbf{K}_{(g)}^x$ iteratively onto $\boldsymbol{\Sigma}$, the following two quantities are constructed:

$$\mathbf{R}_u \in \mathbb{R}^{p\times ug} = \left(\mathbf{K}_g^x,\ \boldsymbol{\Sigma}\mathbf{K}_g^x,\ \boldsymbol{\Sigma}^2\mathbf{K}_g^x, \ldots,\ \boldsymbol{\Sigma}^{u-1}\mathbf{K}_g^x\right)$$

$$\mathbf{R}_u^{\mathrm{F}} \in \mathbb{R}^{p\times u(0.5g(g+1)-1)} = \left(\mathbf{K}_{(g)}^x,\ \boldsymbol{\Sigma}\mathbf{K}_{(g)}^x,\ \boldsymbol{\Sigma}^2\mathbf{K}_{(g)}^x, \ldots, \boldsymbol{\Sigma}^{u-1}\mathbf{K}_{(g)}^x\right).$$

Then, finally, the following two matrices are computed from $\mathbf{R}_u$ and $\mathbf{R}_u^{\mathrm{F}}$:

$$\mathbf{B}_u \in \mathbb{R}^{p\times g} = \mathbf{R}_u\left(\mathbf{R}_u^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R}_u\right)^{-1}\mathbf{R}_u^{\mathrm{T}}\mathbf{K}_g^x$$

$$\mathbf{B}_u^{\mathrm{F}} \in \mathbb{R}^{p\times(0.5g(g+1)-1)} = \mathbf{R}_u^{\mathrm{F}}\left(\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}}\right)^{-1}\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\mathbf{K}_{(g)}^x.$$

According to Cook *et al.* (2007), the columns of both $\mathbf{B}_u$ and $\mathbf{B}_u^{\mathrm{F}}$ span $\mathcal{S}_{Y|\mathbf{X}}$.

Now, we need to discuss two issues in the estimation of $\mathbf{B}_u$ and $\mathbf{B}_u^{\mathrm{F}}$. The first one is the selection of u. The termination index u should be carefully chosen, so that $\mathbf{R}_u$ and $\mathbf{R}_u^{\mathrm{F}}$ are sufficiently informative to $\mathcal{S}_{Y|\mathbf{X}}$ and both $(\mathbf{R}_u^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R}_u)$ and $(\mathbf{R}_u^{\mathrm{FT}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}})$ are invertible. Defining that $\boldsymbol{\Delta}_u = \mathbf{B}_{u+1} - \mathbf{B}_u$, Cook *at al.* (2007) suggest to utilize $n\,\mathrm{trace}(\boldsymbol{\Delta}_u^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Delta}_u)$. Accordingly, $n\,\mathrm{trace}(\boldsymbol{\Delta}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\boldsymbol{\Delta}_u^{\mathrm{F}})$ is defined for $\mathbf{R}_u^{\mathrm{F}}$.

One simple way is to choose the smallest u to make $n\,\mathrm{trace}(\boldsymbol{\Delta}_u^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Delta}_u) < e$, where the cut-off value of e is user-selected. In numerical studies and real data analysis, we set $e = 0.01$.

The second issue is for inverting $\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}}$. Actually, the dimension of $\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}}$ is $u(0.5g(g+1)-1)\times u(0.5g(g+1)-1)$. For example, setting $n = 100$, $g = 6$ and $u = 5$, it becomes a 100 by 100 matrix. Since $n = 100$, the inversion of $\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}}$ is not plausible. Therefore, we need to fix this issue in a practical way. The main reason that $\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\Sigma}\mathbf{R}_u^{\mathrm{F}}$ has high dimension is placed onto fusing many $\mathbf{K}_g^x$s. In practice, the seed matrix $\mathbf{K}_{(g)}^x$ would have much noise in high-dimensional predictors. Therefore, it should be recommended to eliminate unnecessary noise before constructing $\mathbf{K}_{(g)}^x$ from $\mathbf{K}_g$s.

Here, instead of using $\mathbf{K}_{(g)}$ as seed matrix, we propose to use the collection of the first eigenvector of $\mathbf{K}_g^x$ as a seed matrix for FSIR. Then, its dimension is reduced to $p\times(g-1)$ from $p\times(0.5g(g+1)-1)$. This replacement does not always win against the original $\mathbf{K}_{(g)}^x$, but it should be a realistic, efficient and effective alternative to the inversion issue faced in high-dimensional predictors.

Also, to compute the required matrix inversion in seeded dimension reduction, we also considers further reduction of the seed matrices. The $\mathbf{K}_g^x$ for SIR and the collection of the first eigenvector of $\mathbf{K}_g^x$ for FISR are replaced with their q largest eigenvectors based on the cumulative proportions of

the sum of their eigenvalues. Although one can use various proportions, we can use 99% cumulative proportion.

The following algorithm summarizes our proposed seeded dimension reduction for SIR and FISR.

1. Construct $\mathbf{K}_g^x$ and the collection $\mathbf{\Gamma}_\bullet^x$ of the first eigenvectors of $\mathbf{K}_g^x$ for $g = 2, \ldots, h$.

2. Let $\boldsymbol{\nu}_{\mathrm{SIR}}$ be a set of the eigenvectors of $\mathbf{K}_g^x$ corresponding to its largest eigenvalues, whose sum accounts for, at least, 99% of the sum of all the eigenvalues. Similarly, we define $\boldsymbol{\nu}_{\mathrm{FSIR}}$ as a set of the eigenvectors of $\mathbf{\Gamma}_\bullet^x$ corresponding to its largest eigenvalues, whose sum accounts for, at least, 99% of the sum of all the eigenvalues.

3. With $\boldsymbol{\nu}_{\mathrm{SIR}}$ and $\boldsymbol{\nu}_{\mathrm{FSIR}}$ as seed matrices, compute $\mathbf{R}_u = (\boldsymbol{\nu}_{\mathrm{SIR}}, \mathbf{\Sigma}\boldsymbol{\nu}_{\mathrm{SIR}}, \mathbf{\Sigma}^2\boldsymbol{\nu}_{\mathrm{SIR}}, \ldots, \mathbf{\Sigma}^{u-1}\boldsymbol{\nu}_{\mathrm{SIR}})$ and $\mathbf{R}_u^{\mathrm{F}} = (\boldsymbol{\nu}_{\mathrm{FSIR}}, \mathbf{\Sigma}\boldsymbol{\nu}_{\mathrm{FSIR}}, \mathbf{\Sigma}^2\boldsymbol{\nu}_{\mathrm{FSIR}}, \ldots, \mathbf{\Sigma}^{u-1}\boldsymbol{\nu}_{\mathrm{FSIR}})$.

4. Construct $\mathbf{B}_u = \mathbf{R}_u(\mathbf{R}_u^{\mathrm{T}}\mathbf{\Sigma}\mathbf{R}_u)^{-1}\mathbf{R}_u^{\mathrm{T}}\boldsymbol{\nu}_{\mathrm{SIR}}$ and $\mathbf{B}_u^{\mathrm{F}} = \mathbf{R}_u^{\mathrm{F}}(\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\mathbf{\Sigma}\mathbf{R}_u^{\mathrm{F}})^{-1}\mathbf{R}_u^{\mathrm{F}^{\mathrm{T}}}\boldsymbol{\nu}_{\mathrm{FSIR}}$. In the computation of $\mathbf{B}_u$ and $\mathbf{B}_u^{\mathrm{F}}$, the iterative projections are terminated when both $n\,\mathrm{trace}(\mathbf{\Delta}_u^{\mathrm{T}}\mathbf{\Sigma}\mathbf{\Delta}_u)$ and $n\,\mathrm{trace}(\mathbf{\Delta}_u^{\mathrm{F}^{\mathrm{T}}}\mathbf{\Sigma}\mathbf{\Delta}_u^{\mathrm{F}})$ are less than 0.01.

5. The columns of $\mathbf{B}_u$ and $\mathbf{B}_u^{\mathrm{F}}$ spans $\mathcal{S}_{Y|\mathbf{X}}$.

We call the estimation of $\mathcal{S}_{Y|\mathbf{X}}$ through $\mathbf{B}_u$ and $\mathbf{B}_u^{\mathrm{F}}$ constructed through the above algorithm seeded sliced (sSIR) and *s*eeded (sFSIR), respectively.

In samples, all population quantities are replaced with their usual moment estimators.

4. Numerical studies and data analysis

4.1. Numerical studies

For all simulation studies, the sample sizes were 100 with 500 iterations per model. The structural dimension of $\mathcal{S}_{Y|\mathbf{X}}$ was one for the models under consideration. To measure how well $\mathcal{S}_{Y|\mathbf{X}}$ is estimated, the absolute value $|r|$ of the Pearson correlation coefficient between $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ and $\hat{\boldsymbol{\eta}}^{\mathrm{T}}\mathbf{X}$ was computed. As a summary, boxplots of $|r|$ obtained from seeded sliced inverse regression and seeded fused sliced inverse regression were reported along with lining its means and median against various numbers of slices.

To generate variables, we define $\boldsymbol{\beta}_i$, $i = 1, 2, 3$ as follows: $\boldsymbol{\beta}_1 = (p^{-1/2}, p^{-1/2}, \ldots, p^{-1/2})^{\mathrm{T}}$; $\boldsymbol{\beta}_2$ is defined as its first and last of 20% of elements equal to $(0.4p)^{-1/2}$ and all other elements zeros; $\boldsymbol{\beta}_3 = (2^{-1/2}, 0, \ldots, 0, 2^{1/2})$. Also, two $p \times p$ covariance matrix $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ are defined: $\mathbf{\Sigma}_1 = (2/3)\,\mathrm{diag}(2, \ldots, 2, 1, \ldots, 1)$ with equal multiplicity between 1 and 2; $\mathbf{\Sigma}_2 = (2/p+1)\,\mathrm{diag}(p, p-1, p-2, \ldots, 1)$.

Then, we considered two simulated models in Cook *et al.* (2007):

- Model 1: $Y = \boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{X} + 0.3\varepsilon$;
- Model 2: $Y = \boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{X} + (\boldsymbol{\beta}_i^{\mathrm{T}}\mathbf{X})^3/10 + 0.3\varepsilon$,

  where $\mathbf{X} \sim MN(0, \mathbf{\Sigma}_j)$, $j = 1, 2$ and $\varepsilon \sim N(0, 1) \perp\!\!\!\perp \mathbf{X}$.

Following Cook *et al.* (2007), the combination of $(\boldsymbol{\beta}_i, \mathbf{\Sigma}_j)$ was set to $(\boldsymbol{\beta}_1, \mathbf{\Sigma}_1)$, $(\boldsymbol{\beta}_2, \mathbf{\Sigma}_1)$ and $(\boldsymbol{\beta}_3, \mathbf{\Sigma}_2)$ for $p = 10$ and $p = 500$. So, each model has totally 6 scenarios.

The stopping rule for the iterative projection and choices of the seeded matrix for sSIR and sFSIR are directly followed in the algorithm in Section 3.

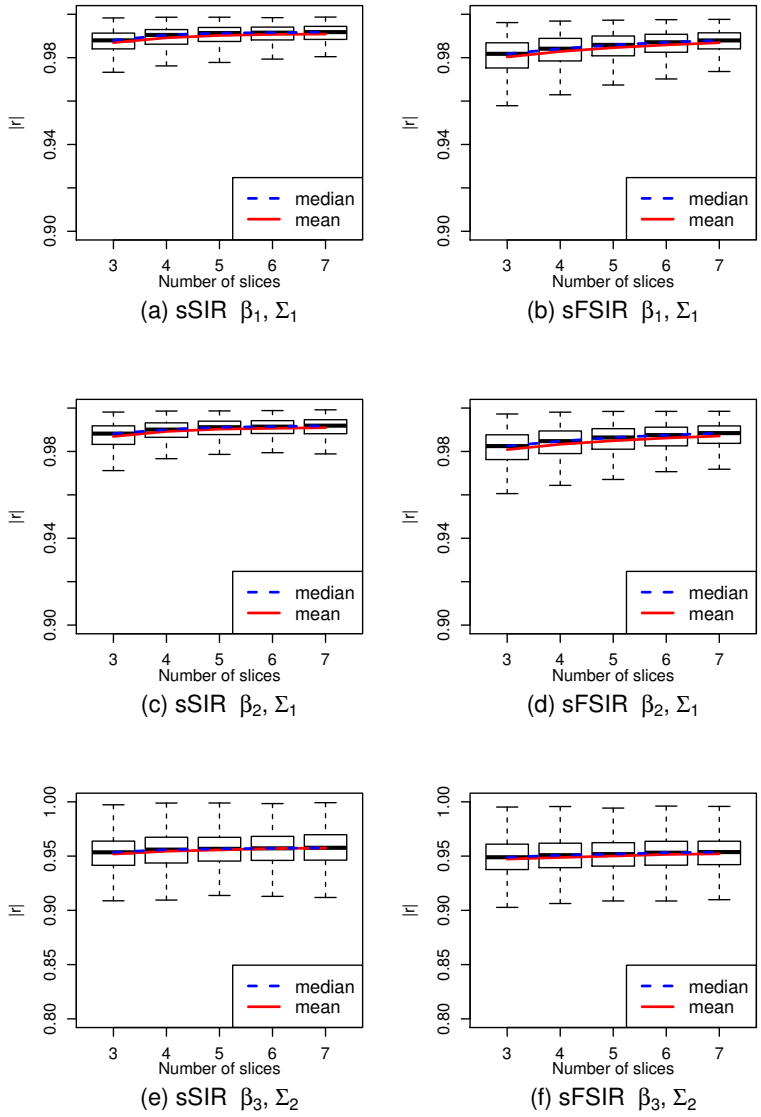

Figure 1: *Side-by-side boxplots of |r| for Model 1 with* $p = 10$.

With $p = 10$ there is no notable difference between sSIR and sFSIR for any scenario of both models, which is summarized in Figures 1 and 3. Figures 2 and 4 show that both sSIR and sFSIR provides good estimation of $\mathcal{S}_{Y|\mathbf{X}}$ for high dimension predictors with $p = 500$. According to Figures 2 and 4, it is observed that sFSIR is consistent to the numbers of slices and often yields better estimation results than sSIR, while sSIR is more sensitive to the number of slices than sFSIR and the estimation accuracy of sSIR gets worse with larger numbers of slices. The numerical studies confirm both sSIR and sFSIR are well-estimate $\mathcal{S}_{Y|\mathbf{X}}$ and that sFSIR is more robust to the numbers of slices than sSIR.

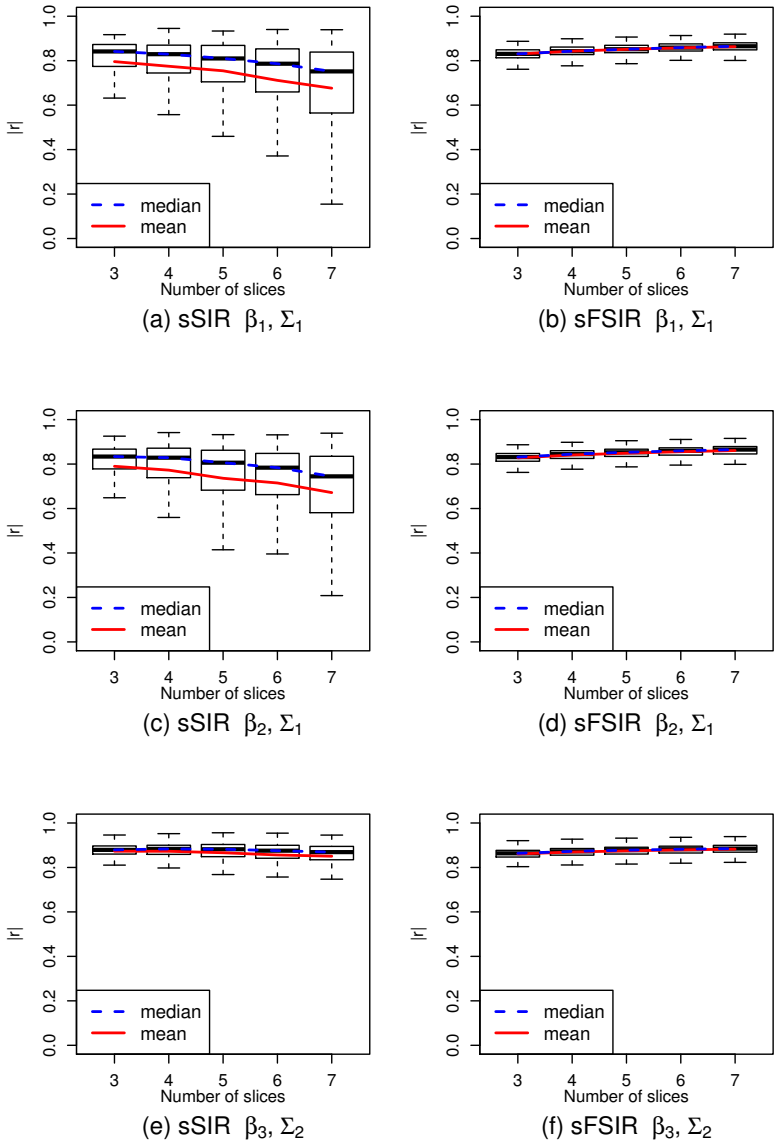

Figure 2: *Side-by-side boxplots of |r| for Model 1 with p* = 500.

4.2. Data analysis: Near-infrared spectroscopy of biscuit doughs data

For real data application, near-infrared spectroscopy of biscuit doughs data (NIR) (Brown *et al.*, 2001), named as COOKIE in PPLS package in R is illustrated. The measurements of the composition of biscuit dough pieces from near-infrared spectroscopy are contained in the data. It is known that near-infrared spectroscopy is one of most preferred tools to analyze constitutions of various materials such as food

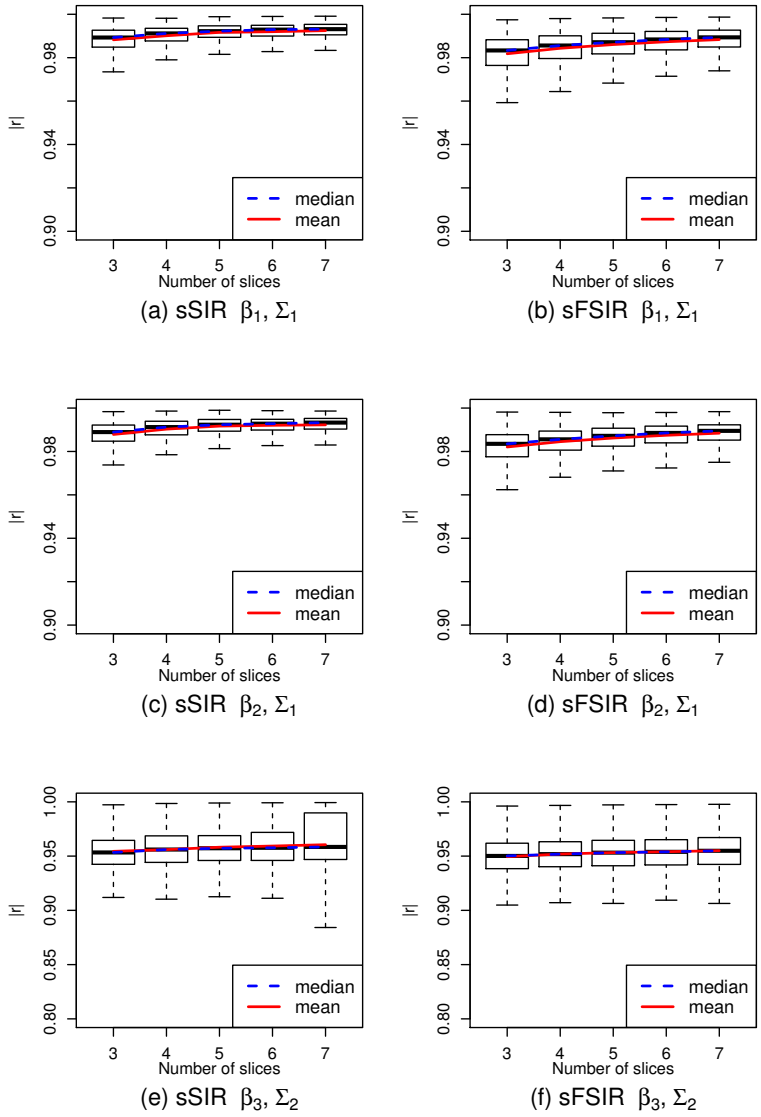

Figure 3: *Side-by-side boxplots of |r| for Model 2 with* $p = 10$.

and drink, pharmaceutical products, and petrochemicals. The data consist of measurements of the composition of biscuit dough pieces and water. A regression analysis of the percentages of water (Y) in dough on the composition of biscuit dough pieces ($\mathbf{X}$) were considered. Predictors of the composition of biscuit dough pieces were measured from 1100 to 2498 nanometers (nm) in steps of 2 nm, so there are total 700 points. Brown *et al.* (2001), however, suggests removing the first 140 and the last 49 wavelengths because of lack of useful information and to increase the steps from 2 to 4 nm. As results, a wavelength ranging from 1380 to 2400 nm was adopted, and hence there are 256 points,

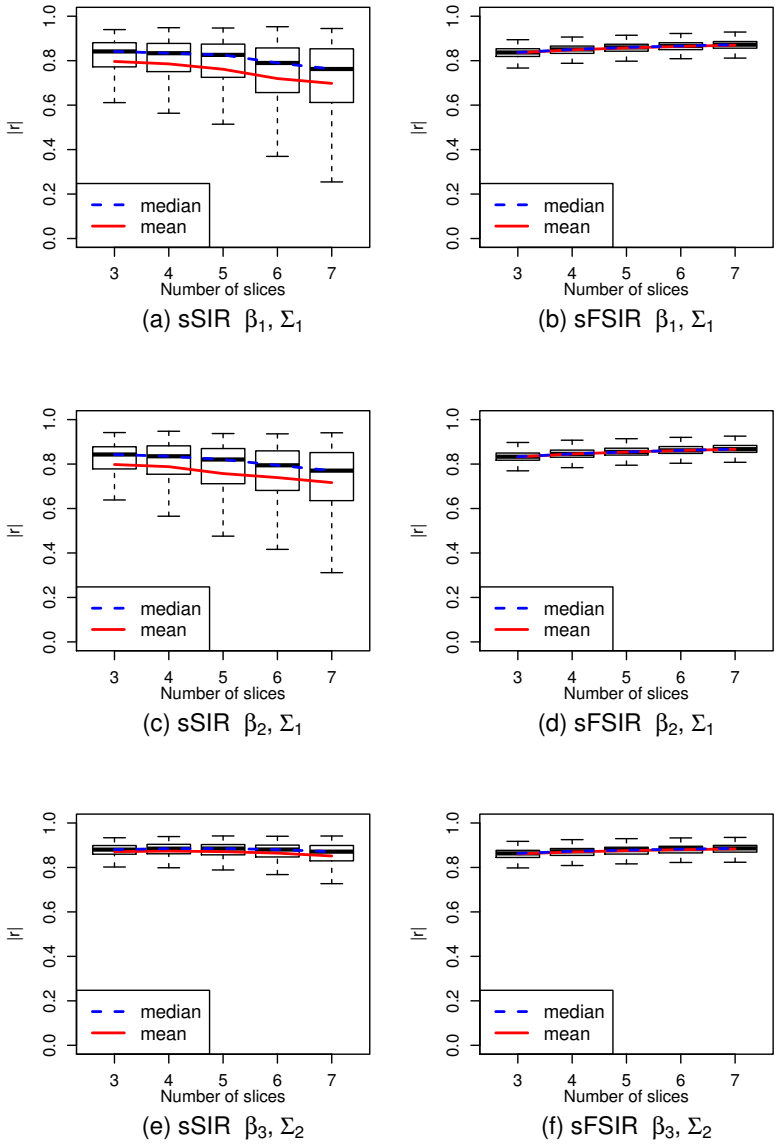

Figure 4: *Side-by-side boxplots of* $|r|$ *for Model 2 with* $p = 500$.

which is the dimension of predictors. Total 72 observations in the data were divided into two groups of 40 training and 32 test sets, and the 23^{rd} and 21^{st} observations in the training test sets, respectively, were eliminated as outliers before analysis. Let $\mathbf{X}_{\text{train}} \in \mathbb{R}^{256\times 40}$ and $\mathbf{X}_{\text{test}} \in \mathbb{R}^{256\times 32}$ stand for the train and test sets, respectively. Accordingly, we define Y_{train} and Y_{test}. The analysis scheme is as follows. First, the dimension of $\mathbf{X}$ is reduced by sSIR and sFSIR with 3, 6, and 9 slices with the training set. The dimension reduced predictors will be denoted as $\hat{\mathbf{B}}^{\mathrm{T}}_{\text{train}}\mathbf{X}$. Following Brown *et al.* (2001), with the

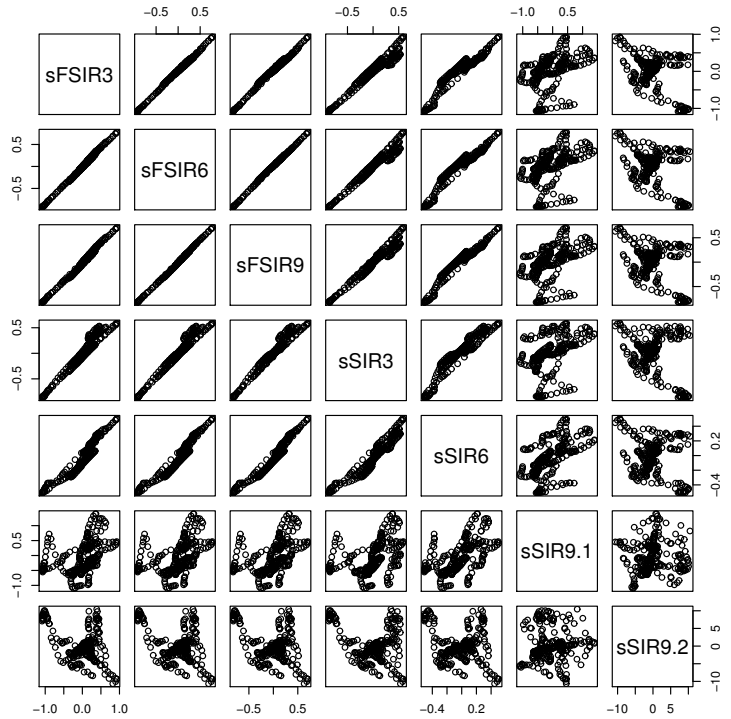

Figure 5: *Scatterplot matrix of dimension reduced predictors from sSIR and sFSIR wih 3, 6, and 9 slices:* sSIR #, *sSIR with # slices;* sFSIR#, *sFSIR with # slices.*

reduced predictors $\hat{\mathbf{B}}_{\text{train}}^{\text{T}}\mathbf{X}_{\text{train}}$, the following linear regression is fitted:

$$Y_{\text{train}} = \alpha + \boldsymbol{\beta}^{\text{T}}\left(\hat{\mathbf{B}}_{\text{train}}^{\text{T}}\mathbf{X}_{\text{train}}\right) + \varepsilon.$$

Then the model is evaluated by the test set $\mathbf{X}_{\text{test}}$ through $\hat{\mathbf{B}}_{\text{train}}^{\text{T}}\mathbf{X}_{\text{test}}$. Mean squared errors of predictions on the test set are computed for the comparison, which is reported in Table 1.

First, we compare the fitted dimension reduction predictors with the application of sSIR and sFSIR with 3, 6 and 9 slices, which is summarized in Figure 5. According to Figure 5, sFSIR provides the essentially same reduction results regardless of the number of slices, and they are also quite close to ones from sSIR with 3 and 6 slices. All the five cases reduce 256 dimension to 1 dimension. However, sSIR with 9 slices report two-dimensional reduction, and its first direction is even different from all the other five. This relationship does not guarantee the results from sFSIR and sSIR with 3 and 6 slices is better than sSIR with 9 slices. To have the insight for this, we compare the computed mean squared error of prediction in Table 1.

Table 1 shows that the mean squared errors of the prediction are somewhat different, although the three results from sFSIR and those of sSIR with 3 and 6 slices are close to each other. The best one should be the application of sFSIR with 9 slices, while the worst one is sSIR with 3 slices. And, the mean squared errors of prediction from sFSIR are less variable than those from sSIR. This confirms that sFSIR is more robust to the number of slices than sSIR in practice.

Table 1: Mean squared errors of prediction from sSIR and sFSIR with 3, 6, and 9 slices: sSIR #, sSIR with # slices; sFSIR#, sFSIR with # slices

sSIR3	sSIR6	sSIR9	sFSIR3	sFSIR6	sFSIR9
0.378	0.245	0.340	0.261	0.213	0.198

5. Discussion

Sufficient dimension reduction (SDR) is useful dimension reduction tool in regression. One of the most popular SDR method should be sliced inverse regression (Li, 1991), but it is known to be sensitive to the number of slices. To overcome this deficit, Cook and Zhang (2014) proposed a fused approach. Both existing methods do dimension reduction, but, ironically, they do not have the direction application to large p-small n regression, in which the dimension reduction is desperately needed.

By adopting seeded dimension reduction approach (Cook *et al.*, 2007), we newly propose seeded sliced inverse regression and seeded fused sliced inverse regression to improve the two existing SDR methods. Numerical studies confirm that both seeded sliced and fused sliced inverse regressions are effective in dimension reduction for large p-small n regression. Also, it is observed that seeded fused sliced inverse regression is more robust to the number of slices than seeded sliced inverse regression. With real data analysis, we observe the same results, which prove their practical usefulness in high-dimensional data analysis.

As further studies, the dimension estimation of seeded sliced and fused sliced inverse regression, adopting information criteria proposed by Zhu *et al.* (2006). This direction is under progress.

Acknowledgements

For Jae Keun Yoo, this work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2019R1F1A1050715).

References

Brown PJ, Fearn T, and Vannucci M (2001). Bayesian wavelet regression on curves with application to a spectroscopic calibration problem, *Journal of the American Statistical Association*, **96**, 398–408.

Cook RD, Li B, and Chiaromonte F (2007). Dimension reduction in regression without matrix inversion, *Biometrika*, **94**, 569–584.

Cook RD and Zhang X (2014). Fused estimators of the central subspace in sufficient dimension reduction, *Journal of the American Statistical Association*, **109**, 815–827.

Li KC (1991). Sliced inverse regression for dimension reduction, *Journal of the American Statistical Association*, **86**, 316–327.

Yoo JK (2016a). Tutorial: Dimension reduction in regression with a notion of sufficiency, *Communications for Statistical Applications and Methods*, **23**, 93–103.

Yoo JK (2016b). Tutorial: Methodologies for sufficient dimension reduction in regression, *Communications for Statistical Applications and Methods*, **23**, 95–117.

Zhu L, Miao B, and Peng H (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association*, **101**, 630-–643.

Received December 28, 2020; Revised January 27, 2021; Accepted February 1, 2021