A note on moment-based sufficient dimension reduction estimators

Yuexiao Dong

The two main groups of moment-based sufficient dimension reduction methods are the estimators for the central space and the estimators for the central mean space. The former group includes methods such as sliced inverse regression, sliced average variance estimation and sliced average third-moment estimation, while ordinary least squares and principal Hessian directions belong to the latter group. We provide unified frameworks for each group of estimators in this short note. The central space estimators can be unified as inverse conditional cumulants, while Stein's Lemma is used to motivate the central mean space estimators.

Keywords and phrases: Central mean space, Central space, Conditional cumulants, Stein's Lemma.

1. INTRODUCTION

Suppose X is a p dimensional predictor and Y is a univariate response. Sufficient dimension reduction (SDR) (Li, 1991; Cook, 1994) aims to find subspace $\mathcal{S}$ such that

$$Y \perp\!\!\!\perp X | P_{\mathcal{S}} X, \tag{1}$$

where $\perp\!\!\!\perp$ means statistical independence, and $P_{(\cdot)}$ denotes the projection operator with respect to the standard inner product. Under mild conditions (Yin, Li and Cook, 2008), the intersection of all $\mathcal{S}$ satisfying (1) still satisfies (1). We call this intersection the central space of regressing Y on X and denote it by $\mathcal{S}_{Y|X}$. In the case when the focus is on the regression mean $E(Y|X)$, sufficient dimension reduction targets at finding the subspace $\mathcal{S}$ satisfying

$$E(Y|X) \perp\!\!\!\perp Y | P_{\mathcal{S}} X. \tag{2}$$

When the intersection of all $\mathcal{S}$ satisfying (2) still satisfies (2), we call it the central mean space (Cook and Li, 2002) and denote it by $\mathcal{S}_{E(Y|X)}$.

Let $\mu = E(X)$ and $\Sigma = \mathrm{Cov}(X)$. Then $Z = \Sigma^{-1/2}(X - \mu)$ denotes the standardized predictor. Due to the invariance property of the central space (Cook, 1998) and a similar property for the central mean space, we have $\mathcal{S}_{Y|X} = \Sigma^{-1/2}\mathcal{S}_{Y|Z}$ and $\mathcal{S}_{E(Y|X)} = \Sigma^{-1/2}\mathcal{S}_{E(Y|Z)}$. Hence one can identify the Z-scale central space or the Z-scale central mean space, and then transform them back to the X-scale spaces.

Without loss of generality, we work with the standardized predictor, and assume throughout this paper that $E(X) = 0$ and $\mathrm{Var}(X) = I$.

Two of the most popular central space estimators are sliced inverse regression (SIR; Li, 1991) and sliced average variance estimator (SAVE; Cook and Weisberg, 1991). To find additional structure in regression mixtures, Yin and Cook (2003) proposed sliced average third-moment estimation (SAT). By re-expressing these moments-based methods in terms of inverse conditional cumulants, we cast them under one unified framework. A nice byproduct is that the seemingly complicated SAT estimator is revealed to have a simple form as inverse conditional third-order cumulant.

Our second contribution is to provide a unified treatment of moment-based central mean space estimators through Stein's Lemma. While it is well-known that first-order and second-order Stein's Lemmas lead to ordinary least squares estimator (OLS; Li and Duan, 1989) and principal Hessian directions estimator (PHD; Li, 1992) respectively, we demonstrate that third-order Stein's Lemma can be used to motivate the illusive central mean space estimator proposed in Yin and Cook (2004).

The rest of the paper is organized as follows. We study inverse conditional cumulants for central space estimators in Section 2. We use Stein's Lemma to motivate central mean space estimators in Section 3. We conclude the paper with some discussions in Section 4.

2. INVERSE CONDITIONAL CUMULANTS AND ESTIMATORS OF THE CENTRAL SPACE

2.1 Review of estimators in $\mathcal{S}_{Y|X}$

Suppose β is a basis of $\mathcal{S}_{Y|X}$ and denote $\mathrm{Span}(\beta)$ as the column space of β. Let $P_\beta = \beta(\beta^T\beta)^{-1}\beta^T$ be the projection matrix onto $\mathrm{Span}(\beta)$ and $Q_\beta = I - P_\beta$ be the projection onto the orthogonal complement of $\mathrm{Span}(\beta)$. For the ease of reference, we state the following conditions commonly seen in the sufficient dimension reduction literature.

C1. $E(X|\beta^T X) = P_\beta X$;

C2. $\mathrm{Var}(X|\beta^T X) = Q_\beta$;

C3. $E(A \otimes AA^T|\beta^T X) = 0$, where $A = X - E(X|\beta^T X)$.

Here $\otimes$ in C3 denotes the Kronecker product. All three conditions are satisfied if the marginal distribution of X is normal.

The next proposition summarizes the common estimators in the central space, and we state it without proof. Please refer to Yin and Cook (2003) for details.

Proposition 1. *Suppose β is a basis of $\mathcal{S}_{Y|X}$.*

1. *If C1 is true, then $E(X|Y) \subseteq \mathcal{S}_{Y|X}$.*
2. *If C1 and C2 are true, then $\mathrm{Span}\{I - \mathrm{Var}(X|Y)\} \subseteq \mathcal{S}_{Y|X}$.*
3. *If C1, C2 and C3 are true, then $\mathrm{Span}\{E^T(B \otimes BB^T|Y)\} \subseteq \mathcal{S}_{Y|X}$, where $B = X - E(X|Y)$.*

The results above lead to the SIR estimator in Li (1991), the SAVE estimator in Cook and Weisberg (1991), and the SAT estimator in Yin and Cook (2003) respectively. Specifically, let $\mathcal{M}_{\mathrm{SIR}} = \mathrm{Var}\{E(X|Y)\}$, $\mathcal{M}_{\mathrm{SAVE}} = E[\{I - \mathrm{Var}(X|Y)\}^2]$ and $\mathcal{M}_{\mathrm{SAT}} = E\{E^T(B \otimes BB^T|Y)E(B \otimes BB^T|Y)\}$. Then the column space of $\mathcal{M}_{\mathrm{SIR}}$, $\mathcal{M}_{\mathrm{SAVE}}$ and $\mathcal{M}_{\mathrm{SAT}}$ can all be used to recover the central space $\mathcal{S}_{Y|X}$.

2.2 Inverse conditional cumulants

We review the definition of unconditional cumulants first. Recall that $X = (X_1, X_2, \ldots, X_p)^T \in \mathbb{R}^p$ is a random vector. Denote $\mu_r = E(X_r)$ as components of the mean vector $E(X)$, $\mu_{rs} = E(X_r X_s)$ as components of the second moment matrix, and $\mu_{rst} = E(X_r X_s X_t)$ as the third moments. For $\xi = (\xi_1, \ldots, \xi_p)^T$, the Taylor expansion of the moment generating function $M_X(\xi) = E\{\exp(\xi^T X)\}$ is

$$
\begin{aligned}
(3) \quad M_X(\xi) = {}& 1 + \sum_{r=1}^p \xi_r \mu_r + \frac{1}{2!} \sum_{r,s=1}^p \xi_r \xi_s \mu_{rs} \\
& + \frac{1}{3!} \sum_{r,s,t=1}^p \xi_r \xi_s \xi_t \mu_{rst} + \cdots.
\end{aligned}
$$

Consider the Taylor expansion of $\log M_X(\xi)$ as follows,

$$
\begin{aligned}
(4) \quad \log M_X(\xi) = {}& \sum_{r=1}^p \xi_r \kappa_r + \frac{1}{2!} \sum_{r,s=1}^p \xi_r \xi_s \kappa_{rs} \\
& + \frac{1}{3!} \sum_{r,s,t=1}^p \xi_r \xi_s \xi_t \kappa_{rst} + \cdots.
\end{aligned}
$$

The cumulants are then defined as the coefficients κ_r, κ_{rs} and κ_{rst} in (4). By comparing (3) and (4), one can follow McCullagh (1987) and establish the relationship between the first few moments and cumulants as

$$
\begin{aligned}
(5) \quad & \kappa_r = \mu_r, \\
& \kappa_{rs} = \mu_{rs} - \mu_r \mu_s, \\
& \kappa_{rst} = \mu_{rst} - \mu_{rs}\mu_t - \mu_{rt}\mu_s - \mu_{st}\mu_r + 2\mu_r \mu_s \mu_t.
\end{aligned}
$$

Conditional cumulants can be defined similarly. First denote inverse conditional moments $\mu_r^{X|Y} = E(X_r|Y)$, $\mu_{rs}^{X|Y} =$

$E(X_r X_s|Y)$, and $\mu_{rst}^{X|Y} = E(X_r X_s X_t|Y)$. Parallel to (5), we define inverse conditional cumulants as combinations of inverse conditional moments

$$
\begin{aligned}
(6) \quad & \kappa_r^{X|Y} = \mu_r^{X|Y}, \\
& \kappa_{rs}^{X|Y} = \mu_{rs}^{X|Y} - \mu_r^{X|Y}\mu_s^{X|Y}, \\
& \kappa_{rst}^{X|Y} = \mu_{rst}^{X|Y} - \mu_{rs}^{X|Y}\mu_t^{X|Y} - \mu_{rt}^{X|Y}\mu_s^{X|Y} \\
& \qquad - \mu_{st}^{X|Y}\mu_r^{X|Y} + 2\mu_r^{X|Y}\mu_s^{X|Y}\mu_t^{X|Y}.
\end{aligned}
$$

Recall that $\mu_r^{X|Y}$, $\mu_{rs}^{X|Y}$ and $\mu_{rst}^{X|Y}$ are essential components of $E(X|Y)$, $\mathrm{Var}(X|Y)$ and $E(B \otimes BB^T|Y)$ in Proposition 1. We now construct $\mathcal{K}_1(X|Y) \in \mathbb{R}^p$, $\mathcal{K}_2(X|Y) \in \mathbb{R}^{p \times p}$ and $\mathcal{K}_3(X|Y) \in \mathbb{R}^{p \times p \times p}$ as follows. Let the rth component of $\mathcal{K}_1(X|Y)$ be $\kappa_r^{X|Y}$, the $(r,s)^{th}$ component of $\mathcal{K}_2(X|Y)$ be $\kappa_{rs}^{X|Y}$, and the $(r,s,t)^{th}$ component $\mathcal{K}_3(X|Y)$ be $\kappa_{rst}^{X|Y}$.

In a parallel fashion, we can replace Y in the definitions above with $\beta^T X$ and define $\mathcal{K}_1(X|\beta^T X)$, $\mathcal{K}_2(X|\beta^T X)$ and $\mathcal{K}_3(X|\beta^T X)$. Before we present our main result, we state conditions C1, C2 and C3 in their equivalent forms

$$
\begin{aligned}
& \text{C1}'.\ \mathcal{K}_1(X|\beta^T X) = P_\beta X; \\
& \text{C2}'.\ \mathcal{K}_2(X|\beta^T X) = Q_\beta; \\
& \text{C3}'.\ \mathcal{K}_3(X|\beta^T X) = 0.
\end{aligned}
$$

For the ease of presentation, we also rearrange the third-order inverse cumulant array $\mathcal{K}_3(X|Y)$ as a $p^2 \times p$ matrix, with the $(t-1)p + s^{th}$ column of $\mathcal{K}_3^T(X|Y)$ being

$$
\begin{aligned}
(7) \quad \mathcal{K}_3^{st}(X|Y) = {}& E(XX_s X_t|Y) - E(XX_s|Y)E(X_t|Y) \\
& - E(XX_t|Y)E(X_s|Y) - E(X_s X_t|Y)E(X|Y) \\
& + 2E(X|Y)E(X_s|Y)E(X_t|Y).
\end{aligned}
$$

In another word, $\mathcal{K}_3^{st}(X|Y) = (\kappa_{1st}^{X|Y}, \ldots, \kappa_{pst}^{X|Y})^T$ for $s, t = 1, 2, \ldots, p$. Similarly, we rearrange the array $\mathcal{K}_3(X|\beta^T X)$ as a $p^2 \times p$ matrix. We now present SIR, SAVE and SAT as inverse conditional cumulants.

Proposition 2. *Suppose β is a basis of $\mathcal{S}_{Y|X}$.*

1. *If C1' is true, then $\mathcal{K}_1(X|Y) \subseteq \mathcal{S}_{Y|X}$.*
2. *If C1' and C2' are true, then $\mathrm{Span}\{I - \mathcal{K}_2(X|Y)\} \subseteq \mathcal{S}_{Y|X}$.*
3. *If C1', C2' and C3' are true, then $\mathrm{Span}\{\mathcal{K}_3^T(X|Y)\} \subseteq \mathcal{S}_{Y|X}$.*

PROOF. From (6), the definition of $\mathcal{K}_1(X|Y)$ and the definition of $\mathcal{K}_2(X|Y)$, we know $\mathcal{K}_1(X|Y) = E(X|Y)$ and $\mathcal{K}_2(X|Y) = E(XX^T|Y) - E(X|Y)E^T(X|Y) = \mathrm{Var}(X|Y)$. After replacing Y with $\beta^T X$, we see immediately that $C1$ is equivalent to $C1'$, and $C2$ is equivalent to $C2'$. The first two parts of Proposition 2 then follow directly from the first two parts of Proposition 1.

For part 3, denote $\mathcal{M}_B = E(B \otimes BB^T|Y)$ with $B = X - E(X|Y)$. From the definition of the Kronecker product, the $(t-1)p + s^{th}$ column of $\mathcal{M}_B^T$ is

$$\mathcal{M}_B^{st} = E[\{X - E(X|Y)\}\{X_s - E(X_s|Y)\}$$
$$(8) \qquad \{X_t - E(X_t|Y)\}|Y].$$

Compare (8) with the definition of $\mathcal{K}_3^{st}(X|Y)$ in (7). By noticing equalities such as

$$E\{XX_sE(X_t|Y)|Y\} = E(XX_s|Y)E(X_t|Y) \text{ and}$$
$$E\{XE(X_s|Y)E(X_t|Y)|Y\} = E(X|Y)E(X_s|Y)E(X_t|Y),$$

it is easy to see that $\mathcal{M}_B^{st} = \mathcal{K}_3^{st}(X|Y)$. It follows that $\mathcal{M}_B = \mathcal{K}_3(X|Y)$. Denote $\mathcal{M}_A = E(A \otimes AA^T|\beta^T X)$ with $A = X - E(X|\beta^T X)$. We can use a similar argument to show $\mathcal{M}_A = \mathcal{K}_3(X|\beta^T X)$, which guarantees the equivalence between $C3$ and $C3'$. Part 3 of Proposition 2 then follows from the last part of Proposition 1. □

The insight of Proposition 2 is that SIR, SAVE and SAT can be viewed as inverse conditional cumulants. In the case of SAT, condition C3 can be replaced with a simpler condition $\mathcal{K}_3(X|\beta^T X) = 0$, and its kernel matrix $E(B \otimes BB^T|Y)$ with $B = X - E(X|Y)$ becomes just the third-order inverse conditional cumulant array $\mathcal{K}_3(X|Y)$.

3. STEIN'S LEMMA AND ESTIMATORS OF THE CENTRAL MEAN SPACE

3.1 Review of estimators in $\mathcal{S}_{E(Y|X)}$

In the seminal paper of Li and Duan (1989), it was demonstrated that without knowing the underlying link function $f(\cdot)$, the classical ordinary least squares (OLS) estimator can recover the index β in the single-index model $Y = f(\beta^T X) + \epsilon$ correctly up to a constant multiplier. Here ϵ is the random error independent of X. In the modern language of Cook and Li (2002), we say that the OLS estimator belongs to the central mean space $\mathcal{S}_{E(Y|X)}$. The principal Hessian directions estimator in Li (1992), as well as the estimator in Yin and Cook (2004), also belong to the central mean space. Let vec(I) be the vectorization of the $p \times p$ identity matrix I. Denote $\mathcal{M}_{OLS} = \mathcal{M}_O \mathcal{M}_O^T$, $\mathcal{M}_{PHD} = \mathcal{M}_P \mathcal{M}_P^T$ and $\mathcal{M}_{YC} = \mathcal{M}_Y^T \mathcal{M}_Y$, where

$$\mathcal{M}_O = E(XY),$$
$$\mathcal{M}_P = E[\{Y - E(Y)\}XX^T] \text{ and}$$
$$(9) \qquad \mathcal{M}_Y = E[\{Y - E(Y)\}X \otimes XX^T] - E(YX) \otimes I$$
$$- I \otimes E(YX) - \text{vec}(I)E^T(YX).$$

We summarize these estimators in the next result without proof. Please refer to Yin and Cook (2004) for details. Note that in Section 2.1, the conditions C1, C2 and C3 are assumed to be true for the basis of the central space. In the next proposition, we assume conditions C1, C2 and C3 to be true for the basis of the central mean space.

Proposition 3. *Suppose β is a basis of $\mathcal{S}_{E(Y|X)}$.*

1. *If $C1$ is true, then* Span$(\mathcal{M}_{OLS}) \subseteq \mathcal{S}_{E(Y|X)}$.
2. *If $C1$ and $C2$ are true, then* Span$(\mathcal{M}_{PHD}) \subseteq \mathcal{S}_{E(Y|X)}$.

3. *If $C1$, $C2$ and $C3$ are true, then* Span$(\mathcal{M}_{YC}) \subseteq \mathcal{S}_{E(Y|X)}$.

Although the OLS and PHD estimators have relatively simple forms, the estimator $\mathcal{M}_{YC}$ is not easy to come up with. The original derivation in Yin and Cook (2004) actually starts from the SAT estimator, and then gets $\mathcal{M}_{YC}$ after carefully delineating the components of SAT that belong to the central space. Just as we have seen in Section 2.2 that the third-order central space estimator SAT can be unified with the first-order method SIR and the second-order method SAVE, we will cast central mean space estimators $\mathcal{M}_{OLS}$, $\mathcal{M}_{PHD}$ and $\mathcal{M}_{YC}$ in one unified framework.

3.2 Stein's Lemma

Recall that for $X = (X_1, X_2, \ldots, X_p)^T$, we have assumed that $E(X) = 0$ and Var$(X) = I$ from the outset. To motivate central mean space estimators via Stein's Lemma, we further assume in this section that $X_i \sim N(0, 1)$ for $i = 1, 2, \ldots, p$ and $X_i \perp\!\!\!\perp X_j$ for $i \neq j$. In the famous paper of Stein (1981), it is demonstrated that for any function $g(X_i)$ with first through third order derivatives denoted by $g'(X_i)$, $g''(X_i)$ and $g'''(X_i)$, we have

$$E\{g'(X_i)\} = E\{X_i g(X_i)\},$$
$$(10) \qquad E\{g''(X_i)\} = E\{(X_i^2 - 1)g(X_i)\}, \text{ and}$$
$$E\{g'''(X_i)\} = E\{(X_i^3 - 3X_i)g(X_i)\}.$$

Consider $g(X) = E(Y|X)$ as a function of X. Denote partial derivatives $d_r = \partial g(X)/\partial X_r$, $d_{rs} = \partial^2 g(X)/\partial X_r \partial X_s$ and $d_{rst} = \partial^3 g(X)/\partial X_r \partial X_s \partial X_t$. We construct $\mathcal{D}_1 \in \mathbb{R}^p$, $\mathcal{D}_2 \in \mathbb{R}^{p \times p}$ and $\mathcal{D}_3 \in \mathbb{R}^{p \times p \times p}$ as follows. Let the rth component of $\mathcal{D}_1$ be $E(d_r)$, the $(r, s)^{th}$ component of $\mathcal{D}_2$ be $E(d_{rs})$, and the $(r, s, t)^{th}$ component $\mathcal{D}_3$ be $E(d_{rst})$. For the ease of presentation, we rearrange the third-order derivatives array $\mathcal{D}_3$ as a $p^2 \times p$ matrix, where the entry in the rth column and $(t-1)p + s^{th}$ row is $E(d_{rst})$.

The next result combines central mean space estimators $\mathcal{M}_{OLS}$, $\mathcal{M}_{PHD}$ and $\mathcal{M}_{YC}$ in one unified framework.

Proposition 4. *Suppose $X \sim N(0, I)$.*

1. $\mathcal{D}_1 = \mathcal{M}_O$.
2. $\mathcal{D}_2 = \mathcal{M}_P$.
3. $\mathcal{D}_3 = \mathcal{M}_Y$.
4. Span$(\mathcal{D}_1) \subseteq \mathcal{S}_{E(Y|X)}$, Span$(\mathcal{D}_2) \subseteq \mathcal{S}_{E(Y|X)}$ and Span$(\mathcal{D}_3^T) \subseteq \mathcal{S}_{E(Y|X)}$.

PROOF. From the first equation in (10), we know $E\{\partial E(Y|X)/\partial X_r\} = E\{X_r E(Y|X)\} = E(X_r Y)$. The left-hand side and the right-hand side are the rth entry of $\mathcal{D}_1$ and $\mathcal{M}_O$ respectively. Thus we have $\mathcal{D}_1 = \mathcal{M}_O$.

To find the $(r, s)^{th}$ component of $\mathcal{D}_2$, we need to treat the two cases $r \neq s$ and $r = s$ separately. If $r \neq s$, we have

$$E\{\partial^2 E(Y|X)/\partial X_r \partial X_s\} = E[X_r\{\partial E(Y|X)/\partial X_s\}]$$
$$= E\{X_r X_s E(Y|X)\} = E(X_r X_s Y)$$
$$= E[\{Y - E(Y)\}X_r X_s],$$

where the last equality is true because $E(X_r X_s) = 0$. If $r = s$, we use the second equation in (10) and get

$$E\{\partial^2 E(Y|X)/\partial X_r^2\} = E\{(X_r^2 - 1)E(Y|X)\}$$
$$= E\{(X_r^2 - 1)Y\} = E[\{Y - E(Y)\}X_r^2],$$

where the last equality is true because $E(X_r^2) = 1$. Together with the definition of $\mathcal{M}_P$ in (9), we have $\mathcal{D}_2 = \mathcal{M}_P$.

To find the $(r, s, t)^{th}$ component of $\mathcal{D}_3$, there are three separate cases. When r, s and t are all distinct, we have

$$(11) \qquad E\{\partial^3 E(Y|X)/\partial X_r \partial X_s \partial X_t\} = E(X_r X_s X_t Y).$$

When two of r, s and t are the same, we assume without loss of generality that $r = s \neq t$. Then we have

$$(12) \qquad E\{\partial^3 E(Y|X)/\partial X_r^2 \partial X_t\} = E\{(X_r^2 - 1)X_t Y\}.$$

When $r = s = t$, we use the third equation in (10) and get

$$(13) \qquad E\{\partial^3 E(Y|X)/\partial X_r^3\} = E\{(X_r^3 - 3X_r)Y\}.$$

Now we turn to $\mathcal{M}_Y = E[\{Y - E(Y)\}X \otimes XX^T] - E(YX) \otimes I - I \otimes E(YX) - \text{vec}(I)E^T(YX) \in \mathbb{R}^{p^2 \times p}$ defined in (9), and denote its entry in the rth column and $(t-1)p + s^{th}$ row as $\mathcal{M}_Y(r, s, t)$. For the ease of presentation, we view $\mathcal{M}_Y$ as a $p \times p \times p$ array, where the $(r, s, t)^{th}$ entry is $\mathcal{M}_Y(r, s, t)$.

When r, s and t are all distinct, the last three terms in the definition of $\mathcal{M}_Y$ all become 0, and the $(r, s, t)^{th}$ component of $\mathcal{M}_Y$ becomes $E[\{Y - E(Y)\}X_r X_s X_t]$, which matches the right-hand side of (11) as $E(X_r X_s X_t) = 0$. When $r = s = t$, all four terms in the definition of $\mathcal{M}_Y$ remain. The $(r, s, t)^{th}$ component of $\mathcal{M}_Y$ becomes $E[\{Y - E(Y)\}X_r^3] - 3E(YX_r)$, which matches the right-hand side of (13) by noting that $E(X_r^3) = 0$. When two of r, s and t are the same, only one of the last three terms in the definition of $\mathcal{M}_Y$ remains. For instance, if we have $r = s \neq t$, the $(r, s, t)^{th}$ component of $\mathcal{M}_Y$ becomes $E[\{Y - E(Y)\}X_r^2 X_t] - E(YX_t)$, which matches the right-hand side of (12) because $E(X_r^2 X_t) = 0$. Together we have shown $\mathcal{D}_3 = \mathcal{M}_Y$.

It remains to prove part 4. Let β be a basis of $\mathcal{S}_{E(Y|X)}$. From the definition of the central mean space in (2), we know $E(Y|X) = E(Y|\beta^T X)$. As a result,

$$\frac{\partial E(Y|X)}{\partial X} = \frac{\partial E(Y|\beta^T X)}{\partial X} = \frac{\partial \beta^T X}{\partial X} \frac{\partial E(Y|\beta^T X)}{\partial \beta^T X}.$$

Because $\partial \beta^T X / \partial X = \beta$, we have $\text{Span}(\mathcal{D}_1) = \text{Span}(E\{\partial E(Y|X)/\partial X\}) \subseteq \text{Span}(\beta) = \mathcal{S}_{E(Y|X)}$. Following similar arguments, we have $\text{Span}(\mathcal{D}_2) \subseteq \mathcal{S}_{E(Y|X)}$ and $\text{Span}(\mathcal{D}_3^T) \subseteq \mathcal{S}_{E(Y|X)}$. $\square$

Although we assume normality to facilitate the connections between the central mean space estimators, we have seen in Proposition 3 that the normality assumption can be replaced with weaker conditions C1, C2 and C3. The insight of Proposition 4 is that just as OLS and PHD can be motivated from first and second-order partial derivatives of $E(Y|X)$, we can use the third-order partial derivatives of $E(Y|X)$ to construct the much more complicated estimator in Yin and Cook (2004).

4. DISCUSSIONS

We provide a general treatment of moment-based sufficient dimension reduction estimators in this note. On one hand, popular central space estimators such as SIR, SAVE and SAT are cast under the unified framework of inverse conditional cumulants. On the other hand, Stein's Lemma is used to motivate central mean space estimators such as OLS, PHD and the estimator in Yin and Cook (2004). Although the focus of the paper is to reveal connections between existing moment-based estimators, we provide possibilities to come up with new sufficient dimension reduction estimators. For example, one can follow the principle of Proposition 4, and construct new estimators in the central mean space based on fourth-order partial derivatives of $E(Y|X)$.

Stein's Lemma is not limited to motivate estimators in the central mean space. Actually one can construct a general family of sufficient dimension reduction estimators through Stein's Lemma. Instead of choosing $g(X)$ to be $E(Y|X)$, we can set $g(X) = E(Y^k|X)$ for any integer k and apply Stein's Lemma. This will lead to the central kth moment space estimators studied in Yin and Cook (2002). An interesting case here is that when we choose $g(X) = E(e^{iY}|X)$ and apply Stein's Lemma, we will get estimators in the central space (Zhu, Zhu and Wen, 2010).

REFERENCES

Cook, R. D. (1994). On the interpretation of regression plots. *Journal of the American Statistical Association* **89**, 177–189. MR1266295

Cook, R. D. (1998). *Regression Graphics*. New York: Wiley. MR1645673

Cook, R. D. and Li, B. (2002). Dimension reduction for conditional mean in regression. *The Annals of Statistics* **30**, 455–474. MR1902895

Cook, R. D. and Weisberg, S. (1991). Discussion of "Sliced inverse regression for dimension reduction". *Journal of the American Statistical Association* **86**, 316–342. MR1137117

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 316–342. MR1137117

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: another application of Stein's Lemma. *Journal of the American Statistical Association* **87**, 1025–1039. MR1209564

Li, K. C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics* **17**, 1009–1052. MR1015136

McCullagh, P. (1987). *Tensor Methods in Statistics*. London: Chapman and Hall. MR0907286

Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics* **9**, 1135–1151. MR0630098

Yin, X. and Cook, R. D. (2002). Dimension reduction for the conditional kth moment in regression. *Journal of the Royal Statistical Society Ser. B* **64**, 159–175. MR1904698

YIN, X. and COOK, R. D. (2003). Estimating the central subspace via inverse third moment. *Biometrika* **90**, 113–125. MR1966554

YIN, X. and COOK, R. D. (2004). Dimension reduction via marginal fourth moments in regression. *Journal of Computational and Graphical Statistics* **13**, 554–570. MR2087714

YIN, X., LI, B. and COOK, R. D. (2008). Successive direction extraction for estimating the central subspace in a multiple-index regression. *Journal of Multivariate Analysis* **99**, 1733–1757. MR2444817

ZHU, L. P., ZHU, L. X. and WEN, S. (2010). On dimension reduction in regressions with multivariate response. *Statistica Sinica* **20**, 1291–1307. MR2730185

Yuexiao Dong
Temple University
Philadelphia, PA, 19122
USA
E-mail address: ydong@temple.edu