Available online at www.sciencedirect.com
ScienceDirect
Statistics & Probability Letters 77 (2007) 817–821
STATISTICS & PROBABILITY LETTERS
www.elsevier.com/locate/stapro

A note on sufficient dimension reduction

Xuerong Meggie Wen

Department of Mathematics and Statistics, University of Missouri, Rolla, MO 65409, USA

Received 3 December 2004; received in revised form 30 October 2006; accepted 19 December 2006
Available online 26 January 2007

Abstract

In this paper, we presented a theoretical result and then discussed possible applications of our result to SDR problems. In addition to providing insights into existing SDR methods when Y is univariate; our theorem also applies to multivariate responses, especially when the response takes the form of (Y, W), where Y is a continuous variable and W is categorical.

Keywords: Sufficient dimension reduction; Slicing; Multivariate dimension reduction; Censoring regression

1. Introduction

In a regression analysis with a response variable Y and a vector of random predictors $\mathbf{X} = (X_1, \ldots, X_p)^{\mathrm{T}} \in \mathbb{R}^p$, we seek a parsimonious characterization of the conditional distribution of $Y|\mathbf{X}$. The goal of sufficient dimension reduction (SDR; Cook, 1994, 1998) is to reduce the dimension of $\mathbf{X}$ by replacing it with a minimal set of linear combinations of $\mathbf{X}$, without loss of information on $Y|\mathbf{X}$ and without requiring a pre-specified parametric model. These linear combinations are called the *sufficient predictors*. More formally, we seek subspaces $\mathscr{S} \subseteq \mathbb{R}^p$ such that

$$Y \perp\!\!\!\perp \mathbf{X}|P_{\mathscr{S}}\mathbf{X},$$

where $\perp\!\!\!\perp$ indicates independence, and $P_{(\cdot)}$ stands for a projection operator with respect to the standard inner product. Such an $\mathscr{S}$ is called a *dimension reduction subspace*. Under mild conditions (Cook, 1998) that almost always hold in practice, the minimal dimension reduction subspace is uniquely defined and coincides with the intersection of all dimension reduction subspaces. This intersection is called the *central subspace* (CS; Cook, 1994, 1998) of the regression and denoted as $\mathscr{S}_{Y|\mathbf{X}}$. SDR is concerned with making inferences for the CS.

When the conditional mean function $\mathrm{E}(Y|\mathbf{X})$ is of special interest, the inquiry of SDR is restricted to the *central mean subspace*, the intersection of all subspaces $\mathscr{S}$ satisfying $Y \perp\!\!\!\perp \mathrm{E}(Y|\mathbf{X})|P_{\mathscr{S}}\mathbf{X}$. Or equivalently, the intersection of all subspaces $\mathscr{S}$ satisfying the conditional independent condition

$$\mathrm{E}(Y|\mathbf{X}) \perp\!\!\!\perp \mathbf{X}|P_{\mathscr{S}}\mathbf{X}. \tag{1.1}$$

Cook and Li (2002)investigated possible approaches to inferring about the central mean subspace.

E-mail address: wenx@umr.edu.

0167-7152/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.spl.2006.12.016

The definition of the central (mean) subspace does not depend on whether Y is univariate or multivariate. Setodji and Cook (2004) presented an estimation method for the multivariate CS using k-means method. Cook and Setodji (2003), Yoo and Cook (2007) studied estimation methods for the multivariate central mean subspace.

In this paper, we present a theoretical result revealing the connections between the multivariate CS and the marginal CSs. We then apply this result to characterize the loss of information due to slicing, and censoring regression. In addition to providing useful insights into current methods in sufficient dimension reduction, our work also suggests the estimation accuracy can be greatly improved by taking slicing effects into account. When the response is bivariate and multivariate, our theoretical results may also open the door to new and better estimation methods.

The rest of this paper is organized as follows. In Section 2, we present our main result. We then explore the applications of our result under different contexts. In Section 3, we discuss how our result relates to the current estimation methods for the multivariate central mean subspace. Section 4 is dedicated to the discussion of the effects of slicing. In Section 5, we focus on applying our theoretical findings to censoring regressions. A brief conclusion is given in Section 6.

2. The connections between multivariate CS and marginal CS

Let $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^r$, where $r > 1$; $\mathbf{Y}_1 \in \mathbb{R}^l$, $l \geqslant 1$; and $\mathbf{Y}_2 \in \mathbb{R}^m$, $m \geqslant 1$. Adopting the definition of the partial CS from Chiaromonte et al. (2002), we define $\mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}}$ as the intersection of all subspaces $\mathscr{S}$ satisfying $\mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X} \mid (P_{\mathscr{S}}\mathbf{X}, \mathbf{Y}_2)$. The following propositions reveal the connection among the multivariate CS, the partial CS and the marginal CS. Notice that the roles of $\mathbf{Y}_1$ and $\mathbf{Y}_2$ are exchangeable. The sum of two subspaces $\mathscr{S}_1$ and $\mathscr{S}_2$ means that the collection of all vectors of the form $\sum_k \boldsymbol{v}_k$, with $\boldsymbol{v}_k \in \mathscr{S}_k$, $k = 1, 2$.

Proposition 1. *Assume that* $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}})^{\mathrm{T}}$, *where* $\mathbf{Y}_1 \in \mathbb{R}^l$, *and* $\mathbf{Y}_2 \in \mathbb{R}^m$. *Then*

$$\mathscr{S}_{\mathbf{Y}|\mathbf{X}} = \mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}} = \mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}}. \tag{2.2}$$

Proof. Let $\mathscr{S}(\boldsymbol{\eta})$ be a dimension reduction subspace for the regression of $(\mathbf{Y}_1, \mathbf{Y}_2)|\mathbf{X}$. Then

$$\begin{aligned}(\mathbf{Y}_1, \mathbf{Y}_2) \perp\!\!\!\perp \mathbf{X} \mid \boldsymbol{\eta}^{\mathrm{T}}\mathbf{X} &\Longrightarrow (\mathbf{Y}_1, \mathbf{Y}_2) \perp\!\!\!\perp \mathbf{X} \mid (\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \mathbf{Y}_2)\\ &\Longrightarrow \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X} \mid (\boldsymbol{\eta}^{\mathrm{T}}\mathbf{X}, \mathbf{Y}_2).\end{aligned}$$

Thus $\mathscr{S}(\boldsymbol{\eta})$ is a partial dimension reduction subspace for the regression of $Y_1|(\mathbf{X}, \mathbf{Y}_2)$, and $\mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}} \subseteq \mathscr{S}_{\mathbf{Y}|\mathbf{X}}$. Since $\mathscr{S}_{\mathbf{Y}_2|\mathbf{X}} \subseteq \mathscr{S}_{\mathbf{Y}|\mathbf{X}}$, we then have

$$\mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}} \subseteq \mathscr{S}_{\mathbf{Y}|\mathbf{X}}.$$

On the other hand, let $\boldsymbol{\beta}$ be an orthonormal basis for $\mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}}$, $\boldsymbol{\zeta}$ be an orthonormal basis for $\mathscr{S}_{\mathbf{Y}_2|\mathbf{X}}$. By definition, $\mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \mathbf{Y}_2)$ and $\mathbf{Y}_2 \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X}$. We then have the following two conditions:

$$\text{(a1) } \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X}, \mathbf{Y}_2), \quad \text{(a2) } \mathbf{Y}_2 \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X}).$$

By Proposition 4.6 from Cook (1998), $(\mathbf{Y}_1, \mathbf{Y}_2) \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{X}, \boldsymbol{\zeta}^{\mathrm{T}}\mathbf{X})$. Hence, $\mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}} \subseteq \mathscr{S}^{(\mathbf{Y}_2)}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}}$. □

Chiaromonte et al. (2002) in their Eq. (10) showed that

$$\mathscr{S}_{Y|\mathbf{X}} \subseteq \mathscr{S}_{W|\mathbf{X}} + \mathscr{S}^{(W)}_{Y|\mathbf{X}}, \tag{2.3}$$

where W is a categorical variable. From Proposition 1, it is straightforward that the equality holds in (2.3) when W is a function of Y.

Proposition 2. *Assume that* $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \mathbf{Y}_2^{\mathrm{T}})^{\mathrm{T}} \in \mathbb{R}^r$, *where* $\mathbf{Y}_1 \in \mathbb{R}^l$, *and* $\mathbf{Y}_2 \in \mathbb{R}^m$. *If* $\mathbf{Y}_1 \perp\!\!\!\perp \mathbf{Y}_2|\mathbf{X}$, *then*

$$\mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}} = \mathscr{S}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}}. \tag{2.4}$$

Proof. Let $\boldsymbol{\rho}$ be an orthonormal basis for $\mathscr{S}_{\mathbf{Y}_1|\mathbf{X}}$, then

$$\begin{aligned}
&\mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X} \quad \text{and} \quad \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{Y}_2|\mathbf{X} \\
&\quad \Leftrightarrow \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X}|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X} \quad \text{and} \quad \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{Y}_2|(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}, \mathbf{X}) \\
&\quad \Leftrightarrow \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{X}|(\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}, \mathbf{Y}_2) \quad \text{and} \quad \mathbf{Y}_1 \perp\!\!\!\perp \mathbf{Y}_2|\boldsymbol{\rho}^{\mathrm{T}}\mathbf{X}.
\end{aligned}$$

Therefore, $\mathscr{S}_{\mathbf{Y}_1|\mathbf{X}}^{(\mathbf{Y}_2)} \subseteq \mathscr{S}_{\mathbf{Y}_1|\mathbf{X}}$. Hence,

$$\mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}} = \mathscr{S}_{\mathbf{Y}_1|\mathbf{X}}^{(\mathbf{Y}_2)} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}} \subseteq \mathscr{S}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}}.$$

Also, $\mathscr{S}_{\mathbf{Y}_1|\mathbf{X}} \subseteq \mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}}$ and $\mathscr{S}_{\mathbf{Y}_2|\mathbf{X}} \subseteq \mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}}$, it is obvious that $\mathscr{S}_{\mathbf{Y}_1|\mathbf{X}} + \mathscr{S}_{\mathbf{Y}_2|\mathbf{X}} \subseteq \mathscr{S}_{(\mathbf{Y}_1,\mathbf{Y}_2)|\mathbf{X}}$. □

We can also prove that $\mathscr{S}_{\mathbf{Y}|\mathbf{X}} = \sum_{i=1}^r \mathscr{S}_{Y_i|\mathbf{X}}$, if given $\mathbf{X}$, Y_i, the ith element of $\mathbf{Y}$, $i = 1, \ldots, r$, are mutually independent of each other.

3. Multivariate central mean subspace

Cook and Setodji (2003), Yoo and Cook (2007) presented the estimation methods for the multivariate central mean subspace. Both methods are based on the following proposition which is included here for reference. Corollary 1 shows that this proposition is a special case of our Proposition 2.

Proposition 3 (*Cook and Setodji, 2003, Proposition 4*). *Assume that* $\mathbf{Y} = (Y_1, \ldots, Y_r)$, *where* Y_k *is the kth coordinate of* $\mathbf{Y}$. *Then*

$$\mathscr{S}_{\mathrm{E}(\mathbf{Y}|\mathbf{X})} = \sum_{k=1}^r \mathscr{S}_{\mathrm{E}(Y_k|\mathbf{X})},$$

where $\mathscr{S}_{\mathrm{E}(\mathbf{Y}|\mathbf{X})}$ *and* $\mathscr{S}_{\mathrm{E}(Y_k|\mathbf{X})}$, *are the central mean subspaces for* $\mathbf{Y}|\mathbf{X}$ *and* $Y_k|\mathbf{X}$, *respectively*.

Corollary 1. *Based on the second definition of the central mean subspace* (1.1), *Proposition* 3 *is a special case of Proposition* 2.

Proof. To easy exposition, let $\mathrm{E}(Y_k|\mathbf{X}) = U_i$, $k = 1, \ldots, r$, $\mathrm{E}(\mathbf{Y}|\mathbf{X}) = \mathbf{U} = (U_1, \ldots, U_r)$. Since U_k is a function of $\mathbf{X}$, $U_k \perp\!\!\!\perp U_j|\mathbf{X}$, for $k, j = 1, \ldots, r$ and $k \neq j$. By Proposition 2,

$$\mathscr{S}_{\mathbf{U}|\mathbf{X}} = \sum_{k=1}^r \mathscr{S}_{U_k|\mathbf{X}}.$$

Based on (1.1), $\mathscr{S}_{\mathbf{U}|\mathbf{X}} = \mathscr{S}_{\mathrm{E}(\mathbf{Y}|\mathbf{X})}$, and $\mathscr{S}_{U_k|\mathbf{X}} = \mathscr{S}_{\mathrm{E}(Y_k|\mathbf{X})}$. Hence, we proved that $\mathscr{S}_{\mathrm{E}(\mathbf{Y}|\mathbf{X})} = \sum_{k=1}^r \mathscr{S}_{\mathrm{E}(Y_k|\mathbf{X})}$. □

4. The loss of information due to slicing

For a many-valued or continuous response Y, a standard treatment in sufficient dimension reduction is to partition the range of Y into a fixed number (h) of slices, and work on the discrete version, $\tilde{Y}$, assuming that the new regression retains all the information, i.e., that $\mathscr{S}_{\tilde{Y}|\mathbf{X}} = \mathscr{S}_{Y|\mathbf{X}}$. However, this assumption is not always true, and the differences between the working and target regressions can be significant when sample size is not large. Moreover, even under the case of equality, we will still face the loss of power since we make use of only the information retained in $\tilde{Y}$, discarding all the intra-slice information.

Let d denote the dimension of $\mathscr{S}_{Y|\mathbf{X}}$. If h is less than d, then the set of sufficient predictors for the regression of $\tilde{Y}$ on $\mathbf{X}$ will necessarily exclude some of the sufficient predictors of Y on $\mathbf{X}$. Experience indicates that good results are often obtained by choosing h to be somewhat larger than $d + 1$, trying a few different values of h as necessary. Since traditional asymptotic results in SDR are based on the number of observations per slice going to infinity, in practice this suggests relatively few slices. Choosing h very much larger than d should generally

be avoided due to the conflicts between the requirements of asymptotic approximations and recovering intra-slice information.

We have noticed that many existing SDR methods are sensitive to the choice of h to some extent (Cook, 1998; Yin and Cook, 2002), especially for small data sets. One of the open questions Kent (1991) asked is what is the effect of changing the number of slices h. Yin and Cook (2002) discussed the connections between the number of slices and the maximum order k of the covariance $\mathrm{E}(Y^k\mathbf{X})$ used to summarize the distribution of $\mathbf{Y}|\mathbf{X}$.

We consider this question from a different aspect. In Proposition 1, setting $\mathbf{Y}_1 = Y \in \mathbb{R}^1$, and $\mathbf{Y}_2 = \tilde{Y}$, we then have:

Corollary 2.

$$\mathscr{S}_{Y|\mathbf{X}} = \mathscr{S}^{(\tilde{Y})}_{Y|\mathbf{X}} + \mathscr{S}_{\tilde{Y}|\mathbf{X}}, \tag{4.5}$$

where $\mathscr{S}^{(\tilde{Y})}_{Y|\mathbf{X}}$ *as the intersection of all subspaces* $\mathscr{S}$ *satisfying* $Y \perp\!\!\!\perp \mathbf{X} \mid (P_{\mathscr{S}}\mathbf{X}, \tilde{Y})$.

Eq. (4.5) gives us the insight about the information we lose on the target regression $Y|\mathbf{X}$ by using $\tilde{Y}$ instead of Y. Most of the existing SDR methods (Cook, 1998; Cook and Ni, 2005) focus on estimating $\mathscr{S}_{\tilde{Y}|\mathbf{X}}$. Since $\mathscr{S}^{(\tilde{Y})}_{Y|\mathbf{X}}$ typically contains more than the origin, slicing will miss relevant intra-slice information. Cook and Ni (2006) used the intra-slice covariances to construct inference methods for the CS, which greatly improved the estimation accuracy.

Note that $\tilde{Y}$ can be replaced with any function of Y. This suggests that we may seek some other functions of Y to replace $\tilde{Y}$ in order to gain better estimation accuracy.

5. Censoring regressions

SDR can be of practical interest to censored regression analysis. As pointed by Zeng (2004), when there are many predictors, nonparametric approaches may be infeasible due to the "curse of dimensionality". Moreover, for semiparametric models, the parametric functions are likely to be misspecified. In contrast, the bivariate SDR methods can be carried out without pre-specifying any parametric model, and it can often avoid the curse of dimensionality. After reduction of $\mathbf{X}$ to the estimated sufficient predictors, many traditional methodologies of survival analysis can be applied.

Proposition 2 is of special interest to us when the response is of the form of made up of continuous and discrete random variables. One special case is censored data. Let T be the true unobservable survival time, and let C be the censoring time. Define $\delta = I\{T \leqslant C\}$, and $Y = T\delta + C(1-\delta)$.

The goal of SDR for survival data is to infer about the CS $\mathscr{S}_{T|\mathbf{X}}$. However, since T is not fully observable, we can estimate only the CS $\mathscr{S}_{(Y,\delta)|\mathbf{X}}$ for the bivariate regression of the observable (Y,δ) on $\mathbf{X}$. Assuming $T \perp\!\!\!\perp C|\mathbf{X}$, the usual independence assumption (Ebrahimi et al., 2003; Tsiatis, 1975), to ensure the identifiability of T, we then have which the following proposition provides a connection between $\mathscr{S}_{(Y,\delta)|\mathbf{X}}$ and $\mathscr{S}_{T|\mathbf{X}}$.

Proposition 4. *If* $T \perp\!\!\!\perp C|\mathbf{B}^{\mathrm{T}}\mathbf{X}$, *then*

$$\mathscr{S}_{(Y,\delta)|\mathbf{X}} = \mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}. \tag{5.6}$$

Proof. By Proposition 2, we have

$$\mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}.$$

Following Proposition 1 from Ebrahimi et al. (2003). Let $S_T(t) = pr(T \geqslant t|\mathbf{X})$, $S_C(t) = pr(C \geqslant t|\mathbf{X})$, $S_Y(t) = pr(Y \geqslant t|\mathbf{X})$, and $S_\delta(t) = pr(Y \geqslant t, \delta = 1|X)$. Following the proof of Ebrahimi et al. (2003), assuming that $T \perp\!\!\!\perp C|\mathbf{X}$, we have

$$S_T(t) = S_Y(t)\exp^{-\int_0^t \mathrm{d}S_\delta(u)/S_Y(u)}.$$

Therefore, $\mathscr{S}_{T|X} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$. In addition, it is straightforward that $S_Y(t) = S_T(t)S_C(t)$. Hence, $\mathscr{S}_{C|X} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$. Thus, $\mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}} \subseteq \mathscr{S}_{(Y,\delta)|\mathbf{X}}$.

Also, note that (Y,δ) is a function of (T,C), therefore, $\mathscr{S}_{(Y,\delta)|\mathbf{X}} \subseteq \mathscr{S}_{(T,C)|\mathbf{X}}$.

We have proved that $\mathscr{S}_{(Y,\delta)|\mathbf{X}} = \mathscr{S}_{(T,C)|\mathbf{X}} = \mathscr{S}_{T|\mathbf{X}} + \mathscr{S}_{C|\mathbf{X}}$. □

6. Summary and future directions

In this paper, we presented a theoretical result and then discussed possible applications of our result to SDR problems. In addition to providing insights into existing SDR methods when Y is univariate; our result also applies to multivariate responses, especially when the repose takes the form of (Y, W), where Y is a continuous variable and W is categorical.

Our theoretical findings give a fresh view over SDR and open the door for new and better estimation methods. We are currently working on developing new theory and methodology using Proposition 1 as a guide.

References

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. Using dimension-reduction subspaces to identify important inputs in models of physical systems. In: Proceedings of Section on Physical and Engineering Sciences. American Statistical Association, Alexandria, VA, pp. 18–25.

Cook, R.D., 1998. Regression Graphics. Wiley, New York.

Cook, R.-D., Li, B., 2002. Dimension reduction for the conditional mean. Ann. Statist. 30, 455–474.

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., Ni, L., 2006. Using intraslice covariances for improved estimation of the central subspace in regression. Biometrika 93, 65–74.

Cook, R.D., Setodji, C.M., 2003. A model-free test for reduced rank in multivariate regression. J. Amer. Statist. Assoc. 98, 340–351.

Ebrahimi, N., Molefe, D., Ying, Z., 2003. Identifiability and censored data. Biometrika 90, 724–727.

Kent, J.T., 1991. Discussion of Li, K.C. (1991). Sliced inverse regression for dimension reduction (with discussion). J. Amer. Statist. Assoc. 86, 316–342. J. Amer. Statist. Assoc. 86, 336–337.

Setodji, C.M., Cook, R.D., 2004. K-means inverse regression. Technometrics 46, 421–429.

Tsiatis, A., 1975. A nonidentifiability aspect of the problem of competing risks. Proc. Nat. Acad. Sci. 72, 20–22.

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

Yoo, P., Cook, R.D., 2007. Optimal sufficient dimension reduction for the conditional mean in multivariate regression. Biometrika, to appear.

Zeng, D., 2004. Estimating marginal survival function by adjusting for dependent censoring using many covariates. Ann. Statist. 32, 1533–1555.