Statistica Sinica **27** (2017), 335-355
doi:http://dx.doi.org/10.5705/ss.202015.0274

SUFFICIENT DIMENSION REDUCTION WITH MIXTURE MULTIVARIATE SKEW-ELLIPTICAL DISTRIBUTIONS

Yu Guan[1], Chuanlong Xie[2] and Lixing Zhu[2]

[1]*Zhejiang Agriculture & Forest University and* [2]*Hong Kong Baptist University*

Abstract: In inverse regression-based methodologies for sufficient dimension reduction, ellipticity (or slightly more generally, the linearity condition) of the predictor vector is a widely used condition, though there is concern over its restrictiveness. In this paper, Stein's Lemma is generalized to the class of mixture multivariate skew-elliptical distributions in different scenarios to identify and estimate the central subspace. Within this class, necessary and sufficient conditions are explored for the simple covariance between the response (or its function) and the predictor vector to identify the central subspace. Further, we provides a way to do adjustments such that the central subspace can still be identifiable when this simple covariance fails to work. Simulations are used to assess the performance of the results and compare with existing methods. A data example is analysed for illustration.

Key words and phrases: Central subspace, mixture multivariate skew-elliptical distributions, Stein's Lemma, sufficient dimension reduction.

1. Introduction

Consider the regression of a response Y on a vector of p predictors $\mathbf{X} = (X_1, \ldots, X_p)^\tau$. Sufficient dimension reduction (SDR) identifies several linear combinations of $\{X_1, \ldots, X_p\}$ to model the regression of $Y|\mathbf{X}$ without losing information. To be specific, for an $p \times K$ matrix $\mathbf{B} = (b_1, \cdots, b_K)$, a conditional independence holds:

$$Y \perp\!\!\!\perp \mathbf{X}|\mathbf{B}^\tau\mathbf{X}, \tag{1.1}$$

where $\perp\!\!\!\perp$ means "independent of". The smallest column spaces $S_{Y|\mathbf{X}}$ spanned by $\mathbf{B}$ is called the central subspace (CS, Cook (1998a)). When the conditional independence is between Y and $E(Y|\mathbf{X})$, $\mathbf{B}^\tau\mathbf{X}$ given, the subspace spanned by $\mathbf{B}$ is called the central mean subspace (CMS, Cook and Li (2002)) which is a subspace of CS. K is called the structural dimension of the central subspace (or the central mean subspace). In SDR, several inverse regression-based methodologies are used for identifying and estimating CMS and CS. First-order methods include least squares (LS, Duan and Li (1991)) and sliced inverse regression (SIR, Li (1991)), second-order methods cover principal Hessian directions (pHd, Li (1992)) and

sliced average variance estimation (SAVE, Cook and Weisberg (1991)). A hybrid of first-order and second-order methods is directional regression (DR, Li and Wang (2007)). Examples of recent improvements are discretization-expectation estimation (DEE, Zhu, Zhu and Wang (2010)) and cumulative slicing estimation (CSE, Zhu, Zhu and Feng (2010)).

Current methods require strong assumptions on the predictor vector $\mathbf{X}$. First-order methods such as SIR require that: the conditional mean $E(\mathbf{X}|\mathbf{B}^\tau\mathbf{X})$ is linear in $\mathbf{B}^\tau\mathbf{X}$, and second-order methods such as SAVE require linearity and the conditional variance $var(\mathbf{X}|\mathbf{B}^\tau\mathbf{X})$ be a constant matrix. It is known that the linearity condition is slightly weaker than the ellipticity of $\mathbf{X}$, and that the constant conditional variance assumption is close to the normality. As Cook and Nachtsheim (1994) pointed out, when the distribution of $\mathbf{X}$ deviates substantially from elliptical symmetry, present methods can produce misleading results. To relax these conditions, Cook and Nachtsheim (1994) suggested a re-weighting approach to achieve elliptically symmetric covariates through trimming off some data. In the framework of SIR and SAVE, Li and Dong (2009) and Dong and Li (2010) proposed the central solution space (CSS) method to relax the linearity condition and/or constant conditional variance condition. They proved that CSS is a subspace of CS, and is equal to CS in some cases where the linearity condition (for SIR) and the constant conditional variance condition (for SAVE) are violated. For some skewed $\mathbf{X}$, CSS can still be identified when SIR or SAVE is used. Yet, it is still unclear what kinds of skew distributions satisfy the conditions for CSS to be contained in CS, even be equal to CS, when the linearity condition or/and the constant conditional variance condition is/are violated. Another relevant reference is Cook and Li (2009). Feng, Wang and Zhu (2014) provided a necessary and sufficient condition for the least squares formulation to identify the single index, involving the inverse regression function. But, to the best of our knowledge, there is no published answer this question.

To attack the problem, we revisit Stein's Lemma. Stein's Lemma and Hessian matrix were applied in pHd for normal $\mathbf{X}$ (Li (1992)). Cook (1998b) extended its use, and LS can also be regarded as an application of Stein's Lemma. The key feature is that Stein's Lemma successfully links the covariance between Y and $\mathbf{X}$ to CS. This makes estimating CS easy. Later, it was shown that the linearity condition and the constant conditional variance condition can, respectively, make LS and pHd feasible, see Yin and Cook (2002), and Zhu and Zhu (2009). In this paper, we explore the use of Stein's Lemma for multivariate skew-elliptical distributions and, more generally, mixture multivariate skew-elliptical distributions. These classes include elliptical distributions and mixture elliptical distributions as special cases. They are important in such fields as Bayesian statistics (Azzalini (1985); O'Hagan and Leonhard (1976)), engineering, environmetrics, economics,

and the biomedical sciences (Genton (2004)). The feasibility of SDR for these distributions is of importance as breakthroughs could make the SDR theory more widely applied. To this end, we investigate a generalized Stein's Lemma. Within these classes of distributions, we provide insights on how much the ellipticity can be violated with the central subspace can still identified. A necessary and sufficient condition is provided in Corollary 2.

We note that pHd (Li (1992)) can only identify CMS though it is based on Stein's Lemma. In the present paper, we see that for any single function $m(Y)$, the generalized Stein's Lemma can only identify one vector in CS. Without using Hessian matrix, the Stein's Lemma-based method can, at most identify one vector in CMS. We suggest a matrix that integrates all matrices according to a class of functions $m(\cdot)$ of Y, each of which can identify one vector. Such classes of functions were discussed in Yin and Cook (2002), Wu and Li (2011). A very brief description of the algorithm is in Section 3.3. There are several proposals that use different families of functions $m(\cdot)$: LS uses the identity function, Zhu and Zhu (2009) uses distribution function of Y, Zhu and Zeng (2006) and Zeng and Zhu (2010) use the characteristic function, and Zhu, Zhu and Wang (2010) use the indicator function. Yin and Li (2011) provide a summary.

This paper is organized as follows. Section 2 discusses mixture multivariate skew-elliptical distribution and their special cases. Section 3 presents the generalized Stein's Lemma for some classes of distributions for recovering CS. Section 4 uses stimulations to compare with existing methods. A data example is analysed as illustration in Section 5. Some concluding remarks are in Section 6. Proofs of theorems and propositions are in the Appendix.

2. Mixture Multivariate Skew-elliptical Distributions

A general class of multivariate skew-elliptical distributions can be written as

$$f_{\mathbf{X}}(\mathbf{x}) = c \cdot f_p(\mathbf{x}) \cdot F_q(t(\mathbf{x})), \tag{2.1}$$

where $f_p(\mathbf{x})$ is the pdf of a p-dimensional elliptical distribution, $F_q(t(\mathbf{x}))$ is the cumulative distribution function (cdf) of $t(\mathbf{x})$, having a q-dimensional elliptical distribution and a symmetric function of $\mathbf{x}$, and c is a positive scalar guaranteeing that $f_{\mathbf{X}}(\mathbf{x})$ is a pdf (Branco and Dey (2001)).

Definition 1 (Multivariate elliptical distribution (ME))**.** A $p-$dimension random vector $\mathbf{X} = (X_1, \ldots, X_p)^\tau$ is said to have multivariate elliptical distribution, $\mathbf{X} \sim ME_p(\mu, \Sigma, g^{(p)})$, if it is continuous with pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \psi_p(\mathbf{x}; \mu, \Sigma, g^{(p)}) = |\Sigma|^{-1/2} g^{(p)}\{Q(\mathbf{x})\}, \qquad x \in \Omega \subseteq R^p,$$

where $Q(\mathbf{x}) = (\mathbf{x} - \mu)^\tau \Sigma^{-1} (\mathbf{x} - \mu)$ and Ω is support set of $f_{\mathbf{X}}(\mathbf{x})$. The density generator function $g^{(p)}(u)$, $u \geq 0$, satisfies

$$\int_0^\infty u^{p/2-1} g^{(p)}(u)du = \frac{\Gamma(p/2)}{\pi^{p/2}}.$$

Multivariate elliptical distributions include the multivariate normal, the multivariate t, and Pearson type II distributions as special cases.

Definition 2 (Mixture multivariate elliptical distribution (MME))**.** If $\mathbf{X}^{(j)} = (X_1^{(j)}, \ldots, X_p^{(j)})^\tau$ is $ME_p(\mu_j, \Sigma_j, g_j^{(p)})$ with density function $\psi_p(\mathbf{x}^{(j)}; \mu_j, \Sigma_j, g_j^{(p)})$ $(j = 1, \ldots, m)$. $\mathbf{X}$ follows the mixture multivariate elliptical distribution if its probability density function (pdf) is

$$f_{\mathbf{X}}(\mathbf{x}) \triangleq \sum_{j=1}^m w_j \psi_p(\mathbf{x}; \mu_j, \Sigma_j, g_j^{(p)}) = \sum_{j=1}^m w_j |\Sigma_j|^{-1/2} g_j^{(p)}\{Q_j(\mathbf{x})\}, \quad x \in \Omega \subseteq R^p,$$

where $Q_j(\mathbf{x}) = (\mathbf{x}-\mu_j)^\tau \Sigma_j^{-1}(\mathbf{x}-\mu_j)$, and weights $w_j \geq 0$ are such that $\Sigma_{j=1}^m w_j = 1$. In particularly, if $\mathbf{X}^{(j)}$ is $N_p(\mu_j, \Sigma_j)$ $(j = 1, \ldots, m)$, then the pdf of a mixture multivariate normal is

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^m w_j |\Sigma_j|^{-1/2} (2\pi)^{-p/2} \exp\{-\frac{1}{2} Q_j(\mathbf{x})\}, \qquad x \in R^p.$$

Remark 1. For $m \geq 2$, an MME distribution is not an ME distribution. For example, it is easy to see that the pdf of a mixture multivariate normal is multimodal, whereas the pdf of a multivariate normal is unimodal.

Definition 3 (Mixture multivariate skew-elliptical distribution (MMSE))**.** Suppose $\mathbf{X}^{(j)} = (X_1^{(j)}, \ldots, X_p^{(j)})^\tau$ is $MSE_{p,q}(\mu_j, \Sigma_j, C_j, g_j^{(p+q)}, \nu_j, D_j)$ $(j = 1, \ldots, m)$ with the density

$$f_{p,q}^{(j)}(\mathbf{x}^{(j)}; MSE) = \psi_p(\mathbf{x}^{(j)}; \mu_j, \Sigma_j, g_j^{(p)}) \frac{\Psi_q(C_j^\tau \mathbf{x}^{(j)} + \nu_j; D_j, g^{(q)}_{Q_j(\mathbf{x}^{(j)})})}{\Psi_q(\nu_j; D_j + C_j^\tau \bar{\Sigma}_j C_j, g_j^{(q)})},$$
$$\mathbf{x}^{(j)} \in \Omega_j \subseteq R^p,$$

where $Q_j(\mathbf{x}^{(j)}) = (\mathbf{x}^{(j)} - \mu_j)^\tau \Sigma_j^{-1}(\mathbf{x}^{(j)} - \mu_j)$ and $\psi_p(\mathbf{x}^{(j)}; \mu_j, \Sigma_j, g_j^{(p)}) = |\Sigma_j|^{-1/2}$ $\times g_j^{(p)}\{Q_j(\mathbf{x}^{(j)})\}$, the correlation matrix $\bar{\Sigma}_j = \sigma_j^{-1}\Sigma_j\sigma_j^{-1}$, σ_j is a diagonal matrix formed by the standard deviations of Σ_j, and Ω_j is the support set of $f_{p,q}^{(j)}(\mathbf{x}^{(j)}; MSE)$. The function $\Psi_q(z; D, g^{(q)}_{Q(\mathbf{x})})$ denotes the q-dimensional centered elliptical cumulative distribution with $q \times q$ dispersion matrix D and density generator $g^{(q)}$, and $g^{(q)}_{Q(\mathbf{x})} = g^{(p+q)}\{z + Q(\mathbf{x})\}/\mathbf{g^{(p)}}\{\mathbf{Q}(\mathbf{x})\}$. A mixture MSE (MMSE) vector $\mathbf{X}$ has the pdf:

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^m w_j \cdot f_{p,q}^{(j)}(\mathbf{x}; MSE), \quad w_j \geq 0, \ \sum_{j=1}^m w_j = 1.$$

We list some special cases.

(1) When $m = 1$, and $C = 0$, $\mathbf{X}$ has a multivariate elliptical distribution $ME_p(\mu, \Sigma, g^{(p)})$.

(2) When $m = 1$, MMSE reduces to a multivariate skew-elliptical distribution (MSE), see Arellano-Valle and Genton (2010a). The mean vector and covariance matrix of $MSE_{p,q}(\mu, \Sigma, C, g^{(p+q)}, \nu, D)$ are usually not equal to μ and Σ in the corresponding ME distribution, unless $C = 0$.

(3) When $m = 1$, a simple multivariate skew normal (MSN) distribution was introduced by Azzalini and Dalla-Valle (1996), where X follows $SN_p(\mu, \Sigma, C)$ with the pdf

$$f_{\mathbf{X}}(\mathbf{x}, SN) = 2\phi_p(\mathbf{x}; \mu, \Sigma)\Phi(C^\tau(\mathbf{x} - \mu)), \qquad \mathbf{x} \in R^p,$$

$\Phi(t)$ being the cdf of a standard normal. The vector C controls the shape and the special case $C = 0$ corresponds to $N_p(\mu, \Sigma)$.

(4) $\mathbf{X}^{(j)} = (X_1^{(j)}, \ldots, X_p^{(j)})^\tau$ is multivariate skew normal $MSN_{p,q}(\mu_j, \Sigma_j, C_j, \nu_j, D_j)$ for $j = 1, \ldots, m$ with the pdf

$$f_{p,q}^j(\mathbf{x}^{(j)}; MSN) = \phi_p(\mathbf{x}^{(j)}; \mu_j, \Sigma_j)\ \frac{\Phi_q(C_j^\tau(\mathbf{x} - \mu_j) + \nu_j; D_j)}{\Phi_q(\nu_j; D_j + C_j^\tau \bar{\Sigma}_j^{-1} C_j)}, \quad \mathbf{x}^{(j)} \in R^p,$$

where $\phi_p(\mathbf{x}^{(j)}; \mu_j, \Sigma_j)$ is the pdf of $N_p(\mu_j, \Sigma_j)$ and $\Phi_q(C_j^\tau(\mathbf{x} - \mu_j) + \nu_j; D_j)$ is the cdf value of the normal distribution $N_q(C_j^\tau(\mathbf{x} - \mu_j) + \nu_j, D_j)$ at point $C_j^\tau(\mathbf{x}^{(j)} - \mu_j) + \nu_j$. Then vector $\mathbf{X}$ is mixture multivariate skew normal (MMSN) with the pdf

$$f_{\mathbf{X}}(\mathbf{x}) = \sum_{j=1}^{m} w_j \cdot f_{p,q}^j(\mathbf{x}^{(j)}; MSN), \qquad w_j \geq 0,\ \sum_{j=1}^{m} w_j = 1.$$

There are other special case. See Arellano-Valle and Genton (2010a,b) for recent developments in this area.

3. Generalized Stein's Lemma

3.1. A brief review of Stein's Lemma

We start with a brief description. Suppose $\mathbf{X}$ and $\mathbf{U}$ are jointly normally, $h(\cdot)$ is a differentiable function satisfying $E\{|[H(\mathbf{X}) - E(H(\mathbf{X}))][\mathbf{U} - E(\mathbf{U})]|\} < \infty$ and $E(|\partial H(\mathbf{X})/\partial \mathbf{X}|) < \infty$. Then Stein's Lemma has

$$Cov(H(\mathbf{X}), \mathbf{U}) = Cov(\mathbf{X}, \mathbf{U}) E\Big[\frac{\partial H(\mathbf{X})}{\partial \mathbf{X}}\Big], \tag{3.1}$$

where the gradient operator $\frac{\partial}{\partial \mathbf{X}} = (\frac{\partial}{\partial X_1}, \ldots, \frac{\partial}{\partial X_p})^\tau$, $\mathbf{X} = (X_1, \ldots, X_p)^\tau$, and the superscript "$\tau$" is the transpose operator. When $\mathbf{U} = \mathbf{X}$, (3.1) reduces to

$$E\{H(\mathbf{X})[\mathbf{X} - E(\mathbf{X})]\} = Cov(\mathbf{X})E\Big[\frac{\partial H(\mathbf{X})}{\partial \mathbf{X}}\Big] = Cov(\mathbf{X})E\Big[\frac{\partial H(\mathbf{X})}{\partial \mathbf{X}}\Big], \quad (3.2)$$

When we use a transformation $m(\cdot)$ of Y, and $H(\mathbf{X}) = E(m(Y)|\mathbf{X}) = h(\mathbf{B}^\tau \mathbf{X})$,

$$E\{m(Y)[\mathbf{X} - E(\mathbf{X})]\} = Cov(\mathbf{X})\mathbf{B}E\Big[\frac{\partial h(\mathbf{B}^\tau \mathbf{X})}{\partial \mathbf{B}^\tau \mathbf{X}}\Big]. \quad (3.3)$$

A similar relationship holds when the underlying distribution of $\mathbf{X}$ is elliptically symmetric, see for example Yin and Cook (2002) and Zhu and Zhu (2009). Therefore, $Cov(\mathbf{X})^{-1}E\{m(Y)[\mathbf{X} - E(\mathbf{X})]\}$ can be used to identify one base vector in CS. Estimating this vector is simple and computationally efficient. We generalize this to mixture multivariate skew-elliptical distributions, and discuss several special scenarios.

3.2. Generalized Stein's Lemma for mixture multivariate skew-elliptical distributions

We generalize Stein's Lemma to handle mixture multivariate skew-elliptical distributions (MMSE). Stein's Lemma for multivariate elliptical distribution was derived by Landsman and Neslehováb (2008).

Theorem 1 (Stein's Lemma for MMSE)**.** *Let* $\mathbf{X}$ *be MMSE,* $G_j^{(p)}(u) = \int_u^{+\infty}(1/2)$ $g_j^{(p)}(t)dt$ *for* $u \in (0, \infty)$ *and* $j = 1, \ldots, m$*. For given differentiable functions* $H(\cdot)$ *and* $m(\cdot)$*, take* $H(\mathbf{X}) = E[m(Y)|\mathbf{X}] \triangleq h(\mathbf{B}^\tau \mathbf{X})$*. If*

$$G_j^{(p)}\{Q_j(x_i^{(j)})\} \cdot H(x_i^{(j)})|_{x_i^{(j)} \in \partial\Omega} = 0,$$
$$\Omega = \bigcup_{j=1}^{m} \Omega_j \subseteq R^p, \ i = 1, \ldots, p, \ j = 1, \ldots, m, \quad (3.4)$$

then

$$\begin{aligned}
&\sum_{j=1}^{m} w_j E_{\mathbf{X}^{(j)}}[\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}} \frac{\partial H(\mathbf{X}^{(j)})}{\partial \mathbf{X}^{(j)}}] \\
&= \sum_{j=1}^{m} w_j \Sigma_j \mathbf{B} E_{\mathbf{X}^{(j)}}[\frac{\partial h(\mathbf{B}^\tau \mathbf{X}^{(j)})}{\partial \mathbf{B}^\tau \mathbf{X}^{(j)}} \frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}] \\
&= E_{\mathbf{X}}[m(Y)(\mathbf{X} - \mu^*)] \\
&\quad - \sum_{j=1}^{m} w_j \Sigma_j E_{\mathbf{X}^{(j)}}\Big[m(Y)\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}} \frac{\partial\{\ln \Psi_q(C_j^\tau \mathbf{X}^{(j)} + \nu_j; D_j, g_{Q_j(\mathbf{X}^{(j)})}^{(q)})\}}{\partial \mathbf{X}^{(j)}}\Big],
\end{aligned} \quad (3.5)$$

where $\mu^* = \sum_{j=1}^m E_{\mathbf{X}^{(j)}}[H(\mathbf{X}^{(j)})]/E_{\mathbf{X}}[H(\mathbf{X})]w_j\mu_j = \sum_{j=1}^m w_j\mu_j$. *When* $\mathbf{X}^{(j)} \sim SN_{p,q}(\mu_j, \Sigma_j, C_j, \nu_j, D_j)$ *for* $j = 1, \ldots, m$, *we have*

$$\sum_{j=1}^m w_j\Sigma_j\mathbf{B}E_{\mathbf{X}^{(j)}}\left[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial\mathbf{B}^\tau\mathbf{X}^{(j)}}\right] \tag{3.6}$$
$$= E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu^*)] - \sum_{j=1}^m w_j\Sigma_j E_{\mathbf{X}^{(j)}}\left[m(Y)\frac{\partial\{\ln\Phi_q(C_j^\tau\mathbf{X}^{(j)}+\nu_j;0,D_j)\}}{\partial\mathbf{X}^{(j)}}\right].$$

Here $\mu^* = E(\mathbf{X})$ when the distribution is elliptical or mixture elliptical. From this result, we have several formulas under special cases.

Corollary 1. (1) *If* $\mathbf{X}^{(j)} \sim SN_{p,q}(\mu_j, \Sigma_j, C_j, \nu_j, D_j)$ $(j = 1, \ldots, m)$, *then* $G^{(p)}\{Q(\mathbf{X}^{(j)})\}/g^{(p)}\{Q(\mathbf{X}^{(j)})\} = 1$ *and*

$$E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu^*)] = \sum_{j=1}^m w_j\Sigma_j\mathbf{B}E_{\mathbf{X}^{(j)}}\left[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial\mathbf{B}^\tau\mathbf{X}^{(j)}}\right]$$
$$+\sum_{j=1}^m w_j\Sigma_j E_{\mathbf{X}^{(j)}}\left[m(Y)\frac{\partial\{\ln\Phi_q(C_j^\tau\mathbf{X}^{(j)}+\nu_j;0,D_j)\}}{\partial\mathbf{X}^{(j)}}\right]; \tag{3.7}$$

(2) *if* $\mathbf{X}^{(j)} \sim ME_p(\mu_j, \Sigma_j, g_j^{(p)})$ $(j = 1, \ldots, m)$, *then*

$$E_{\mathbf{X}}[m(Y)(\mathbf{X}-E(\mathbf{X}))] = \sum_{j=1}^m w_j\Sigma_j\mathbf{B}E_{\mathbf{X}^{(j)}}\left[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial\mathbf{B}^\tau\mathbf{X}^{(j)}}\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\right]; \tag{3.8}$$

(3) *if* $\mathbf{X}^{(j)} \sim N_p(\mu_j, \Sigma_j)$ $(j = 1, \ldots, m)$, *then*

$$E_{\mathbf{X}}[m(Y)(\mathbf{X}-E(\mathbf{X}))] = \sum_{j=1}^m w_j\Sigma_j\mathbf{B}E_{\mathbf{X}^{(j)}}\Big[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial\mathbf{B}^\tau\mathbf{X}^{(j)}}\Big]. \tag{3.9}$$

If all $\Sigma_j = c_j\Sigma$ *for constants* c_j *and a matrix* Σ, (3.9) *reduces to*

$$\Sigma^{-1}E_{\mathbf{X}}[m(Y)(\mathbf{X}-E(\mathbf{X}))] = \mathbf{B}\sum_{j=1}^m w_jc_jE_{\mathbf{X}^{(j)}}\Big[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial\mathbf{B}^\tau\mathbf{X}^{(j)}}\Big]. \tag{3.10}$$

(4) *For a MSE distribution* (*with* $m = 1$), *we have*

$$\Sigma^{-1}E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu)]$$
$$= \mathbf{B}\cdot E_{\mathbf{X}}\left[\frac{G^{(p)}\{Q(x)\}}{g^{(p)}\{Q(x)\}}\frac{\partial h(\mathbf{B}^\tau\mathbf{X})}{\partial\mathbf{B}^\tau\mathbf{X}}\right]$$

$$+E_{\mathbf{X}}\left[m(Y)\frac{G^{(p)}\{Q(x)\}}{g^{(p)}\{Q(x)\}}\frac{\partial\{\ln\Psi_q(C^\tau\mathbf{X}+\nu;D,g^{(q)}_{Q(\mathbf{X})})\}}{\partial\mathbf{X}}\right]. \tag{3.11}$$

Particularly, for the ME distibution $ME_p(\mu,\Sigma,g^{(p)})$*, we have*

$$\Sigma^{-1}E_{\mathbf{X}}[m(Y)(\mathbf{X}-E(\mathbf{X}))]=\mathbf{B}\cdot E_{\mathbf{X}}\left[\frac{G^{(p)}\{Q(x)\}}{g^{(p)}\{Q(x)\}}\frac{\partial h(\mathbf{B}^\tau\mathbf{X})}{\partial\mathbf{B}^\tau\mathbf{X}}\right]. \tag{3.12}$$

Corollary 2. *From Case* (4) *of Corollary* 1*, for a MSE distribution,* $\Sigma^{-1}E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu)]$ *can identify a vector in CS if and only if either* $C=0$ *or* $C=\mathbf{B}\times O$ *for a* $K\times K$ *matrix* O*. From Theorem* 1*, for a MMSE distribution, there exists a positive definite matrix* Σ *such that* $\Sigma^{-1}E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu^*)]$ *identifies a vector in CS if and only if* $\Sigma_j=c_j\Sigma$ *for constants* c_j *and a matrix* Σ*,or* C_i *are equal to either* 0 *or* $\mathbf{B}\times O_j$ *where* O_j*'s are* $K\times K$ *matrices.*

Remark 2. For a MSE distribution, that the distribution is either elliptical or skewed towards the directions CS contains. For a MMSE distribution, Corollary 2 has all the variance matrices Σ_j's in the elliptical components proportional, and the distribution is either mixture elliptical or skewed toward the directions CS contains.

Remark 3. When the necessary and sufficient conditions are not satisfied, we use (3.5) to identify the CS. For this purpose, the following terms must be estimable: μ^* and

$$\sum_{j=1}^{m}w_j\Sigma_jE_{\mathbf{X}^{(j)}}\left[m(Y)\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\frac{\partial\{\ln\Psi_q(C_j^\tau\mathbf{X}^{(j)}+\nu_j;D_j,g^{(q)}_{Q_j(\mathbf{X}^{(j)})})\}}{\partial\mathbf{X}^{(j)}}\right].$$

Write the right side of (3.5) as $M(\alpha)$, where α records all the unknown parameters, except for C_j if it is equal to $\mathbf{B}O_j$. When μ^* and α are estimable, one vector in CS is estimable. We discuss in Subsection 3.3 the cases where this is realistic.

3.3. Exhaustive Identification of $S_{Y|\mathbf{X}}$

Although the generalized Stein's Lemma can only identify a base vector in $S_{Y|\mathbf{X}}$ for any function $m(\cdot)$, it still can be used to identify the central subspace $S_{Y|\mathbf{X}}$ by the following integrated approach.

By Yin and Li (2011), we can use a class of functions such as the characteristic function family $\Im=\{e^{ity}:t\in R\}$ or the indicator function family $\Im=\{I_{Y\le t}:t\in R\}$ (see Zhu, Zhu and Wang (2010)) to identify CS. Under certain regularity conditions, an integral of matrices can exhaustively identify CS. Write $m(\cdot,t)$ as a

function in the family for any t, and the corresponding $h(\mathbf{X}, t) = E(m(Y, t)|\mathbf{X})$. We use the result in Case (4) of Corollary 1 for the MSE distribution as an example.

Theorem 2. *Write* $z(t) = E_{\mathbf{X}}[G^{(p)}\{Q(\mathbf{X})\}/g^{(p)}\{Q(\mathbf{X})\}\ \partial h(\mathbf{B}^\tau\mathbf{X}, t)/\partial\mathbf{B}^\tau\mathbf{X}]$ *and* $Z = \int z(t)z(t)^\tau d\ F_Y(t)$. *Write the right side of (3.5) as* $M(t, \alpha)$ *for any function* $m(Y, t)$, *where* α *records all the unknown parameters. Then the subspace* $span(M(\alpha))$, *spanned by* $M(\alpha) = \int M(t, \alpha)M(t, \alpha)^\tau d\ F_Y(t)$, *lies in the central subspace* $S_{Y|\mathbf{X}}$, *where* F_Y *is the marginal distribution of* Y. *If* Z *is non-singular,* $span(M(\alpha)) = S_{Y|\mathbf{X}}$.

3.4. Estimation and implementation

Consider the multivariate skew-elliptical distribution. When the conditions in Corollary 2 are satisfied, Theorem 2 shows that CS can be identified through the integrated matrix M by all $\Sigma^{-1}E_{\mathbf{X}}[m(Y, t)(\mathbf{X} - \mu)]$ for all t. Its sample version can serve as an estimate.

Estimation of $M(\alpha)$**.** When the density and distribution functions are known up to some unknown parameter α, this matrix and then the base vectors in $span(M(\alpha))$ can be estimated by simply replacing the distribution of Y by its empirical distribution to get $M_n(\hat{\alpha}) = 1/n\sum_{j=1}^n M_n(Y_j, \hat{\alpha})M_n(Y_j, \hat{\alpha})^\tau$ where $M_n(t, \hat{\alpha})$ is the sample average of $M(t, \alpha)$ with an estimate $\hat{\alpha}$ of α. For parametric distributions, $\hat{\alpha}$ can often have root-n consistency. In general, for MMSE distributions with given density and distribution functions, the parameter α can be estimated by the expectation-maximization (EM) algorithm, see Dempster, Laird and Rubin (1977), Lin (2009) and Cabral, Lachos and Prates (2012) for detail.

We give two examples of multivariate skew-elliptical distributions in which the term $v(\mathbf{X}) = [G^{(p)}\{Q(\mathbf{X})\}/g^{(p)}\{Q(\mathbf{X})\}]\ [\partial\{\ln\Psi_q(C^\tau\mathbf{X} + \nu; D, g^{(q)}_{Q(x)})\}/\partial\mathbf{X}]$ can be estimated by the method of moments, without the help from the EM algorithm. The root-n consistency is then easy to derive.

Example 1. $\mathbf{X}$ is multivariate skew-normal, $SN_p(\Sigma, C)$, with the density

$$f_{\mathbf{X}}(\mathbf{x}) = 2\phi_p(\mathbf{x}; \Sigma)\Phi(C^\tau\mathbf{x}), \qquad (\mathbf{x} \in R^p). \tag{3.13}$$

The mean vector and variance matrix of $\mathbf{X}$ are

$$E(\mathbf{X}) = \sqrt{\frac{2}{\pi}}\ [1 + C^\tau\Sigma C]^{-1/2}\Sigma C,\ Var(\mathbf{X}) = \Sigma - E(\mathbf{X})E^\tau(\mathbf{X}),$$

and it is easy to compute that $v(\mathbf{X}) = \Sigma^{-1}\mathbf{X} + [\phi(C^\tau\mathbf{X})/\Phi(C^\tau\mathbf{X})]\ C$. With $\bar{\mathbf{X}}$ and $\overline{\mathbf{X}\mathbf{X}^\tau}$ the sample mean vector and second order moment matrix of $\mathbf{X}$, the moment estimates of Σ and C are:

$$\hat{\Sigma} = \overline{\mathbf{X}\mathbf{X}^\tau}, \qquad \hat{C} = \sqrt{\frac{\pi}{2 - \pi\bar{\mathbf{X}}^\tau\hat{\Sigma}^{-1}\bar{\mathbf{X}}}}\ \hat{\Sigma}^{-1}\bar{\mathbf{X}}.$$

Then $v(\mathbf{X})$ can be estimated by

$$\hat{v}(\mathbf{X}) = \hat{\Sigma}^{-1}\mathbf{X} + \frac{\phi(\hat{C}^\tau\mathbf{X})}{\Phi(\hat{C}^\tau\mathbf{X})}\ \hat{C}.$$

Example 2. $\mathbf{X}$ is multivariate skew-t distribution, $St_p(\Sigma, C, \gamma)$, with the density:

$$f_{\mathbf{X}}(\mathbf{x}) = 2t_p(;\Sigma,\gamma)T\bigg(C^\tau\Big[\frac{\gamma+p}{\mathbf{x}^\tau\Sigma^{-1}+p}\Big]^{1/2};\gamma+p\bigg), \qquad (\mathbf{x}\in R^p), \tag{3.14}$$

where $t(\cdot)$ is a t density and $T(\cdot)$ is a cumulative t distribution function, γ is the degrees of freedom, and the $p\times p$ matrix Σ has 1 as diagonal elements. Then

$$E(\mathbf{X}) = b_\gamma C,\ Var(\mathbf{X}) = \frac{\gamma}{\gamma-2}\Sigma - E(\mathbf{X})E^\tau(\mathbf{X}),$$

where $b_\gamma = (\gamma/\pi)^{1/2}[\Gamma((\gamma-1)/2)/\Gamma(\gamma/2)]$ $(\gamma>1)$. The matrix $[\gamma/(\gamma-2)]\Sigma$ has trace $p[\gamma/(\gamma-2)]$. Let $\bar{\mathbf{X}}$ and $\overline{\mathbf{X}\mathbf{X}^\tau}$ be the sample mean vector and second order moment matrix of $\mathbf{X}$. Then an estimate of γ is

$$\hat{\gamma} = round\Big(\frac{2}{1-p/[trace(\overline{\mathbf{X}\mathbf{X}^\tau})]}\Big),$$

where $round(c)$ rounds the element c to the nearest integer, and

$$\hat{\Sigma} = (1-\frac{2}{\hat{\gamma}})\overline{\mathbf{X}\mathbf{X}^\tau}, \qquad \hat{C} = \Big(\frac{\pi}{\hat{\gamma}}\Big)^{1/2}\frac{\Gamma(\hat{\gamma}/2)}{\Gamma((\hat{\gamma}-1)/2)}\bar{\mathbf{X}}.$$

Eventually,

$$v(\mathbf{X}) = \Sigma^{-1}\mathbf{X} + \frac{\gamma+Q(\mathbf{X})}{\gamma+p-2}\Big(\frac{\gamma+p}{Q(\mathbf{X})+p}\Big)^{1/2}\frac{t(C^\tau\mathbf{X}((\gamma+p)/(Q(\mathbf{X})+p))^{1/2};\gamma+p)}{T(C^\tau y((\gamma+p)/(Q(y)+p))^{1/2};\gamma+p)}C$$

is estimable.

3.5. Determination of structural dimension

A target matrix $M(\alpha)$ has been constructed and estimated. Based on existing methodologies, such as a criterion of BIC type (see, e.g., Zhu, Miao and Peng (2006)) that work through determining the non-zero eigenvalues of the estimated $M(\alpha)$, we can determine the structural dimension K. We omit these details here.

4. Simulation Studies

In this section, the generalized Stein's Lemma-based method, StI, that uses the indicator function family $\{I_{Y\leq t} : t \in R\}$ identifies CS, as discussed in Section 3.3. Several simulation studies were conducted to assess the performance of StI, with methods such as SIR, DEE-SIR, SAVE, DR, and pHd used for comparison. Li and Dong (2009) and Dong and Li (2010) showed that under some cases, SIR and SAVE could be still applicable when the linearity condition or/and constant conditional variance condition is violated. Therefore, another purpose of the simulations was to explore their robustness to the ellipticity violation.

The paper mainly concerns the application of the generalized Stein's Lemma, so here the structural dimension is assumed to be given. To measure estimation accuracy, we adopted the distance criterion proposed by Ferré (1998). The distance between spaces spanned by $\mathbf{A}$ and $\mathbf{B}$ is defined as $D(\mathbf{A}, \mathbf{B}) = tr[(\mathbf{A}(\mathbf{A}^\tau\mathbf{A})^{-1}\mathbf{A}^\tau)(\mathbf{B}(\mathbf{B}^\tau\mathbf{B})^{-1}\mathbf{B}^\tau)/K]$, with values in $[0, 1]$. The larger the $D(\mathbf{A}, \mathbf{B})$ value is, the better the similarity between $S(\mathbf{A})$ and $S(\mathbf{B})$).

Of the several proposals of distributions that extend the normal by introducing skewness, we chose distribution with density

$$SN_p(\mu, \Sigma, C) = 2\phi_p(\mu, \Sigma)\Phi(C^\tau(y - \mu)),$$

where μ is a $p \times 1$ location vector, Σ is a $p \times p$ positive definite dispersion matrix and C is a $p \times 1$ skewness parameter vector. Let $\lambda = \Sigma^{1/2}C$. We generated predictors from:

$$\begin{pmatrix} X \\ Z \end{pmatrix} \sim N_{p+1}\left(\begin{pmatrix} \mu \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma^{1/2}\delta \\ \delta^\tau\Sigma^{1/2} & 1 \end{pmatrix}\right)$$

with $X \triangleq (X|Z > 0)$ and $\delta = \lambda/(1 + \lambda^\tau\lambda)$. For a mixture multivariate skew normal distribution, we used the skew-normal distribution for mixture components and adopted the EM-type algorithm for maximum likelihood estimation as proposed by Cabral, Lachos and Prates (2012).

Example 3 (Single-index models)**.** Consider the models:

$$Y = \beta^\tau\mathbf{X} + \varepsilon, \tag{4.1}$$

$$Y = \exp\Big(\frac{\beta^\tau\mathbf{X}}{2} + 2 * \varepsilon\Big), \tag{4.2}$$

$$Y = \sin\Big(\frac{\beta^\tau\mathbf{X}}{3} + \varepsilon\Big), \tag{4.3}$$

where $\beta = (1, 1, 1, 1, 0, \ldots, 0)^\tau$, and $\varepsilon \sim N(0, 1)$. Let $p = 10$ and the number of slices $h = 10$ for slicing estimation in SIR, SAVE, and DR. We took $\mathbf{X}$ as follows.

(1.1) $\mathbf{X} \sim N_p(0, I_p)$.

Table 1. The means and standard deviations (in parentheses)of $100 \cdot D(A, B)$ for Example 3 (single-index model).

X	model	StI	SIR	DEE-SIR	SAVE	DR	pHd
Case (1.1)	(4.1)	84.56(6.64)	82.52(9.16)	85.12(7.08)	6.28(7.61)	49.52(26.73)	13.64(15.45)
normal	(4.2)	95.37(2.18)	95.56(2.02)	95.14(2.30)	7.10(11.97)	93.14(3.55)	48.90(20.96)
	(4.3)	87.64(5.36)	85.06(6.68)	85.93(6.27)	6.13(7.97)	51.31(25.33)	9.53(13.28)
Case (1.2)	(4.1)	97.88(1.07)	98.15(0.92)	97.26(1.16)	3.31(4.47)	94.04(3.44)	25.43(20.85)
chi-square	(4.2)	99.06(0.51)	99.38(0.33)	98.43(0.69)	9.61(20.34)	97.45(1.56)	51.48(16.00)
	(4.3)	61.42(18.58)	45.22(24.72)	54.76(20.25)	10.97(14.27)	11.37(13.25)	8.21(14.10)
Case (1.3)	(4.1)	87.07(6.66)	77.59(16.47)	71.64(13.75)	15.77(16.29)	37.67(29.62)	33.52(29.57)
skew normal	(4.2)	96.23(1.73)	95.81(2.09)	90.97(4.37)	11.72(14.60)	89.10(12.75)	73.39(22.84)
parallel	(4.3)	89.19(6.23)	75.77(18.69)	68.26(14.13)	14.38(15.74)	33.79(27.39)	18.62(18.54)
Case (1.4)	(4.1)	82.64(10.17)	79.94(13.31)	81.59(8.60)	7.53(7.95)	41.77(27.47)	15.32(15.49)
skew normal	(4.2)	94.12(4.94)	95.11(2.54)	92.61(3.22)	8.18(11.86)	91.60(5.21)	54.25(20.94)
unparallel	(4.3)	85.61(8.64)	83.97(9.63)	81.70(8.33)	4.99(6.24)	43.50(27.88)	12.23(14.54)
Case (1.5)	(4.1)	82.57(8.73)	91.01(4.38)	90.39(4.12)	5.96(7.68)	63.74(25.71)	13.04(14.60)
MMN	(4.2)	91.42(5.11)	97.72(1.12)	96.94(1.31)	40.51(30.70)	96.34(1.65)	64.33(17.81)
	(4.3)	84.61(8.42)	92.32(3.79)	90.08(4.96)	7.65(9.73)	62.71(25.78)	12.06(14.94)
Case (1.6)	(4.1)	98.73(0.57)	90.88(4.57)	70.76(7.64)	23.88(23.55)	80.55(14.78)	52.64(20.06)
MMSN	(4.2)	98.45(0.62)	97.49(1.14)	74.86(4.10)	63.21(25.66)	95.40(2.52)	57.65(24.86)
	(4.3)	98.12(1.10)	90.52(4.79)	68.96(8.43)	19.60(18.11)	72.22(20.88)	36.84(21.58)

(1.2) $\mathbf{X} = (X_1, \ldots, X_p)^\tau$, $X_1, \ldots, X_p$ i.i.d. $\sim \chi^2(5) - 5$, $i = 1, 2, \ldots, p$. This is not an elliptical distribution, but it is not very seriously skewed as the degrees of freedom is 5.

(1.3) $\mathbf{X} \sim SN_p(I_p, C)$ where $C = \beta$. Here the skewness parameter is in the space spanned by β;

(1.4) $\mathbf{X} \sim SN_p(I_p, C)$ where $C = [1, 1, 0, 0, 1, 1, 0, \ldots]$. The skewness parameter is not in the space spanned by β, but not orthogonal to it.

(1.5) $\mathbf{X} \sim 0.5N(\mu_1, I_p) + 0.3N(\mu_2, I_p) + 0.2N(\mu_3, I_p)$ is a mixture multivariate normal, where $\mu_1 = [3, 3, 3, 0, \ldots, 0]^\tau$, $\mu_2 = [3, 0, 3, 3, \ldots, 0]^\tau$, and $\mu_3 = [3, 0, 0, 3, 3, 0, \ldots, 0]^\tau$.

(1.6) $\mathbf{X} \sim 0.5SN_p(\mu_1, I_p, C_1) + 0.3SN_p(\mu_2, I_p, C_2) + 0.2SN_p(\mu_3, I_p, C_3)$ is a mixture multivariate skew normal, where μ_1, μ_2, μ_3 are same as in Case(1.5); $C_1 = \beta$, $C_2 = 3\beta$, and $C_3 = [1, 1, 0, 0, 0, 0, 1, 1, 0, 0]^\tau$.

Based on experience with single-index models, the sample size $n = 250$ can be regarded as large enough to alleviate random variability in the simulation. We took $n = 250$ for Cases (1.1), (1.2), (1.3) and (1.4), and $n = 400$ for Cases (1.5) and (1.6). We assume the number of mixture components known, though it is a crucial parameter in practice. We did 100 replications. The results are in Table 1.

In Case (1.1), SIR and DEE-SIR work well, sometimes better than the new method StI. In Case (1.5), SIR and DEE-SIR outperform StI. It seems that, in normal cases, the traditional methods have good performance even when the ellipticity is violated slightly. However, in most cases, StI is the winner and is robust to non-ellipticity. In Case (1.1), StI performs well, and SIR and DEE-SIR are competitive. In Case (1.2), StI, SIR, DEE-SIR, and DR all have good performance for Models (4.1) and (4.2), suggesting robustness to non-ellipticity. For Model (4.3), StI does much better than the others. In Case (1.3), StI shows its strong adaptability for non-elliptical distributions, while SIR and DEE-SIR lag in both estimation accuracy and stability. Case (1.6) is much the same. Case (1.4) was to examine the robustness of StI when the skew parameter is not in the space spanned by β. SIR and DEE-SIR also work here, but for Models (4.1) and (4.3), they are not better than StI. For Case (1.5), before performing StI we need to complement an EM algorithm to estimate proportion, mean, covariance, and skew parameters in each simple distribution. The performance of StI is hampered by estimating too many parameters, a poor estimator will notably weaken the power of StI. Traditional methods avoid this question. In Case (1.6), skewness causes no unified effects on the performance of traditional methods, and here StI performs better than its competitors.

Example 4. Consider the bi-index models

$$Y = \exp\Big(\frac{\beta_1^\tau \mathbf{X}}{2}\Big) + \sin\Big(\frac{\beta_2^\tau \mathbf{X}}{2}\Big) + 0.5\varepsilon, \tag{4.4}$$

$$Y = \frac{\beta_1^\tau \mathbf{X} + 0.3\varepsilon}{2 + |\beta_2^\tau \mathbf{X} - 4 + \varepsilon|}, \tag{4.5}$$

$$Y = \sin\Big(\frac{\beta_1^\tau \mathbf{X}}{4}\Big) + \exp\Big(\frac{\beta_2^\tau \mathbf{X}}{2}\Big)\varepsilon, \tag{4.6}$$

$$Y = \beta_1^\tau \mathbf{X} + \Big|\frac{\beta_2^\tau \mathbf{X}}{2}\Big| + 0.5\varepsilon, \tag{4.7}$$

where $\beta_1 = (1,1,1,1,0,\ldots,0)^\tau$, $\beta_2 = (1,1,0,0,1,1,0,\ldots,0)^\tau$, $\varepsilon \sim N(1,0)$. Let $p = 10$, the number of slices for SIR, SAVE and DR was $h = 10$. We took $\mathbf{X}$ as follows:

(2.1) $\mathbf{X} \sim N_p(0, I_p)$.

(2.2) $\mathbf{X} = (X_1,\ldots,X_p)^\tau$, $X_1,\ldots,X_p$ i.i.d $X_1 \sim \chi^2(1) - 1$.

(2.3) $\mathbf{X} \sim SN_p(I_p, C)$ where $C_1 = (\beta_1 + \beta_2)/2$. Here C is in the space spanned by $\mathbf{B}$.

(2.4) $\mathbf{X} \sim SN_p(I_p, C)$, where $C = [1,1,0,0,1,1,0,0,1,1]^\tau/2$. The skewness parameter C is not in the space spanned by $\mathbf{B}$, but not orthogonal to it.

Table 2. The means and standard deviations (in parentheses)of $100 \cdot D(A, B)$ for Example 4 (two directions model).

X	model	StI	SIR	DEE-SIR	SAVE	DR	pHd
Case (2.1)	(4.4)	79.08(9.94)	73.27(12.22)	75.97(10.02)	54.23(6.96)	67.75(12.21)	44.39(8.17)
normal	(4.5)	87.08(6.11)	82.23(10.48)	88.37(4.91)	54.51(6.03)	77.50(10.97)	50.71(15.51)
	(4.6)	86.14(5.46)	83.08(8.55)	82.85(7.64)	28.66(13.51)	72.15(12.62)	36.84(9.32)
	(4.7)	73.28(11.21)	67.35(12.98)	76.69(10.00)	59.24(11.25)	68.83(13.14)	48.06(13.92)
Case (2.2)	(4.4)	77.89(8.44)	75.15(10.04)	71.40(11.32)	11.79(5.60)	53.44(4.99)	39.83(9.38)
chi-square	(4.5)	81.83(7.90)	76.34(10.75)	80.39(6.27)	11.12(6.38)	53.55(7.62)	34.26(8.46)
	(4.6)	81.74(8.03)	80.90(8.53)	73.99(13.16)	13.87(7.64)	54.35(10.27)	35.40(8.66)
	(4.7)	65.02(10.85)	59.91(9.72)	62.98(10.59)	16.38(7.40)	56.59(6.08)	36.25(10.02)
Case (2.3)	(4.4)	81.28(7.99)	66.86(12.09)	75.74(9.36)	55.62(7.83)	63.03(12.33)	53.11(8.27)
skew normal	(4.5)	84.10(8.32)	85.99(7.81)	87.51(5.51)	55.73(8.06)	82.53(10.06)	63.82(14.49)
parallel	(4.6)	77.19(10.41)	65.83(12.82)	66.78(12.21)	25.24(12.98)	59.68(12.20)	42.57(11.01)
	(4.7)	61.46(9.97)	62.73(12.29)	71.46(10.32)	59.43(11.24)	66.29(12.58)	55.13(10.76)
Case (2.4)	(4.4)	86.41(6.65)	76.87(12.79)	79.16(8.61)	54.88(6.66)	69.77(13.24)	49.69(10.51)
skew normal	(4.5)	85.49(7.37)	81.49(10.41)	84.31(6.26)	56.95(8.75)	77.03(12.92)	50.75(13.48)
unparallel	(4.6)	78.84(7.78)	70.36(11.89)	72.92(10.48)	25.18(14.42)	58.65(10.53)	39.20(9.76)
	(4.7)	69.18(11.71)	62.23(11.84)	68.95(11.45)	58.57(10.26)	65.26(12.17)	45.99(12.59)
Case (2.5)	(4.4)	75.60(8.63)	73.64(4.25)	68.06(10.29)	69.87(10.61)	76.11(10.59)	55.87(6.71)
MMN	(4.5)	51.28(1.58)	51.59(2.39)	50.91(1.36)	76.20(10.47)	50.26(2.11)	45.49(13.81)
	(4.6)	60.09(11.63)	65.56(8.08)	59.44(11.03)	52.95(8.29)	66.06(8.60)	31.22(9.40)
	(4.7)	56.31(5.03)	53.76(4.88)	60.78(8.50)	56.07(8.17)	55.00(6.44)	24.16(9.85)
Case (2.6)	(4.4)	95.60(2.12)	73.15(3.61)	51.48(1.91)	65.06(5.30)	63.12(4.84)	49.73(7.21)
MMSN	(4.5)	95.83(2.10)	56.20(5.68)	55.82(2.50)	80.42(10.48)	53.16(4.48)	50.06(8.95)
	(4.6)	98.60(0.60)	52.52(3.75)	51.21(4.79)	53.97(13.15)	59.51(6.12)	36.97(10.47)
	(4.7)	93.19(3.18)	76.47(4.37)	73.90(9.14)	83.58(10.29)	84.55(4.77)	32.37(12.21)

(2.5) $\mathbf{X} \sim 0.5N(\mu_1, I_p) + 0.3N(\mu_2, I_p) + 0.2N(\mu_3, I_p)$ is a mixture multivariate normal, where $\mu_1 = [3, 3, 3, 3, 0, \ldots, 0]^\tau$, $\mu_2 = [3, 0, 0, 3, 3, 0, \ldots, 0]^\tau$, and $\mu_3 = [3, 3, 0, 0, 3, 3, 0, \ldots, 0]^\tau$.

(2.6) $\mathbf{X} \sim 0.5SN_p(\mu_1, I_p, C_1) + 0.3SN_p(\mu_2, I_p, C_2) + 0.2SN_p(\mu_3, I_p, C_3)$ is a mixture multivariate skew normal, where μ_1, μ_2, μ_3 are same as in (2.5), $C_1 = 2\beta$, $C_2 = 3\beta$, and $C_3 = [-1, -1, 0, 0, 0, 0, -1, -1, 0, 0]^\tau$.

We again took $n = 250$ for the simple distributions and $n = 400$ for the mixed distributions to alleviate random variability in the simulation. We did 100 replications. The results are reported in Table 2. In Case (2.1), StI performs well and DEE-SIR also works well in some cases. As to the slightly skewed distribution in Case (2.2), StI begins to show its superiority when compared to the other methods. In Case (2.3), StI is dominant in two models while SIR and DEE-SIR work better under the other two models. For the skewed distributions with nonparallel locations, StI works best. Comparing the results of Case (2.5) to that of Case (1.5), we find that the results for the synthetic model are not as distinguishable as those for the single-index model. In this case, SIR, DEE-SIR, and DR work better than StI, though slightly. For Case (2.6), the results

indicate that StI works much better than its competitors. These observations suggest that StI has the advantages in handling non-elliptical distributions, while SIR-based methods are also robust to non-ellipticity to some extent. SAVE is not a good method in these settings, and, consequently DR is not being in effect, a combination of SIR and SAVE.

Sliced inverse regression and directional regression are, to some extent, robust against non-elliptical distributions, as demonstrated in Li and Dong (2009) and Dong and Li (2010). But the Stein's Lemma-based method generally works well, and in most cases better when compared with these methods for mixture multivariate skew-elliptical distributions, even when there are parameters to be estimated. Unlike traditional methods, StI is widely tolerant of both elliptical and non-elliptical distribution. The results of Cases (1.6) and (2.6) suggest that StI has advantages for distributions beyond the elliptical.

5. Data Example: Handwritten Digital Data

The University of California at Irvine machine-learning repository (`ftp://ftp.ics.uci.edu/pub/machine-learning-databases/pendigits/`) contains 10,992 samples of handwritten digits $(0, 1, \ldots, 9)$ from 44 different writers. Each digit is stored as a 16-dimensional vector, regarded as the predictor, with 16 digits $0, 1, \ldots, 9$ as the response variable values. Zhu and Hastie (2003) divided the dataset into a learning set (7,494 cases) and a testing set (3,498 cases), then investigated and found important discriminant directions. Later, Li and Wang (2007), and Dong and Li (2010) studied the data for sufficient dimension reduction.

Among digits 0 to 9, we consider 0, 6 and 9 because they are easily confused. We used the dimension reduction methods StI, SIR, SAVE, and DR to identify dimension reduction directions. The learning set contains $780, 720$, and 719 sample points for digits $0, 6, 9$ respectively, and the testing set contains $363, 336$, and 336 sample points accordingly. With $p = 16$, $(16+5)^2 = 441$, the sample size was taken to be $n = 450$. We randomly drew $n = 450$ points from the learning set, then applied StI, SIR, SAVE, and DR to identify central subspaces B when the structural dimension was set to be $K = 2$, and 3. Figures 1 and 2 show the data structure of X when one random sampling was performed. It is clear that the data have three groups and are not elliptically symmetric. When the data were projected onto the respective central subspaces B, SIR and DR got data clouds with a certain symmetry: three groups put the data cloud in a triangular pattern. This was seen before, see Dong and Li (2010). SAVE did not separate the three groups well. In Figure 1, the data cloud determined by StI shows that the three groups are significantly separated, and that the points with the response value 0 form a half-circle around the other two groups. Previous studies did not

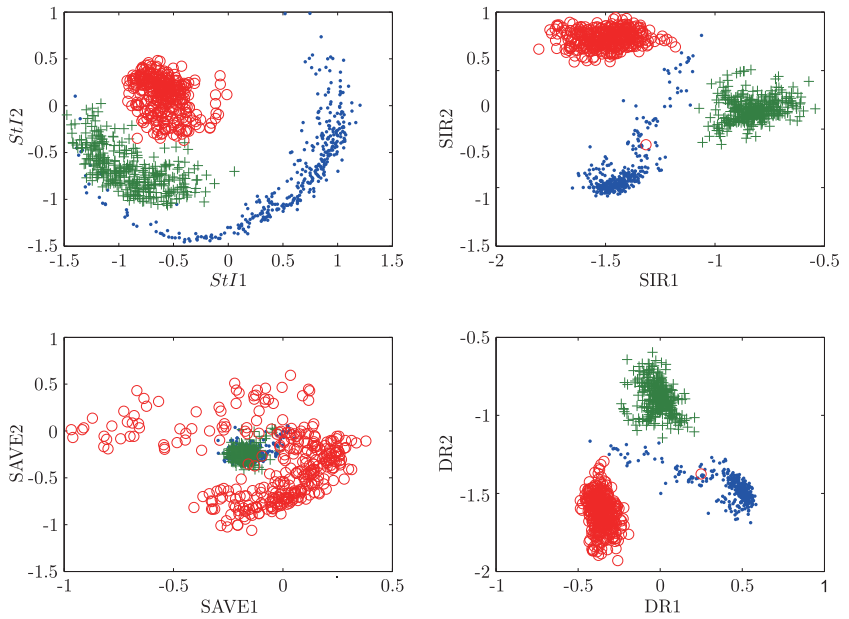

Figure 1. Plots for the handwritten digits data with methods StI, SIR, SAVE and DR and $K = 2$. $\cdot$, $+$ and $\circ$ respectively denote the digits 0, 6 and 9.

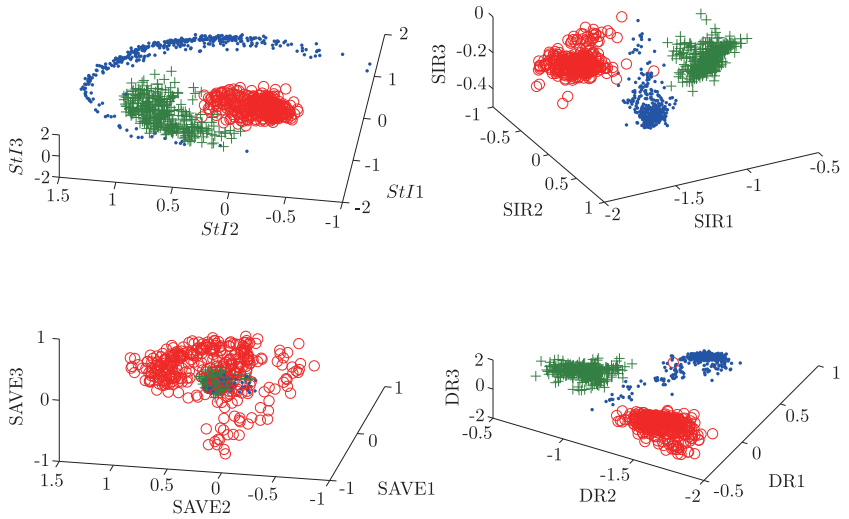

Figure 2. Plots for the handwritten digits data with methods StI, SIR, SAVE and DR and $K = 3$. $\cdot$, $+$ and $\circ$ respectively denote the digits 0, 6 and 9.

find this. In Figure 2 with $K = 3$, we see that the points with the response value 0 form a crescent apart from the other two groups. Together with the

Table 3. The means and standard deviations (in parentheses) of $100R^2$ and $100R_w^2$ for handwritten digits.

Fitted Model	K	StI	SIR	SAVE	DR
Nonparametric regression	$K=2$	99.91(0.13)	99.96(0.11)	28.96(5.09)	86.84(2.15)
	K=3	99.99(0.01)	99.99(0.06)	37.42(3.65)	88.38(2.05)
Logistic regression	$K=2$	99.43(0.72)	99.69(0.34)	1.70(1.69)	99.44(0.64)
	K=3	99.41(0.89)	99.78(0.36)	2.07(1.96)	99.65(0.57)

data structure that SIR and DR found, classification can work better. To show the fitting effect, we first blindly estimated a nonparametric regression function without taking the central subspaces into account. To perform this, $n = 450$ observations were randomly sampled from the test set, then projected onto the corresponding central subspaces B. A kernel smoother was used to estimate the nonparametric regression function $E(Y|X) = G(B^T X)$,

$$\hat{y}_i = \frac{\sum_{j=1}^n y_j K_h(\hat{\mathbf{B}}^\tau x_j - \hat{\mathbf{B}}^\tau x_i)}{\sum_{j=1}^n K_h(\hat{\mathbf{B}}^\tau x_j - \hat{\mathbf{B}}^\tau x_i)}, \qquad (i = 1, 2, \ldots, n). \tag{5.1}$$

Here the kernel function is the standard Gaussian function with bandwidth h selected by cross-validation, and (x_j, y_j)'s the sample points from the test set. Then the R^2 value was computed,

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \tag{5.2}$$

where $\bar{y} = n^{-1}\Sigma_{i=1}^n y_i$. The responses y_i are discrete, we applied logistic regression to these test samples projected onto the corresponding central subspaces B. The Matlab function "mnrfit(x,y)" was called to obtain the predicted values $\hat{y}_i$. Then R^2 was computed.

The procedure was repeated 100 times to compute the mean and standard deviation of R^2. The results are reported in Table 3. We see that the central subspaces determined by StI, SIR and DR aid good model fitting, whereas SAVE did not work well. SIR again shows its robustness to the underlying distribution.

6. Concluding Remarks

In the paper, Stein's Lemma is revisited to see how it handles multivariate mixture skew-elliptical distributions. As a by-product, we give some results to see how the ellipticity (or the linearity condition) is close to a necessary and sufficient condition for a Stein's Lemma-based method to identify the central subspace. This idea could be extended to consider a pHd-type method related to the second derivative of conditional expectation of Y given X. Also, it would

be possible to study for which skew distributions the linearity could be removed when SIR or MAVE is applied. These researches are ongoing.

In practice, the testing of distributions is in needed. From the second part of Corollary 2, when its conditions are not satisfied, the covariance between Y and $\mathbf{X}$ is not enough for identifying the central subspace and the term

$$\sum_{j=1}^{m} w_j \Sigma_j E_{\mathbf{X}^{(j)}} \left[m(Y) \frac{\partial \{\ln \Phi_q(C_j^\tau \mathbf{X}^{(j)} + \nu_j; 0, D_j)\}}{\partial \mathbf{X}^{(j)}} \right]$$

needs to be estimated. When the generalized Stein's Lemma is applied in this case, we should check whether the underlying distribution is a MMSE distribution. This is an issue for existing SDR methods, for which the testing of ellipticity is required. In our case, only skewness testing is needed; this is a ongoing project.

Acknowledgement

The research described here was supported by a grant from the Research Council of Hong Kong, and a grant from Hong Kong Baptist University, Hong Kong. The authors thank the Editor, an Associate editor and two referees for their constructive comments that led to an improvement of an early manuscript. The first two coauthors share the first authorship due to their contributions in the original and later revisions, respectively.

Appendix: Proofs of Theorems and Propositions

Lemma A.1 (Integration by parts)**.** *Let* $\mathbf{X} = (X_1, X_2, \cdots, X_p)^\tau \in R^p$. *Suppose that functions* $H(\mathbf{x})$ *and* $J(\mathbf{x})$ *are weakly differentiable. Then*

$$\int_\Omega H(\mathbf{x}) \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} d\,\mathbf{x} + \int_\Omega J(\mathbf{x}) \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} d\,\mathbf{x} = \int_\Omega \frac{\partial \{H(\mathbf{x}) J(\mathbf{x})\}}{\partial \mathbf{x}} d\,\mathbf{x}, \tag{A.1}$$

provided that all integrals exist, where the set $\Omega \subseteq R^p$ *is a given domain. When* $H(x)J(x)$ *satisfies that, at the boundary* Ω^* *of* Ω, $H(x_i)J(x_i) \mid_{x_i \in \Omega^*} = 0$ *for all* $i = 1, \ldots, p$, *then*

$$-\int_\Omega H(\mathbf{x}) \frac{\partial J(\mathbf{x})}{\partial \mathbf{x}} d\,\mathbf{x} \int_\Omega J(\mathbf{x}) \frac{\partial H(\mathbf{x})}{\partial \mathbf{x}} d\,\mathbf{x}. \tag{A.2}$$

Proof of Theorem 1. First, consider an MSE distribution with $dG^{(p)}(u)/du = (-1/2)g^{(p)}(u)$ for $u \in (0, +\infty)$ and $dQ(\mathbf{x})/d\mathbf{x} = 2\Sigma^{-1}(\mathbf{x} - \mu)$. We then have, invoking Lemma A.1,

$$\begin{aligned} &\Sigma^{-1} E_{\mathbf{X}}[m(Y)(\mathbf{X} - \mu)] = \Sigma^{-1} E_{\mathbf{X}}[H(\mathbf{X})(\mathbf{X} - \mu)] \\ &= \int_\Omega H(\mathbf{x})\ \Sigma^{-1}(\mathbf{x} - \mu)\ c_E |\Sigma|^{-1/2} g^{(p)}\{Q(\mathbf{x})\} \Psi_q(C\mathbf{x} + \nu; D, g^{(q)}_{Q(\mathbf{x})})\ d\,\mathbf{x} \end{aligned}$$

$$
\begin{aligned}
&= -c_E|\Sigma|^{-1/2}\int_\Omega H(\mathbf{x})\ \Psi_q(C\mathbf{x}+\nu; D, g^{(q)}_{Q(x)})\ dG^{(p)}\{Q(\mathbf{x})\}\\
&= c_E|\Sigma|^{-1/2}\int_\Omega \Big[\frac{\partial H(\mathbf{x})}{\partial \mathbf{x}}\ \Psi_q(C\mathbf{x}+\nu; D, g^{(q)}_{Q(x)}) + H(\mathbf{x})\frac{\partial \Psi_q(C\mathbf{x}+\nu; D, g^{(q)}_{Q(\mathbf{x})})}{\partial \mathbf{x}}\Big]\\
&\quad \times G^{(p)}\{Q(\mathbf{x})\}d\,\mathbf{x}\\
&= E_{\mathbf{X}}\left[\frac{\partial H(\mathbf{X})}{\partial \mathbf{X}}\cdot\frac{G^{(p)}\{Q(\mathbf{X})\}}{g^{(p)}\{Q(\mathbf{X})\}}\right]\\
&\quad + E_{\mathbf{X}}\left[H(\mathbf{X})\cdot\frac{G^{(p)}\{Q(\mathbf{X})\}}{g^{(p)}\{Q(\mathbf{X})\}}\cdot\frac{\partial\{\ln\Psi_q\{C\mathbf{X}+\nu; D, g^{(q)}_{Q(\mathbf{X})}\}}{\partial \mathbf{X}}\right]\\
&= E_{\mathbf{X}}\left[\frac{\partial h(\mathbf{B}^\tau\mathbf{X})}{\partial \mathbf{B}^\tau\mathbf{X}}\cdot\frac{G^{(p)}\{Q(\mathbf{X})\}}{g^{(p)}\{Q(\mathbf{X})\}}\right]\\
&\quad + E_{\mathbf{X}}\left[m(Y)\cdot\frac{G^{(p)}\{Q(\mathbf{X})\}}{g^{(p)}\{Q(\mathbf{X})\}}\cdot\frac{\partial\{\ln\Psi_q\{C\mathbf{X}+\nu; D, g^{(q)}_{Q(\mathbf{X})}\}}{\partial \mathbf{X}}\right],
\end{aligned}
$$

provided $H(\mathbf{X}) = E(m(Y)|\mathbf{X})$, where $c_E = [\Psi_q(\nu; D + C\bar{\Sigma}C^\tau, g^{(q)})]^{-1}$. Since the distribution function $\Psi_q(C\mathbf{X}+\nu; D, g^{(q)}_{Q(x)}) \in [0,1]$, the boundary conditions $G^{(p)}\{Q(x_i)\}\cdot\Psi_q(C^\tau\mathbf{X}+\nu; D, g^{(q)}_{Q(\mathbf{X})})\cdot H(\mathbf{x})|_{x_i\in\partial\Omega} = 0$ simplified to $G^{(p)}\{Q(x_i)\}\cdot H(x_i)|_{x_i\in\partial\Omega} = 0$ for all $i = 1,\ldots,p$.

Now we show (3.5) for MMSE distributions. From their definition and the above proof for MSE distributions, we see that

$$
\begin{aligned}
&E_{\mathbf{X}}[m(Y)(\mathbf{X}-\mu^*)]\\
&= \sum_{j=1}^m w_j E_{\mathbf{X}^{(j)}}\left[\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\frac{\partial H(\mathbf{X}^{(j)})}{\partial \mathbf{X}^{(j)}}\right]\\
&\quad + \sum_{j=1}^m w_j\Sigma_j E_{\mathbf{X}^{(j)}}\left[m(Y)\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\frac{\partial\{\ln\Psi_q(C_j^\tau\mathbf{X}^{(j)}+\nu_j; D_j, g^{(q)}_{Q_j(\mathbf{X}^{(j)})})\}}{\partial \mathbf{X}^{(j)}}\right]\\
&= \sum_{j=1}^m w_j\Sigma_j\mathbf{B}E_{\mathbf{X}^{(j)}}\left[\frac{\partial h(\mathbf{B}^\tau\mathbf{X}^{(j)})}{\partial \mathbf{B}^\tau\mathbf{X}^{(j)}}\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\right]\\
&\quad + \sum_{j=1}^m w_j\Sigma_j E_{\mathbf{X}^{(j)}}\left[m(Y)\frac{G_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}{g_j^{(p)}\{Q_j(\mathbf{X}^{(j)})\}}\frac{\partial\{\ln\Psi_q(C_j^\tau\mathbf{X}^{(j)}+\nu_j; D_j, g^{(q)}_{Q_j(\mathbf{X}^{(j)})})\}}{\partial \mathbf{X}^{(j)}}\right].
\end{aligned}
$$

Here $\mu^* = \sum_{j=1}^m\{E_{\mathbf{X}^{(j)}}[H(\mathbf{X}^{(j)})]/E_{\mathbf{X}}[H(\mathbf{X})]\}w_j\mu_j = \sum_{j=1}^m w_j\mu_j$, because $E_{\mathbf{X}^{(j)}}[H(\mathbf{X}^{(j)})] = E_{\mathbf{X}}[H(\mathbf{X})] = E(m(Y))$ for any $1 \le j \le m$. The proof is finished.

Proof of Theorem 2. Write the right side of (4.4) as $M(t)$ for any function $m(Y,t)$. We know that $M(t)=\mathbf{B}z(t)$ where $z(t)=E_{\mathbf{X}}[\{G^{(p)}\{Q(\mathbf{X})\}/g^{(p)}\{Q(\mathbf{X})\}\}\times\{\partial h(\mathbf{B}^\tau\mathbf{X},t)/\partial\mathbf{B}^\tau\mathbf{X}\}]$ is a $K\times 1$ vector. Thus $M(t)$ lies in $S_{Y|\mathbf{X}}$, the corresponding eigenvector associated with the nonzero eigenvalue of $M(t)M(t)^\tau$ lies in $S_{Y|\mathbf{X}}$. Then the corresponding eigenvectors associated with the nonzero eigenvalues of $M=\int M(t)M(t)^\tau d\ F_Y(t)$ lie in $S_{Y|\mathbf{X}}$, $span(M)\subseteq S_{Y|\mathbf{X}}$. On the other hand, when $Z=\int z(t)z(t)^\tau d\ F_Y(t)$ is non-singular, $M=\mathbf{B}Z\mathbf{B}^\tau$ has K nonzero eigenvalues, so $span(M)=S_{Y|\mathbf{X}}$.

References

Arellano-Valle, R. B. and Genton, M. G. (2010a). Multivariate unified skew-elliptical distributions. *Chilean J. Statist.* **1**, 17-33

Arellano-Valle, R. B. and Genton, M. G. (2010b). Multivariate extended skew-t distributions and related families. *Metron-International Journal of Statistics* **LXVIII**, 201-234.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scand. J. Statist.* **12**, 171-178.

Azzalini, A. and Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715-726.

Branco, M. D. and Dey, D. K. (2001). A general class of multivariate skew-elliptical distributions. *J. Multivariate Anal.* **79**, 99-113.

Cabral, C. R. B., Lachos, V. H. and Prates, M. O. (2012). Multivariate mixture modeling using skew-normal independent distributions. *Comput. Statist. Data Anal.* **56**, 124-142.

Cook, R. D. (1998a). *Regression Graphics: Ideas for Studying Regressions through Graphics.* Wiley, New York.

Cook, R. D. (1998b). Principal Hessian directions revisited. *J. Amer. Statist. Assoc.* **93**,84-94

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *Ann. Statist.* **30**, 455-474.

Cook, R. D. and Li, L. (2009). Dimension reduction in regression with exponential family predictors. *J. Comput. Graph. Statist.* **18**, 774-791.

Cook, R. D. and Nachtsheim, C. J. (1994). Re-weighting to achieve elliptically contoured covariates in regression. *J. Amer. Statist. Assoc.* **89**, 592-599.

Cook, R. D. and Weisberg, S. (1991). Sliced inverse regression for dimension reduction: Comment. *J. Amer. Statist. Assoc.* **86**, 328-332.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM Algorithm. *J. Roy. Statist. Soc. Ser. B* **39**, 1-38.

Dong, Y. X. and Li, B. (2010). Dimension reduction for non-elliptically distributed predictors: second-order methods. *Biometrika* **97**, 279-294.

Duan, N. and Li, K. C. (1991). A bias bound for least squares linear regression. *Statist. Sinica* **1**, 127-136.

Feng, Z. H., Wang, T. and Zhu, L. (2014). Transformation-based estimation. *Comput. Statist. Data. Anal.* **78**, 186-205.

Ferré, L. (1998). Determing the dimension in sliced inverse regression and related methods. *J. Amer. Statist. Assoc.* **93**, 132-140.

Genton, M. G. (ed.) (2004). *Skew-elliptical Distributions and Their Applications: A Journey Beyond Normality.* Chapman & Hall/CRC, Boca Raton, Florida.

Landsman, Z. and Neslehováb, J. (2008). Stein's Lemma for elliptical random vectors. *J. Multivariate Anal.* **99**, 912-927.

Li, B. and Dong, Y. X. (2009). Dimension reduction for non-elliptically distributed predictors. *Ann. Statist.* **37**, 1272-1298.

Li, B. and Wang, S. (2007). On directional regression for dimension reduction. *J. Amer. Statist. Assoc.* **102**, 997-1008.

Li, K. C. (1991). Sliced inverse regression for dimension reduction. *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.

Lin, T. I. (2009). Maximum likelihood estimation for multivariate skew normal mixture models. *J. Multivariate Anal.* **100**, 257-265.

O'Hagan, A. and Leonhard, T. (1976). Bayes estimation subject to uncertainty about parameter constraints. *Biometrika* **63**, 201-202.

Wu, Y. and Li, L. (2011). Asymptotic properties of sufficient dimension reduction with a diverging number of predictors. *Statist. Sinica* **21**, 707-730.

Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression. *J. Roy. Statist. Soc. Ser. B* **64**, 159-175.

Yin, X. and Li, B. (2011). Sufficient dimension reduction based on an ensemble of minimum average variance estimators. *Ann. Statist.* **39**, 3392-3416.

Zeng, P. and Zhu, Y. (2010). An integral transform method for estimating the central mean and central subspaces. *J. Multivariate Anal.* **101**, 271-290.

Zhu, M. and Hastie, T. J. (2003). Feature extraction for nonparametric discriminant analysis. *J. Comput. Graph. Statist.* **12**, 101-120.

Zhu, L. P. and Zhu, L. X. (2009). Dimension reduction for conditional variance in regressions. *Statist. Sinica* **19**, 869-883.

Zhu, L. P., Zhu, L. X. and Feng, Z. (2010). Dimension reduction in regressions through cumulative slicing estimation. *J. Amer. Statist. Assoc.* **105**, 1455-1466.

Zhu, L. P., Zhu, L. X. and Wang, S. Q. (2010). On dimension reduction in regressions with multivariate responses. *Statist. Sinica* **20**, 1291-1307.

Zhu, L. X., Miao, B. Q. and Peng, H. (2006). On sliced inverse regression with high-dimensional covariates. *J. Amer. Statist. Assoc.* **100**, 630-643.

Zhu, Y. and Zeng, P. (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *J. Amer. Statist. Assoc.* **101**, 1638-1651.

Department of Statistics, Zhejiang Agriculture & Forest University, Zhejiang, China.

E-mail: guanyu@zafu.edu.cn

Department of Mathematics, Hong Kong Baptist University, Hong Kong, China.

E-mail: 14485435@life.hkbu.edu.hk

Department of Mathematics, Hong Kong Baptist University, Hong Kong, China.

E-mail: lzhu@hkbu.edu.hk

(Received August 2015; accepted December 2015)