Regression Graphics[*]

R. D. Cook

Department of Applied Statistics

University of Minnesota

St. Paul, MN 55108

Abstract

This article, which is based on an Interface tutorial, presents an overview of regression graphics, along with an annotated bibliography. The intent is to discuss basic ideas and issues without delving into methodological or theoretical details, and to provide a guide to the literature.

1 Introduction

One of the themes for this Interface conference is *dimension reduction,* which is also a leitmotif of statistics. For instance, starting with a sample $z_1, \ldots, z_n$ from a univariate normal population with mean μ and variance 1, we know the sample mean $\bar{z}$ is sufficient for μ. This means that we can replace the original n-dimensional sample with the one-dimensional mean $\bar{z}$ without loss of information on μ.

In the same spirit, *dimension reduction without loss of information* is a key theme of regression graphics. The goal is to reduce the dimension of the predictor vector $\mathbf{x}$ without loss of information on the regression. We call this *sufficient dimension reduction*, borrowing terminology from classical statistics. Sufficient dimension reduction leads naturally to *sufficient summary plots* which contain all of the information on the regression that is available from the sample.

Many graphically oriented methods are based on the idea of dimension reduction, but few formally incorporate the notion of sufficient dimension reduction. For example, projection pursuit involves searching for "interesting" low dimensional projections by using a user-selected projection index that serves to quantify "interestingness" (Huber 1995). Such methods may produce interesting results in a largely exploratory context, but one may be left wondering what they have to do with overarching statistical issues like regression.

We assume throughout that the scalar response y and the $p \times 1$ vector of predictors $\mathbf{x}$ have a joint distribution. This assumption is intended to focus the discussion and is not crucial

for the fundamental ideas. The general goal of the regression is to infer about the conditional distribution of $y|\mathbf{x}$ as far as possible with the available data: How does the conditional distribution of $y|\mathbf{x}$ change with the value assumed by $\mathbf{x}$? It is assumed that the data $\{y_i, \mathbf{x}_i\}$, $i = 1, \ldots, n$, are iid observations on $(y, \mathbf{x})$ with finite moments.

The notation $U \perp\!\!\!\perp V$ means that the random vectors U and V are independent. Similarly, $U \perp\!\!\!\perp V|Z$ means that U and V are independent given any value for the random vector Z. Subspaces will be denoted by $\mathcal{S}$, and $\mathcal{S}(B)$ means the subspace of R^t spanned by the columns of the $t \times u$ matrix B. Finally, P_B denotes the projection operator for $\mathcal{S}(B)$ with respect to the usual inner product, and $Q_B = I - P_B$.

2 The Central Subspace

Let B denote a fixed $p \times q$, $q \leq p$, matrix so that

$$y \perp\!\!\!\perp \mathbf{x}|B^T\mathbf{x} \qquad (1)$$

This statement is equivalent to saying that the distribution of $y|\mathbf{x}$ is the same as that of $y|B^T\mathbf{x}$ for all values of $\mathbf{x}$ in its marginal sample space. It implies that the $p \times 1$ predictor vector $\mathbf{x}$ can be replaced by the $q \times 1$ predictor vector $B^T\mathbf{x}$ without loss of regression information, and thus represents a potentially useful reduction in the dimension of the predictor vector. Such a B always exists, because (1) is trivially true when $B = I_p$, the $p \times p$ identity matrix.

Statement (1) holds if and only if

$$y \perp\!\!\!\perp \mathbf{x}|P_{\mathcal{S}(B)}\mathbf{x}$$

Thus, (1) is appropriately viewed as a statement about $\mathcal{S}(B)$, which is called a *dimension-reduction subspace for the regression of y on $\mathbf{x}$* (Li 1991, Cook 1994a). The idea of a dimension-reduction subspace is useful because it represents a "sufficient" reduction in the dimension of the predictor vector. Clearly, knowledge of the smallest dimension-reduction subspace would be useful for parsimoniously characterizing how the distribution of $y|\mathbf{x}$ changes with the value of $\mathbf{x}$.

Let $\mathcal{S}_{y|\mathbf{x}}$ denote the intersection of all dimension-reduction subspaces. While $\mathcal{S}_{y|\mathbf{x}}$ is always a subspace, it is not necessarily a dimension-reduction subspace. Nevertheless, $\mathcal{S}_{y|\mathbf{x}}$ is a dimension-reduction subspace under various reasonable conditions (Cook 1994a, 1996, 1998b). In this article, $\mathcal{S}_{y|\mathbf{x}}$ is assumed to be a dimension-reduction subspace and, following Cook (1994b, 1996, 1998b), is called the *central dimension-reduction subspace*, or simply the *central subspace*. The dimension $d = \dim(\mathcal{S}_{y|\mathbf{x}})$ is the *structural dimension* of the regression; we will identify regressions as having $0D, 1D, \ldots, dD$ structure.

The central subspace, which is taken as the inferential object for the regression, is the smallest dimension-reduction subspace such that $y \perp\!\!\!\perp \mathbf{x} | \eta^T \mathbf{x}$, where the columns of the matrix η form a basis for the subspace. In effect, the central subspace is a super parameter that will be used to characterize the regression of y on $\mathbf{x}$. If $\mathcal{S}_{y|\mathbf{x}}$ were known, the *minimal sufficient summary plot* of y versus $\eta^T \mathbf{x}$ could then be used to guide subsequent analysis. If an estimated basis $\hat{\eta}$ of $\mathcal{S}_{y|\mathbf{x}}$ were available then the summary plot of y versus $\hat{\eta}^T \mathbf{x}$ could be used similarly.

2.1 Structural Dimension

The simplest regressions have structural dimension $d = 0$, in which case $y \perp\!\!\!\perp \mathbf{x}$ and a histogram or density estimate based on $y_1, \ldots, y_n$ is a minimal sufficient summary plot. The idea of $0D$ structure may seem to limiting to be of much practical value. But it can be quite important when using residuals in model checking. Letting r denote the residual from a fit of a target model in the population, the correctness of the model can be checked by looking for information in the data to contradict $0D$ structure for the regression of r on $\mathbf{x}$. Sample residuals $\hat{r}$ need to be used in practice.

Many standard regression models have $1D$ structure. For example, letting ε be a random error with mean 0, variance 1 and $\varepsilon \perp\!\!\!\perp \mathbf{x}$, the following additive-error models each have $1D$ structure provided $\beta \neq 0$:

$$y|\mathbf{x} = \beta_0 + \beta^T \mathbf{x} + \sigma\varepsilon \qquad (2)$$
$$y|\mathbf{x} = \mu(\beta^T \mathbf{x}) + \sigma\varepsilon \qquad (3)$$
$$y|\mathbf{x} = \mu(\beta^T \mathbf{x}) + \sigma(\beta^T \mathbf{x})\varepsilon \qquad (4)$$
$$y^{(\lambda)}|\mathbf{x} = \mu(\beta^T \mathbf{x}) + \sigma(\beta^T \mathbf{x})\varepsilon \qquad (5)$$

Model (2) is just the standard linear model. It has $1D$ structure because since $y \perp\!\!\!\perp \mathbf{x} | \beta^T \mathbf{x}$. Models (3) and (4) allow for a nonlinear mean function and a nonconstant standard deviation function, but still have $1D$ structure. Model (5) allows for a transformation of the response. Generalized linear models with link functions that correspond to those in (2)–(4) also have $1D$ structure. In each case, the vector β spans the central subspace. Clearly, the class of regressions with $1D$ structure is quite large.

In regressions with $1D$ structure and central subspace $\mathcal{S}_{y|\mathbf{x}}(\eta)$, a 2D plot of y versus $\eta^T \mathbf{x}$ is a minimal sufficient summary plot. In practice, $\mathcal{S}_{y|\mathbf{x}}(\eta)$ would need to be estimated.

The following two models each have $2D$ structure provided the nonzero vectors β_1 and β_2 are not collinear:

$$y|\mathbf{x} = \mu(\beta_1^T \mathbf{x}, \beta_2^T \mathbf{x}) + \sigma\varepsilon$$
$$y|\mathbf{x} = \mu(\beta_1^T \mathbf{x}) + \sigma(\beta_2^T \mathbf{x})\varepsilon$$

In this case, $\mathcal{S}_{y|\mathbf{x}}(\eta) = \mathcal{S}(\beta_1, \beta_2)$ and a $3D$ plot of y over $\mathcal{S}_{y|\mathbf{x}}(\eta)$ is a sufficient summary plot. Again, an estimate of $\mathcal{S}_{y|\mathbf{x}}$ would be required in practice.

In effect, the structural dimension can be viewed as an index of the complexity of the regression. Regressions with $2D$ structure are usually more difficult to summarize than those with $1D$ structure, for example.

2.2 Issues

There are now two general issues: how can we estimate the central subspace, including its dimension, and what might we gain in practice by doing so? We will consider the second issue first by using an example in the next section to contrast the results of a traditional analysis with those based on an estimate of the central subspace. Estimation methods are discussed in Section 4.

3 Naphthalene Data

Box and Draper (1987, p. 392) describe the results of an experiment to investigate the effects of three process variables (predictors) in the vapor phase oxidation of naphthalene. The response variable y is the percentage mole conversion of naphthalene to naphthoquinone, and the three predictors are $x_1 = $ logarithm of the air to naphthalene ratio (L/pg), $x_2 = $ logarithm of contact time (seconds) and $x_3 = $ bed temperature. There are 80 observations.

3.1 Traditional analysis

Box and Draper suggested an analysis based on a full second-order response surface model in the three predictors. In the ordinary least squares fit of this model, the magnitude of each of the 10 estimated regression coefficients is at least 4.9 times its standard error, indicating that the response surface may be quite complicated. It is not always easy to develop a good scientific understanding of such fits. In addition, the standard plot of residuals versus fitted values of Figure 1 shows curvature, so that a more complicated model may be necessary. Further analysis from this starting point may also be complicated by at least three influential observations that are
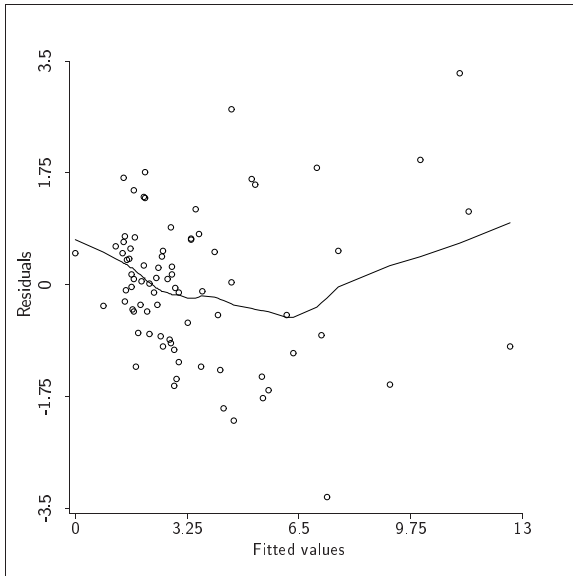
Figure 1: Residuals versus fitted values from the fit of the full quadratic model to the naphthalene data.

Figure 2: Cook's distances versus case numbers from the fit of the full quadratic model to the naphthalene data.

apparent in the plot of Cook's distance (Cook 1997) versus case numbers shown in Figure 2.

While there are many ways to continue the analysis, most involve iterating between fitted models and diagnostics until a satisfactory solution is found. This analysis paradigm can yield useful results, but the process is often laborious and the resulting models can be difficult to interpret.

3.2 Graphical analysis

We next turn to the analysis of the naphthalene data based on the central subspace. Using methods to be discussed a bit later, the dimension of the central subspace was inferred to be $\dim(\mathcal{S}_{y|\mathbf{x}}) = 1$ with

$$\hat{\mathcal{S}}_{y|\mathbf{x}} = \mathcal{S}(\hat{\eta})$$

where $\hat{\eta} = (0.38, 0.46, 0.8)^T$. The corresponding 2D summary plot, which is estimated to contain all the regression information, is shown in Figure 3 along with a lowess smooth and a fitted quadratic. The quadratic matches the lowess smooth over the upper half of the range on the horizontal axis, but otherwise fails to provide a good fit. This failure accounts for the curvature in the residual plot of Figure 1. The influential observations that would likely complicate an analysis based on response surface models fall mostly along directions orthogonal to $\hat{\eta}$ and contributed to the estimation of the central subspace only by hinting at the possibility that $\dim(\mathcal{S}_{y|\mathbf{x}}) = 2$.

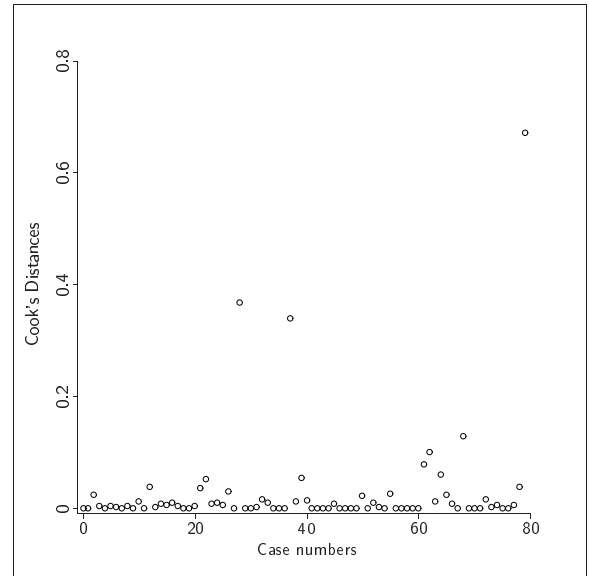There are many ways to continue the analysis, but all can be based on the summary plot of Figure 3. For example, we

might transform the response in an effort to straighten the relationship. In this way, the estimated central subspace along with the corresponding summary plot provide a starting point for traditional modeling that is often more informative than an initial fit based on a handy model.

One role that regression graphics can play in an analysis is represented schematically in Figure 4. Many traditional analyses follow the paradigm represented by the lower two boxes. We begin in the lower left box studying the problem and the data, and formulating an initial model. This is followed by estimation, leading to the fitted model in the right box. Diagnostic methods are then applied to the fitted model, leading back to the model formulation and the data when remedial action is required. Depending on the adequacy of the initial model, many iterations may be required to reach a satisfactory solution. Our brief look at the naphthalene data in Section 3 represents just one iteration. This paradigm for a regression analysis has been discussed by several authors, including Cook and Weisberg (1982, p. 7) and Box (1980).

Regression graphics is not intended to replace this traditional paradigm. Rather, as illustrated in the upper left box, its role is to provide a firm starting point through investigation of the central subspace and sufficient summary plots. We think of the graph in Figure 3 as providing information for the formulation of an initial model for the naphthalene data, which is the first step in the traditional paradigm. The hope is that this starting point will be close to a satisfactory answer and that only a few iterations will be required.

Regression graphics can also be of value in the diagnostic portion of an analysis. Letting r denote the population resid-
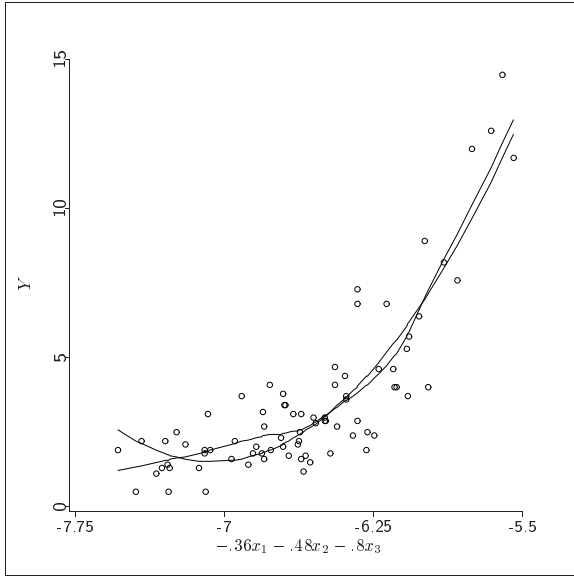
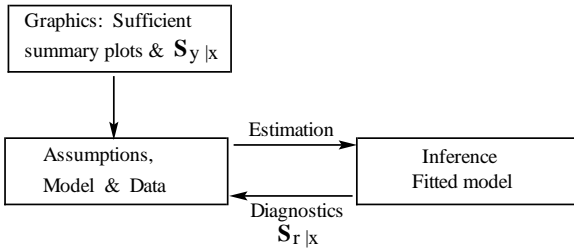Figure 3: Estimated sufficient summary plot for the naphthalene data.

Figure 4: Analysis paradigm incorporating central subspaces.

ual from the fitted model in the lower right box, we can use the central subspace $\mathcal{S}_{r|\mathbf{x}}$ for the regression of r on $\mathbf{x}$ to guide diagnostic considerations (Cook 1998b).

4 Estimating the Central Subspace

There are both numerical and graphical methods for estimating subspaces of $\mathcal{S}_{y|\mathbf{x}}(\eta)$. All methods presently require that the conditional expectation $E(\mathbf{x}|\eta^T\mathbf{x})$ is linear when $p > 2$. We will refer to this as the *linearity condition*. In addition, methods based on second moments work best when the conditional covariance matrix $\text{var}(\mathbf{x}|\eta^T\mathbf{x})$ is constant. Both the linearity condition and the *constant variance condition* apply to the marginal distribution of the predictors and do not involve the response. Hall and Li (1993) show that the linearity condition will hold to a reasonable approximation in many problems. In addition, these conditions might be induced by using predictor transformations and predictor weighting

(Cook and Nachtsheim 1994).

We next consider graphical estimates of the central subspace using a procedure called *graphical regression* (GREG).

4.1 Graphical Regression

4.1.1 Two predictors

With $p = 2$ predictors, $\mathcal{S}_{y|\mathbf{x}}(\eta)$ can be estimated visually from a 3D plot with y on the vertical axis and (x_1, x_2) on the horizontal axes. Visual assessment of $0D$ can be based on the result that $y \perp\!\!\!\perp \mathbf{x}$ if and only if $y \perp\!\!\!\perp b^T\mathbf{x}$ for all $b \in R^2$ (Cook 1996, 1998b). As the 3D plot of y versus (x_1, x_2) is rotated about its vertical axis, we see 2D plots of y versus all linear combinations $b^T\mathbf{x}$. Thus, if no clear dependence is seen in any 2D plot while rotating we can infer that $y \perp\!\!\!\perp \mathbf{x}$. If dependence is seen in a 2D plot, then the structural dimension must be 1 or 2. Visual methods for deciding between $1D$ and $2D$, and for estimating $\mathcal{S}_{y|\mathbf{x}}$ when $d = 1$, were discussed in detail by Cook (1992a, 1994a, 1998b), Cook and Weisberg (1994a), and Cook and Wetzel (1993).

4.1.2 More than two predictors

The central subspace can be estimated when $p > 2$ by assessing the structural dimension in a series of 3D plots. Here is the general idea. Let x_1 and x_2 be two of the predictors, and collect all of the remaining predictors into the $(p - 2) \times 1$ vector $\mathbf{x}_3$. We seek to determine if there are constants c_1 and c_2 so that

$$y \perp\!\!\!\perp \mathbf{x}|(c_1x_1 + c_2x_2, \mathbf{x}_3) \qquad (6)$$

If (6) is true and $c_1 = c_2 = 0$ then x_1 and x_2 can be deleted from the regression without loss of information, reducing the dimension of the predictor vector by 2. If (6) is true and either $c_1 \neq 0$ or $c_2 \neq 0$ then we can replace x_1 and x_2 with the single predictor $x_{12} = c_1x_1 + c_2x_2$ without loss of information, reducing the dimension of the predictor by 1. In which case we would have a new regression problem with predictors $(x_{12}, \mathbf{x}_3)$. If (6) is false for all values of (c_1, c_2) then the structural dimension is at least 2, and we could select two other predictors to combine. This methodology was developed by Cook (1992a, 1994a, 1996, 1998b), Cook and Weisberg (1994a), and Cook and Wetzel (1993), where it is shown that various 3D plots, including 3D added-variable plots, can be used to study the possibilities associated with (6) and to estimate (c_1, c_2) visually. Details are available from Cook (1998b).

The summary plot of Figure 3 was produced by this means (Cook 1998b, Section 9.1). It required a visual analysis of 2 3D plots and took only a minute or two to construct. Graphical regression is not time intensive with a modest number of predictors. Even with many predictors, it often takes less time

than more traditional methods which may require that a battery of numerical and graphical diagnostics be applied after each fit. Graphical regression has the advantage that nothing is hidden since all decisions are based only on visual inspection. In effect, diagnostic considerations are an integral part of the graphical regression process. Outliers and influential observations are usually easy to identify, for example.

4.2   Numerical methods

Several numerical methods are available for estimating the central subspace. We begin with standard methods and then turn to more recent developments.

4.2.1   Standard fitting methods and residual plots.

Consider summarizing the regression by minimizing a convex objective function of a linear kernel $a + b^T\mathbf{x}$:

$$(\hat{a}, \hat{b}) = \arg\min_{a,b} \frac{1}{n} \sum_{i=1}^{n} L(a + b^T\mathbf{x}_i, y_i) \qquad (7)$$

where $L(\cdot, \cdot)$ is a convex function of its first argument. This class of objective functions includes most of the usual methods including ordinary least squares with $L = (y_i - a - b^T\mathbf{x}_i)^2$, Huber's M-estimate, and various maximum likelihood estimates based on generalized linear models. This setup is being used only as a summarization device and does not imply the assumption of a model. Next, let

$$(\alpha, \beta) = \arg\min_{a,b} \mathrm{E}(L(a + b^T\mathbf{x}, y))$$

denote the population version of $(\hat{a}, \hat{b})$. Then under the linearity condition, β is in the central subspace (Li and Duan 1989, Cook 1994a). Under some regularity conditions, $\hat{b}$ converges to β at the usual root-n rate.

There are many important implications of this result. First, if $\dim(\mathcal{S}_{y|\mathbf{x}}) = 1$ and the linearity condition holds then a plot of y versus $\hat{b}^T\mathbf{x}$ is an estimated sufficient summary plot, and is often all that is needed to carry on the analysis. Second, it is probably a reason for the success of standard method like ordinary least squares. Third, the result can be used to facilitate other graphical methods. For example, it can be used to facilitate the analysis of the 3D plots encountered during a GREG analysis.

Finally, under the linearity condition, this result can be used to show that $\mathcal{S}_{r|\mathbf{x}} \subset \mathcal{S}_{y|\mathbf{x}}$ (Cook 1994a, 1998b). Consequently, the analysis of residual plots provides information that is directly relevant to the regression of y on $\mathbf{x}$. On the other hand, if the linearity condition does not hold it is probable that $\mathcal{S}_{r|\mathbf{x}} \not\subset \mathcal{S}_{y|\mathbf{x}}$ and thus that a residual analysis will be complicated by irrelevant information (the part of $\mathcal{S}_{r|\mathbf{x}}$ that is not in $\mathcal{S}_{y|\mathbf{x}}$). For example, if $\dim(\mathcal{S}_{y|\mathbf{x}}) = 1$ and the linearity

condition fails then we may well have $\dim(\mathcal{S}_{r|\mathbf{x}}) = 2$. In this case the residual regression is intrinsically more complicated than the original regression and a residual analysis may do more harm than good.

The objective function (7) can also be used with quadratic kernels $a + b^T\mathbf{x} + \mathbf{x}^T B\mathbf{x}$ where B is a $p \times p$ symmetric matrix of quadratic coefficients. When the predictors are normally distributed, $\mathcal{S}(\hat{B})$ converges to a subspace that is contained in the central subspace (Cook 1992b; 1998b, Section 8.2).

4.2.2   Inverse regression methods

In addition to the standard fitting methods discussed in the previous section, there are at least three other numerical methods that can be used to estimate subspaces of the central subspace:

- sliced inverse regression, SIR (Li 1991),

- principal Hessian directions, pHd (Li 1992, Cook 1998a), and

- sliced average variance estimation, SAVE (Cook and Weisberg 1991, Cook and Lee 1998).

All require the linearity condition. The constant variance condition facilitates analyses with all three methods, but is not essential for progress. Without the constant variance condition, the subspaces estimated will likely still be dimension-reduction subspaces, but they may not be central. All of these methods can produce useful results, but recent work (Cook and Lee 1998) indicates that SAVE is the most comprehensive.

SAVE, SIR, pHd, the standard fitting methods mentioned in the previous section and other methods not discussed here are all based on the following general procedure. First, let

$$\mathbf{z} = \Sigma^{-1/2}(\mathbf{x} - \mathrm{E}(\mathbf{x}))$$

denote the standardized version of $\mathbf{x}$, where $\Sigma = \mathrm{var}(\mathbf{x})$. Then $\mathcal{S}_{y|\mathbf{x}} = \Sigma^{-1/2}\mathcal{S}_{y|\mathbf{z}}$. Thus there is no loss of generality working on the $\mathbf{z}$ scale because any basis for $\mathcal{S}_{y|\mathbf{z}}$ can be back-transformed to a basis for $\mathcal{S}_{y|\mathbf{x}}$. Suppose we can find a consistent estimate $\hat{M}$ of a $p \times k$ procedure-specific population matrix M with the property that $\mathcal{S}(M) \subset \mathcal{S}_{y|\mathbf{z}}$. Then inference on at least a part of $\mathcal{S}_{y|\mathbf{z}}$ can be based on $\hat{M}$. SAVE, SIR, pHd and other numerical procedures differ on M and the method of estimating it. For example, for the ordinary least squares option of (7), set $k = 1$ and $M = \mathrm{cov}(y, \mathbf{z})$.

Let $\hat{s}_1 \geq \hat{s}_2 \geq \ldots \geq \hat{s}_q$ denote the singular values of $\hat{M}$, and let $\hat{u}_1, \ldots, \hat{u}_q$ denote the corresponding left singular vectors, where $q = \min(p, k)$. Assuming that $d = \mathcal{S}(M)$ is known,

$$\hat{\mathcal{S}}(M) = \mathcal{S}(\hat{u}_1, \ldots, \hat{u}_d)$$

is a consistent estimate of $\mathcal{S}(M)$. For use in practice, d will typically need to be replaced with an estimate $\hat{d}$ based on inferring about which of the singular values $\hat{s}_j$ is estimating 0, or on using GREG to infer about the regression of y on the q linearly transformed predictors $\hat{u}_j^T \mathbf{z}$, $j = 1, \ldots, q$.

References

[1] Box, G. E. P. (1980). Sampling and Bayes inference in scientific modeling and robustness (with discussion). *Journal of the Royal Statistical Society, Ser. A*, 142, 383–430.

[2] Box, G. E. P. and Draper, N. (1987). *Empirical Model Building and Response Surfaces*. New York: Wiley.

[3] Bura, E. (1997). Dimension reduction via parametric inverse regression. In Y. Dodge (ed.), *L1 Statistical Procedures and Related topics: IMS Lecture Notes, Vol* 31, 215–228: Investigates a parametric version of SIR [30].

[4] Chen, C-H, and Li, K. C. (1998). Can SIR be as popular as multiple linear regression? *Statistica Sinica*, 8, 289–316: Explores applied aspects of SIR [30], including standard errors for SIR directions.

[5] Cheng, C-S. and Li, K. C. (1995). A study of the method of principal Hessian directions for analysis of data from designed experiments. *Statistica Sinica*, 5, 617–637.

[6] Chiaromonte, F and Cook, R. D. (1997). On foundations of regression graphics. University of Minnesota, School of Statistics Technical Report 616: Consolidates and extends a theory for regression graphics, including graphical regression (GREG), based on dimension-reduction subspaces. Report is available from http://www.stat.umn.edu/PAPERS/tr616.html.

[7] Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, 19, 15–18.

[8] Cook, R. D. (1992a). Graphical regression. In Dodge, Y. and Whittaker, J. (eds), *Computational Statistics*, Vol. 1. New York: Springer-Verlag, 11–22: Contains introductory ideas for graphical regression (GREG) which is a paradigm for a "complete" regression analysis based on visual estimation of dimension-reduction subspaces.

[9] Cook, R. D. (1992b). Regression plotting based on quadratic predictors. In Dodge, Y. (ed), *L1-Statistical Analysis and Related Methods*. New York: North-Holland, 115-128: Proposes numerical estimates of the central subspace based on fitting quadratic kernels.

[10] Cook, R. D. (1993). Exploring partial residual plots. *Technometrics*, 35, 351–362: Uses selected foundations of regression graphics developed in [11] to reinterpret and extend partial residual plots, leading to a new class of plots called CERES plots.

[11] Cook, R. D. (1994a). On the interpretation of regression plots. *Journal of the American Statistical Association*, 89, 177–190: Basic ideas for regression graphics, including minimum dimension-reduction subspaces, the interpretation of 3D plots, GREG, direct sums of subspaces for constructing a sufficient summary plot; extends [8].

[12] Cook, R. D. (1994b). Using dimension-reduction subspaces to identify important inputs in models of physical systems. In *1994 Proceedings of the Section on Physical Engineering Sciences*, Washington: American Statistical Association: Case study of graphics in a regression with 80 predictors using GREG [8, 11], SIR [30], pHd [31], CERES plots [10]; introduction of central dimension-reduction subspaces.

[13] Cook, R. D. (1995). Graphics for studying the net of regression predictors. *Statistica Sinica*, 5, 689–709: Net effect plots for studying the contribution of a predictor to a regression problem; generalizes added-variable plots.

[14] Cook, R. D. (1996). Graphics for regressions with a binary response. *Journal of the American Statistical Association* 91, 983–992: Adapts the foundations of regression graphics, including GREG [8, 11], and develops graphical methodology for regressions with binary response variables; includes results on the existence of central dimension-reduction subspaces.

[15] Cook, R. D. (1998a). Principal Hessian directions revisited (with discussion). *Journal of the American Statistical Association*, 93, 84–100: Rephrases the foundations for pHd [31] in terms of central dimension-reduction subspaces and proposes modifications of the methodology.

[16] Cook, R. D. (1998b). *Regression Graphics: Ideas for studying regressions thru graphics*. New York: Wiley. (to appear in the Fall 1998). This is a comprehensive treatment of regression graphics suitable for Ph.D. students and others who may want theory along with methodology and application.

[17] Cook, R. D. and Croos-Dabrera, R. (1998). Partial residual plots in generalized linear models. *Journal of the American Statistical Association*, 93, 730–739: Extents [10] to generalized linear models.

[18] Cook, R. D. and Lee, H. (1998). Dimension reduction in regressions with a binary response. Submitted: Develops and compares dimensions-reduction methodology – SIR [30] pHd [31], SAVE [21] and a new method based on the difference of the conditional covariance matrices of the predictors given the response – in regressions with a binary response.

[19] Cook, R. D. and Nachtsheim, C. J. (1994). Re-weighting to achieve elliptically contoured covariates in regression. *Journal of the American Statistical Association*, 89, 592–599: Methodology for weighting the observations so the predictors satisfy the linearity condition needed to ensure that certain estimates fall in the central subspace asymptotically.

[20] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression.* London: Chapman and Hall.

[21] Cook, R. D. and Weisberg, S. (1991). Discussion of Li [30]. *Journal of the American Statistical Association*, 86, 328–332: Introduction of sliced average variance estimation, SAVE, that can be used to estimate the central dimension-reduction subspace.

[22] Cook, R. D. and Weisberg, S. (1994a). *An Introduction to Regression Graphics*. New York: Wiley. This book integrates selected graphical ideas at the introductory level. It comes with the R-code, a computer program for regression graphics written in LISP-STAT [34]. The book serves as the manual for R-code which works under Macintosh, Windows and Unix systems. See http://www.stat.umn.edu/ and follow "Projects & Papers".

[23] Cook, R. D. and Weisberg, S. (1994b). Transforming a response variable for linearity. *Biometrika*, 81, 731–737: Proposes inverse response plots as a means of estimating response transformations graphically.

[24] Cook, R. D. and Weisberg, S. (1997). Graphics for assessing the adequacy of regression models. *Journal of the American Statistical Association*, 92, 490–499: Proposes general graphical methods for investigating the adequacy of almost any regression model, including linear regression, nonlinear regression, generalized linear models and additive models.

[25] Cook, R. D. and Wetzel, N. (1993). Exploring regression structure with graphics (Invited with discussion). *TEST*, 2 ,1–57: Develops practical aspects of GREG.

[26] Ferr, L. (1998). Determining the dimension of sliced inverse regression and related methods. *Journal of the American Statistical Association*, 93, 132-140: Introduces model selection methods for estimating in the context of SIR.

[27] Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. The *Annals of Statistics*, 21, 867–889: Shows that the linearity condition will be reasonable in many problems.

[28] Hsing, T. and Carroll (1992). An asymptotic theory of sliced inverse regression. *Annals of Statistics*, 20, 1040–1061: Develops the asymptotic theory for the "two-slice" version of SIR [30].

[29] Huber, P. (1985). Projection pursuit (with discussion). *Annals of Statistics*, 13, 435–474.

[30] Li, K. C. (1991). Sliced inverse regression for dimension reduction (with discussion). *Journal of the American Statistical Association*, 86, 316–342: Basis for sliced inverse regression, SIR.

[31] 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: Basis for pHd.

[32] Li, K. C. (1997). Nonlinear confounding in high-dimensional regression. *Annals of Statistics*, 25, 577–612.

[33] Li, K. C. and Duan, N. (1989). Regression analysis under link violation. *Annals of Statistics*, 17, 1009–1052: Shows conditions under which usual estimation methods (OLS, M-estimates, ..) produce estimates that fall in a one-dimensional dimension-reduction subspace asymptotically.

[34] Tierney, L. (1990). *LISP-STAT*. New York: Wiley: Resource for the parent statistical language of the R-code.

[35] Zhu, L. X. and Fang, K. T. (1996). Asymptotics for kernel estimates of sliced inverse regression. *Annals of Statistics*, 24, 1053–1068: Uses smoothing rather than slicing to estimate the kernel matrix for SIR [30].

[36] Zhu, L. X. and Ng, K. W. (1995). Asymptotics of sliced inverse regression. *Statistica Sinica*, 5, 727–736.