Using Dimension-Reduction Subspaces to Identify Important Inputs in Models of Physical Systems*

R. Dennis Cook, University of Minnesota
Department of Applied Statistics, 1994 Buford Ave., St. Paul, MN 55108

Key Words: CERES Plots, Generalized Additive Models, pHd, Regression Graphics, SIR

Abstract

Graphical methods based on dimension-reduction subspaces for regression problems (Cook 1994) may be useful for studying the relative importance of inputs in computer models of physical systems. Sliced inverse regression (Li 1991), principal Hessian directions (Li 1992), CERES plots (Cook 1993), and inverse response plots (Cook and Weisberg 1994) are recent methods that can identify characteristics of dimension-reduction subspaces and facilitate graphical analyses of the important variables. These methods work well in the problem under consideration and may serve as important graphical methodology more generally, particularly when combined into a single paradigm for graphical regression analysis.

1 Introduction

A large simulation code was developed at Los Alamos National Laboratories (LANL) to aid in a study of the fate of an environmental contaminant introduced into an ecosystem. Environmental contaminants have the potential for ecological and human health effects due to their toxicological properties. A good appreciation of the ecological risk associated with contamination requires an understanding of the dynamics of contamination.

The LANL code is essentially a compartment model with 8 compartments: vegetation interior, vegetation surface, terrestrial invertebrates, small herbivores, large herbivores, insectivores, predators and litter. The litter compartment is a sink that only receives the contaminant. The model consists of a set of coupled differential equations representing the various compartments in the ecosystem. It is based on the assumption that the contaminant enters the ecosystem by dissolution to water and then moves through the food web by one organism consuming another. The concentration of the contaminant in water is assumed to be constant for any run of the code. In total, the model requires 84 inputs and, for the purposes of this study, the response y is the amount of contamination in the terrestrial invertebrates at day 5000. The specific issue addressed in this article is how to assess the relative importance of the inputs for explaining the variation in y. In short, which inputs are most important?

Any assessment of relative importance will depend on the joint distribution of the inputs. The scientists who developed the LANL model provided ranges for each of the 84 input variables and an estimate of the nominal value. Based on this and other information, Dick Beckman and Mike McKay, members of the LANL Statistics Group, developed beta distributions for each of the 84 inputs. No J-shaped or U-shaped beta distributions were allowed, but the distributions could be symmetric or highly skewed. The inputs were regarded as mutually independent so their joint distribution is the product of the individual marginal beta densities. This general setup will be called the *beta sampling plan*. Let x denote the 84×1 vector of inputs. I will refer to the sample space for the inputs as the *operability region* and use p as a generic representation for the number of inputs.

2 Assessment rationale

The notion of relative importance is perhaps easiest to quantify locally, where measures of importance might be based on rates of change around selected points. The results of such an analysis could be elusive, however, particularly if the response surface is flat on the average but has many local peaks and

*The content of this article was delivered in 1994 during an invited paper session at the Annual Meetings of the American Statistical Association, and subsequently appeared in the 1994 Proceedings of the Section on Physical and Engineering Sciences.

valleys. The local relative importance of the inputs near the nominal value of x need not be the same as that near the edges of the operability region, for example.

Alternatively, relative importance might be approached globally, looking for inputs or functions of inputs that account for most of the variation in the response across the operability region. Such an approach might yield useful results if there are clear global trends, but it could well fail if the primary variation in the response is local. In this article I concentrate on the global approach since experience with a distinct but similar simulation code indicated that there may well be clear global trends.

Let $F(y|x)$ denote the cdf of the response given the inputs. Since the simulation code itself is deterministic, $F(y|x)$ places mass 1 at the value of y corresponding to the value of x. The basic idea in a global analysis is to find a low dimensional function $g(x)$ that can serve as a substitute for x itself without important loss of information. In particular, $F(y|g(x))$ should be a good approximation of $F(y|x)$,

$$F(y|g(x)) \approx F(y|x) \tag{1}$$

for all values of x in the operability region. If a low dimensional function g can be found so that (1) holds to a useful approximation, then the problem of assessing relative importance might be eased considerably. The success of this approach depends in part on the specific types of functions allowed for g. In this article I concentrate on linear functions and linear functions after coordinate transformations of the inputs.

In the next section I briefly review the methodology that will be adapted to study the relative importance of the inputs.

3 Methodology

Though out this section I consider a generic regression problem with response y and $p \times 1$ vector of predictors x. The cdf for the regression problem will still be denoted as $F(y|x)$. The notation $u \perp\!\!\!\perp v$ will be used to indicate that the random variables u and v are independent. Similarly, $u \perp\!\!\!\perp v|z$ means that u and v are independent given any value for the random variable z. The subspace of $\Re^p$ spanned by the columns of a $p \times k$ matrix B will be indicated by $S(B)$.

3.1 Dimension-reduction subspaces

Let $B = (b_1, \ldots, b_k)$ denote a $p \times k$, $k \leq p$, matrix so that

$$F(y|B^T x) = F(y|x) \tag{2}$$

for all values of x in the relevant sample space. In reference to (1), $g(x) = B^T x$. If (2) holds then $y \perp\!\!\!\perp x|B^T x$ and x can be replaced with $B^T x$ in the regression problem without loss of information. Additionally, if (2) holds then it also holds with B replaced by BA for any full rank $k \times k$ matrix A. Thus, (2) is really a statement about $S(B)$ rather than about a particular basis B.

If (2) holds then $S(B)$ is called a *dimension-reduction subspace for* $F(y|x)$. Dimension-reduction subspaces with the minimum dimension are of particular importance since they provide the maximum reduction of the predictor dimension without loss of information: If $S_{y|x}$ is a dimension-reduction subspace that is contained in all dimension-reduction subspaces S_{drs},

$$S_{y|x} = \cap S_{drs}, \tag{3}$$

then $S_{y|x}$ is called a *central dimension-reduction subspace for* $F(y|x)$. Clearly, a central dimension-reduction subspace has minimum dimension. However, a dimension-reduction subspace with minimal dimension need not be a central dimension-reduction subspace because such subspaces need not exist. In this article central dimension-reduction subspaces are assumed to exist and will be denoted by $S_{y|x}$ with the subscripts indicating the particular regression.

While central dimension-reduction subspaces do not always exist, conditions can be imposed to guarantee their existence: If x has a density that is positive everywhere on $\Re^p$ then central dimension-reduction subspaces always exist (Cook 1994). This result does not apply in the LANL problem because the density of x is zero outside the unit cube. However, the following lemma shows that if $\mathrm{E}(y|x)$ is well-behaved then central dimension-reduction subspaces again exist.

Lemma 1 *Let $S(\alpha)$ and $S(\phi)$ be dimension-reduction subspaces for a regression problem that is characterized by its regression function, $y \perp\!\!\!\perp x|E(y|x)$. If x has a density $f(x) > 0$ for $x \in \Omega \subset \Re^p$ and if $E(y|x)$ can be expressed as a convergent power series in the coordinates of $x = (x_k)$,*

$$E(y|x) = \sum_{k_1,\ldots,k_p}^{\infty} a_{k_1,\ldots,k_p} x_1^{k_1} \cdots x_p^{k_p}$$

then $S(\alpha) \cap S(\phi)$ is a dimension-reduction subspace.

In view of this lemma it seems reasonable to assume that central dimension-reduction subspaces exist in the LANL problem. For further discussion of dimension-reduction subspaces, see Cook (1994), Cook and Wetzel (1993) and Li (1991).

3.2 SIR and pHd

Several methods have been proposed in recent years to estimate dimension-reduction subspaces. I will confine attention to two methods in this section. The first is sliced inverse regression (SIR), as proposed by Li (1991) and Duan and Li (1991). Letting η denote a basis for $S_{y|x}$, SIR is based on the result that if $\mathrm{E}(x|\eta^T x)$ is a linear function of the value of $\eta^T x$ then

$$\mathrm{E}[\Sigma^{-1}(x - \mathrm{E}(x))|y] \in S_{y|x} \tag{4}$$

for all values of y. The SIR procedure for estimating $S_{y|x}$ employs a principal components analysis of an estimate of the covariance matrix

$$\mathrm{var}\{\mathrm{E}[\Sigma^{-1}(x - \mathrm{E}(x))|y]\}$$

obtained by slicing on the observed values of y. Although SIR requires no conditions on $F(y|x)$, circumstances can arise in which the manifold

$$\mathrm{E}[\Sigma^{-1}(x - \mathrm{E}(x))|y]$$

obtained by allowing y to vary over its sample space fails to span $S_{y|x}$. In the extreme, it is possible to have

$$\mathrm{E}[\Sigma^{-1}(x - \mathrm{E}(x))|y] = 0$$

for all values of y while $F(y|x) \neq F(y)$. The existence of central dimension-reduction subspaces helps avoid such non-informative possibilities but provides no universal guarantees. For example, experience has shown that SIR may miss directions in $S_{y|x}$ when the regression function $\mathrm{E}(y|x)$ is curved with little linear trend (Cook and Weisberg 1991). Tests to assist in estimating the dimension of $S_{y|x}$ are a part of the SIR procedure but require x to be normally distributed.

The second method is based on finding the principal Hessian directions (pHd), as proposed by Li (1992). Let $\Sigma = \mathrm{var}(x)$ and

$$\Sigma_{yxx} = \mathrm{E}[y - \mathrm{E}(y)][x - \mathrm{E}(x)][x - \mathrm{E}(x)]^T$$

Next, define the eigenvectors ν_k and eigenvalues λ_k as the solutions to

$$\Sigma_{yxx}\nu_k = \lambda_k \Sigma \nu_k$$

If x is normally distributed then (Li 1992)

$$\nu_k \in S_{y|x} \tag{5}$$

for all $\lambda_k \neq 0$. Thus, $S_{y|x}$ can be estimated via an eigenvalue decomposition of Σ_{yxx} relative to Σ. pHd seems to work well for finding directions in $S_{y|x}$ when $\mathrm{E}(y|x)$ is curved and has little linear tend. For this reason it may serve as a useful compliment to SIR. On the down side, pHd can be sensitive to non-normality in x. The weighting procedure proposed by Cook and Nachtsheim (1994) might be used to reduce the importance of predictor normality. The results of Hall and Li (1993) suggest that the various conditions for SIR and pHd may hold to a reasonable approximation in many problems.

3.3 Predictor transformations

In many regression problems, coordinate-wise transformations of the predictors $t(x) = (t_j(x_j))$ can facilitate an analysis. Specifically, transformations might be found so that $F(y|x) = F(y|t(x))$ for all values of x in the relevant sample space while $\dim(S_{y|t(x)}) < \dim(S_{y|x})$ where "dim" denotes dimension.

For example, suppose that $p = 2$ and that $F(y|x)$ can be described by the linear model

$$y|x = x_1 + \log(x_2) + \epsilon$$

where $\epsilon \perp\!\!\!\perp x$. For that regression problem, $S_{y|x} = I_2$ and $\dim(S_{y|x}) = 2$. However, defining $t_1(x_1) = x_1$ and $t_2(x_2) = \log(x_2)$ gives a new regression problem with $S_{y|t(x)} = S((1,1)^T)$ and $\dim(S_{y|t(x)}) = 1$ so that the dimension of the central dimension-reduction subspace has been reduced by 1.

CERES plots as described by Cook (1993) will be used to determine predictor transformations that may result in a simplified regression. CERES plots are a generalization of partial residual plots that do not require linear regression functions among the predictors.

3.4 Response transformations

Let $t_y(y)$ denote a strictly monotonic transformation of the response. Then it is not difficult to see that

$$S_{t_y(y)|x} = S_{y|x} \tag{6}$$

Accordingly, central dimension-reduction subspaces are invariant under monotonic transformations of the response and thus such transformations have essentially no technical role in the identification of $S_{y|x}$. Response transformations may be important

for increasing the resolution of various plots and during the traditional modeling phase of an analysis, however.

3.5 An analysis paradigm

In broad terms, one paradigm for combining the tools of the previous sections in a regression analysis is as follows. The first step is to explain variation in the y by using response and predictor transformations to linearize the regression function as far as practically reasonable. Specifically, write the regression function as

$$\mathrm{E}(t_y(y)|t(x)) = \beta_0 + \beta^T t(x) + d(t(x)) \qquad (7)$$

where

$$d(t(x)) = [\mathrm{E}(t_y(y)|t(x)) - \beta_0 - \beta^T t(x)],$$

t_y is a monotonic transformation of the response and $t(x)$ is a coordinate-wise transformation of the predictors as described in Section 3.3. The general idea is to choose t_y and $t(x)$ so that the variation in the deviations from linearity $d(t(x))$ is small relative to the variation in the linear component $\beta^T t(x)$. Once the transformations have been determined, the linear component might be estimated by using SIR, ordinary least squares or a number of other procedures, depending on the behavior of the predictors.

Let

$$e_t = t_y(y) - \beta_0 - \beta^T t(x)$$

denote the population residuals from a linear regression of $t_y(y)$ on $t(x)$. Under fairly general conditions on the marginal distribution of $t(x)$, $\beta \in S_{t_y|t(x)}$ (Li and Duan 1989, Hall and Li 1993, Cook 1994, Cook and Nachtsheim 1994). It then follows that $S_{e_t|t(x)} \subset S_{t_y|t(x)}$. pHd might now be applied to the regression of e_t on $t(x)$ to estimate vectors ν_k in $S_{e_t|t(x)}$. Plots of t_y versus the q important pHd linear combinations $\nu_k^T t(x)$, $k = 1, \ldots, q$, of the transformed predictors can then be smoothed, and the smooth

$$g_1(V_1^T t(x)) = g_1(\nu_1^T t(x), \ldots, \nu_q^T t(x))$$

included in the regression function

$$\begin{aligned} \mathrm{E}(t_y(y)|t(x)) &= \beta_0 + \beta^T t(x) + \\ & \quad g_1(V_1^T t(x)) + d_1(t(x)) \end{aligned}$$

where $V_1 = (\nu_k)$ is the $p \times q$ matrix with columns ν_k. Hopefully q will typically be small, say 1 or 2.

Iteration can be used at this point, recalculating residuals and reapplying pHd until the residual subspaces $S_{e_t|t(x)}$ has apparently been exhausted. The final regression function will be of the form,

$$\begin{aligned} \mathrm{E}(t_y(y)|t(x)) &= \beta_0 + \beta^T t(x) + g_1(V_1^T t(x)) + \\ & \quad \cdots + g_m(V_m^T t(x)) + d_m(t(x)) \end{aligned}$$

where $g_1, \ldots, g_m$ are the smooths deriving from application of pHd. With luck, the variation in the final deviations $d_m(t(x))$ will be small relative to that in the rest of the regression function.

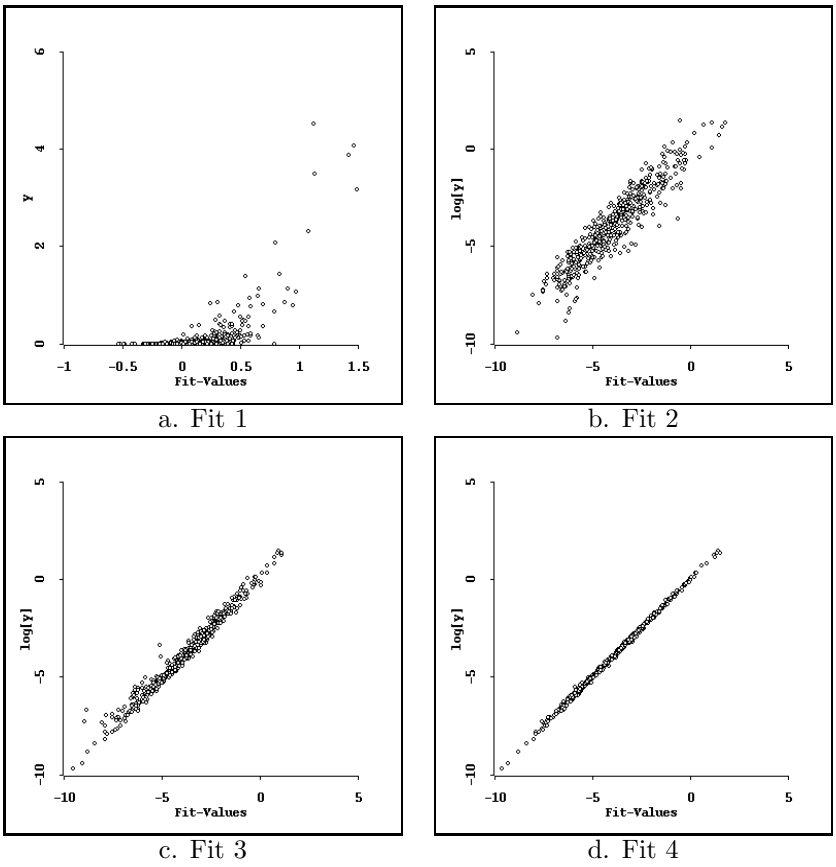
4 Analysis

The analysis described in this section is based on the beta sampling plan as described in Section 1. More specifically, 500 Latin hypercube samples (McKay, Beckman and Conover 1979) were taken at the midpoints of equal probable slices across the ranges of the inputs. Although more observations could have been taken easily, it was felt that 500 would be sufficient to gain useful information about the inputs and to guide further sampling as necessary. All sampling was carried out by members of the LANL Statistics Group.

As a baseline, Figure 1a gives a plot of the fitted values versus the response from the OLS regression of the response on the first-order model containing all 84 inputs. Letting η denote a basis for $S_{y|x}$, this starting point is based on the rationale that if $\mathrm{E}(x|\eta^T x)$ is a linear function of the value of $\eta^T x$ and certain regularity conditions hold then the 84×1 vector of coefficient estimates $\hat{b}$ converges to a vector in $S_{y|x}$ (Li and Duan 1989). OLS estimation was used for convenience only. Many other methods based on convex objective functions could be used as well.

Figure 1a indicates that the response variable is highly skewed, ranging over several orders of magnitude from 2×10^{-5} to 4.5. Figure 1b was constructed as Figure 1a, except that y was replaced by $\log(y)$. There is fairly strong linearity in at least one direction in the log scale with $R^2 = 0.86$. Figure 1b seems to be a reasonable starting point for further analysis. The transformation from y to $\log(y)$ does not change the objective of the analysis because $S_{\log(y)|x} = S_{y|x}$.

Application of SIR with 100 slices of 5 observations each to the regression with $\log(y)$ did not provide anything beyond the projection in Figure 1b. SIR strongly indicated that $\dim(S_{\log(y)|x}) = 1$. The correlation between the linear combination of the inputs implied by SIR and the OLS fitted values is 0.99, so both procedures indicate essentially the same solution. pHd did not produce any information that was felt relevant to $S_{\log(y)|x}$ beyond that provided by OLS.



a. Fit 1 b. Fit 2

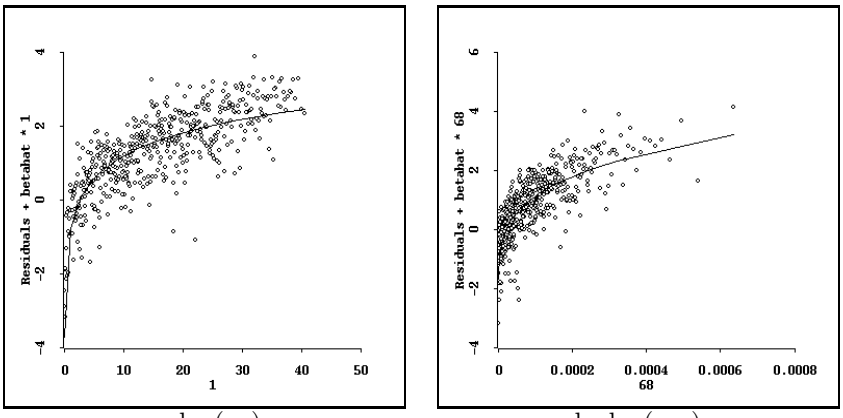
c. Fit 3 d. Fit 4

Figure 1: Plots of the log response versus the OLS fitted values from four stages in the analysis of the data from the beta sampling plan. (a) Original data. (b) $\log(y)$ and original inputs. (c) $\log(y)$ and transformed predictors following the removal of nonessential inputs. (d) Final fit.

The plot in Figure 1b shows a strong linear trend and thus attempting to find an improved approximation through input transformations as described in Sections 3.3 and 3.5 may yield useful results. Because of the way in which the inputs were generated, conditional expectations of the form $\mathrm{E}(x_j|x_k)$ should be essentially constant, implying that partial residual plots may be used to suggest appropriate input transformations (Cook 1993). The two panels of Figure 2 show illustrative partial residual plots for inputs 1 and 68 along with superimposed estimated power curves. The implications from Figure 2 are that $\dim(S_{\log(y)|x}) > 1$ and that improvements may result by replacing x_1 and x_{68} with their logarithms. Partial residual plots were used iteratively, restricting the class of transformations to power transformations. This procedure essentially fits a generalized additive model (Hastie and Tibshirani 1990) with a restricted class of smooths. Because powers seemed to work well in each case, there was little reason to allow more general smoothing of the partial residual plots. In total, 13 inputs were transformed in this way and, as it turned out, the only transformations used were the logarithm and the square root.

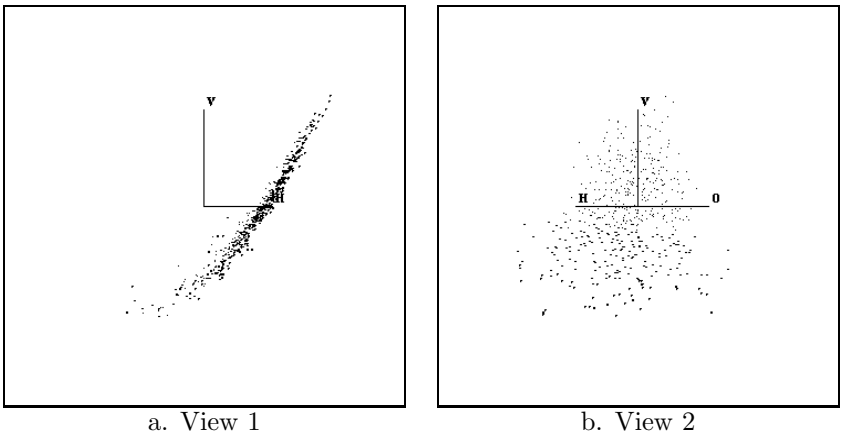
Following the transformation there seemed to be many inputs that had little if any effect on the log response under the beta sampling plan. Using a backward elimination procedure, 54 input variables were removed in the hope of increasing the power of subsequent analysis. The variables removed were judged to be relatively unimportant. Let X_{30} denote the vector of the 30 remaining inputs with the transformations applied. A plot of the log response versus the OLS fitted values for the regression on X_{30} is shown in Figure 1c. A reduction in variance relative to Figure 1b is evident and $R^2 = 0.976$. Figure 1d, which is the result of further analysis based on the reduced model with transformed predictors, will be discussed a bit later.

Following the rationale in Section 3.5, pHd was next applied to the residuals from the OLS regression of $\log(y)$ on X_{30}. The results indicated one clear di-



a. $\log(x_1)$ b. $\log(x_{68})$

Figure 2: Partial residual plots for the first and 68th input variables with response $\log(y)$. The superimposed curves, (a) $\log(x_1)$ and (b) $\log(x_{68})$, are the estimated input transformations.



a. View 1 b. View 2

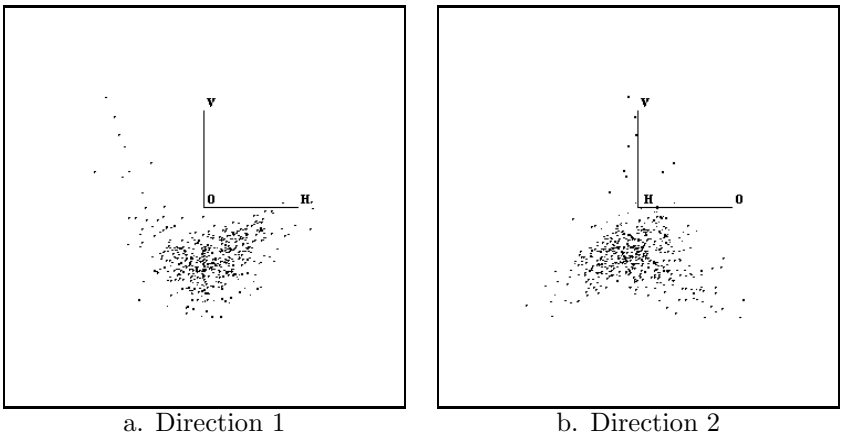
Figure 3: Two views of the 3D partial residual plot for t_{68} and t_{69} in the regression with the transformed and reduced inputs.

rection, with the possibility of one or two more. Evidently, the associated central dimension-reduction subspace is at least 2 and possibly 3 dimensional. An inspection of the coefficients associated with the first principal Hessian direction indicated that the nonlinearity rests almost exclusively with the sum of the two transformed inputs $t_{68} = \log(x_{68})$ and $t_{69} = \log(x_{69})$. Two views of a 3D partial residual plot (Cook 1993, Cook and Weisberg 1994) for t_{68} and t_{69} are shown in Figure 3. View 2 is provided for contrast. View 1 gives a 2D projection of best visual fit and strongly suggests that the structural dimension of the 3D plot is 1 (Cook 1994, Cook and Weisberg 1994, Cook and Wetzel 1993). The 3D partial residual plot then confirms the indication of pHd that the sum $t_{68} + t_{69}$ is relevant. Because the coefficients of t_{68} and t_{69} are nearly identical in the visual fit of Figure 1c and because the projection in Figure 3a is well fit with a quadratic, these two predictors were replaced with linear and quadratic terms in their sum, $t_{68} + t_{69} = \log(x_{68}x_{69})$. Let X^1_{30} denote the resulting set of 30 transformed inputs and let e^1_{30} denote the OLS residuals from the regression of $\log(y)$ on X^1_{30}. Finally, let X^1_{29} denote the vector of 29 predictors constructed from X_{30} by replacing (t_{68}, t_{69}) with the sum $t_{68} + t_{69}$. The input vectors X^1_{30} and X^1_{29} differ only by the quadratic term $(t_{68} + t_{69})^2$

pHd was now applied to the regression of e^1_{30} on X^1_{29}. Two significant linear combinations phd_1 and phd_2 were indicated, implying that

$$\dim(S_{e^1_{30}|X^1_{29}}) = 2$$

The 3D plot of e^1_{30} versus phd_1 and phd_2 seems fairly complicated, resembling a saddle with a high back. Plots of e^1_{30} versus phd_1 and phd_2 are shown in Figure 4. These plots were smoothed and the smooths



a. Direction 1 b. Direction 2

Figure 4: Plots of the response versus the first two principal Hessian directions from the regression on X^1_{29} of the residuals from $\log(y)$ on X^1_{30}.

added to X^1_{30}, giving a new vector of 32 predictors X^2_{32}. Using a two-dimensional smooth may have produced somewhat better results.

Beyond this point little improvement seemed possible. A plot of $\log(y)$ versus the OLS fitted values from the regression of $\log(y)$ on X^2_{32} is shown in Figure 1d. The corresponding $R^2 = 0.999$.

The final regression function is of the form

$$\begin{aligned} \mathrm{E}(\log(y)|X) \quad \approx \quad & \beta_0 + \beta^T X^1_{29} \\ & +\alpha_1 \log^2(x_{68}x_{69}) \\ & +g_1(\alpha_2^T X^1_{29}) \\ & +g_2(\alpha_3^T X^1_{29}) \end{aligned} \tag{8}$$

where g_1 and g_2 represent the smooths resulting from the second application of pHd. Further refinement may be desirable for a parsimonious predictive regression function but the present analysis seems sufficient for a reasonable idea about the important inputs. Figure 1d suggests that there is a dominant linear trend cross the operability region in the transformed scale. This combined with the relative simplicity of the nonlinear terms in (8) and the approximate independence of the inputs can be used to assess the relative importance of the transformed inputs by partitioning $\mathrm{var}(y)$ in terms of the various components of the final model.

Based on a partition of $\mathrm{var}(y)$, the sum $\log(x_{68}) + \log(x_{69})$ was judged to be the most important input, accounting for about 35 percent of the variation. The second most important predictor $\log(x_1)$ accounts for about 25 percent of $\mathrm{var}(y)$. Other predictors in X_{30} were partially ordered in 3 sets as follows: the three inputs

$$[\log(x_{24}), \log(x_{63}), \log(x_{84})]$$

account for about 17 percent of the variation and are more important than the four inputs

$$[\log(x_{35}), x_{48}, x_{67}, x_{83}]$$

which are in turn more important that the five inputs

$$[x_{19}, x_{54}, \log(x_{55}), \log(x_{61}), x_{65}]$$

The remaining inputs were judged to of relatively minor importance.

5 Discussion

Guided by central dimension-reduction subspaces, the basic procedure employed here consists of using generalized additive modeling followed by iterative application of pHd to the residuals. SIR did not provide any information beyond that provided by OLS. As indicated by the sequence of plots in Figure 1, the method seemed to produce good results and may represent a useful basis for graphical analyses more generally.

It turns out that the decision to replace the pair of predictors (t_{68}, t_{69}) with their sum was exactly right: The LANL code depends only on the product $x_{68}x_{69}$, as pointed out by Dick Beckman in response to the first draft of this article.

The inputs were generated to be independent and many steps in the analysis were taken with this in mind. Linear relationships among the inputs may have made the analysis somewhat harder, but would not have required changing the underlying methodology. However, nonlinear relationships among the predictors could have made assessing importance much more difficult.

All techniques used in this analysis are easily accessible in the *R-code* (Cook and Weisberg 1994).

6 Acknowledgment

This work was supported in part by grant DMS-9212413 from the National Science Foundation.

7 References

Cook, R. D. (1993). Exploring partial residual plots. *Technometrics* **35**, 351-362.

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

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

Cook, R. D. and Weisberg, S. (1991). Comment on Li (1991), *op. cit. Journal of the American Statistical Association* **86**, 328-332.

Cook, R. D. and Weisberg, S. (1994). *An Introduction to Regression Graphics.* New York:Wiley.

Cook, R. D. and Wetzel, N. (1993). Exploring regression structure with graphics (with discussion). *TEST* **2**, 1-57.

Duan, N.and Li, K. C. (1991). Slicing regression: A link-free regression method. *Annals of Statistics* **19**, 505-530.

Hall, P. and Li, K. C. (1993). On almost linearity of low dimensional projections from high dimensional data. *Annals of Statistics* **21**, 867-889.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models.* New York: Chapman & Hall.

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.

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

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

McKay, M. D., Beckman, R. J. and Conover, W. J. (1979). A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21**, 239-245.