Journal of the Korean Statistical Society 43 (2014) 559–566

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

journal homepage: www.elsevier.com/locate/jkss

Multivariate seeded dimension reduction

Jae Keun Yoo *, Yunju Im

Department of Statistics, Ewha Womans University, Seoul 120-750, Republic of Korea

ARTICLE INFO

Article history:
Received 22 March 2013
Accepted 14 March 2014
Available online 6 April 2014

AMS 2000 subject classifications:
primary 62G08
secondary 62H05

Keywords:
Large p small n
Multivariate regression
Seed matrix
Sufficient dimension reduction

ABSTRACT

A recently introduced seeded dimension reduction approach enables existing sufficient dimension reduction methods to be used in regressions with $n < p$. The dimension reduction is accomplished through successive projections of seed matrices on a subspace to contain the central subspace. In the article, we will develop a seeded dimension reduction for multivariate regression, whose responses are multi-dimensional. For this we suggest two conditions that the dimension reduction is attained without the loss of information of the central subspace. Based on this, we construct possible candidate seed matrices. Numerical studies and two data analyses are presented.

1. Introduction

Sufficient dimension reduction (SDR) in the univariate regression of $Y \in \mathbb{R}^1 | \mathbf{X} \in \mathbb{R}^p$ reduces the dimension of the original predictors $\mathbf{X}$ through a lower-dimensional linear projection predictor without loss of information about the conditional distribution of $\mathbf{Y}|\mathbf{X}$ such that

$$Y \perp\!\!\!\perp \mathbf{X} | \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{X}, \tag{1}$$

where $\perp\!\!\!\perp$ stands for independence and $q \leq p$.

Statement (1) is equivalently rephrased that the conditional distributions of $Y|\mathbf{X}$ and $Y|\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}$ are the same, and hence the dimension reduction of $\mathbf{X}$ through $\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{X}$ is achieved without loss of information about $Y|\mathbf{X}$. A subspace spanned by the columns of such $\boldsymbol{\alpha}$ is called a dimension reduction subspace, and SDR typically seeks for the intersection of all dimension reduction subspaces, which is called the *central subspace* $\mathcal{S}_{Y|\mathbf{X}}$. The true dimension and an orthonormal basis matrix of $\mathcal{S}_{Y|\mathbf{X}}$ will be denoted as d and $\boldsymbol{\eta} \in \mathbb{R}^{p \times d}$, respectively. And the dimension reduced predictor of $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ is called *sufficient predictors*.

For the multivariate regression of $\mathbf{Y} \in \mathbb{R}^r | \mathbf{X} \in \mathbb{R}^p$, the idea of SDR is the same as univariate regression, and the central subspace is defined accordingly. To recover $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, two popular approaches of inverse regression and forward regression are widely used. The inverse regression-based methods construct a subspace spanned by the conditional moments of the inverse regression of $\mathbf{X}|\mathbf{Y}$. Methods of K-means inverse regression (Setodji & Cook, 2004) and K-means average variance estimation (Yoo, Lee, & Wu, 2010) estimate $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ through investigating $E(\mathbf{X}|\mathbf{Y})$ and $\mathrm{cov}(\mathbf{X}|\mathbf{Y})$ respectively. In the inverse regression approach, the range of $\mathbf{Y}$ into h clusters through the K-means clustering algorithm, called *slicing*, is the key-step for methodological implementation.

For the latter method, Yoo and Cook (2007) developed a method done by usual ordinary least squares (OLS) application in the regression of $\mathbf{Y}|\mathbf{X}$ such that $\boldsymbol{\beta} = \boldsymbol{\Sigma}^{-1}\mathrm{cov}(\mathbf{X}, \mathbf{Y})$, where $\boldsymbol{\Sigma} = \mathrm{cov}(\mathbf{X})$. To recover more information on $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ through the OLS,

* Corresponding author.
E-mail addresses: peter.yoo@ewha.ac.kr (J.K. Yoo), limyunju825@gmail.com (Y. Im).

http://dx.doi.org/10.1016/j.jkss.2014.03.002
1226-3192/

Yoo (2008) proposed a method to utilize information from polynomial regression of $(\mathbf{Y}, \mathbf{Y}^2, \ldots, \mathbf{Y}^k)|\mathbf{X}$ through constructing $\boldsymbol{\beta}(k) = \boldsymbol{\Sigma}^{-1}\text{cov}\{\mathbf{X}, (\mathbf{Y}, \mathbf{Y}^2, \ldots, \mathbf{Y}^k)\}$, where $\mathbf{Y}^k = (Y_1^k, \ldots, Y_r^k)$.

In order that subspaces spanned by $E(\mathbf{X}|\mathbf{Y})$, $\text{cov}(\mathbf{X}|\mathbf{Y})$, $\boldsymbol{\beta}$, and $\boldsymbol{\beta}(k)$ are proper subsets of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ or equal to $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, the following condition is required: $E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})$ is linear in $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$. This condition is call *linearity condition*, which is very common in the SDR literature. Since the linearity condition is for that of the marginal distribution of $\mathbf{X}$, it is much weaker than a modeling condition usually imposed in $\mathbf{Y}|\mathbf{X}$. Elliptically contoured distributions of $\mathbf{X}$ guarantee that the condition holds. If the linearity condition does not hold, the predictors are often power-transformed for normality.

Although the idea of SDR and the introduced SDR methods for multivariate regression do not have limitation for large p-small n multivariate regression in theory, its practical implementation is not possible, because the inverse of $\mathbf{X}$ is needed to be computed. Recently Cook, Li, and Chiaromonte (2007) introduced a seeded dimension reduction, which provide a general paradigm to use existing SDR methods in such cases. In the seeded dimension reduction, a seed matrix is successively projected to recover $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. We will discuss this dimension reduction method in detail in later sections.

The purpose of the article is to develop a seeded dimension reduction for multivariate regression, called *multivariate seeded dimension reduction*. For this we assume that all information of the regression of $\mathbf{Y}|\mathbf{X}$ is given in $E(\mathbf{Y}|\mathbf{X})$. Based on this, we construct possible candidate seed matrices and withdraw certain conditions to guarantee that the seed matrices reduce the dimension of $\mathbf{X}$ without loss of information on $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

The article is organized as follows. Section 2 is devoted to explain seeded dimension reduction. We develop multivariate seeded dimension reduction in Section 3. Numerical studies and two data analyses are presented in Section 4. We summarize our work in Section 5.

We will define the notations frequently used throughout the rest of the paper. For $\mathbf{B} \in \mathbb{R}^{q\times p}$ and a subspace $\mathcal{S}$ of $\mathbb{R}^p$, $\mathbf{B}\mathcal{S}$ and $\mathcal{S}(\mathbf{B})$ represent the set of $\{\mathbf{B}x : x \in \mathcal{S}\}$ and a subspace spanned by the columns of $\mathbf{B}$, respectively. For a symmetric and positive definite matrix $\boldsymbol{\Sigma}$, a $\boldsymbol{\Sigma}$ inner-product in $\mathbb{R}^p$ is defined as $\langle a, b\rangle_{\boldsymbol{\Sigma}} = a^{\mathrm{T}}\boldsymbol{\Sigma}b$. An orthogonal projection operator onto $\mathcal{S}(\mathbf{B})$ relative to $\langle a, b\rangle_{\boldsymbol{\Sigma}}$ will be defined as $\mathbf{B}(\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{B})^{\dagger}\mathbf{B}^{\mathrm{T}}\boldsymbol{\Sigma}$, where $\dagger$ stands for the Moore–Penrose inverse.

2. Seeded dimension reduction

Popular SDR methods, including ones introduced in Section 1, typically require the inversion of $\boldsymbol{\Sigma}$. When $n < p$, the inversion is not possible, and hence practical application is not plausible any more to such regressions. To overcome this issue in SDR, Cook et al. (2007) proposed a paradigm of sufficient dimension reduction without matrix inversion. To do this, a $p \times q$ seed matrix $\boldsymbol{\nu}$ is needed to be defined for a regression of $Y \in \mathbb{R}^1|\mathbf{X} \in \mathbb{R}^p$ such that $\mathcal{S}(\boldsymbol{\nu}) \subseteq \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$. One important requirement for the seed matrix is that it should be constructed without inverting $\boldsymbol{\Sigma}$. To give some examples for seed matrices, we assume the linearity condition that $E(\mathbf{X}|\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})$ is linear in $\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$. The linearity condition is common in the SDR literature. If $\mathbf{X}$ has an elliptically contoured distribution, the condition is automatically satisfied. In the case that the linearity condition does not hold, $\mathbf{X}$ can often be one-to-one transformed to satisfy this condition. Hereafter we will assume that the linearity condition holds, unless stated otherwise. Under the linearity condition, popular choices for seed matrices are as follows.

2.a When Y is a categorical predictor, $E(\mathbf{X}|Y = y) - E(\mathbf{X}) \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$ for $Y = 1, \ldots, h$.
2.b When Y is many-valued or continuous, the range of Y is divided into h partitions $J_s(Y)$, $s = 1, \ldots, h$ so that $J_s(Y) = 1$, if $Y \in J_s(Y)$ and 0, otherwise. Then $E\{\mathbf{X}|J_s(Y) = 1\} - E(\mathbf{X}) \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$.
2.c $\text{cov}(\mathbf{X}, Y) \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$.
2.d $\text{cov}\{\mathbf{X}, U(k)\} \in \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$, where $U = \{Y - E(Y)\}/\sqrt{\text{var}(Y)}$ and $U(k) = (U, U^2, \ldots, U^k)$, $k = 1, 2, \ldots$.

For simplicity, we will assume that $\mathcal{S}(\boldsymbol{\nu}) = \boldsymbol{\Sigma}\mathcal{S}_{Y|\mathbf{X}}$ throughout the rest of paper.

For a known subspace $\mathcal{M}_{Y|\mathbf{X}}$ of $\mathbb{R}^p$ such that $\mathcal{S}_{Y|\mathbf{X}} \subseteq \mathcal{M}_{Y|\mathbf{X}}$, it is obvious that $\boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}) \subseteq \mathcal{M}_{Y|\mathbf{X}}$. Let $\mathbf{P}_{\mathcal{M}_{Y|\mathbf{X}}(\boldsymbol{\Sigma})} = \mathbf{R}(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}$ be an orthogonal projection operator $\mathbf{P}_{\mathcal{M}_{Y|\mathbf{X}}(\boldsymbol{\Sigma})}$ onto $\mathcal{M}_{Y|\mathbf{X}}$ relative to $\langle a, b\rangle_{\boldsymbol{\Sigma}}$, where $\mathbf{R}$ is a $p \times q$ matrix such that $\mathcal{S}(\mathbf{R}) = \mathcal{M}_{Y|\mathbf{X}}$. Since the projection of $\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}$ onto $\mathcal{M}_{Y|\mathbf{X}}$ returns itself, the following equivalences are derived:

$$\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu} = \mathbf{P}_{\mathcal{M}_{Y|\mathbf{X}}(\boldsymbol{\Sigma})}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu} = \mathbf{R}(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu} = \mathbf{R}(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{\nu}. \tag{2}$$

Since $\boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}) = \mathcal{S}_{Y|\mathbf{X}}$, the columns of $\mathbf{R}(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{\nu}$ span $\mathcal{S}_{Y|\mathbf{X}}$ by the last equivalence in (2). Here, one crucially notable thing is that $\boldsymbol{\Sigma}^{-1}$ is not required in $\mathbf{R}(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{-1}\mathbf{R}^{\mathrm{T}}\boldsymbol{\nu}$. If $\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R}$ is not invertible, $(\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R})^{\dagger}$ is used instead.

Then, naturally, the matrix $\mathbf{R}$ is needed to be constructed so that its column spans a subspace large enough to contain $\mathcal{S}_{Y|\mathbf{X}}$ but reasonably estimable from data. For this, iterative projections of $\boldsymbol{\nu}$ onto $\boldsymbol{\Sigma}$ were proposed in Cook et al. (2007):

$$\mathbf{R}_u \equiv (\boldsymbol{\nu}, \boldsymbol{\Sigma}\boldsymbol{\nu}, \ldots, \boldsymbol{\Sigma}^{u-1}\boldsymbol{\nu}), \quad u = 1, 2, \ldots, u^*. \tag{3}$$

The sufficient dimension reduction through the successive projection of seed matrices is called *seeded dimension reduction*.

The letter u in (3) is called a termination index of projections. It is noted that $\mathcal{S}(\mathbf{R}_{u-1}) \subseteq \mathcal{S}(\mathbf{R}_u)$ for any $u \geq 2$. Since $\mathcal{S}(\mathbf{R}_u)$ forms a nondecreasing sequence, it is important to make a proper choice of the termination index u, small enough to guarantee that $\mathcal{S}(\mathbf{R}_u) = \mathcal{S}_{Y|\mathbf{X}}$. Recently Yoo (2013) suggests bootstrap coefficients of variations to determine the termination

index, which does not require any asymptotics and is implemented in a simple way. Since the determination is not a purpose of the paper, readers can refer to Yoo (2013) for more detail.

In practice, first, $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}$ are replaced by their sample quantities and then a proper value of u, saying u^*, is determined. Then the sample version $\hat{\mathbf{R}}_{u^*}$ is constructed, and finally $\mathcal{S}_{Y|\mathbf{X}}$ is estimated by the columns of $\hat{\mathbf{R}}_{u^*}(\hat{\mathbf{R}}_{u^*}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}\hat{\mathbf{R}}_{u^*})^{-1}\hat{\mathbf{R}}_{u^*}^{\mathrm{T}}\hat{\boldsymbol{\nu}}$.

3. Multivariate seeded dimension reduction

Multivariate regression of $\mathbf{Y} = (Y_1, \ldots, Y_r)^{\mathrm{T}}|\mathbf{X}$ involves multi-dimensional responses with $r \geq 2$. To distinguish between univariate and multivariate responses, bold fonts are used for multivariate responses $\mathbf{Y}$. The definitions of a dimension reduction subspace and the central subspace are the same as univariate-response regression. Let $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ stand for the central subspace of $\mathbf{Y}|\mathbf{X}$.

When considering a seeded dimension reduction in multivariate regression, the most important thing is to find a seed matrix $\boldsymbol{\nu}_M$ such that $\mathcal{S}(\boldsymbol{\nu}_M) \subseteq \boldsymbol{\Sigma}\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, which should be constructed as simple as possible. As candidates of the seed matrices for multivariate regression, we consider the following population quantities under the linearity condition.

a. Use only with $r = 2$, that is, $\mathbf{Y} = (Y_1, Y_2)$. Slice one response Y_1 first. Let $h_{(1)}$ be the number of slices for Y_1. Construct $h_{(1,2)}$ slices J_s, $s = 1, \ldots, h_{(1,2)}$, through dividing another response Y_2 within each of $h_{(1)}$ slices. If one of the response is categorical, then the categorical one should be sliced first. Then $\boldsymbol{\nu}_M = E(\mathbf{X}|J_s) - E(\mathbf{X})$.
b. Let K_s be a cluster indicator acquired from K-means clustering for $s = 1, \ldots, h$ so that $K_s = 1$, if $\mathbf{Y} \in K_s$ and 0 otherwise. Then $\boldsymbol{\nu}_M = E(\mathbf{X}|K_s) - E(\mathbf{X})$.
c. $\boldsymbol{\nu}_M = \mathrm{cov}(\mathbf{X}, \mathbf{Y})$.

For all candidate seed matrices, $\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}_M \subseteq \mathcal{S}_{\mathbf{Y}|\mathbf{X}}$. In the first candidate, we only consider bivariate response regression. The inverse mean $\boldsymbol{\Sigma}^{-1}E(\mathbf{X}|\mathbf{Y})$ is an element of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$, so $E(\mathbf{X}|\mathbf{Y})$ can be used as a seed matrix. It is constructed, however, by slicing $\mathbf{Y}$ in practice. For instance, if $r = 4$, the minimum total number of slices will be $2^4 = 16$ which may not be effective for a small sample of size 100 or less. However, for bivariate responses, we may be able to obtain more than 4 cells in the hypercubes with sufficient observations in each of them for the estimation of the first moments of an inverse regression. Because of it, for higher-dimensional responses, K-means clustering replaces the hierarchical slicing. The second candidate seed matrix is the inverse mean of $E(\mathbf{X}|\mathbf{Y})$ with slices constructed through the K-means clustering algorithm. A similar approach to using the clustering algorithm to restore the second conditional moments of $\mathbf{X}|\mathbf{Y}$ is suggested in Yoo et al. (2010).

The last candidate seed is the covariance matrix of $\mathbf{X}$ and $\mathbf{Y}$, which successfully reduces the dimension in multivariate regression in Yoo and Cook (2007).

The conditions to attain dimension reduction of $\mathbf{X}$ without loss of the information of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ with the three candidate seed matrices are summarized in the next proposition.

Proposition 1. *Assume that* $\mathcal{S}_{Y_k|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M)$ *for* $k = 1, \ldots, r$ *and* $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \oplus_{k=1}^{r}\mathcal{S}_{Y_k|\mathbf{X}}$. *Then* $\boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M) = \mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Proof. Showing that $\mathcal{S}(\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}_M) \subseteq \mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is easy, because the linearity condition guarantees it. On the other hand, since $\mathcal{S}_{Y_k|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M)$ for $k = 1, \ldots, r$, it is guaranteed that $\oplus_{k=1}^{r}\mathcal{S}_{Y_k|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M)$. By the condition that $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \oplus_{k=1}^{r}\mathcal{S}_{Y_k|\mathbf{X}}$, it is straightforward that $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M)$. This completes the proof. □

When an underlying regression satisfies a condition that $\mathbf{Y} \perp\!\!\!\perp \mathbf{X}|E(\mathbf{Y}|\mathbf{X})$, the regression is called *location regression*. Then the information of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is completely characterized in $E(\mathbf{Y}|\mathbf{X})$, so we have that $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \mathcal{S}\{E(\mathbf{Y}|\mathbf{X})\}$. That is, all information of the regression is placed onto the first conditional moment of $\mathbf{Y}|\mathbf{X}$. This assumption holds in multivariate linear regression and related reduced-rank regression, which are the most popular in multivariate regression analysis. And, the goal of dimension reduction in high-dimensional data is often placed onto reasonable simplification of data as much as possible to conduct proper statistical inference. Therefore, a class of the location regression is considered as mild restriction in various regression problems in practice, and it satisfies the two conditions of $\mathcal{S}_{Y_k|\mathbf{X}} \subseteq \boldsymbol{\Sigma}^{-1}\mathcal{S}(\boldsymbol{\nu}_M)$, $k = 1, \ldots, r$, and $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \oplus_{k=1}^{r}\mathcal{S}_{Y_k|\mathbf{X}}$ stated in Proposition 1. It is noted that the two conditions in Proposition 1 are weaker than the location regression. Hence, Proposition 1 should not be an obstacle in practice to conduct seeded dimension reduction in multivariate regression.

Next we need to define $\mathcal{M}_{\mathbf{Y}|\mathbf{X}}$ to contain $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ so that it is straightforward that $\boldsymbol{\Sigma}^{-1}\boldsymbol{\nu}_M \in \mathcal{M}_{\mathbf{Y}|\mathbf{X}}$. And, as a basis matrix $\mathbf{R}_{M,u}$ for $\mathcal{M}_{\mathbf{Y}|\mathbf{X}}$, we will use the following matrix

$$\mathbf{R}_{M,u} \equiv (\boldsymbol{\nu}_M, \boldsymbol{\Sigma}\boldsymbol{\nu}_M, \ldots, \boldsymbol{\Sigma}^{u-1}\boldsymbol{\nu}_M), \quad u = 1, 2, \ldots, u^*.$$

Then, for a proper value u^* of u, $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is spanned by the columns of

$$\mathbf{B} = \mathbf{R}_{M,u^*}(\mathbf{R}_{M,u^*}^{\mathrm{T}}\boldsymbol{\Sigma}\mathbf{R}_{M,u^*})^{-1}\mathbf{R}_{M,u^*}^{\mathrm{T}}\boldsymbol{\nu}_M.$$

In practice, first, $\boldsymbol{\Sigma}$ and $\boldsymbol{\nu}_M$ are replaced by their sample quantities and then a proper value of u, saying u^*, is determined, and the sample versions of $\hat{\mathbf{R}}_{M,u^*}$ and $\hat{\mathbf{B}}$ are constructed accordingly. Then $\mathcal{S}(\hat{\mathbf{B}})$ is an estimator of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Table 1
Comparisons with a projective resampling based on Example 1 with in Section 4.1: cov(**X**, **Y**), seeded dimension reduction; PR, a projective resampling method by Li et al. (2008).

	$n = 100$	$n = 200$	$n = 400$	$n = 800$	$n = 1600$
cov(**X**, **Y**)	0.193	0.135	0.097	0.069	0.048
PR	0.276	0.185	0.133	0.095	0.067

4. Numerical studies and data analysis

4.1. Numerical studies

We considered four different regression models. The first three examples for multivariate regression with additive errors, and the fourth one is a survival regression.

Example 1. A resampling method developed by Li, Wen, and Zhu (2008) is shown to have potential advantages over other dimension reduction methods available in multivariate regression. The purpose of the example is to show usefulness of the multivariate seeded dimension reduction through methodological comparisons with the existing method. So we followed Model 4.4 in Li et al. (2008).

In the example, six-dimensional predictors of $\mathbf{X} = (X_1, \ldots, X_6)$ were independently generated from $N(0, 1)$. Then, five dimensional responses of $\mathbf{Y} = (Y_1, \ldots, Y_5)$ are constructed as follows: $Y_1 = X_2 + (3X_2)/(0.5 + (X_1 + 1.5)^2) + \varepsilon_1$; $Y_2 = X_1 + \exp(0.5X_2) + \varepsilon_2$; $Y_3 = X_1 + X_2 + \varepsilon_3$; $Y_4 = \varepsilon_4$; $Y_5 = \varepsilon_5$, where $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_5) \sim MN(0, \boldsymbol{\Delta}) \perp\!\!\!\perp \mathbf{X}$, with $\boldsymbol{\Delta} = \mathrm{diag}(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2)$, in which $\boldsymbol{\Delta}_1 = \begin{bmatrix} 1 & -1/2 \\ -1/2 & 1 \end{bmatrix}$ and $\boldsymbol{\Delta}_2 = \mathrm{diag}(1/2, 1/3, 1/4)$. And, a notation of *MN* stands for multivariate normal distribution.

In the example, the central subspace is spanned by $\boldsymbol{\eta} = \{(1, 0, 0, 0, 0, 0)^{\mathrm{T}}, (0, 1, 0, 0, 0, 0)^{\mathrm{T}}\}$. As a seed, we consider cov(**X**, **Y**) with $u^* = 4$. To measure how well $\mathcal{M}_{Y|\mathbf{X}}$ is estimated, following the distance used in Li et al. (2008), we computed $m = \|\boldsymbol{\eta}\boldsymbol{\eta}^{\mathrm{T}} - \hat{\mathbf{B}}(\hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{B}})^{-1}\hat{\mathbf{B}}^{\mathrm{T}}\|$. The resulted *m*s are reported in Table 1 along with the results from a projective resampling method.

According to Table 1, the seeded dimension reduction shows slightly better performances in estimating $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ than the projective resampling method. Here, we are not going to conclude that the proposed method is superior to the existing one. Rather, we are telling that the multivariate seeded dimension reduction can have good applications to regressions with $n > p$ where various other dimension reduction methods are applicable.

For Examples 2–3, the following variable configuration for predictors of $\mathbf{X} \in \mathbb{R}^p$ was commonly used: predictors $\mathbf{X} \in \mathbb{R}^p$ for either $p = 10$ or $p = 500$ were independently generated from $MN(0, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = (2/3)\mathrm{diag}(2, \ldots, 2, 1, \ldots, 1)$ with equal multiplicity between 1 and 2. And, for both the examples, the number of iteration was 100, and $u^* = 4$ was used to simplify simulations. Define that $\boldsymbol{\eta}_1 \in \mathbb{R}^p = (p^{-1/2}, p^{-1/2}, \ldots, p^{-1/2})$ and $\boldsymbol{\eta}_2 \in \mathbb{R}^p = \{(0.4{*}p)^{-1/2}, (0.4{*}p)^{-1/2}, \ldots, (0.4{*}p)^{-1/2}, 0, \ldots, 0\}$. In $\boldsymbol{\eta}_1$, all coordinate values are equal and are normalized for its length to be one. For $\boldsymbol{\eta}_2$, all of the first 40% coordinate values are equal to $(0.4 * p)^{-1/2}$ and otherwise zeros. Either one or both were commonly used to define $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$.

Example 2. Here we consider a case that $\dim(\mathcal{S}_{\mathbf{Y}|\mathbf{X}}) = 1$, which is spanned by the column of either $\boldsymbol{\eta}_1$ or $\boldsymbol{\eta}_2$, and constructs the following bivariate regressions: $Y_1 = \sin(\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}) + 0.3\varepsilon_1$ and $Y_2 = \exp(\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}) + 0.3\varepsilon_2$ for $i = 1, 2$, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 1) \perp\!\!\!\perp \mathbf{X}$. As seed matrices, we used the three candidates of cov(**X**, **Y**) and $E(\mathbf{X}|J_s)$ and $E(\mathbf{X}|K_s)$, $s = 1, 2, 3, 4$.

Let $\hat{\mathbf{B}}$ stand for the estimate of $\boldsymbol{\eta}_i$s through the proposed multivariate seeded dimension reduction. Then, to summarize how well $\boldsymbol{\eta}_i$s are estimated, we considered averages of $|\sqrt{R_i^2}|$s for $i = 1, 2$, where R_i^2 represent the coefficient of determination computed from a regression of $\boldsymbol{\eta}_i^{\mathrm{T}}\mathbf{X}|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$. The simulation results are reported in Table 2.

Table 2 represents the characteristic behaviors in the estimation of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ observed in other numerical studies regarding bivariate response regressions. According to the table, for the estimation of either $\boldsymbol{\eta}_1$ of $\boldsymbol{\eta}_2$, the seed of $E(\mathbf{X}|K_s)$ performs relatively worse than the other two seed matrices, and it turns out to be sensitive to sizes of n and p. The reason why $E(\mathbf{X}|J_s)$ shows better performances than $E(\mathbf{X}|K_s)$ is because the slicing scheme is better in the construction of inverse mean of **X** than the K-means algorithm with bivariate responses, as Yoo et al. (2010) indicates. This implies that, with bivariate responses, one is recommended to use either of cov(**X**, **Y**) and $E(\mathbf{X}|J_s)$ over $E(\mathbf{X}|K_s)$. And, the seed of cov(**X**, **Y**) shows a quite good robust estimation of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ regardless of the true basis matrices in either case of $n > p$ and $n < p$.

Example 3. As the second simulation example, the following model was considered: $Y_1 = \boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X} + (\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X})^3/10 + 0.3\varepsilon_1$; $Y_2 = \boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X} + (\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X})^3/10 + 0.3\varepsilon_2$; $Y_3 = (\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X})^2 + 0.3\varepsilon_3$, where $\varepsilon_i \overset{\text{i.i.d.}}{\sim} N(0, 1) \perp\!\!\!\perp \mathbf{X}$.

In the example, the dimension of responses is three, and the central subspace $\mathcal{S}_{\mathbf{Y}|\mathbf{X}}$ is spanned by the two columns of $(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$. As seed matrices, cov(**X**, **Y**) and $E(\mathbf{X}|K_s)$, $s = 1, 2, 3, 4$, were considered, following that 3 or higher dimensional

Table 2
Averages of $|R^2|$s for Example 2 with in Section 4.1.

	From $\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X}\|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$				From $\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X}\|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
	$p = 10$	$p = 500$	$p = 10$	$p = 500$	$p = 10$	$p = 500$	$p = 10$	$p = 500$
$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.985	0.909	0.992	0.911	0.985	0.911	0.992	0.903
$E(\mathbf{X}\|J_s)$	0.962	0.962	0.980	0.980	0.961	0.752	0.980	0.774
$E(\mathbf{X}\|K_s)$	0.923	0.722	0.975	0.786	0.923	0.706	0.975	0.779

Table 3
Averages of $|R^2|$s for Example 3 with in Section 4.1.

	From $\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X}\|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$				From $\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X}\|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$			
	$n = 50$		$n = 100$		$n = 50$		$n = 100$	
	$p = 10$	$p = 500$	$p = 10$	$p = 500$	$p = 10$	$p = 500$	$p = 10$	$p = 500$
$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.985	0.927	0.990	0.929	0.992	0.959	0.995	0.958
$E(\mathbf{X}\|K_s)$	0.918	0.805	0.951	0.830	0.895	0.754	0.929	0.793

slicing does not work very well according Yoo et al. (2010). For $Y_3|\mathbf{X}$ alone, both seed matrices are never informative to $\boldsymbol{\eta}_1$, but the relation of $\mathcal{S}_{\mathbf{Y}|\mathbf{X}} = \oplus_{i=1}^{3} \mathcal{S}_{Y_i|\mathbf{X}}$ holds, because $\boldsymbol{\eta}_1$ can be restored from $Y_1|\mathbf{X}$ and $Y_2|\mathbf{X}$. To measure how well $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$ is estimated, we computed averages of $|\sqrt{R_i^2}|$s, $i = 1, 2$, from $\boldsymbol{\eta}_1^{\mathrm{T}}\mathbf{X}|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$ and $\boldsymbol{\eta}_2^{\mathrm{T}}\mathbf{X}|\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$, where $\hat{\mathbf{B}}$ is the estimate of $\boldsymbol{\eta}$ through the multivariate seeded dimension reduction. The simulation results are reported in Table 3.

From Table 3, we can see similar patterns to Table 2. Again, in the example, the seed of $\mathrm{cov}(\mathbf{X}, \mathbf{Y})$ usually works better than $E(\mathbf{X}|K_s)$.

Example 4. For the survival application of the proposed method, we considered the Cox-proportional hazard model used in Yoo (2013). With $n = 300$ with either $p = 100$ or 1000, the first 30% predictors and all the other predictors, respectively, were independently generated from $N(0, 5)$ and $N(0, 1)$. Next we defined that $\boldsymbol{\eta}$ had the first 30% elements equal to $(0.3p)^{-0.5}$ and the remaining ones equal to zeros. The score function of $f(x) = 5\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}$ was examined. For baseline hazard, the Weibull distribution with the shape and scale parameters varied to 1 and 10 respectively. Censoring time C, which is independent of $\mathbf{X}$, was generated from $U(0, \upsilon)$ for $\upsilon = 4, 8, 12$. Then the observed survival time ranges between 0 and 10 years depending on choices of υ. Then true survival time T is defined as follows: $T = \{-\log(U_1)\exp(-5\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X})\}^{1/10}$, where $U_1 \sim U(0, 1) \perp\!\!\!\perp (\mathbf{X}, C)$. By this the observed survival time Y and censoring status δ are defined as follows: $Y = \min(T, C)$ and $\delta = 0$, if $Y = T$ and 1, otherwise. In the model, observed censoring percentages were around 35%, 20%, and 14% for $\upsilon = 4$, 8, 12 in order.

Here we adopted $\mathrm{cov}\{\mathbf{X}, \mathbf{Y} = (Y, \delta, Y * \delta)\}$ as a seed matrix with $u^* = 4$ for simplicity. The usefulness of $\mathrm{cov}(\mathbf{X}, \mathbf{Y})$ in dimension reduction in survival regression is well discussed in Yoo and Lee (2011). Among total 300 samples, 200 samples were randomly selected as training samples, which were used to get estimates $\hat{\mathbf{B}}$ of $\boldsymbol{\eta}$ through the proposed seeded dimension reduction. Using the remaining 100 test samples, areas under ROC curves of predicted values were computed from $\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$ to measure estimation performances. This example was already done in Yoo (2013) with $E(\mathbf{X}|J_s)$ via the bivariate slicing of Y and δ as a seed matrix. Table 4 reports the areas under the ROC curves from the two seed matrices.

It shows that the seeded dimension is robust to choices of censoring and provides reliable prediction with larger p. Comparing performances between $\mathrm{cov}(\mathbf{X}, \mathbf{Y})$ and $E(\mathbf{X}|J_s)$, there are no notable differences, although $\mathrm{cov}(\mathbf{X}, \mathbf{Y})$ is slightly better with $C \sim U(0, 4)$ and $C \sim U(0, 8)$ than $E(\mathbf{X}|J_s)$ and vice versa with $C \sim U(0, 12)$.

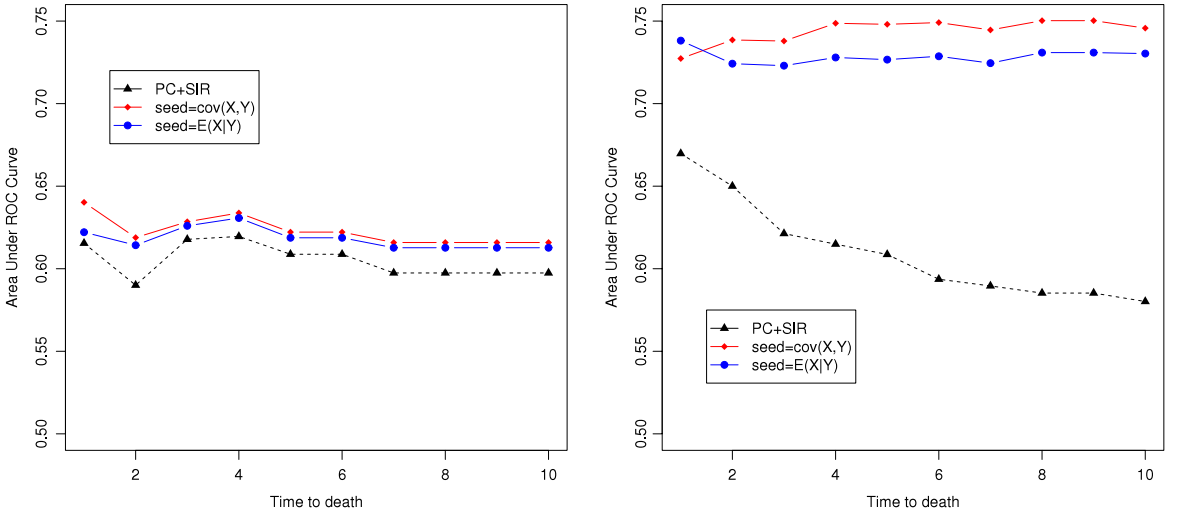
4.2. Diffuse large-B-cell lymphoma data

For the purpose of illustration of the proposed method in high-dimensional data analysis, diffuse large-B-cell lymphoma data (DLBCL; Rosenwald et al., 2002) are analyzed. The DLBCL data of Rosenwald et al. (2002) contain measurements of 7399 genes from 240 patients obtained from customized cDNA microarrays. For each patients, survival time was recorded and varied from 0 to 21.8 years. The total uncensored cases (deceased) are 138 among 240 patients. The DLBCL dataset is available at http://llmpp.nih.gov/DLBCL.

The DLBCL dataset was analyzed by Li (2004) using gene expression information. First the dataset was randomly divided into a training set of 148 cases and a test set of the remaining 74 cases. Then a two-step procedure was employed to reduce dimensions of 7399 genes for the training set. The genes were initially replaced with their 40 principal components through principal component analysis, and then bivariate sliced inverse regression (Cook, 2003) was conducted in a survival regression of bivariate responses of the observed survival time and censoring status given 40 selected principal components. Finally the Cox-proportional hazard model was fitted with the second dimension-reduced gene expressions. For model-validation, predicted scores for both the training and testing sets were computed.

Table 4
Area under ROC curves predicted from $\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{X}$ for Example 4 in Section 4.1.

Year		1	2	3	4	5	6	7	8	9	10
Censoring $C \sim U(0, 4)$											
$p = 100$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.746	0.715	0.717	0.658	–	–	–	–	–	–
	$E(\mathbf{X}\|J_s)$	0.738	0.714	0.714	0.634	–	–	–	–	–	–
$p = 1000$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.703	0.684	0.685	0.654	–	–	–	–	–	–
	$E(\mathbf{X}\|J_s)$	0.699	0.679	0.678	0.637	–	–	–	–	–	–
Censoring $C \sim U(0, 8)$											
$p = 100$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.744	0.712	0.713	0.712	0.713	0.699	–	–	–	–
	$E(\mathbf{X}\|J_s)$	0.728	0.702	0.701	0.707	0.716	0.707	–	–	–	–
$p = 1000$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.705	0.688	0.682	0.689	0.684	0.651	–	–	–	–
	$E(\mathbf{X}\|J_s)$	0.699	0.679	0.680	0.678	0.663	0.619	–	–	–	–
Censoring $C \sim U(0, 12)$											
$p = 100$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.737	0.711	0.712	0.725	0.736	0.737	0.743	0.721	0.688	0.671
	$E(\mathbf{X}\|J_s)$	0.733	0.709	0.710	0.716	0.727	0.728	0.728	0.725	0.690	0.669
$p = 1000$	$\mathrm{cov}(\mathbf{X}, \mathbf{Y})$	0.696	0.685	0.689	0.693	0.698	0.699	0.685	0.652	0.626	0.614
	$E(\mathbf{X}\|J_s)$	0.700	0.684	0.684	0.683	0.686	0.700	0.690	0.679	0.655	0.622



(a) Testing data. (b) Training data.

Fig. 1. Area under ROC curves at time 1–10 years; PC + SIR, Li's approach; $\mathrm{cov}(X, Y)$, a seed with $\mathrm{cov}\{\mathbf{X}, (Y, \delta, Y * \delta)\}$; $E(X|Y)$, a seed with $E(\mathbf{X}|J_s)$.

One possibly arguable issue in the analysis should be the initial reduction of genes through principal component analysis, because it is done only based on the marginal information of the genes, ignoring the conditional dependence of the survival time and censoring status given in the genes.

We apply multivariate seeded dimension reduction and replace the two-step dimension reduction of the genes with the one-step seeded dimension reduction. This analysis is expected to be potentially better in prediction than Li's analysis, because the former can attain more informative dimension reduction of the genes during a whole process by considering the conditional dependency.

As candidate seed matrices, $E(\mathbf{X}|J_s)$ and $\mathrm{cov}\{\mathbf{X}, (Y, \delta, Y * \delta)\}$ were considered and their performances were compared with Li's approach. To decide u^*, we applied bootstrap determination criteria by Yoo (2013), and it turned out that $u^* = 2$ for both cases. (not reported).

Evaluation of the performances was done by computing areas under ROC curves of predicted values from both the training and testing sets, which is reported in Fig. 1.

According to Fig. 1, the application of the proposed methods with two different seed matrices yields a better prediction of survival time for both test and training sets than Li's analysis, although the differences are much larger for training sets.

The difference from the analysis in Yoo (2013, Section 5.2) is placed onto utilizing information about the international prognostic index. The index is regarding clinical characteristics such as age, tumor stage, serum lactate dehydrogenase concentration, performance status and a number of extranodal disease sites. The usage of information on the index improved area under ROC curves in the testing set from the analysis presented here.

Table 5
Mean squared errors of prediction on the test set in NIR data: cov($\mathbf{X}$, $\mathbf{Y}$), seeded dimension reduction; mPLS, multivariate partial least squares.

	Fat	Sugar	Flour	Water
cov($\mathbf{X}$, $\mathbf{Y}$)	0.119	0.605	0.422	0.114
mPLS	0.123	1.117	0.594	0.099

4.3. Near-infrared spectroscopy of biscuit doughs data

For another illustration of the proposed method, near-infrared spectroscopy of biscuit doughs data (Brown, Fearn, & Vannucci, 2001; NIR) is analyzed. This dataset includes the measurements of the composition of biscuit dough pieces from near-infrared spectroscopy, which is one of most favorable methods to analyze constitutions of various materials such as food and drink, pharmaceutical products, and petrochemicals. The data were collected to measure the composition of biscuit dough pieces and the four constituents are under investigation: fat, sucrose, dry flour and water. The calculated percentages of the four ingredients are four-dimensional responses. As predictors, in the original dataset, there are 700 points measured from 1100 to 2498 nanometers (nm) in steps of 2 nm. Following Brown et al. (2001), the number of points was reduced by removing the first 140 and the last 49 wavelengths, which were believed to have little useful information and increasing the steps from 2 to 4 nm. Then a wavelength ranging from 1380 to 2400 nm is used and there are 256 points, representing the dimension of predictors equal to 256. The data can be acquired from `ppls-package` of statistical language R and is named `cookie`. There are total 72 samples in the data, and they were divided into two groups of 40 training and 32 test samples, and two observations of sample 23 in the training set and sample 21 in the test set were eliminated as outliers before analysis.

For data analysis, we consider a classical multivariate linear regression of

$$\mathbf{Y} \in \mathbb{R}^4 | \mathbf{X} \in \mathbb{R}^{256} = \boldsymbol{\alpha} + \boldsymbol{\beta}^{\mathrm{T}} \mathbf{X} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\alpha} \in \mathbb{R}^4$, $\boldsymbol{\beta} \in \mathbb{R}^{256\times 4}$ and $\boldsymbol{\varepsilon} \in \mathbb{R}^4$ is a random vector with mean 0 and the covariance matrix $\boldsymbol{\Sigma}$ and is independent of $\mathbf{X}$.

With the training set, the multivariate linear regression was estimated and then evaluated by the test set. Since the direct estimation of α and $\boldsymbol{\beta}$ through the ordinary least squares was not possible due to $p = 256 > n = 40$, we constructed dimension-reduced predictors $\hat{\mathbf{B}}_{\mathrm{tr}}^{\mathrm{T}} \mathbf{X}$ by the proposed method with cov($\mathbf{X}$, $\mathbf{Y}$) as a seed and $u^* = 4$. Then the new predictors $\hat{\mathbf{B}}^{\mathrm{T}} \mathbf{X}$ replaced the original 256-dimensional predictors, and the multivariate linear regression was fitted through the ordinary least squares on the regression of $\mathbf{Y} | \hat{\mathbf{B}}^{\mathrm{T}} \mathbf{X}$.

For the purpose of comparison, the data was fitted through multivariate partial least squares, which was adopted in Brown et al. (2001) as one of the standard statistical analyses. In Brown et al. (2001), five components were considered for partial least squares, and we followed the guidance.

After fitting the data via the two approaches, as comparison criteria, mean squared errors of predictions on the test set were computed for each response variable and are reported in Table 5.

As we can see from Table 5, the model building through multivariate seeded dimension reduction provides potential advantages over the standard analysis, and we can again confirm practical usefulness in high dimensional data analysis.

5. Discussions

In this paper, we present a seeded dimension approach for multivariate regression, which is applicable with a case of $n < p$. Also we provide two conditions to guarantee the dimension reduction of predictors without loss of information about the regression. Since the conditions can cover a large class of multivariate regression, called location regression, they should not be heavy in practice.

Numerical studies confirm that the proposed dimension reduction method is theoretically well supported. To show practical usefulness of the proposed methodology, it is applied to diffuse large-B-cell lymphoma data and near-infrared spectroscopy of biscuit dough data. In both data analyses, the proposed method produces better predictions than corresponding existing analyses.

The proposed approach will provide a possible neat solution to big data analysis such as high-dimensional classification or functional data analysis that the dimensionality of datasets dramatically increases over time. The computer codes for the paper is available upon request.

Acknowledgments

The authors are grateful to the associate editor and the two referees for many insightful and helpful comments.

For Jae Keun Yoo, this works was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (NRF-2012R1A1A1040077).

For Yunju Im, this work was supported by the BK21 Plus Project through the National Research Foundation of Korea (NRF) funded by the Korean Ministry of Education (22A20130011003).

References

Brown, P. J., Fearn, T., & 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, R. D. (2003). Dimension reduction and graphical exploration in regression including survival analysis. *Statistics in Medicine*, *22*, 1399–1413.
Cook, R. D., Li, B., & Chiaromonte, F. (2007). Dimension reduction in regression without matrix inversion. *Biometrika*, *94*, 569–584.
Li, L. (2004). Dimension reduction methods for microarrays with application to censored survival data. *Bioinformatics*, *20*, 3406–3412.
Li, B., Wen, S., & Zhu, L. (2008). On a projective resampling method for dimension reduction with multivariate responses. *Journal of the American Statistical Association*, *103*, 1177–1186.
Rosenwald, A., et al. (2002). The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *New England Journal of Medicine*, *346*, 1937–1947.
Setodji, C. M., & Cook, R. D. (2004). *K*-means inverse regression. *Technometrics*, *46*, 421–429.
Yoo, J. K. (2008). A novel moment-based dimension reduction approach in multivariate regression. *Computational Statistics and Data Analysis*, *52*, 3843–3851.
Yoo, J. K. (2013). Advances in seeded dimension reduction: bootstrap criteria and extensions. *Computational Statistics and Data Analysis*, *60*, 70–79.
Yoo, J. K., & Cook, R. D. (2007). Sufficient dimension reduction for the conditional mean in multivariate regression. *Biometrika*, *94*, 231–242.
Yoo, J. K., & Lee, K. (2011). Model-free predictor tests in survival regression through sufficient dimension reduction. *Lifetime Data Analysis*, *17*, 433–444.
Yoo, J. K., Lee, K., & Wu, S. (2010). On the extension of sliced average variance estimation to multivariate regression. *Statistical Methods and Applications*, *19*, 529–540.