Statistica Sinica Preprint No: SS-2019-0388	
Title	A METHOD OF LOCAL INFLUENCE ANALYSIS IN SUFFICIENT DIMENSION REDUCTION
Manuscript ID	SS-2019-0388
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202019.0388
Complete List of Authors	Fei Chen, Lei Shi, Lin Zhu and Lixing Zhu
Corresponding Author	Lixing Zhu
E-mail	lzhu@hkbu.edu.hk

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

A METHOD OF LOCAL INFLUENCE ANALYSIS IN SUFFICIENT DIMENSION REDUCTION

Fei Chen[1], Lei Shi[1], Lin Zhu[1,2] and Lixing Zhu[3*]

[1] *Yunnan University of Finance and Economics,*

[2] *Tiantai County Branch of the People's Bank of China,*

[3] *Hong Kong Baptist University*

Abstract: A general framework for a local influence analysis is developed for sufficient dimension reduction when the data likelihood is absent and the inference result is a space rather than a vector. A clear and intuitive interpretation of this approach is described. Its application to the sliced inverse regression is presented, together with its invariance properties. A data trimming strategy is also suggested, based on the influence assessment for observations provided by our method. A simulation study and a real-data analysis are presented. The results indicate that the local influence analysis avoids the masking effect, and that the data trimming provides a substantial increase in the inference accuracy.

Key words and phrases: Central subspace, displacement function, influence measure, perturbation scheme.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

1. Introduction

In nonparametric regression modeling, if the response is related to all predictors through a small number of their linear combinations, determining these linear combinations without loss of information on the response will help enhance inference efficiency and model visualization. This is the goal of sufficient dimension reduction (Li, 1991; Cook, 1998a), which includes methods such as the sliced inverse regression (Li, 1991), sliced average variance estimation (Cook and Weisberg, 1991), principal Hessian directions (Li, 1992), directional regression (Li and Wang, 2007), discretization-expectation estimation (Zhu et al., 2010a), cumulative slicing estimation (Zhu et al., 2010b), gradient-based methods (Xia, 2007; Zhu and Zeng, 2006; Wang and Xia, 2008; Yin and Li, 2011), and semiparametric method proposed by Ma and Zhu (2012), among others.

In contrast to the issue of fitting a regression function, the majority of the above dimension-reduction methods strongly depend on assumptions about the distribution of the predictor vector. For example, the method of principal Hessian directions needs the normality assumption, and the sliced inverse regression is based on the linearity condition, which is only slightly weaker than the elliptical symmetry of distribution. To handle more general distributions, new methods have been proposed by, among others, Li

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

and Dong (2009), Dong and Li (2010), and Guan et al. (2017). Note that these methods are designed to deal with an overall violation of the desired distributional structure, for example, skewness of the joint distribution of the predictors. In practice, there is a challenging issue related to observations with outlying values of the predictor vector, called high-leverage data points. These outliers may be extreme values that still follow the designed distribution, or could be genuine outliers that make the underlying distribution violate the distributional assumption. Even the former may break the symmetry of the distribution of the sample data, which is important for many sufficient dimension reduction methods (e.g., the sliced inverse regression). Although high-leverage data points are not always "bad" data points in other regression modeling procedures, they are in terms of sufficient dimension reduction. Moreover, outlying values of errors in the regression may impair the inference accuracy on dimension reduction. The effects of an observation may be due to one of the above causes, or both. In general, the threat of outliers is non-ignorable. As we show in our simulation studies, changing the values of several observations can lead to a sharp decrease in the estimation accuracy for the dimension reduction subspace and structural dimension when a sliced inverse regression is used in some classical scenarios. This issue has received some attention; see Cook and

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Critchley (2000), Gather et al. (2001), and Gather et al. (2002), as well as Cook and Nachtsheim (1994), who proposed a re-weighting method to practically achieve elliptically contoured predictors, and Zhou et al. (2015), who dealt with contaminated data. We focus on detecting and handling outliers during the sufficient dimension reduction procedure.

By simultaneously considering the above two causes of outlyingness, an influence analysis provides us with an integrated assessment of the effects of observations. Influential observations, whether outlying on predictors or errors, are problematic in sufficient dimension reduction. Thus, finding and removing them from the data set may be a feasible and parsimonious strategy to avoid the problems caused by a small proportion of outliers. For example, for the method of principal Hessian directions, Prendergast (2008) and Lue (2001) proposed useful strategies of trimming influential observations based on case-deletion diagnostics. Of course, trimming data may result in some information loss, owing to the risk of misspecification, which may be the price of pursuing robustness. In sufficient dimension reduction, existing influence assessment methods are primarily case-deletion methods based on the influence function; see Prendergast (2006, 2007) and Prendergast and Smith (2010), among others. However, case-deletion methods are not efficient in some cases, particularly when a masking effect exists.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Here, a masking effect means that the effect of an influential observation may be underestimated because of the existence of other influential observations, especially when their positions are close to each other. Hence, a local influence analysis for sufficient dimension reduction is required. This analysis introduces a perturbation vector, with each entry perturbing an object of interest, say an observation, and investigates the change of inference. In general, the method avoids masking effects due to simultaneous perturbations.

Numerous works have examined local influence analyses; see, for example, Cook (1986), Shi (1997), Zhu and Lee (2001) and Zhu et al. (2007). However, no existing method can be used directly for sufficient dimension reduction, owing to two of its features. First, there is no data likelihood. Second, the statistic of interest is a space rather than a vector. Therefore, we attempt to develop a general framework for local influence analysis for sufficient dimension reduction. A clear and intuitive interpretation of this approach is described. As an application, we implement an influence analysis of a sliced inverse regression and show the invariance properties. We also propose a strategy for data trimming based on our influence assessment. A simulation study and a real-data analysis are presented to illustrate the proposed methodologies.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

The remainder of the paper is organized as follows. Section 2 introduces sufficient dimension reduction and the sliced inverse regression method. The main methodology and results are presented in Sections 3 to 6. Section 7 reports a real-data example for illustration. Section 8 concludes the paper. All technical proofs and simulation studies are provided in the Supplementary Material. The assumptions are labeled.

2. Central subspace and sliced inverse regression

Let Y and $\mathbf{x}$ denote the response and a $p \times 1$ random predictor vector, respectively, in a regression. A subspace $\mathcal{M}(\mathbf{A})$ spanned by the columns of a matrix $\mathbf{A}$ is called a dimension reduction subspace if $F(y|\mathbf{x}) = F(y|\mathbf{A}^{\mathrm{T}}\mathbf{x})$, for $y \in \mathcal{R}$, where $F(y|\mathbf{x})$ denotes the conditional distribution function of Y given $\mathbf{x}$. The intersection of all such subspaces is called the central subspace (Cook, 1998b), $\mathcal{B} \subset \mathcal{R}^p$, if it is still a dimension reduction subspace. The dimension K of the central subspace and the vectors it contains are called the structural dimension and dimension reduction vectors, respectively. Henceforth, we focus on local influence analyses for the methods in order to estimate the central subspace. However, the proposed methodology is also applicable to other sufficient dimension reduction methods for the so-called central mean subspace (Cook and Li, 2002) if we are interested

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

in the dimension reduction subspace spanned by the columns in the mean function of a regression model.

We now very briefly describe the sliced inverse regression proposed by Li (1991), one of the most promising methods for estimating the central subspace $\mathcal{B}$.

Assumption 1. (Linear design condition) For any $\mathbf{b} \in \mathcal{R}^p$, there exist some constants $c_0, c_1, \ldots, c_K$, such that $\mathrm{E}(\mathbf{b}^\top \mathbf{x} \mid \boldsymbol{\beta}_1^\top \mathbf{x}, \ldots, \boldsymbol{\beta}_K^\top \mathbf{x}) = c_0 + c_1 \boldsymbol{\beta}_1^\top \mathbf{x} + \cdots + c_K \boldsymbol{\beta}_K^\top \mathbf{x}$, where $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_K$ denote a basis of $\mathcal{B}$.

The sliced inverse regression is based on the key fact that $\mathrm{E}(\mathbf{x} \mid Y) - \mathrm{E}(\mathbf{x})$ is contained in $\boldsymbol{\Sigma}_{\mathbf{x}}\mathcal{B}$ under assumption 1, where $\mathbf{C}\mathcal{A}$ denotes $\{\mathbf{C}\boldsymbol{\zeta} : \boldsymbol{\zeta} \in \mathcal{A}\}$ for any matrix $\mathbf{C}$ and subspace $\mathcal{A}$. Divide the range of Y into τ slices, $\mathcal{S}_1, \ldots, \mathcal{S}_\tau$, and let $\boldsymbol{\Sigma}_{\mathbf{x}} = \mathrm{cov}(\mathbf{x})$ and $\boldsymbol{\Sigma}_\eta = \sum_{l=1}^{\tau} \mathrm{P}(Y \in \mathcal{S}_l)\mathrm{E}(\mathbf{x} \mid Y \in \mathcal{S}_l)\mathrm{E}(\mathbf{x} \mid Y \in \mathcal{S}_l)^{\mathrm{T}} - \mathrm{E}(\mathbf{x})\mathrm{E}(\mathbf{x})^{\mathrm{T}}$. Then, $\mathcal{M}(\boldsymbol{\Sigma}_\eta) \subset \boldsymbol{\Sigma}_{\mathbf{x}}\mathcal{B}$, from the above fact and the equality $\mathrm{E}(\mathbf{x} \mid Y \in \mathcal{S}_l) = \mathrm{E}\{\mathrm{E}(\mathbf{x} \mid Y) \mid Y \in \mathcal{S}_l\}$. Let K^* be $\mathrm{rk}(\boldsymbol{\Sigma}_{\mathbf{x}}^{-1}\boldsymbol{\Sigma}_\eta)$ and $\mathbf{b}_1, \ldots, \mathbf{b}_{K^*}$ denote orthonormal eigenvectors of $\boldsymbol{\Sigma}_\eta$ with respect to $\boldsymbol{\Sigma}_{\mathbf{x}}$ associated with nonzero eigenvalues, where $\mathrm{rk}(\mathbf{A})$ denotes the rank of $\mathbf{A}$. That is, $\boldsymbol{\Sigma}_\eta \mathbf{b}_k = \lambda_k \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{b}_k$, for $k = 1, \ldots, K^*$, and $\mathbf{B}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathbf{x}} \mathbf{B} = \mathbf{I}$, where $\lambda_1 \geq \cdots \geq \lambda_{K^*} > 0$, $K^* \leq K$, $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_{K^*})$, and $\mathbf{I}$ denotes a $K^* \times K^*$ identity matrix. Then, $\mathbf{b}_1, \ldots, \mathbf{b}_{K^*}$ are dimension reduction vectors. To estimate these vectors, we need to estimate $\boldsymbol{\Sigma}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_\eta$. Let

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

$(y_1, \mathbf{x}_1^{\mathrm{T}}), \ldots, (y_n, \mathbf{x}_n^{\mathrm{T}})$ be n observations of $(Y, \mathbf{x}^{\mathrm{T}})$. The matrix $\boldsymbol{\Sigma}_{\mathbf{x}}$ can be estimated as $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}$, where $\bar{\mathbf{x}} = n^{-1}\sum_{i=1}^{n}\mathbf{x}_i$. Let $\mathcal{I}_l$ and n_l denote the index set and the number of y_i in $\mathcal{S}_l$, respectively. Then, we estimate $\boldsymbol{\Sigma}_{\eta}$ as $\hat{\boldsymbol{\Sigma}}_{\eta} = n^{-1}\sum_{l=1}^{\tau} n_l(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})(\bar{\mathbf{x}}_l - \bar{\mathbf{x}})^{\mathrm{T}}$, where $\bar{\mathbf{x}}_l = n_l^{-1}\sum_{i\in\mathcal{I}_l}\mathbf{x}_i$ is the lth slice mean of $\mathbf{x}$. Hence, $\mathbf{b}_k$ can be estimated using $\hat{\mathbf{b}}_k$, which satisfies $\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\mathbf{b}}_k = \hat{\lambda}_k\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\hat{\mathbf{b}}_k$, with $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_p$, and $\mathcal{B}$ can be estimated as $\hat{\mathcal{B}} = \mathcal{M}(\hat{\mathbf{B}})$, where $\hat{\mathbf{B}} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_{\hat{K}})$ and $\hat{K}$ is the estimate of K.

For estimating K, Li (1991) proposed a sequential testing procedure in which the hypothesis $K \leq k$ is rejected if $\sum_{i=k+1}^{p}(n\hat{\lambda}_i) > \chi^2_{\alpha_k}\{(p-k)(\tau-k-1)\}$, where α_k is the test level given beforehand, and the cutoff point is the α_k upper quantile of the χ^2 distribution with degrees of freedom $(p-k)(\tau-k-1)$. Zhu et al. (2006) proposed a Bayesian information criterion method for constructing a consistent estimate. Following their idea, we define the estimator of K as $\hat{K}$, which satisfies $G(\hat{K}) = \max_{0\leq k\leq p-1} G(k)$, where $G(k) = \log L_k - P(k)$, with $\log L_k = \sum_{i=1+\min(\nu,k)}^{p} n\{\log(\hat{\lambda}_i + 1) - \hat{\lambda}_i\}/2$ and $P(k) = C_n k(2p-k+1)/2$. Here, ν denotes the number of $\hat{\lambda}_i$ that are positive, and C_n is a penalty constant given beforehand.

Hereinafter, regardless of the method used, we always let $\hat{K}$ and $\hat{\mathcal{B}}$ be the estimates of K and $\mathcal{B}$, respectively, and let $\hat{\mathbf{B}} = (\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_{\hat{K}})$, with

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

$\hat{\mathbf{b}}_1, \ldots, \hat{\mathbf{b}}_{\hat{K}}$ being the dimension reduction vector estimates, a basis of $\hat{\mathcal{B}}$.

3. A general framework for local influence analysis

Let the vector $\boldsymbol{\omega}_{s\times 1}$ denote the perturbation introduced to the model, and let $\hat{\mathbf{b}}_1(\boldsymbol{\omega}), \ldots, \hat{\mathbf{b}}_{\hat{K}}(\boldsymbol{\omega})$ be the estimates of the dimension reduction vectors under the perturbed model. The scheme to introduce $\boldsymbol{\omega}$ is called a perturbation scheme, where each entry of $\boldsymbol{\omega}$ is associated with an influence measure, say, an observation. We specify perturbation schemes for influence measures of observations in Section 5. The following general framework applies to any appropriate perturbation scheme. Let $\hat{\mathcal{B}}(\boldsymbol{\omega}) = \mathcal{M}\{\hat{\mathbf{B}}(\boldsymbol{\omega})\}$ denote the estimated central subspace under a perturbation, where $\hat{\mathbf{B}}(\boldsymbol{\omega}) = (\hat{\mathbf{b}}_1(\boldsymbol{\omega}), \ldots, \hat{\mathbf{b}}_{\hat{K}}(\boldsymbol{\omega}))$. Moreover, let $\boldsymbol{\omega}_{(0)}$ stand for no perturbation; that is, the model is not perturbed when $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)}$. Clearly, $\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)}) = \hat{\mathbf{B}}$ and $\hat{\mathcal{B}}(\boldsymbol{\omega}_{(0)}) = \hat{\mathcal{B}}$. First, to measure the discrepancy between the subspaces $\hat{\mathcal{B}}$ and $\hat{\mathcal{B}}(\boldsymbol{\omega})$, we construct a space displacement function $D(\boldsymbol{\omega})$ based on the trace correlation (Hooper, 1959):

$$D(\boldsymbol{\omega}) = 1 - \frac{1}{\hat{K}}\mathrm{tr}(\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathcal{B}}}\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathcal{B}}(\boldsymbol{\omega})}), \tag{3.1}$$

where $\mathbf{Z}$ is a $p\times n$ matrix, with the ith column $\mathbf{z}_i = \mathbf{x}_i - \bar{\mathbf{x}}$, and $\mathbf{P}_{\mathcal{A}}$ denotes the orthogonal projection matrix on $\mathcal{A}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Let $\bar{r}^2(\boldsymbol{\omega}) = \sum_{i=1}^{\hat{K}} r_i^2/\hat{K}$ denote the square of the trace correlation (Hooper, 1959) of $\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{x}$, as explained by $\hat{\mathbf{B}}(\boldsymbol{\omega})^{\mathrm{T}}\mathbf{x}$, where r_i denotes the ith canonical correlation between $\hat{\mathbf{B}}^{\mathrm{T}}\mathbf{x}$ and $\hat{\mathbf{B}}(\boldsymbol{\omega})^{\mathrm{T}}\mathbf{x}$. Then, $D(\boldsymbol{\omega}) = 1 - \bar{r}^2(\boldsymbol{\omega})$. Because the purpose of sufficient dimension reduction is to find the variables $\mathbf{b}^{\mathrm{T}}\mathbf{x}, \mathbf{b} \in \mathcal{B}$, the space displacement $D(\boldsymbol{\omega})$ is designed to take the variance-covariance structure of $\mathbf{x}$ into account. It possesses the following two properties: (i) $0 \leq D(\boldsymbol{\omega}) \leq 1$; and (ii) $D(\boldsymbol{\omega})$ attains its minimum at $\boldsymbol{\omega}_{(0)}$. Property (i) is illustrated by Hooper (1959), and (ii) can be derived from (i) with the fact that $D(\boldsymbol{\omega}_{(0)}) = 0$.

In our methodology, the space displacement $D(\boldsymbol{\omega})$ plays an important role, similar to that of the likelihood displacement in Cook (1986). The geometric surface $(\boldsymbol{\omega}^{\mathrm{T}}, D(\boldsymbol{\omega}))^{\mathrm{T}}$ is called an influence graph, from which we attempt to draw information on the local influence of the perturbation $\boldsymbol{\omega}$ around $\boldsymbol{\omega}_{(0)}$. The bottom of the influence graph is the point $(\boldsymbol{\omega}_{(0)}^{\mathrm{T}}, D(\boldsymbol{\omega}_{(0)}))^{\mathrm{T}}$. Let $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, with $\mathbf{h}^{\mathrm{T}}\mathbf{h} = 1$. For a given standardized vector $\mathbf{h}$, the graph of $\mathcal{L}(\mathbf{h}) = \{((\boldsymbol{\omega}_{(0)} + t\mathbf{h})^{\mathrm{T}}, D(\boldsymbol{\omega}_{(0)} + t\mathbf{h}))^{\mathrm{T}} : t \in \mathcal{R}^1\}$ is a curve on the influence graph, called the lifted line of the influence graph along direction $\mathbf{h}$, passing through $(\boldsymbol{\omega}_{(0)}^{\mathrm{T}}, D(\boldsymbol{\omega}_{(0)}))^{\mathrm{T}}$. Note that $(\boldsymbol{\omega}_{(0)}^{\mathrm{T}}, D(\boldsymbol{\omega}_{(0)}))^{\mathrm{T}}$ is the common bottom of all lifted lines. The local behavior of the lifted line $\mathcal{L}(\mathbf{h})$ around $\boldsymbol{\omega}_{(0)}$ reveals information about the local influence of the

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

perturbation $\boldsymbol{\omega}$ along $\mathbf{h}$ on the estimate of the central subspace, although $D(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is not necessarily second-order differentiable at $t = 0$ for each direction $\mathbf{h}$. We now attempt to find a statistic that represents the influence of the perturbation along $\mathbf{h}$.

By the coordinate system rotation, the lifted line $\mathcal{L}(\mathbf{h})$ can be regarded as a plain curve with the expression $\{(t, D(\boldsymbol{\omega}_{(0)} + t\mathbf{h}))^{\mathrm{T}} : t \in \mathcal{R}^1\}$ in rotated coordinates. Inspired by Cook (1986), we now investigate the local behavior of the function $D(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ at $t = 0$ using an expansion of $D(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$. The following expanding expressions, in which the random observations $(\mathbf{x}_1^\top, y_1)^\top, \ldots, (\mathbf{x}_n^\top, y_n)^\top$ are regarded as given, are not asymptotic, but instead are perturbation expressions with $o(t)$ and $o(t^2)$ unrelated to the sample size.

Assumption 2. (1)$\mathrm{rk}(\mathbf{Z}) = p$; (2)for any given $\mathbf{h}$, $\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})$ is continuous in a neighborhood of $t = 0$; (3)there is a matrix $\mathbf{F}_{\mathbf{B},\mathbf{h}}$, such that

$$\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \hat{\mathbf{B}} + t\mathbf{F}_{\mathbf{B},\mathbf{h}} + o(t). \tag{3.2}$$

Then, the following holds (see part S1 of the Supplementary Material for the proof):

$$D(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \frac{1}{2}t^2 \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}})^{\mathrm{T}} \left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A}) \partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}) + o(t^2), \tag{3.3}$$

where $d(\mathbf{A}) = 1 - \mathrm{tr}(\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\mathbf{A}})/\hat{K}$, $\mathbf{P}_{\mathbf{C}} = \mathbf{C}(\mathbf{C}^\top\mathbf{C})^{-1}\mathbf{C}^\top$, and $\mathrm{vec}(\mathbf{A}) =$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

$(\mathbf{a}_1^{\mathrm{T}}, \ldots \mathbf{a}_u^{\mathrm{T}})^{\mathrm{T}}$, with $\mathbf{a}_i$ being the ith column vector of $\mathbf{A}$. Recall that in Cook (1986), the lifted line

$$\textsc{ld}(\boldsymbol{\omega}_{(0)} + t\mathbf{h}) = \frac{1}{2}t^2 C_{\mathbf{h}} + o(t^2),$$

where the displacement function $\textsc{ld}(\boldsymbol{\omega})$ is based on the likelihood and assumed to be second-order differentiable with respect to $\boldsymbol{\omega}$, and $C_{\mathbf{h}}$ is the normal curvature employed for influence assessment. Hence, for $D(\boldsymbol{\omega})$, we define the quasi-curvature of the lifted line along $\mathbf{h}$ at $\boldsymbol{\omega}_{(0)}$ as

$$\textsc{qc}_{\mathbf{h}} = \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}})^{\mathrm{T}} \left. \frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A}) \partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}} \right|_{\mathbf{A}=\hat{\mathbf{B}}} \mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}}),$$

which is a statistic that measures the influence of the perturbation along $\mathbf{h}$. This quasi-curvature is an analogue of the normal curvature $C_{\mathbf{h}}$ defined in Cook (1986). In fact, $\textsc{qc}_{\mathbf{h}}$ is exactly the curvature of the lifted line $\mathcal{L}(\mathbf{h})$ at $t = 0$ if $D(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is second-order differentiable at $t = 0$. However, we do not assume the existence of the curvature because $\hat{\mathbf{B}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is not necessarily smooth enough for each $\mathbf{h}$. That is why we call $\textsc{qc}_{\mathbf{h}}$ quasi-curvature, rather than curvature. Furthermore, we define the influential direction as

$$\mathbf{h}_{\max} = \arg \max_{\mathbf{h}^{\mathrm{T}}\mathbf{h}=1} \textsc{qc}_{\mathbf{h}}.$$

This direction is an important diagnostic that indicates how to perturb the model to produce the strongest local influence on the central subspace

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

estimate. Hence, for $i = 1, \ldots, s$, the absolute value of the ith element of $\mathbf{h}_{\max}$ is used as the influence measure for the aspect perturbed by ω_i in the model. For example, to assess the influence of observations, we design a perturbation scheme in which the ith entry of $\boldsymbol{\omega}$ is associated with the ith observation, and we use $|h_{\max,i}|$ as its influence measure.

For the cutoff value, Zhu and Lee (2001) proposed a benchmark that takes the sample mean and variation of influence measures into account. Inspired by this work, we take the benchmark for the influence measures to be $\bar{M} + 1.645 s_M$, where $\bar{M}$ and s_M are their sample mean and standard deviation, respectively. An observation is called influential if its influence measure is larger than the benchmark.

When we find a matrix $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}$ such that $\mathrm{QC}_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}} \ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}} \mathbf{h}$, the influential direction is the eigenvector of $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}$ associated with its largest eigenvalue. Inspired by Zhu and Lee (2001), we construct another option for the influence measure vector, called the aggregate contribution vector, defined as $\mathbf{M}_0 = \sum_{i=1}^{s} \tilde{\lambda}_i \mathbf{e}_i^{(s)}$, where $\mathbf{e}_i^{(s)} = (e_{i1}^2, \ldots, e_{is}^2)^{\mathrm{T}}$ and $\{(\tilde{\lambda}_i, \mathbf{e}_i)\}_{i=1}^{s}$ are pairs of eigenvalues and orthonormal eigenvectors of $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}$.

4. Specification and interpretation for the quasi-curvature

Lemma 1 presents the expression of $\partial^2 d(\mathbf{A}) / \partial \mathrm{vec}(\mathbf{A}) \partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}} |_{\mathbf{A}=\hat{\mathbf{B}}}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Lemma 1. *Let $\otimes$ denote the Kronecker product. Then, it holds that*

$$\left.\frac{\partial^2 d(\mathbf{A})}{\partial \mathrm{vec}(\mathbf{A}) \partial \mathrm{vec}(\mathbf{A})^{\mathrm{T}}}\right|_{\mathbf{A}=\hat{\mathbf{B}}} = \frac{2}{\hat{K}} (\hat{\mathbf{B}}^{\mathrm{T}} \mathbf{Z}\mathbf{Z}^{\mathrm{T}} \hat{\mathbf{B}})^{-1} \otimes \{\mathbf{Z}(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}})\mathbf{Z}^{\mathrm{T}}\}.$$

Note that $\mathbf{Z}\mathbf{Z}^{\mathrm{T}}/n = \hat{\mathbf{\Sigma}}_{\mathbf{x}}$. It is natural to take $\hat{\mathbf{B}}$ satisfying $\hat{\mathbf{B}}^{\mathrm{T}} \hat{\mathbf{\Sigma}}_{\mathbf{x}} \hat{\mathbf{B}} = \mathbf{I}$ in sufficient dimension reduction methods such as the sliced inverse regression, principal Hessian directions, and so on. The above lemma indicates that when $\hat{\mathbf{B}}^{\mathrm{T}} \hat{\mathbf{\Sigma}}_{\mathbf{x}} \hat{\mathbf{B}} = \mathbf{I}$, the expression of the quasi-curvature can be written as

$$\mathrm{QC}_{\mathbf{h}} = \frac{2}{\hat{K}} \sum_{k=1}^{\hat{K}} \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)\top} (\hat{\mathbf{\Sigma}}_{\mathbf{x}} - \hat{\mathbf{\Sigma}}_{\mathbf{x}} \hat{\mathbf{B}} \hat{\mathbf{B}}^{\top} \hat{\mathbf{\Sigma}}_{\mathbf{x}}) \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)},$$

where $\mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}$ denotes the kth column of $\mathbf{F}_{\mathbf{B},\mathbf{h}}$. This expression is useful for calculation. The quasi-curvature can also be written as

$$\mathrm{QC}_{\mathbf{h}} = \frac{2}{n\hat{K}} \sum_{k=1}^{\hat{K}} \|(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}\|^2, \tag{4.1}$$

where $\|\cdot\|$ denotes the Euclidean norm. This expression provides us with an intuitive interpretation of $\mathrm{QC}_{\mathbf{h}}$. We begin with the interpretation of $(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}$. From equality (3.2), it holds that for $k = 1, \ldots, \hat{K}$,

$$\begin{aligned}(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \hat{\mathbf{b}}_k (\boldsymbol{\omega}_{(0)} + t\mathbf{h}) &= (\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \hat{\mathbf{b}}_k + t\{(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}\} \\ &\quad + o(t).\end{aligned}$$

Because $\mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}$ is the projection matrix on $\mathcal{M}(\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}})$, this equality implies that the vector $(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}}) \mathbf{Z}^{\mathrm{T}} \mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}$ represents the local change of the pro-

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

jection of $\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_k$ on the orthogonal complement of $\mathcal{M}(\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}})$ under the local perturbation along $\mathbf{h}$. Here, $\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_k$ is the centralized sample vector of $\hat{\mathbf{b}}_k^{\mathrm{T}}\mathbf{x}$, called the kth dimension reduction variate, and $\mathcal{M}(\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}})$ is spanned by $\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_1, \ldots, \mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_{\hat{K}}$. The projection plays a key role. It separates the local change of the kth dimension reduction direction into two parts, with one in the estimated central subspace, and the other in its orthogonal complement; only the latter part is used in QC_h. This appears to be reasonable because $\mathrm{QC}_{\mathbf{h}}$ is supposed to describe the local change of this subspace estimate. In addition, $\hat{\mathbf{B}}^{\mathrm{T}}\hat{\mathbf{\Sigma}}_{\mathbf{x}}\hat{\mathbf{B}} = \mathbf{I}$ means that the $\hat{\mathbf{b}}_k^{\mathrm{T}}\mathbf{x}$ are uncorrelated, which is why $\|(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}})\mathbf{Z}^{\mathrm{T}}\mathbf{f}_{\mathbf{B},\mathbf{h}}^{(k)}\|^2$s are additive in (4.1).

5. Properties of the quasi-curvature and perturbation schemes

Irrespective of the perturbation scheme, the influence assessment provided by the quasi-curvature method is invariant when the basis of $\mathcal{B}$ changes in the influence analysis for $\mathcal{B}$.

Theorem 1. *When* $\hat{\mathbf{B}}$ *and* $\hat{\mathbf{B}}(\boldsymbol{\omega})$ *in (3.1) are substituted by* $\hat{\mathbf{B}}\mathbf{A}$ *and* $\hat{\mathbf{B}}(\omega)\mathbf{C}$*, respectively, with* $\mathbf{A}$ *and* $\mathbf{C}$ *invertible matrices, the space displacement function* $D(\boldsymbol{\omega})$ *is invariant, indicating that the quasi-curvature and the influential direction are also both invariant under Assumption 2.*

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

However, selecting an appropriate perturbation scheme is still crucial. In a local influence analysis, the perturbation scheme is not an assumption that the data should satisfy, because the perturbation itself is artificial. Actually, under any perturbation scheme that is smooth enough as a function of $\boldsymbol{\omega}$, the quasi-curvature method can always give an influence assessment. However, a reasonable perturbation scheme should fairly perturb all the considered aspects. Moreover, for a specific sufficient dimension reduction method, we need to carefully design a perturbation scheme to ensure some properties. For example, in a sliced inverse regression, $\hat{\mathbf{b}}_i^{\mathrm{T}}\mathbf{x}$ is invariant under the transformation $\mathbf{A}\mathbf{x}$, for any $p \times p$ invertible matrix $\mathbf{A}$ (Li, 1991). Thus, $\hat{\mathbf{b}}_i^{*\mathrm{T}}\mathbf{x}^* = \hat{\mathbf{b}}_i^{\mathrm{T}}\mathbf{x}$, where $\mathbf{x}^* = \mathbf{A}\mathbf{x}$ and the $\hat{\mathbf{b}}_i^*$ are estimates of the dimension reduction directions for $\mathbf{x}^*$. Hence, it is natural to assume that the influence assessment remains invariant under the transformation of observations $\mathbf{x}_i^* = \mathbf{A}\mathbf{x}_i$, for $i = 1, \ldots, n$.

We now design a perturbation scheme to assess the influence of observations in a sliced inverse regression, which satisfies the invariance property under the transformation $\mathbf{A}\mathbf{x}$. For this purpose, we adopt the so-called multiplicative scheme (Shi, 1997; Lee and Tang, 2004). Specifically, we directly perturb the observations $(\mathbf{x}_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$ to $(\omega_i\mathbf{x}_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$, for $i = 1, \ldots, n$, and obtain the estimates $\hat{\mathbf{b}}_1(\boldsymbol{\omega}), \ldots, \hat{\mathbf{b}}_{\hat{K}}(\boldsymbol{\omega})$, with $(\mathbf{x}_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$ replaced by $(\omega_i\mathbf{x}_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Note that y_i is not perturbed, because it is used only for slicing, and the local influence analysis is the same, regardless of whether y_i is perturbed. The reason is as follows. In general, slicing should be based on the distributional information of the response values for the observations. There is always a small neighborhood of $\boldsymbol{\omega}_{(0)}$ such that, for $\boldsymbol{\omega}$ in it, $\omega_1 y_1, \ldots, \omega_n y_n$ keep the same order as $y_1, \ldots, y_n$ when sorted. Hence, a small perturbation on y will not change the inference result because the slicing remains unchanged. Because the impact of the observations on the central subspace estimate depends on slicing, the influence assessment of observations can be obtained only when slicing is given. This scheme can be expressed as

$$\mathbf{X}(\boldsymbol{\omega}) = \mathbf{X}\mathrm{diag}(\boldsymbol{\omega}), \tag{5.1}$$

where $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_n)^{\mathrm{T}}$, $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$, and $\mathrm{diag}(\boldsymbol{\omega})$ denotes a diagonal matrix with the ith diagonal element ω_i. Under scheme (5.1), we have the following property for the sliced inverse regresson.

Theorem 2. *Let $\mathbf{x}_1^*, \ldots, \mathbf{x}_n^*$ denote the sample of $\mathbf{x}^*$ under the invertible affine transformation $\mathbf{x}^* = \mathbf{A}\mathbf{x}$, and let $D^*(\boldsymbol{\omega})$ be the space displacement function under the model in which Y is regressed on $\mathbf{x}^*$. Then, under the sliced inverse regression and scheme (5.1), where the data before and after the transformation are perturbed to $(\omega_i \mathbf{x}_i^{\mathrm{T}}, y_i)^{\mathrm{T}}$ and $(\omega_i \mathbf{x}_i^{*\mathrm{T}}, y_i)^{\mathrm{T}}$, respectively, for $i = 1, \ldots, n$, it holds that $D^*(\boldsymbol{\omega}) = D(\boldsymbol{\omega})$, which indicates that the*

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

quasi-curvature and the influential direction are both invariant under Assumption 2.

This theorem also indicates that under (5.1), $D(\boldsymbol{\omega})$ remains invariant when the measuring units of the predictors change, which is a special case in which $\mathbf{A}$ is diagonal. Not every scheme possesses this invariance property. For example, it does not hold under the scheme $\mathbf{x}_i(\boldsymbol{\omega}) = \mathbf{x}_i + (\omega_i, \ldots, \omega_i)^{\mathrm{T}}$, for $i = 1, \ldots, n$. Moreover, under scheme (5.1), Assumption 1 holds when some mild conditions, specified in Theorem 3, are satisfied.

Remark 1. Another natural option is the so-called re-weighting-case perturbation scheme, which also possesses the invariance property. We include some results under this scheme in part S10 of the Supplementary Material, but this is not the focus of the present study.

6. Assessing the joint influence of the observations in a sliced inverse regression

To derive the expression of $\mathrm{vec}(\mathbf{F}_{\mathbf{B},\mathbf{h}})$, the following lemma is useful. Without loss of generality, we assume that the data points $(y_1, \mathbf{x}_1^{\mathrm{T}})^{\mathrm{T}}, \ldots, (y_n, \mathbf{x}_n^{\mathrm{T}})^{\mathrm{T}}$ have been sorted by Y. Some of the following concepts can be found in Kato (2013).

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Lemma 2. *Let $\hat{\lambda}$ be a simple eigenvalue of $\hat{\boldsymbol{\Sigma}}_\eta$ with respect to $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$, and let $\hat{\mathbf{b}}$ be an eigenvector associated with $\hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}\hat{\mathbf{b}} = 1$, where $\hat{\boldsymbol{\Sigma}}_\eta$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ are, respectively, symmetric and positive-definite matrices, and a simple eigenvalue means the dimension of its eigen-subspace is one. Let $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega})$ and $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega})$ be the estimates of $\boldsymbol{\Sigma}_{\mathbf{x}}$ and $\boldsymbol{\Sigma}_\eta$, respectively, under the perturbation, with $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ and $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ both being symmetric for t in a real neighborhood of $t = 0$, for any standardized $\mathbf{h}$. Suppose $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is well defined and holomorphic in a complex neighborhood of $t = 0$, and $\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)} + t\mathbf{h})$ is differentiable in a real neighborhood of $t = 0$. Then, the dimension of the total eigenspace for the $\hat{\lambda}$-group is one for t in a real neighborhood of $t = 0$, and the corresponding perturbations of $\hat{\lambda}$ and $\hat{\mathbf{b}}$, denoted by $\hat{\lambda}(t)$ and $\hat{\mathbf{b}}(t)$, respectively, are continuous in this neighborhood and differentiable at $t = 0$ in the real space:*

$$\hat{\lambda}(t) = \hat{\lambda} + t\hat{\lambda}_{*,1} + o(t), \qquad \hat{\mathbf{b}}(t) = \hat{\mathbf{b}} + t\mathbf{f} + o(t), \tag{6.1}$$

where

$$\hat{\lambda}_{*,1} = \hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\eta,1}\hat{\mathbf{b}} - \hat{\lambda}\hat{\mathbf{b}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} \tag{6.2}$$

$$\begin{aligned}\mathbf{f} = {} & \frac{1}{2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} \\ & + \hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}} - 2\hat{\boldsymbol{\Sigma}}_{\eta,1}\hat{\mathbf{b}}) - \frac{1}{2}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}},\end{aligned} \tag{6.3}$$

in which $\mathbf{A}^{+}$ denotes the Moore–Penrose inverse of matrix $\mathbf{A}$, $\hat{\boldsymbol{\Sigma}}_{\eta,1}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

denote $\partial\hat{\boldsymbol{\Sigma}}_\eta(\boldsymbol{\omega}_{(0)}+t\mathbf{h})/\partial\{t\}|_{t=0}$ *and* $\partial\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})/\partial\{t\}|_{t=0}$*, respectively, and* $\partial\mathbf{A}(t)/\partial\{t\}$ *denotes the matrix with its* (i,j)*th element* $\partial a_{ij}(t))/\partial t$*.*

Remark 2. (1) The expression of $\hat{\mathbf{b}}(t)$ given by (6.1) means $\hat{\mathbf{b}}(t)$ can be chosen in this way because it can also be $-\hat{\mathbf{b}}-t\mathbf{f}+o(t)$, but this does not change $\mathrm{QC}_{\mathbf{h}}$ (see Theorem 1); (2) What we need is the perturbation theory in the real space, and $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})$ is real for real t. However, we still require that $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}(\boldsymbol{\omega}_{(0)}+t\mathbf{h})$ be well defined and holomorphic in a neighborhood of $t=0$ in the complex plane. This ensures the differentiability of $\hat{\lambda}(t)$ and $\hat{\mathbf{b}}(t)$ at $t=0$ in the real space, considering that the smoothness of the eigen-projections can be lost if the holomorphy in the complex plane of a matrix $\mathbf{A}(t)$ is replaced by differentiability in the real space (Kato, 2013). Actually, this requirement is easy to satisfy. For example, it is satisfied under the scheme (5.1) in a sliced inverse regression. (3) The condition of simple eigenvalues is not so demanding; see the comment under the sliced inverse regression in the Supplementary Material.

Remark 3. A brief expression of $\mathbf{f}$ in Lemma 2 is as follows:

$$\mathbf{f}=-\mathbf{P}_{\Sigma,b}(\hat{\boldsymbol{\Sigma}}_\eta-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}})^{+}\mathbf{P}_{\Sigma,b}^{\top}(\hat{\boldsymbol{\Sigma}}_{\eta,1}-\hat{\lambda}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1})\hat{\mathbf{b}}-\frac{1}{2}(\hat{\mathbf{b}}^{\top}\hat{\mathbf{b}})\mathbf{P}_{\mathbf{b}}\hat{\boldsymbol{\Sigma}}_{\mathbf{x},1}\hat{\mathbf{b}}, \quad (6.4)$$

where $\mathbf{P}_{\Sigma,b}=\mathbf{I}-\hat{\mathbf{b}}\hat{\mathbf{b}}^{\top}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ is the projection matrix along $\mathcal{M}(\hat{\mathbf{b}})$ to the orthogonal complement of $\mathcal{M}(\boldsymbol{\Sigma}_{\mathbf{x}}\hat{\mathbf{b}})$, and $\mathbf{P}_{\mathbf{b}}$ denotes the orthogonal projection

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

matrix on $\mathcal{M}(\hat{\mathbf{b}})$. This expression is intuitively clearer.

For the proof of Lemma 2, see the Supplementary Material, Part S5. We obtain (6.2) and (6.3) by generalizing the analysis of the perturbation effects on individual eigenvalues and eigenvectors (Sibson, 1979), with a concise proof provided. Equality (6.3) can also be derived from theorem 1 in Prendergast and Smith (2010). Equality (6.2) is a by-product that is not used here, but we expect it to be useful in future, because the $\hat{\lambda}_i$ play key roles in estimates of K.

Under the sliced inverse regression and scheme (5.1) with $\boldsymbol{\omega} = \boldsymbol{\omega}_{(0)} + t\mathbf{h}$, the conditions in Lemma 2 are obviously all satisfied. Combining Lemmas 1, 2, and 3 (see the Supplementary Material, part S7) provides Theorem 3.

Theorem 3. *Under the sliced inverse regression and perturbation scheme (5.1), if* $\mathrm{rk}(\mathbf{Z}) = p$ *and the eigenvalues* $\hat{\lambda}_1, \ldots, \hat{\lambda}_{\hat{K}}$ *of* $\hat{\boldsymbol{\Sigma}}_\eta$ *with respect to* $\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}$ *are all simple, the quasi-curvature of the lifted line along* $\mathbf{h}$ *at* $\boldsymbol{\omega}_{(0)}$ *can be expressed as* $\mathrm{QC}_{\mathbf{h}} = \mathbf{h}^{\mathrm{T}}\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}\mathbf{h}$, *where* $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}$ *denotes*

$$\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}} = \frac{2}{n\hat{K}} \sum_{k=1}^{\hat{K}} \boldsymbol{\Delta}_{\mathbf{B},k}^{\mathrm{T}} \{\mathbf{Z}(\mathbf{I} - \mathbf{P}_{\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{B}}})\mathbf{Z}^{\mathrm{T}}\} \boldsymbol{\Delta}_{\mathbf{B},k},$$

in which

$$\begin{aligned} \boldsymbol{\Delta}_{\mathbf{B},i} = {} & \frac{1}{2}(\boldsymbol{\Sigma}_{\eta,\mathbf{x},i}\hat{\boldsymbol{\Sigma}}_\eta\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1} + \hat{\lambda}_i\boldsymbol{\Sigma}_{\eta,\mathbf{x},i} - \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1})\{\frac{1}{n}\mathbf{X}\mathrm{diag}(\mathbf{Z}^{\mathrm{T}}\hat{\mathbf{b}}_i) \\ & + \frac{1}{n}\mathbf{Z}\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_i)\} - \boldsymbol{\Sigma}_{\eta,\mathbf{x},i}\{\frac{1}{n}\mathbf{X}\mathrm{diag}(\mathbf{Z}_\eta^{\mathrm{T}}\hat{\mathbf{b}}_i) + \frac{1}{n}\mathbf{Z}_\eta\mathrm{diag}(\mathbf{X}^{\mathrm{T}}\hat{\mathbf{b}}_i)\}, \end{aligned}$$

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

for $i = 1, \ldots, \hat{K}$, *and* $\boldsymbol{\Sigma}_{\eta,\mathbf{x},i} = \hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}(\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}\hat{\boldsymbol{\Sigma}}_{\eta}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2} - \hat{\lambda}_i\mathbf{I})^{+}\hat{\boldsymbol{\Sigma}}_{\mathbf{x}}^{-1/2}$. *The influential direction* $\mathbf{h}_{\max}$ *is the eigenvector of* $\ddot{\mathbf{D}}_{\boldsymbol{\omega}_{(0)}}$ *associated with its largest eigenvalue.*

The above quasi-curvature method for an influence assessment of observations supports a simple strategy of data trimming. When the influence measures for a small proportion of observations are outstanding, clipping them out of the data set before the sufficient dimension reduction may be a feasible and parsimonious way to avoid the risk of data contamination. We illustrate this further in the simulation and real-data analysis.

The above procedure can be easily extended to many other sufficient dimension reduction methods. Lemma 2 is shared by other methods that obtain dimension reduction directions by calculating the eigenvectors of a kernel matrix, which is similar to $\hat{\boldsymbol{\Sigma}}_{\eta}$, with respect to the covariance matrix of $\mathbf{x}$ (e.g., the principal Hessian direction and sliced average variance estimation methods). Then, the only work needed by the extension for these methods is a matrix differentiation similar to Lemma 3, which depends on the specific form of the perturbed kernel matrix and the covariance matrix of $\mathbf{x}$.

Remark 4. Under the re-weighting-case perturbation scheme, $\mathrm{QC}_{\mathbf{h}}$ can also be expressed in the quadratic form $\mathbf{h}^{\mathrm{T}}\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}\mathbf{h}$, with $\ddot{\mathbf{D}}^{(R)}_{\boldsymbol{\omega}_{(0)}}$ and related

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

quantities given in the Supplementary Material (part S10). In addition, we have shown that the quasi-curvature method under the re-weighting-case scheme is similar to the case-deletion method.

Remark 5. We also considered local influence analyses for the cumulative mean estimation proposed by Zhu et al. (2010b) and MAVE based on conditional density function (dMAVE) proposed by Xia (2007). Both the theoretical results and the simulation studies are available in the Supplementary Material.

7. A real-data example

Here,we examine the relationship between the ambient nitrate concentration and the predictors; see, for example, Bondell et al. (2010) and Chen et al. (2015). We conduct a sliced inverse regression for the visualization. The response is the total ambient nitrate concentration (y), and the predictors are the mean ambient particulate ammonium concentration (x_1), mean ambient particulate sulfate concentration (x_2), relative humidity (x_3), ozone (x_4), precipitation (x_5), temperature difference between 9 m and 2 m probes (x_6), and scalar wind speed (x_7). The original data are obtained from the Clean Air Status and Trends Network (www.epa.gov/castnet), provided by the United States Environmental Protection Agency. The data are seasonal

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

for y, x_1, and x_2 and hourly for x_3–x_7. The hourly data are transformed to be seasonal using the method of Chen et al. (2015), and all predictors are standardized. We use the data from BEL116 and BWR139, two sites in Maryland, the United States, from 2001 to 2009. Suppose that two observations from another site (WSP144) are contained in our data set by mistake, with case numbers 68 and 69; then we have 69 observations, excluding those with missing entries.

We conduct a sliced inverse regression and its influence analysis for this data set. For the slicing strategy, we obtain $[n/v_s]$ slices, with each of the first $[n/v_s]-1$ slices containing v_s observations, and the last slice containing the remaining observations, where $[\xi]$ denotes the integer closest to ξ. For comparison, three methods are used, including our quasi-curvature approach, denoted by QC, and two sample influence functions proposed by Prendergast (2006, 2007) and Prendergast and Smith (2010), denoted by SIFB and SIFC, respectively. The latter two are both case-deletion methods. We denote the influence measures that they provide for the ith observation as SIFB(i) and SIFC(i). In both cases, the slices are kept unchanged after the deletion of each observation. For the quasi-curvature method, the influential direction $\mathbf{h}_{\max}$ under the perturbation scheme (5.1) is used, with $|h_{\max,i}|$ as the influence measure of the ith observation.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

We show the results obtained when v_s is six, a moderate value for data slicing. Without data trimming, $\hat{K} = 1$ and $\hat{\mathbf{b}}_1 = (1.432, -1.035, -0.360, 0.175, -0.071, 0.026, 0.660)$ where $\hat{K}$ is obtained using the sequential tests, with $\alpha_T = 0.05$ and the test level being $\alpha_T/7$ in each step. We obtain the influence measures of the observations using the quasi-curvature method ($\mathbf{h}_{\max}$) under scheme (5.1), and identify five of them as influential with the benchmark $\bar{M} + 1.645 s_M$; see Figure 1 (a1) for the index plot. In the scatter plot of y versus $\hat{\mathbf{b}}_1^{\mathrm{T}}\mathbf{x}$ (see Figure 1 (a2)), these five observations, which are marked by circles, appear to be outliers. For comparison, we also obtain the influence measures of observations using Prendergast's two case-deletion methods, with their index plots given in Figures 1 (b1) and (c1), and identified influential observations marked by circles in Figures 1 (b2) and (c2). It appears that the 68th observation, which is quite outlying in the scatter plot of y versus $\hat{\mathbf{b}}_1^{\mathrm{T}}\mathbf{x}$, may not receive sufficient attention from the two case-deletion methods. Considering that its position is somewhat close to the 69th observation, this lack of attention may be due to the masking effect. Under the re-weighting-case scheme, the aggregate contribution vector based on the quasi-curvature is still similar to that of the case-deletion method (Prendergast and Smith, 2010). The cosine of the angle between $\mathbf{M}_0$ and $(\mathrm{SIFC}(1), \ldots, \mathrm{SIFC}(n))^{\mathrm{T}}$ is 0.980, and the former identifies the 3rd, 20th, 31st and 69th observations

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

as influential.

The following fact appears to indicate that data trimming can help make a more definite and unified conclusion about $\hat{K}$ under different data slicing patterns. We have tried all the values of v_s from 4 to 15. Based on the overall data set, we have $\hat{K} = 2$ when $v_s = 10$, and when v_s is taken to be any other value, we always obtain $\hat{K} = 1$. In the tests for $K < 1$ and $K < 2$, the means of the p-values for $v_s = 4, \ldots, 15$ are 0.00045 and 0.0568, respectively. We now employ the data trimming. For each v_s, we delete the influential observations identified by the quasi-curvature method ($\mathbf{h}_{\max}$) under scheme (5.1) from the data set. After the data trimming, we have $\hat{K} = 1$ for all values of $v_s = 4, \ldots, 15$, and the means of the p-values for $K < 1$ and $K < 2$ are 0.00007 and 0.0711, respectively. When $v_s = 6$, the estimate of b_1 becomes $\tilde{\mathbf{b}}_1 = (1.697, -1.441, -0.358, 0.110, -0.036, -0.050, 0.511)^{\mathrm{T}}$ after data trimming.

8. Discussion

The following offer potential further research directions for the proposed method. The first is its application. For instance, we expect it to perform well on the method of principal Hessian directions, because Prendergast (2008) and Lue (2001) have already shown the usefulness of data trimming

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

based on case-deletion diagnostics.

Supplementary Material

The online Supplementary Material includes the simulations, proofs of the theoretical results, and some other technical details.

Acknowledgments

The authors gratefully acknowledge the grants from the National Natural Science Foundation of China (Grant Nos. 11661078, 11261064, 11161053, U1302267, and 11301463) and a grant from the University Grants Council of Hong Kong.

References

Bondell, H. D., A. Krishna, and S. K. Ghosh (2010). Joint variable selection for fixed and random effects in linear mixed-effects models. *Biometrics* **66**(4), 1069–1077.

Chen, F., Z. Li, L. Shi, and L. Zhu (2015). Inference for mixed models of ANOVA type with high-dimensional data. *Journal of Multivariate Analysis* **133**, 382–401.

Cook, R. D. (1986). Assessment of local influence. *Journal of the Royal Statistical Society, Series B,* **48**, 133–169.

Cook, R. D. (1998a). Principal Hessian directions revisited. *Journal of the American Statistical Association* **93**, 84–94.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Cook, R. D. (1998b). *Regression Graphics: Ideas for Studying regressions through graphics*. New York: Wiley.

Cook, R. D. and F. Critchley (2000). Identifying regression outliers and mixtures graphically. *Journal of the American Statistical Association* **95**(451), 781–794.

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

Cook, R. D. and C. J. Nachtsheim (1994). Reweighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association* **89**, 592–599.

Cook, R. D. and S. Weisberg (1991). Sliced inverse regression for dimension reduction: comment. *Journal of the American Statistical Association* **86**, 328–332.

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

Gather, U., T. Hilker, and C. Becker (2001). A robustified version of sliced inverse regression. In *Statistics in Genetics and in the Environmental Sciences* (Edited by L. T. Fernholz, S. Morgenthaler and W. Stahel), 147–157. Birkhäuser, Basel.

Gather, U., T. Hilker, and C. Becker (2002). A note on outlier sensitivity of sliced inverse regression. *Statistics: A Journal of Theoretical and Applied Statistics* **36**(4), 271–281.

Guan, Y., C. Xie, and L. Zhu (2017). Sufficient dimension reduction with mixture multivariate skew-elliptical distributions. *Statistica Sinica* **27**, 335–355.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Hooper, J. W. (1959). Simultaneous equations and canonical correlation theory. *Econometrica* **27**, 245–256.

Kato, T. (2013). *Perturbation theory for linear operators*. Springer Science & Business Media.

Lee, S. Y. and N. S. Tang (2004). Local influence analysis of nonlinear structural equation models. *Psychometrika* **69**, 573–592.

Li, B. and Y. Dong (2009). Dimension reduction for non-elliptically distributed predictors. *The Annals of Statistics* **37**, 1272–1298.

Li, B. and S. Wang (2007). On directional regression for dimension reduction. *Journal of the American Statistical Association* **102**, 997–1008.

Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association* **86**, 312–342.

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.

Lue, H. H. (2001). A study of sensitivity analysis on the method of principal Hessian directions. *Computational Statistics* **16**, 109–130.

Ma, Y. and L. P. Zhu (2012). A semiparametric approach to dimension reduction. *Journal of the American Statistical Association* **107**, 168–179.

Prendergast, L. A. (2006). Detecting influential observations in sliced inverse regression analysis.

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

Australian and New Zealand Journal of Statistics **48**, 285–304.

Prendergast, L. A. (2007). Implications of influence function analysis for sliced inverse regression and sliced average variance estimation. *Biometrika* **94**, 585–601.

Prendergast, L. A. (2008). Trimming influential observations for improved single-index model estimated sufficient summary plots. *Computational Statistics and Data Analysis* **52**, 5319–5327.

Prendergast, L. A. and J. A. Smith (2010). Influence functions for dimension reduction methods: an example influence study of principal Hessian direction analysis. *Scandinavian Journal of Statistics* **37**, 588–611.

Shi, L. (1997). Local influence in principal components analysis. *Biometrika* **84**, 175–186.

Sibson, R. (1979). Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling. *Journal of the Royal Statistical Society, Series B* **41**, 217–229.

Wang, H. and Y. Xia (2008). Sliced regression for dimension reduction. *Journal of the Royal Statistical Society, Series B* **103**, 811–821.

Xia, Y. (2007). A constructive approach to the estimation of dimension reduction directions. *The Annals of Statistics* **35**, 2654–2690.

Yin, X. and B. Li (2011). Sufficient dimension reduction based on an ensemble of minimum average variance estimators. *The Annals of Statistics* **39**(6), 3392–3416.

Zhou, J., W. Xu, and L. Zhu (2015). Robust estimating equation-based sufficient dimension

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

reduction. *Journal of Multivariate Analysis* **134**, 99–118.

Zhu, H. T., J. G. Ibrahim, S. Y. Lee, and H. P. Zhang (2007). Perturbation selection and influence measures in local influence analysis. *The Annals of Statistics* **35**, 2565–2588.

Zhu, H. T. and S. Y. Lee (2001). Local influence for incomplete data models. *Journal of the Royal Statistical Society, Series B* **63**, 111–126.

Zhu, L. P., T. Wang, L. X. Zhu, and L. Ferré (2010a). Sufficient dimension reduction through discretization- expectation estimation. *Biometrika* **97**, 295–304.

Zhu, L. P., L. X. Zhu, and Z. H. Feng (2010b). Dimension reduction in regressions through cumulative slicing estimation. *Journal of the American Statistical Association* **105**, 1455–1466.

Zhu, L. X., B. Q. Miao, and H. Peng (2006). On sliced inverse regression with high-dimensional covariates. *Journal of the American Statistical Association* **101**(474), 630–643.

Zhu, Y. and P. Zeng (2006). Fourier methods for estimating the central subspace and the central mean subspace in regression. *Journal of the American Statistical Association* **101**(476), 1638–1651.

Fei Chen
School of Statistics and Mathematics,
Yunnan University of Finance and Economics,
237 Longquan Road, 650221, Kunming, China
E-mail: chen_fei_77@hotmail.com

Lei Shi
School of Statistics and Mathematics,
Yunnan University of Finance and Economics,
237 Longquan Road, 650221, Kunming, China

Statistica Sinica: Preprint
doi:10.5705/ss.202019.0388

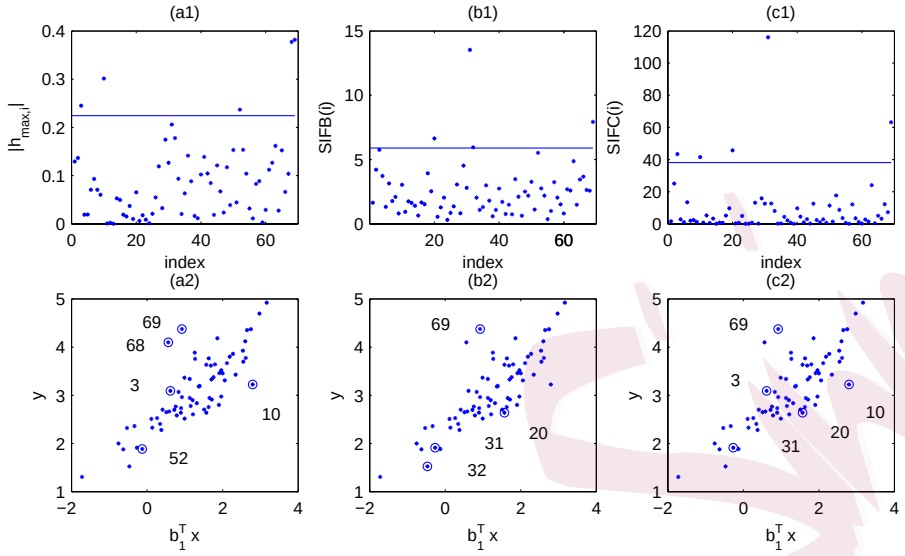

Figure 1: Index plots of influence measures for observations and scatter plots of y versus $\hat{\mathbf{b}}_1^{\mathrm{T}}\mathbf{x}$, with identified influential points marked by circles.

E-mail: shi_lei65@hotmail.com

Lin Zhu
Tiantai County Branch of the People's Bank of China
No. 1 West Renmin Road, Tiantai County, Taizhou, Zhejiang, China
E-mail: 783268587@qq.com

Lixing Zhu (corresponding author)
Department of Mathematics,
Hong Kong Baptist University,
Kowloon, Hong Kong, China
E-mail: lzhu@hkbu.edu.hk
Phone: (852) 3411-7016
Fax: (852) 3411-5811