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