Journal of Statistical Planning and Inference 139 (2009) 2656 – 2664

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

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

Sliced inverse regression for multivariate response regression

Heng-Hui Lue*

Department of Statistics, Tunghai University, Taiwan

ARTICLE INFO

Article history:
Received 25 January 2008
Received in revised form
11 December 2008
Accepted 11 December 2008
Available online 24 December 2008

Keywords:
Canonical correlation
Dimension reduction
Most predictable variates
Multivariate response
Sliced inverse regression

ABSTRACT

We consider a regression analysis of multivariate response on a vector of predictors. In this article, we develop a sliced inverse regression-based method for reducing the dimension of predictors without requiring a prespecified parametric model. Our proposed method preserves as much regression information as possible. We derive the asymptotic weighted chi-squared test for dimension. Simulation results are reported and comparisons are made with three methods—most predictable variates, k-means inverse regression and canonical correlation approach.

1. Introduction

A major interest in analyzing multivariate datasets is the reduction of dimension for visualizing the patterns of data structure. Li (1991) addressed this issue for regression problems based on the notion of the effective dimension reduction (EDR) space. Directions in that space are used to project a p-dimensional predictor vector $\mathbf{X}$ for effectively viewing and studying the relationship with a univariate response y. When the EDR space is minimum, the maximum reduction of dimension exists. As shown by Cook in a series of articles, suitable regularity conditions must be imposed for obtaining a unique EDR space with minimum dimension (Cook, 1998). Cook and his collaborators further clarified the notion of the EDR space in their work, leading to the central subspace (CS) (Cook, 1994) and the central mean subspace (Cook and Li, 2002). The intersection of all EDR spaces is the CS (Cook, 1998).

With a response vector $\mathbf{Y} = (Y_w) \in \mathrm{R}^q$, high-dimensional nonlinear regression becomes much more complicated. The complication caused by the curse of dimensionality on both predictors and response variables is compounded by the uncertainty in choosing the correct form for the regression function. Without requiring a prespecified parametric regression model, one seeks to reduce the dimension of $\mathbf{X}$ without loss of information by identifying the smallest number of linear combinations $\beta_1'\mathbf{X}, \ldots, \beta_\kappa'\mathbf{X}$ such that

$$\mathbf{Y} \perp \mathbf{X} | (\beta_1'\mathbf{X}, \ldots, \beta_\kappa'\mathbf{X}) \tag{1}$$

where $\perp$ means independence and β_j's are unknown vectors of interest. Model (1) means that conditionally on $\beta_j'\mathbf{X}, j = 1, \ldots, \kappa$, $\mathbf{Y}$ is independent of $\mathbf{X}$. This indicates that useful information about $\mathbf{Y}$ from $\mathbf{X}$ can be retrieved from κ projected variables $\beta_j'\mathbf{X}$. Hence this expression is called sufficient dimension reduction (SDR) and the variates $(\beta_1'\mathbf{X}, \ldots, \beta_\kappa'\mathbf{X})$ are then called sufficient predictors.

Several model-free methods are available to estimate the SDR space, for instance, sliced inverse regression (SIR) (Li, 1991), sliced average variance estimation (SAVE) (Cook and Weisberg, 1991), principal Hessian directions (PHD) (Li, 1992), parametric inverse regression (Bura and Cook, 2001a), iterative Hessian transformation (Cook and Li, 2002), minimum average variance

* Tel.: +886 4 23590121; fax: +886 4 23594710.
E-mail address: hhlue@thu.edu.tw.

0378-3758/$ - see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2008.12.006

estimation (Xia et al., 2002), inverse regression estimator (Cook and Ni, 2005) and contour regression (Li et al., 2005), etc. Most work in this area has focused on reducing the dimension of $\mathbf{X}$ only in a univariate response regression. In the context of inverse regression-based dimension reduction, multivariate response models have received relatively less attention until recently when some approaches have been proposed to deal with such models. Aragon (1997) suggested using the first principal component of a response vector for marginal slicing. Hsing (1999) provided an idea of nearest neighbors for determining the slices. Bura and Cook (2001a) introduced parametric inverse regression that may adapt to multivariate response. Li et al. (2003) developed the most predictable (MP) variates method for dimension reduction on multivariate response data. Setodji and Cook (2004) extended SIR to multivariate response by introducing a way of k-means clusters for slicing. Saracco (2005) considered a multivariate SIR_α version of the pooled marginal slicing estimator, which combines SIR-I and SIR-II approaches (Li, 1991; Li and Duan, 1989), though it lacks asymptotic results from dimension tests. Yin and Bura (2006) proposed a moment-based dimension reduction approach.

In this article we propose an SIR-based approach for multivariate response regression data which aids in finding the multivariate CS. The basic idea is to integrate all univariate response kernel matrices by forming a weighted average. With the chosen weighted matrix, we can follow the regular procedure of SIR and conduct the reduction of dimension for $\mathbf{X}$.

The CS in the context of the multivariate regression of $\mathbf{Y}$ on $\mathbf{X}$ and some SDR methods are briefly presented in Section 2. The theoretical foundation of our method is summarized in Section 3. We derive the asymptotic weighted chi-squared test for dimension in Section 4. In Section 5 one simulation example is used to contrast our method against the MP variates of Li et al. (2003), k-means inverse regression (KIR) of Setodji and Cook (2004) and classical canonical correlation approach (CCA) of Hotelling (1935). Section 6 is a conclusion.

Herein, subspace is denoted by $\mathscr{S}$ and $\mathscr{S}(A)$ represents the subspace of R^p spanned by the columns of matrix A. Let $\mathbf{Z} = \Sigma_{\mathbf{X}}^{-1/2}(\mathbf{X} - E\mathbf{X})$, where $\Sigma_{\mathbf{X}}$ denotes the covariance matrix of $\mathbf{X}$ and is assumed to be positive definite. Throughout the rest of this article, discussion on $\mathbf{Z}$ scale is used without loss of generality because of $\mathscr{S}_{Y_w|\mathbf{X}} = \Sigma_{\mathbf{X}}^{-1/2}\mathscr{S}_{Y_w|\mathbf{Z}}$ for any w.

2. General regression context

2.1. Multivariate CS

In multivariate regression of $\mathbf{Y}$ on $\mathbf{Z}$, the CS for $\mathbf{Y}|\mathbf{Z}$, represented by $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$, is defined as the intersection of all dimension-reduction subspaces $\mathscr{S}$ of R^p with the property $\mathbf{Y} \perp \mathbf{Z}|P_{\mathscr{S}}\mathbf{Z}$, where $P_{\mathscr{S}}$ is the orthogonal projection onto $\mathscr{S}$ in the usual inner product. $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ is an effective population meta-parameter for pursuing SDR in regression. Henceforth, this subspace referred to in this article is assumed to exist.

2.2. Sliced inverse regression

Li (1991) introduced SIR (also call SIR-I) to estimate the EDR space. Li showed that if the linearity condition $E(\mathbf{Z}|P_{\mathscr{S}_{\mathrm{y}|\mathbf{Z}}}\mathbf{Z}) = P_{\mathscr{S}_{\mathrm{y}|\mathbf{Z}}}\mathbf{Z}$ holds, then $E(\mathbf{Z}|\mathrm{y}) \in \mathscr{S}_{\mathrm{y}|\mathbf{Z}}$, or equivalently $\mathscr{S}(\Theta) \subseteq \mathscr{S}_{\mathrm{y}|\mathbf{Z}}$, where $\Theta = \mathrm{var}\{E(\mathbf{Z}|\mathrm{y})\}$ denotes the covariance matrix of the inverse regression curve. If also the coverage condition $\mathrm{y} \perp \mathbf{Z}|P_\Theta\mathbf{Z}$ holds, then $\mathscr{S}(\Theta) = \mathscr{S}_{\mathrm{y}|\mathbf{Z}}$ (Chiaromonte et al., 2002). Li defined SIR estimates as the eigenvectors $v_1, \ldots, v_p$ for eigenvalue decomposition

$$\Theta v_i = \tilde{\lambda}_i v_i$$

for $i = 1, \ldots, p$ and $\tilde{\lambda}_1 \geqslant \cdots \geqslant \tilde{\lambda}_p$ (Li, 1991; Li and Duan, 1989). On the other hand, Li (1991) considered the curve of the conditional covariance, $\mathrm{cov}(\mathbf{Z}|\mathrm{y})$, for determining the EDR space and defined the SIR-II directions as the eigenvectors of eigenvalue decomposition of the matrix $\Sigma_{\mathrm{II}} = E\{[\mathrm{cov}(\mathbf{Z}|\mathrm{y}) - \mathrm{airII}]^2\}$, where $\mathrm{airII} = E[\mathrm{cov}(\mathbf{Z}|\mathrm{y})]$ is the average of the inverse regression second-moment curve. If the predictors are normal, Li (1991) showed that the asymptotic distribution of the test statistic

$$\hat{\tilde{L}}_d = n \sum_{j=d+1}^{p} \hat{\tilde{\lambda}}_j$$

is asymptotically $\chi^2_{(p-\kappa)(\tilde{H}-\kappa-1)}$, where κ is the dimension of the EDR space and $\tilde{H}$ is the number of slices on y (Li, 1991; Bura and Cook, 2001b).

2.3. k-Means inverse regression

Setodji and Cook (2004) developed KIR to estimate the $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$. They took the k-means algorithm to cluster the observed response vectors. KIR uses the SIR algorithm except that the slices are replaced with the k-means clusters. The KIR is applied in the following way: within each cluster, compute the mean vectors as $\bar{\mathbf{Z}}_s = \sum_{\mathbf{Y}_i \in C_s} \mathbf{Z}_i/n_s$, $s = 1, \ldots, c_k$, where n_s is the number of observations in cluster C_s. Construct the sample covariance matrix, $\hat{M} = \sum_{s=1}^{c_k} n_s \bar{\mathbf{Z}}_s \bar{\mathbf{Z}}_s'/n$. Then conduct the eigenvalue decomposition, giving

$$\hat{M}\hat{w}_i = \hat{\theta}_i \hat{w}_i$$

for $i=1,\dots,p$ and $\hat{\theta}_1 \geqslant \cdots \geqslant \hat{\theta}_p$. To infer the dimension κ of $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$, they proposed the test statistic, similar to the one proposed by Li (1991),

$$\hat{\Lambda}_d = n \sum_{j=d+1}^{p} \hat{\theta}_j$$

which is chi-squared asymptotically with $(p-\kappa)(c_k-\kappa-1)$ degrees of freedom under certain conditions (Cook, 1998).

2.4. MP variates

Li et al. (2003) confined to a slightly restrictive model $\mathbf{Y} = g(\beta_1'\mathbf{Z},\dots,\beta_\kappa'\mathbf{Z}) + \varepsilon$, which is quite different from (1), and defined MP estimates to be the eigenvectors by conducting the eigenvalue decomposition

$$\mathrm{var}\{E(\mathbf{Y}|\mathbf{Z})\}v_i = \check{\lambda}_i \mathrm{var}(\mathbf{Y})v_i$$

for $i=1,\dots,q$ and $\check{\lambda}_1 \geqslant \cdots \geqslant \check{\lambda}_q$. By observing the similarity of this decomposition to SIR, they used the SIR algorithm given by Li (1991) to find the MP variates. When both $\mathbf{Y}$ and $\mathbf{Z}$ are high-dimensional, they suggested using the iterative procedure for finding leading directions. The iterative procedure is briefly described as follows: first apply SIR to the first few canonical variates of $\mathbf{Y}$ for finding $\mathbf{Z}$ variates, then use the variates to find MP variates; replace the canonical variates of $\mathbf{Y}$ with the MP variates. Repeat the procedure until there is little change in the results; for more details, see Li et al. (2003, Section 5). To infer the dimension κ of predictors, they used the test statistic

$$\hat{L}_d = n \sum_{j=d+1}^{p} \hat{\check{\lambda}}_j$$

which is asymptotically $\chi^2_{(p-\kappa)(H-\kappa-1)}$, where H is the number of slices on MP variates, if the predictors are normal (Li, 1991).

3. Multivariate response SIR

We develop a new method for reducing the dimension of $\mathbf{Z}$ in CS. The following proposition connects $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ to $\mathscr{S}_{Y_w|\mathbf{Z}}$, for $w=1,\dots,q$. Let ϕ be a basis matrix for $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ and φ_w be a basis matrix for $\mathscr{S}_{Y_w|\mathbf{Z}}$. Let $\mathbf{\Theta}_w = \mathrm{var}\{E(\mathbf{Z}|Y_w)\}$ and $\kappa = \dim(\mathscr{S}_{\mathbf{Y}|\mathbf{Z}})$. Denote $\oplus$ as the direct sum between two subspaces (say, $V_1 \oplus V_2 = \{\mathrm{v}_1+\mathrm{v}_2 : \mathrm{v}_1 \in V_1, \mathrm{v}_2 \in V_2\}$).

Proposition 1. $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}} = \bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$.

Proof. To prove this result, let φ be a basis matrix for $\bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$. It is straightforward to see that $\mathbf{Y} \perp \mathbf{Z}|\phi'\mathbf{Z}$ implies $Y_w \perp \mathbf{Z}|\phi'\mathbf{Z}$, for all w. This in turn implies $\mathscr{S}_{Y_w|\mathbf{Z}} \subseteq \mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ for all w, and therefore $\bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}} \subseteq \mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$. Since $\mathscr{S}_{Y_w|\mathbf{Z}} \subseteq \bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$ for all w, it implies $Y_w \perp \mathbf{Z}|\varphi'\mathbf{Z}$ for all w. Hence $\mathbf{Y} \perp \mathbf{Z}|\varphi'\mathbf{Z}$, which implies $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}} \subseteq \bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$.

Although the subspaces $\mathscr{S}_{Y_w|\mathbf{Z}}$ can overlap in any fashion, $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ always coincides with their direct sum. Proposition 1, which plays a key role in the development of our methodology, suggests that $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$ can be estimated by combining dimension reduction within $\mathscr{S}_{Y_w|\mathbf{Z}}$ for all w. Under the condition of $E(\mathbf{Z}|\varphi_w'\mathbf{Z}) = P_{\varphi_w}\mathbf{Z}$, we obtain $E(\mathbf{Z}|Y_w) \in \mathscr{S}_{Y_w|\mathbf{Z}}$ (Li, 1991), or equivalently $\mathscr{S}(\mathbf{\Theta}_w) \subseteq \mathscr{S}_{Y_w|\mathbf{Z}}$ for $w=1,\dots,q$, and therefore $\bigoplus_{w=1}^{q} \mathscr{S}(\mathbf{\Theta}_w) \subseteq \bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$. This relationship suggests combining coordinate-wise SIR to recover $\bigoplus_{w=1}^{q} \mathscr{S}_{Y_w|\mathbf{Z}}$for inference on $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$.

With the covariance matrix $\mathbf{\Theta}_w$ for $w=1,\dots,q$, we simply take a weighted average $\bar{\mathbf{\Theta}}$ of $\mathbf{\Theta}_w$ for combining,

$$\bar{\mathbf{\Theta}} = \sum_{w=1}^{q} (\bar{\lambda}_w/\bar{\lambda}_\cdot)\mathbf{\Theta}_w \tag{2}$$

where $\bar{\lambda}_w$ is the proportion of eigenvalues corresponding to significant eigenvectors for the eigenvalue decomposition $\mathbf{\Theta}_w b_{iw} = \lambda_{iw} b_{iw}$, for $i=1,\dots,p$ and $\lambda_{1w} \geqslant \cdots \geqslant \lambda_{pw}$, (i.e. $\bar{\lambda}_w = \Lambda_w/\sum_{i=1}^{p}\lambda_{iw}$, where Λ_w is the sum of the first k_w eigenvalues and k_w is the dimension of $\mathscr{S}(\mathbf{\Theta}_w)$) and $\bar{\lambda}_\cdot = \sum_{w=1}^{q}\bar{\lambda}_w$. Because this method applies SIR to analyze multivariate response regression data, we call it mrSIR method. We define the mrSIR estimates to be the eigenvectors $b_1,\dots,b_p$ for eigenvalue decomposition, i.e.

$$\bar{\mathbf{\Theta}} b_i = \rho_i b_i \tag{3}$$

for $i=1,\dots,p$ and $\rho_1 \geqslant \cdots \geqslant \rho_p$.

In summary, the population basis for mrSIR is as follows:

Proposition 2. *If either* $E(\mathbf{Z}|\varphi_w'\mathbf{Z}) = P_{\varphi_w}\mathbf{Z}$ *for* $w = 1, \ldots, q$ *or* $E(\mathbf{Z}|\phi'\mathbf{Z}) = P_\phi\mathbf{Z}$ *holds, then* $\bar{\mathbf{\Theta}}$ *as defined in* (2) *has* $\mathscr{S}(\bar{\mathbf{\Theta}}) \subseteq \bigoplus_{w=1}^q \mathscr{S}_{Y_w|\mathbf{Z}}$. *If also* $Y_w \perp \mathbf{Z}|P_{\bar{\mathbf{\Theta}}}\mathbf{Z}$ *holds, then* $\mathscr{S}(\bar{\mathbf{\Theta}}) = \bigoplus_{w=1}^q \mathscr{S}_{Y_w|\mathbf{Z}}$.

The first condition, which is referred to as the linearity condition, may hold to a reasonable approximation in many problems (Hall and Li, 1993). Since the condition involves only the marginal distribution of predictors, we are free to use coordinate-wise predictor transformations or predictor weighting to induce the condition (Cook and Nachtsheim, 1994).

Suppose that an iid sample $(\mathbf{Z}_i, Y_{iw})$ for $i = 1, \ldots, n$ and $w = 1, \ldots, q$ is available, use mrSIR is applied as follows: divide the sample range of Y_w into h_w slices and compute the intraslice mean vectors as

$$\bar{\mathbf{Z}}_{sw} = \frac{1}{n_{sw}} \sum_{i|s} \mathbf{Z}_i, \quad s = 1, \ldots, h_w$$

where the sum is over indexes i of Y_w that fall into slice s and n_{sw} is the number of observations in slice s. Now construct sample versions of $\mathbf{\Theta}_w$, for $w = 1, \ldots, q$,

$$\hat{\mathbf{\Theta}}_w = \sum_{s=1}^{h_w} \frac{n_{sw}}{n} \bar{\mathbf{Z}}_{sw} \bar{\mathbf{Z}}_{sw}'$$

and a sample version of $\bar{\mathbf{\Theta}}$

$$\hat{\bar{\mathbf{\Theta}}} = \sum_{w=1}^q (\hat{\bar{\lambda}}_w / \hat{\bar{\lambda}}_\cdot) \hat{\mathbf{\Theta}}_w$$

where $\hat{\bar{\lambda}}_w$ is the sample version of $\bar{\lambda}_w$ for eigenvalue decomposition of $\hat{\mathbf{\Theta}}_w$ (i.e. $\hat{\bar{\lambda}}_w = \hat{\Lambda}_w / \sum_{i=1}^p \hat{\lambda}_{iw}$, where $\hat{\Lambda}_w$ is the sum of the first $\hat{k}_w$ estimated eigenvalues and $\hat{k}_w$ is the estimated dimension of $\mathscr{S}(\hat{\mathbf{\Theta}}_w)$ obtained by using the test statistic $\hat{\bar{L}}_d$ in dimension estimation at a significance level of α, say 0.05) and $\hat{\bar{\lambda}}_\cdot = \sum_{w=1}^q \hat{\bar{\lambda}}_w$. Then conduct the eigenvalue decomposition, giving

$$\hat{\bar{\mathbf{\Theta}}} \hat{b}_i = \hat{\rho}_i \hat{b}_i \quad \text{for } i = 1, \ldots, p, \ \hat{\rho}_1 \geqslant \cdots \geqslant \hat{\rho}_p \tag{4}$$

The eigenvalue decomposition (4) is just a sample version of (3).

4. Weighted chi-squared test for rank of $\bar{\mathbf{\Theta}}$

A test statistic of the form suggested by Li (1991)

$$\hat{T}_m = n \sum_{j=m+1}^p \hat{\rho}_j \tag{5}$$

where $\hat{\rho}_1 \geqslant \cdots \geqslant \hat{\rho}_p$ are the eigenvalues of the eigenvalue decomposition in (4), is used iteratively to estimate the rank κ of $\bar{\mathbf{\Theta}}$. The general weighted chi-squared test is derived in this section (also see, Bura and Cook, 2001b; Chiaromonte et al., 2002).

4.1. Development

Denote $\hat{f}_{sw} = \sqrt{n_{sw}/n}$, $\hat{a}_w = \sqrt{\hat{\bar{\lambda}}_w / \hat{\bar{\lambda}}_\cdot}$ and $h = \sum_{w=1}^q h_w$, for $s = 1, \ldots, h_w$, $w = 1, \ldots, q$. Let $\bar{\mathbf{Z}}_{\cdot w} = (\hat{f}_{1w}\bar{\mathbf{Z}}_{1w}, \ldots, \hat{f}_{h_w w}\bar{\mathbf{Z}}_{h_w w})$ be a $p \times h_w$ matrix of weighted slice means and let $\bar{\mathbf{Z}}_{\cdot\cdot} = (\hat{a}_1\bar{\mathbf{Z}}_{\cdot 1}, \ldots, \hat{a}_q\bar{\mathbf{Z}}_{\cdot q})$ be a $p \times h$matrix of weighted matrices. Then the $\bar{\mathbf{Z}}_{\cdot w}$ converges in probability to $C_{\cdot w} = (f_{1w}E(\mathbf{Z}|\tilde{Y}_w = 1), \ldots, f_{h_w w}E(\mathbf{Z}|\tilde{Y}_w = h_w))$, where $\hat{f}_{sw} \to f_{sw} = [\Pr(\tilde{Y}_w = s)]^{1/2}$ and $\tilde{Y}_w$ denotes as a discrete version of Y_w, and the $\bar{\mathbf{Z}}_{\cdot\cdot}$ converges in probability to $C_{\cdot\cdot} = (a_1 C_{\cdot 1}, \ldots, a_q C_{\cdot q})$, where $\hat{a}_w \to a_w = \sqrt{\bar{\lambda}_w / \bar{\lambda}_\cdot}$, with singular value decomposition

$$C_{\cdot\cdot} = \Gamma' \begin{pmatrix} \mathrm{D} & 0 \\ 0 & 0 \end{pmatrix} \Psi$$

where Γ' and Ψ are orthonormal matrices with order $p \times p$ and $h \times h$, and D is a $\kappa \times \kappa$ diagonal matrix of singular values. Partition $\Gamma' = (\Gamma_1, \Gamma_0)$ and $\Psi' = (\Psi_1, \Psi_0)$, where Γ_0 is $p \times (p - \kappa)$ and Ψ_0 is $h \times (h - \kappa)$.

According to Eaton and Tyler (1994), the asymptotic distribution of the smallest $\min(p-\kappa, h-\kappa)$ singular values of $\sqrt{n}\bar{\mathbf{Z}}_{\cdot\cdot}$ is the same as that of the singular values of the $(p-\kappa)\times(h-\kappa)$ matrix $\sqrt{n}U = \sqrt{n}\Gamma_0'(\bar{\mathbf{Z}}_{\cdot\cdot} - C_{\cdot\cdot})\Psi_0 = \sqrt{n}\Gamma_0'\bar{\mathbf{Z}}_{\cdot\cdot}\Psi_0$. Thus, the asymptotic distribution of $\hat{T}_\kappa$ is the same as that of

$$T = n \cdot \mathrm{trace}[\Gamma_0'\bar{\mathbf{Z}}_{\cdot\cdot}\Psi_0(\Gamma_0'\bar{\mathbf{Z}}_{\cdot\cdot}\Psi_0)'] = n \cdot \mathrm{vec}(U)'\mathrm{vec}(U)$$

where $\mathrm{vec}(U)$ is the $(p-\kappa)(h-\kappa)\times 1$ vector constructed by stacking the columns of U. Partition $\Psi_0 = (\Psi_{01}', \ldots, \Psi_{0q}')'$, where Ψ_{0w} has order $h_w \times (h-\kappa)$. Then, because $\Gamma_0' C_{\cdot w} = 0$ for all w, we have

$$\sqrt{n}U = \sum_{w=1}^{q} \hat{a}_w \sqrt{n}\Gamma_0'\bar{\mathbf{Z}}_{\cdot w}\Psi_{0w} \equiv \sum_{w=1}^{q} \sqrt{n}U_w$$

Let $\bar{\mathbf{X}}_{\cdot w} = (\bar{\mathbf{X}}_{1w}, \ldots, \bar{\mathbf{X}}_{h_w w})$ be a $p \times h_w$ matrix of sliced means, $w = 1, \ldots, q$, let f_w be a $h_w \times 1$ vector with elements f_{sw}, $s = 1, \ldots, h_w$, and let $\hat{f}_w$ be the corresponding sample version in terms of $\hat{f}_{sw}$. Denote D_v as a diagonal matrix with elements v, let $P_v = vv'/v'v$ be the projection on the span of v and $Q_v = \mathrm{I} - P_v$. Then, we can derive (Chiaromonte et al., 2002)

$$\sqrt{n}U_w = \hat{a}_w \sqrt{n}\Gamma_0'\Sigma^{-1/2}[\bar{\mathbf{X}}_{\cdot w} - E(\bar{\mathbf{X}}_{\cdot w})]\mathrm{D}_{f_w} Q_{f_w}\Psi_{0w} + \mathrm{o_p}(1)$$

By the central limit theorem and the multivariate version of Slutsky's theorem, as $n \to \infty$, $\sqrt{n}\,\mathrm{vec}[\bar{\mathbf{X}}_{\cdot w} - E(\bar{\mathbf{X}}_{\cdot w})]$ converges in distribution to a normal random vector with mean 0 and $ph_w \times ph_w$ covariance matrix $(\mathrm{D}_{f_w}^{-1} \otimes \mathrm{I}_p)V_w^*(\mathrm{D}_{f_w}^{-1} \otimes \mathrm{I}_p)$, where V_w^* is a $ph_w \times ph_w$ block diagonal matrix with diagonal blocks $\mathrm{cov}(\mathbf{X}|\tilde{Y}_w = s), s = 1, \ldots, h_w$, and $\otimes$ denotes the Kronecker product. It then follows that $\sqrt{n}\,\mathrm{vec}(U_w)$ converges in distribution to a normal random vector with mean 0 and $(p-\kappa)(h-\kappa)\times(p-\kappa)(h-\kappa)$ covariance matrix

$$\Omega_w = a_w^2(\Psi_{0w}'Q_{f_w} \otimes \mathrm{I}_{p-\kappa})V_w(Q_{f_w}\Psi_{0w} \otimes \mathrm{I}_{p-\kappa})$$

where V_w is a $(p-\kappa)h_w \times (p-\kappa)h_w$ block diagonal matrix with diagonal blocks $\Gamma_0'\,\mathrm{cov}(\mathbf{Z}|\tilde{Y}_w = s)\Gamma_0$, for $s = 1, \ldots, h_w$. Moreover, the covariance matrix of $\sqrt{n}\,\mathrm{vec}(U_w)$ and $\sqrt{n}\,\mathrm{vec}(U_{w'})$ for $w \neq w'$ is

$$\Omega_{ww'} = a_w a_{w'}(\Psi_{0w}'Q_{f_w} \otimes \mathrm{I}_{p-\kappa})V_{ww'}(Q_{f_{w'}}\Psi_{0w'} \otimes \mathrm{I}_{p-\kappa})$$

where $V_{ww'}$ is a $(p-\kappa)h_w \times (p-\kappa)h_{w'}$ block matrix with the (s,s')th block $\Gamma_0'\,\mathrm{cov}(\mathbf{Z}|\tilde{Y}_w = s,\ \mathbf{Z}|\tilde{Y}_{w'} = s')\Gamma_0$, for $s = 1, \ldots, h_w$, $s' = 1, \ldots, h_{w'}$, and $\mathbf{Z}|\tilde{Y}_w = s$ denotes predictors of $\tilde{Y}_w$ that fall into slice s. Thus, the limiting distribution of the vector version of $\sqrt{n}\,U$ is

$$\sqrt{n}\,\mathrm{vec}(U) \xrightarrow{\mathscr{D}} N_{(p-\kappa)(h-\kappa)}(0, \Omega)$$

where $\Omega = \sum_{w=1}^{q}\Omega_w + \sum_{w\neq w'}\Omega_{ww'}$. For ease of reference, in the following theorem we summarize the results for inference about κ.

Theorem 1. *Let $\kappa = \mathrm{rank}(\bar{\boldsymbol{\Theta}})$. The asymptotic distribution of $\hat{T}_\kappa$ is the same as the distribution of*

$$\chi = \sum_{j=1}^{(p-\kappa)(h-\kappa)} c_j\chi_j$$

where χ_j's are independent chi-squared random variables, each with 1 degree of freedom, and $c_1 \geqslant \cdots \geqslant c_{(p-\kappa)(h-\kappa)}$ are the eigenvalues of Ω.

This theorem allows for a general method for inferring about dimension of $\mathscr{S}_{E(\mathbf{X}|\tilde{\mathbf{Y}})}$ provided that we can obtain a consistent estimate of Ω. The linearity condition, an essential ingredient for SIR to establish the connection between the subspace being estimated and $\mathscr{S}_{\mathbf{Y}|\mathbf{Z}}$is not required for this theorem. When the number of observations per slice is small, the variation in $\hat{c}_j$ might alleviate the effectiveness of the estimated distribution of $\hat{T}_\kappa$. For this reason, the next corollary gives a simplified result on the distribution of $\hat{T}_\kappa$ under some conditions. The proof is then presented in Appendix.

Corollary 1. *Let $\kappa = \mathrm{rank}(\bar{\boldsymbol{\Theta}})$. If* (a) $Y_w \perp \mathbf{Z}|P_{\bar{\boldsymbol{\Theta}}}\mathbf{Z}$ *for* $w = 1, \ldots, q$, (b) $E(\mathbf{Z}|P_{\bar{\boldsymbol{\Theta}}}\mathbf{Z}) = P_{\bar{\boldsymbol{\Theta}}}\mathbf{Z}$, (c) $\mathrm{var}(\mathbf{Z}|P_{\bar{\boldsymbol{\Theta}}}\mathbf{Z}) = Q_{\bar{\boldsymbol{\Theta}}}$ *and* (d) $\mathrm{cov}(\mathbf{Z}|\tilde{Y}_w = s, \mathbf{Z}|\tilde{Y}_{w'} = s') = 0$ *for all s, s' and $w \neq w'$, then the asymptotic distribution of $\hat{T}_\kappa$ is $\chi^2_{(p-\kappa)(h-\kappa-1)}$.*

Here (a) corresponds to the coverage condition, (b) corresponds to the linearity condition under coverage and (c) corresponds to the constant covariance condition required to derive the asymptotic distribution χ^2. Since the covariance matrix $\boldsymbol{\Theta}_w$ associated with zero weight does not contribute to the asymptotic distribution of $\hat{T}_\kappa$, one may use the sum of slices corresponding to nonzero weights in computing the degree of freedom for testing the rank in practice. The weighted chi-squared test of Theorem 1 for mrSIR is used throughout this article.

4.2. mrSIR algorithm

1. Let $\hat{\Gamma}'$ be the orthonormal $p \times p$ matrix of left singular vectors and $\hat{\Psi}'$ be the orthonormal $h \times h$ matrix of right singular vectors in the singular value decomposition of $\bar{\mathbf{Z}}_{\cdot\cdot}$ (i.e. there exist $\hat{\Gamma}$, $\hat{\mathrm{D}}$ and $\hat{\Psi}$ such that $\bar{\mathbf{Z}}_{\cdot\cdot} = \hat{\Gamma}'\hat{\mathrm{D}}\hat{\Psi}$).
2. For a hypothesized value of κ, let $\hat{\Gamma}_0$ be the matrix of the last $p-\kappa$ columns of $\hat{\Gamma}'$ and Ψ_0 be the matrix of the last $h-\kappa$ columns of $\hat{\Psi}'$.
3. Use the results from step 2 to compute $\hat{\Omega}$ by adding $\sum_{w=1}^{q}\hat{\Omega}_w$ and $\sum_{w\neq w'}\hat{\Omega}_{ww'}$, and then find its eigenvalues $\hat{c}_1 \geqslant \cdots \geqslant \hat{c}_{(p-\kappa)(h-\kappa)}$.
4. Construct the p-value for the weighted chi-squared test as

$$\Pr(\hat{\chi} > \text{observed } \hat{T}_\kappa)$$

where $\hat{\chi} = \sum_{j=1}^{(p-\kappa)(h-\kappa)} \hat{c}_j\chi_j$ and the χ_j's are as described in Theorem 1. A substantial literature on computing tail probabilities of distributions of linear combinations of chi-squared random variables are available (Wood, 1989; Field, 1993).
5. Once the p-value for $\kappa = 0, \ldots, p-1$ are available, inference in κ can proceed as follows: beginning with $j=0$, if p-value $> \alpha$ (i.e. the null hypothesis is not rejected) conclude that $\kappa = j$; if p-value $< \alpha$, conclude that $\kappa > j$, set $j = j+1$ and repeat the procedure (Li, 1991; Bura and Cook, 2001a; Setodji and Cook, 2004).

5. A simulation study

This section contains a simulation-based comparison of power of test for dimension reduction methods. Three sample sizes are used: $n = 100, 200, 400$. For each sample size, the p-values corresponding to the test statistics for selected dimensions were collected over 1000 replications. We use an affine invariant criterion proposed by Li (1991)

$$R^2(b) = \max_{\beta\in\mathscr{B}} \frac{(b'\beta)^2}{b'b \cdot \beta'\beta}$$

where $\mathscr{B}$ is the true dimension-reduction space, to evaluate the performance of estimation. The performance is then compared with CCA, KIR and MP variates. We emphasize the effectiveness of estimated SDR directions. For the k-means algorithm, the number of cluster centers is set to 10. In the iterative procedure of MP variates, we choose the first two canonical variates to reduce the dimension of $\mathbf{Y}$ tentatively and then apply a double slicing approach to the reduced $\mathbf{Y}$ with slices (h_1, h_2) in each coordinate to obtain $\mathbf{Z}$ variates. We repeat the procedure several times on each dataset for more convergent estimation. The number of slices for mrSIR here is the same (i.e. $h_w = c$, $\forall w$). In our simulations, we take $c = 5, 8$ or 10 in order that our sample size n would be sufficient.

We generate an iid example from the following model with

$$\begin{aligned} Y_1 &= z_1 + z_2 + 0.25\varepsilon_1 \exp\{2(1-z_3)\} \\ Y_2 &= z_1 + z_2 - 0.25\varepsilon_2 \exp\{2z_3\} \end{aligned} \tag{6}$$

where the coordinates of $\mathbf{Z}$, $Y_w, w = 3, 4, 5$, and ε_i's are independent standard normal random variables. Set $p = 6$ and $q = 5$. Denote $\beta_1 = (1, 1, 0, 0, 0, 0)'$ and $\beta_2 = (0, 0, 1, 0, 0, 0)'$. The data contain linear and exponential features. With proper guidance from mrSIR, the dimension of $\mathbf{Z}$ can be reduced to just two.

Tables 1–3 summarize the performance of the true dimension estimation for methods according to (6). The numerical entries of the rows of those tables corresponding to the test statistics indexed by 0 are empirical estimates of the power of the corresponding

Table 1
Empirical power and size for mrSIR applied to model (6).

h_w	5	8	10
$n = 100$			
$\hat{T}_0$	0.954 (0.892)	0.131 (0.009)	0 (0)
$\hat{T}_1$	0.821 (0.689)	0.026 (0)	0 (0)
$\hat{T}_2$	0.043 (0.004)	0 (0)	0 (0)
$n = 200$			
$\hat{T}_0$	0.999 (0.999)	0.984 (0.974)	0.938 (0.830)
$\hat{T}_1$	0.998 (0.992)	0.939 (0.854)	0.502 (0.038)
$\hat{T}_2$	0.103 (0.030)	0.009 (0)	0 (0)
$n = 400$			
$\hat{T}_0$	1 (1)	1 (1)	0.997 (0.996)
$\hat{T}_1$	1 (1)	0.998 (0.996)	0.992 (0.988)
$\hat{T}_2$	0.203 (0.101)	0.103 (0.018)	0.015 (0.002)

Table 2
Empirical power and size for MP variates applied to model (6).

(h_1, h_2)	(4, 5)	(5, 5)
$n = 100$		
$\hat{\Lambda}_0$	0.513 (0.326)	0.415 (0.223)
$\hat{\Lambda}_1$	0.070 (0.017)	0.044 (0.010)
$\hat{\Lambda}_2$	0.004 (0.001)	0.003 (0)
$n = 200$		
$\hat{\Lambda}_0$	0.698 (0.594)	0.646 (0.579)
$\hat{\Lambda}_1$	0.201 (0.094)	0.199 (0.088)
$\hat{\Lambda}_2$	0.006 (0.001)	0.015 (0.003)
$n = 400$		
$\hat{\Lambda}_0$	0.820 (0.764)	0.812 (0.741)
$\hat{\Lambda}_1$	0.393 (0.261)	0.420 (0.289)
$\hat{\Lambda}_2$	0.008 (0.001)	0.012 (0.001)

Table 3
Empirical power and size for KIR applied to model (6).

c_k	5	10	15
$n = 100$			
$\hat{L}_0$	0.826 (0.683)	0.909 (0.776)	0.899 (0.724)
$\hat{L}_1$	0.242 (0.093)	0.325 (0.124)	0.295 (0.096)
$\hat{L}_2$	0.008 (0.001)	0.016 (0.003)	0.011 (0.002)
$n = 200$			
$\hat{L}_0$	0.976 (0.954)	0.987 (0.967)	0.988 (0.961)
$\hat{L}_1$	0.524 (0.369)	0.672 (0.471)	0.677 (0.472)
$\hat{L}_2$	0.027 (0.004)	0.025 (0.006)	0.025 (0.002)
$n = 400$			
$\hat{L}_0$	0.979 (0.971)	0.987 (0.972)	0.991 (0.979)
$\hat{L}_1$	0.545 (0.425)	0.697 (0.616)	0.717 (0.646)
$\hat{L}_2$	0.021 (0.004)	0.025 (0.004)	0.018 (0.001)

Table 4
Mean and standard deviation of $R^2(\hat{b}_1)$ and $R^2(\hat{b}_2)$ for model (6).

	mrSIR			CCA	
	$h_w = 5$	8	10		
$n = 100$	0.966 (0.02) 0.927 (0.05)	0.969 (0.02) 0.946 (0.04)	0.967 (0.02) 0.947 (0.04)	0.508 (0.34) 0.665 (0.25)	
$n = 200$	0.986 (0.01) 0.969 (0.02)	0.985 (0.01) 0.976 (0.01)	0.985 (0.01) 0.978 (0.01)	0.539 (0.35) 0.686 (0.25)	
$n = 400$	0.993 (0.005) 0.985 (0.01)	0.993 (0.005) 0.989 (0.008)	0.993 (0.005) 0.990 (0.007)	0.611 (0.35) 0.689 (0.25)	
	KIR			MP variates	
	$c_k = 5$	10	15	$(h_1, h_2) = (4, 5)$	(5, 5)
$n = 100$	0.648 (0.32) 0.691 (0.25)	0.759 (0.25) 0.754 (0.22)	0.768 (0.23) 0.796 (0.19)	0.583 (0.36) 0.522 (0.28)	0.538 (0.35) 0.524 (0.27)
$n = 200$	0.805 (0.26) 0.767 (0.22)	0.844 (0.23) 0.854 (0.16)	0.853 (0.21) 0.878 (0.14)	0.681 (0.35) 0.611 (0.28)	0.665 (0.35) 0.628 (0.27)
$n = 400$	0.753 (0.34) 0.834 (0.17)	0.785 (0.33) 0.900 (0.13)	0.784 (0.32) 0.926 (0.10)	0.811 (0.29) 0.705 (0.27)	0.803 (0.30) 0.736 (0.26)

test. They express the proportion of times that the null hypothesis of dimension 0 is rejected when the nominal significance level is 0.05; the numbers in parentheses are the analogous proportions for a significance level of 0.01. The estimated power entries show that the performance of mrSIR test degenerates as the number of slices increases when the sample size is small, say $n = 100$. Since the degree of freedom $(p - \kappa)(h - \kappa)$ of Theorem 1 clearly depends on h (or $\sum_{w=1}^{q} h_w$), this may mean the null hypothesis of the dimension test is difficult to be rejected when the number of slices h_w is large and the sample size is small. To temper this inflation, we recommend using smaller h_w when n is small or larger sample should be used due to a combination effect of sample

Table 5
Mean for the coefficients in the leading directions for model (6) by mrSIR with $h_w = 5$ and $n = 200$; standard deviation is in round.

$\hat{b}_1$	0.708 (0.05)	0.706 (0.05)	−0.003 (0.09)	0.001 (0.04)	−3e−4 (0.05)	−0.001 (0.05)
$\hat{b}_2$	0.001 (0.09)	0.004 (0.08)	0.999 (0.02)	−0.003 (0.05)	0.001 (0.06)	0.004 (0.06)

Table 6
Empirical power and size along with $R^2(\hat{b}_i)$ for mrSIR applied to model (6) with number of slices $h_w = 5$.

	$n = 100$	$n = 200$	$n = 400$
$\hat{T}_0$	0.769 (0.620)	0.988 (0.981)	1 (1)
$\hat{T}_1$	0.533 (0.262)	0.964 (0.932)	0.998 (0.998)
$\hat{T}_2$	0.012 (0.003)	0.089 (0.006)	0.189 (0.105)
$R^2(\hat{b}_1)$	0.954 (0.02)	0.978 (0.01)	0.990 (0.006)
$R^2(\hat{b}_2)$	0.907 (0.05)	0.954 (0.02)	0.978 (0.01)

size and number of slices. From these tables, we can see that mrSIR performs much better, but KIR and MP variates fail to detect dimension 2 across sample sizes and choices of the number of slices or clusters. Table 4 summarizes the results about $R^2(\hat{b}_i)$ for measuring performance in estimation. Our results are very good, with means ranging from 0.927 to 0.993, which seems mild for the sensitivity to the slicing choice.

To illustrate the application of mrSIR via data visualization, a single run is taken. We conduct an mrSIR analysis with $h_w = 5$ and $n = 200$ for data generated by (6) to reduce the dimension of $\mathbf{Z}$. The p-value sequence (1e−14, 9e−8, 0.74, 0.86, 0.94, 0.98) suggests two $\mathbf{Z}$ variates. Two leading directions are $(0.709, 0.692, 0.084, -0.041, 0.045, 0.079)'$ and $(-0.035, -0.095, 0.982, 0.007, 0.155, 0.028)'$, which are approximately proportional to $0.707\beta_1$ and β_2. Turning to the sampling performance, the results for 1000 datasets according to (6) are summarized in Table 5, which reports statistics related to $\hat{b}_1$ and $\hat{b}_2$. Our estimates are consistently close to the true directions, though all other methods produce considerable bias in estimation. Similarly a result with choice of $p = 8$ and $q = 7$ for mrSIR is shown in Table 6.

6. Conclusion

Most recently developed dimension-reduction methods are concerned with univariate response data. We here propose an adaption of SIR to multivariate response regression for estimating the dimension of predictors. The implementation of mrSIR is as simple as that of ordinary methods, and the approach of slicing for mrSIR is intuitively simple compared to the existing methods. Choosing a parametric function to be fitted or a prior choice of $\mathbf{Y}$ variates are not required for mrSIR. We suggest using mrSIR to sidestep sizable variation in estimation, which results from the multiple slicing procedure when the dimension of $\mathbf{Y}$ is large or the sample size is relatively small. We develop an asymptotic weighted chi-squared test for dimension, which is a simplified chi-squared test under certain conditions. The simulation example was generated to demonstrate the usefulness of mrSIR compared to CCA, KIR and MP variates.

Acknowledgments

We are grateful to the editor and two referees for insight suggestions. We thank Professor Ker-Chau Li for comments and Dr. C.M. Setodji for providing the program of KIR for k-means inverse regression. Lue's research was supported in part by a Grant 92-2118-M-029-005 from the National Science Council of Taiwan.

Appendix

Proof of Corollary 1. Using conditions (a)–(c), we obtain $\Gamma_0' \operatorname{cov}(\mathbf{Z}|\tilde{Y}_w)\Gamma_0 = \mathrm{I}_{p-\kappa}$. Thus, $V_w = \mathrm{I}$ and $\Omega_w = a_w^2(\Psi_{0w}' Q_{f_w} \Psi_{0w} \otimes \mathrm{I}_{p-\kappa})$. Under (d), we simplify

$$\begin{aligned}\Omega = \sum_{w=1}^{q} \Omega_w &= \sum_{w=1}^{q} a_w^2[\mathrm{I}_{(p-\kappa)(h-\kappa)} - (\Psi_{0w}' f_w f_w' \Psi_{0w} \otimes \mathrm{I}_{p-\kappa})] \\ &= \mathrm{I}_{(p-\kappa)(h-\kappa)} - \left(\Psi_0' \sum_{w=1}^{q} a_w^2 F_w F_w' \Psi_0\right) \otimes \mathrm{I}_{p-\kappa}\end{aligned}$$

where $F_w = (0_{h_1}', \ldots, f_w', \ldots, 0_{h_q}')'$ is an $h \times 1$ vector with 0_m denoting as an $m \times 1$ vector of 0's. Because $C_{\cdot} F_w = 0$, we have $F_w \in \operatorname{Span}(\Psi_0)$ and $F_w F_w'$ is a projection onto a subspace of $\operatorname{Span}(\Psi_0)$. It is straightforward to show that Ω is a symmetric idempotent matrix with trace $(p - \kappa)(h - \kappa - 1)$.

References

Aragon, Y., 1997. A gauss implementation of multivariate sliced inverse regression. Comput. Statist. 12, 355–372.
Bura, E., Cook, R.D., 2001a. Estimating the structural dimension of regressions via parametric inverse regression. J. Roy. Statist. Soc. B 63, 393–410.
Bura, E., Cook, R.D., 2001b. Extending sliced inverse regression: the weighted chi-squared test. J. Amer. Statist. Assoc. 96, 996–1003.
Chiaromonte, F., Cook, R.D., Li, B., 2002. Sufficient dimension reduction in regressions with categorical predictors. Ann. Statist. 30, 475–497.
Cook, R.D., 1994. On the interpretation of regression plots. J. Amer. Statist. Assoc. 89, 177–189.
Cook, R.D., 1998. Regression Graphics: Ideas for Studying Regressions through Graphics. Wiley, New York.
Cook, R.D., Li, B., 2002. Dimension reduction for the conditional mean in regression. Ann. Statist. 30, 455–474.
Cook, R.D., Nachtsheim, J.C., 1994. Reweighting to achieve elliptically contoured covariates in regression. J. Amer. Statist. Assoc. 89, 592–599.
Cook, R.D., Ni, L., 2005. Sufficient dimension reduction via inverse regression: a minimum discrepancy approach. J. Amer. Statist. Assoc. 100, 410–428.
Cook, R.D., Weisberg, S., 1991. Discussion of sliced inverse regression. J. Amer. Statist. Assoc. 86, 328–332.
Eaton, M.L., Tyler, D.E., 1994. The asymptotic distribution of singular values with applications to canonical correlations and correspondence analysis. J. Multivariate Anal. 50, 238–264.
Field, C., 1993. Tail areas of linear combinations of chi-squares and noncentral chi-squares. J. Statist. Comput. Simulation 45, 243–248.
Hall, P., Li, K.C., 1993. On almost linearity of low dimensional projections from high dimensional data. Ann. Statist. 21, 867–889.
Hotelling, H., 1935. The most predictable criterion. J. Educ. Psych. 26, 139–142.
Hsing, T., 1999. Nearest-neighbor inverse regression. Ann. Statist. 27, 697–731.
Li, K.C., 1991. Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342.
Li, K.C., 1992. On principal Hessian directions for data visualization and dimension reduction: another application of Stein's lemma. J. Amer. Statist. Assoc. 87, 1025–1039.
Li, K.C., Duan, N., 1989. Regression analysis under link violation. Ann. Statist. 17, 1009–1052.
Li, K.C., Aragon, Y., Shedden, K., Agnan, C.T., 2003. Dimension reduction for multivariate response data. J. Amer. Statist. Assoc. 98, 99–109.
Li, B., Zha, H., Chiaromonte, F., 2005. Contour regression: a general approach to dimension reduction. Ann. Statist. 33, 1580–1616.
Saracco, J., 2005. Asymptotics for pooled marginal slicing estimator based on SIR_z approach. J. Multivariate Anal. 96, 117–135.
Setodji, C.M., Cook, R.D., 2004. K-means inverse regression. Technometrics 46, 421–429.
Wood, A.T.A., 1989. An F approximation to the distribution of a linear combination of chi-squared variables. Comm. Statist. Simulation Comput. 18, 1439–1456.
Xia, Y., Tong, H., Li, W.K., Zhu, L.X., 2002. An adaptive estimation of dimension reduction space. J. Roy. Statist. Soc. B 64, 363–410.
Yin, X., Bura, E., 2006. Moment-based dimension reduction for multivariate response regression. J. Statist. Plann. Inference 136, 3675–3688.