Statistica Sinica **5**(1995), 617-639

A STUDY OF THE METHOD OF PRINCIPAL HESSIAN DIRECTION FOR ANALYSIS OF DATA FROM DESIGNED EXPERIMENTS

Ching-Shui Cheng and Ker-Chau Li

University of California at Berkeley and University of California at Los Angeles

Abstract. Methods for high dimensional regression are often discussed in studies where regressors follow continuous distributions. How well do they perform when applied to data collected by designed experiments where design points are rather isolated from each other? We initiate such a study by focusing on the method of principal Hessian direction (pHd) (Li (1992)). Quadratic regression surfaces are considered first and then extended to general nonlinear surfaces. Special attention is given to factorial designs and rotatable response surface designs. Both theoretic and empirical results are presented.

Key words and phrases: ANOVA, dimension reduction, graphics, factorial designs, principal Hessian directions, residual plots, rotatable designs, sliced inverse regression, visualization.

1. Introduction

Analysis of designed experiments is usually carried out via simple parametric modeling. This practice is in part due to the mathematical tractability of design theory (Fedorov (1972), Kiefer (1959, 1974), Steinburg and Hunter (1984)). In realty, however, the collected data may not always follow the model speculated at the design stage. The purpose of this paper is to enrich the analysis by bringing in recent dimension reduction and data visualization techniques from general regression problems.

To see how restrictive parametric modeling might be, consider a full two-level factorial design with p factors $\boldsymbol{x} = (x_1, \ldots, x_p)$ and y the output variable. Using traditional ANOVA analysis, a response function with the form

$$E(y|\boldsymbol{x}) = h(x_1 + x_2 + x_3 + x_4) \tag{1.1}$$

for example, will lead to four main effects, six two-term interactions, four three-term interactions, and one four-term interaction, if $h(\cdot)$ is a general nonlinear function. This is hardly useful because of the well-known poor behavior in interpolation for high order polynomials.

Such situations are not uncommon in designed experiments. The well-known Box and Cox transformation was applied first in a 3^3 full factorial design (Box and Cox (1964)). But sometimes the shape of h could be far from the power family to validate the transformation; for example, $h(x) = |x|$.

This important issue is seldom discussed in the literature of designed experiments, although similar such questions have been the focus of several recent studies on high dimensional data analysis. To approximate a general nonlinear surface, computer intensive methods such as projection pursuit regression (Friedman and Stuetzle (1981), Huber (1985), Hall (1989), Chen (1991)), ACE (Breiman and Friedman (1985)), CART (Breiman, Friedman, Olshen and Stone (1984)), MARS (Friedman (1991)), SUPPORT (Chaudhuri, Huang, Loh and Yao (1994)), etc., generally take good advantage of the machine's great computing power. They search through several classes of well-motivated functions iteratively (and adaptively) for an optimal solution. Alternatively, instead of being driven by functional fitting, SIR (Li (1991), Duan and Li (1991), Hsing and Carroll (1992)) and pHd (Li (1992)) rely on a rather different strategy which attempts to make good use of graphical facilities in modern statistical software. They find a small number of critical directions for projecting the high dimensional data on computer screens. Such graphical information provides good insight about the shape of the response function and other data structure. Model diagnosis/selection and low-dimensional nonparametric smoothing can then be more fruitfully carried out.

These high dimensional regression techniques have been applied to a variety of data sets. But the discussion often stems from situations where the regressor variables follow continuous distributions. How well do they perform in designed experiments where the design points are rather isolated from each other? In this paper, we initiate such a study by focusing on the method of pHd, owing to its nice connection to traditional analysis.

Equation (1.1) is a special case of the dimension reduction model in general regression problems (Li (1991)): $y = f(\beta_1'\boldsymbol{x}, \ldots, \beta_k'\boldsymbol{x}, \epsilon)$. Here f is completely unspecified and so is the distribution of ϵ which is independent of $\boldsymbol{x}$. A vector b in the space spanned by the β vectors is called an e.d.r. (effective dimension reduction) direction, and the linear combination $b'\boldsymbol{x}$ of $\boldsymbol{x}$ is called an e.d.r. variate. Usually k is much smaller than p and the structure of f is easier to study after finding e.d.r. variates.

The method of pHd aims at finding the directions along which the regression surface shows the largest curvature in a suitable average sense. Let $g(\boldsymbol{x})$ be the response function:

$$g(\boldsymbol{x}) = E(y|\boldsymbol{x}) = h(\beta_1'\boldsymbol{x}, \ldots, \beta_k'\boldsymbol{x}), \tag{1.2}$$

where h is a function with k arguments. The Hessian matrix of the regression function, $H(\boldsymbol{x}) =$ the p by p matrix of second partial derivatives $\frac{\partial^2 g(\boldsymbol{x})}{\partial x_i \partial x_j}$, is used to convey the curvature information. Li (1992) treats $\boldsymbol{x}$ as random with a covariance matrix Σ_x and defines the principal Hessian directions to be the eigenvectors v_i for the matrix $(EH(\boldsymbol{x}))\Sigma_x$:

$$(EH(\boldsymbol{x}))\Sigma_x v_i = \lambda_i v_i, \quad |\lambda_1| \geq \cdots \geq |\lambda_p|.$$

The right-multiplication of the covariance matrix Σ_x to the average Hessian $EH(\boldsymbol{x})$ is needed to achieve affine invariance which is desirable for visualization. We are interested in the first few pHd directions. As a matter of fact, under (1.2), the Hessian matrix at each point $\boldsymbol{x}$ is of rank at most k, because along the directions orthogonal to the e.d.r. space, the regression surface is flat. Exploiting this simple property, we can verify that all but the first k eigenvalues λ_i are zero and that the nonzero eigenvectors are e.d.r. directions.

There are three implementation versions of pHd, the y-based, the r-based, and the q-based methods, depending on different ways of estimating $EH(\boldsymbol{x})$ under the normal distribution assumption on $\boldsymbol{x}$. However, the discussion in Li (1992) extends to elliptic distributions for $\boldsymbol{x}$, and even to a much weaker condition of linearity (Condition 3.1 of Li (1991)) which is reasonable for many data sets (Hall and Li (1993)).

The numerical results for different versions of pHd are often close to each other. As shown in Section 2, for two-level factorial designs and rotatable designs, they are indeed equivalent. This simplifies our discussion somewhat and we shall concentrate on the behavior of the q-based version in Section 3 and following sections.

Whether the estimated pHd directions are close to the e.d.r. space or not depends on three factors: the design, the true response function, and the distribution of random errors in y. The impact of the third factor on eigenvalue decomposition can be studied via perturbation techniques following Li (1992). Setting aside this factor, an immediate question is about the Fisher consistency property: will pHd find e.d.r. directions in the population version? This is the main focus of this paper.

It is natural to begin with a quadratic response function. We show that q-based pHd is consistent in Section 3. Applications are then made to a response surface example and to two-level factorial designs in Sections 3.1 and 3.2 respectively. Plots of y against pHd variates (i.e. projections of $\boldsymbol{x}$ along the pHd directions) are studied.

Section 4 deals with general nonlinear surfaces. For general designs, a sufficient condition for consistency is derived in Section 4.1. A practical implication is that if the non-quadratic part of the function h in (1.2), which is uncorrelated

with any quadratic functions of the e.d.r. variates, also has no (or low) correlation with other quadratic polynomials of $\boldsymbol{x}$, then we can expect pHd to perform well. This is more likely to happen for designs with higher degrees of orthogonality. The discussion is specialized to two-level factorial designs in Section 4.2.

Section 5 concludes this paper by summarizing the merits and the limitations of pHd.

Here are some notations to be used later: $\boldsymbol{x}_i$ = the ith design point; y_i = the response at $\boldsymbol{x}_i$; x_i = the ith coordinate of regressor $\boldsymbol{x}$; $\bar{y} = n^{-1}\sum_{i=1}^n y_i$; $\bar{\boldsymbol{x}} = n^{-1}\sum_{i=1}^n \boldsymbol{x}_i$; $\Sigma_x = n^{-1}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'$. We also abbreviate the averages over quantities associated with design points by "E"; for example, $E\boldsymbol{x} = \bar{\boldsymbol{x}} = (\bar{x}_1, \ldots, \bar{x}_p)'$, $E(x_i - \bar{x}_i)(x_j - \bar{x}_j)(x_k - \bar{x}_k)$ = a third design moment, etc.

2. Equivalence

The q-based pHd method begins with fitting y by quadratic polynomials $Q_p(\boldsymbol{x})$:

$$\min \sum_{i=1}^n (y_i - Q_p(\boldsymbol{x}_i))^2, \tag{2.1}$$

$$Q_p(\boldsymbol{x}) = \mu + \alpha'\boldsymbol{x} + (1/2)\boldsymbol{x}'B\boldsymbol{x}, \tag{2.2}$$

where μ is the intercept, α is a p-vector, and the p by p symmetric matrix B is the Hessian matrix of $Q_p(\boldsymbol{x})$. Let $\hat{B}$ be the least squares solution for B. Then the q-based pHd directions $\hat{v}_{qi}$ are the eigenvectors for the eigenvalue decomposition:

$$\hat{B}\Sigma_{\boldsymbol{x}}\hat{v}_{qi} = \hat{\lambda}_{qi}\hat{v}_{qi}. \tag{2.3}$$

The y-based method relies on another eigenvalue decomposition:

$$\hat{\Sigma}_y\hat{\beta}_{yi} = \hat{\lambda}_{yi}\Sigma_{\boldsymbol{x}}\hat{\beta}_{yi}, \quad i = 1, \ldots, p,$$

where

$$\hat{\Sigma}_y = n^{-1}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})'(y_i - \bar{y}).$$

The r-based method explores the residuals from the linear regression of y on $\boldsymbol{x}$:

$$r_{li} = y_i - \bar{y} - \hat{b}'(\boldsymbol{x}_i - \bar{\boldsymbol{x}}), \quad i = 1, \ldots, n,$$

where $\hat{b}$ is the least squares slope estimate. It proceeds in the same way as the y-based method:

$$\hat{\Sigma}_r = n^{-1}\sum_{i=1}^n (\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{x}_i - \bar{\boldsymbol{x}})' r_{li},$$

$$\hat{\Sigma}_r\hat{\beta}_{ri} = \hat{\lambda}_{ri}\Sigma_{\boldsymbol{x}}\hat{\beta}_{ri}, \quad i = 1, \ldots, p.$$

If the third design moments are zero, then $\hat{\Sigma}_y = \hat{\Sigma}_r$ and the y-based and r-based methods are identical.

For two-level full or fractional factorial designs, Σ_x is an identity matrix I. Moreover, if all two-factor interactions are estimable, then $\hat{B} = \hat{\Sigma}_y = \hat{\Sigma}_r =$ the matrix of two-factor interactions; the diagonal elements are identically zero and the ijth element is the estimated interaction between factors i and j. All three versions of pHd are seen to be identical.

Now consider the second-order rotatable designs (Box and Hunter (1957)). The design moments of $\boldsymbol{x}$ satisfy

$$Ex_i^2 = a, \quad Ex_i^4 = 3c, \quad Ex_i^2 x_j^2 = c, \quad \text{for} \quad i \neq j = 1, \ldots, p \tag{2.4}$$

and

all expectations of other powers and products, up to and including the fourth order, are zero.

Theorem 2.1. *For second-order rotatable designs, we have* $\hat{\Sigma}_y = \hat{\Sigma}_r$ *and*

$$\hat{B} = (2c)^{-1}\hat{\Sigma}_r + dI,$$

where I *stands for an identity matrix and the scalar* $d = \frac{a^2 - c}{2c(2c + p(c - a^2))} \cdot \operatorname{trace}(\hat{\Sigma}_r)$. *All three versions of pHd are equivalent in the sense that they find the same eigenvectors (the eigenvalues may be different, however).*

The proof of this Theorem is given in Appendix A. For designs with high rotatability indices (Draper and Pukelsheim (1990)), we can also expect all three versions to be approximately the same. For other designs, their relative merits still need to be explored further. Computationally, the r-based pHd is simpler than the q-based pHd. Moreover, the q-based method needs at least $1 + 2p + \frac{1}{2}p(p-1)$ runs in order to estimate all quadratic coefficients. When there are not enough runs in the data set, the r-based pHd becomes the natural version to use.

From now on, we shall concentrate on the q-based method, owing to its close relationship with classic parametric analysis. To study the Fisher consistency property, we may replace $\hat{B}$ by $E\hat{B}$ in the eigenvalue decomposition of (2.3):

$$(E\hat{B})\Sigma_x v_{qi} = \lambda_{qi} v_{qi}. \tag{2.5}$$

An eigenvector v_{qi} is consistent if it falls into the e.d.r. space.

3. Quadratic Surfaces

Suppose the true response function (1.2) is a quadratic polynomial (2.2). By the chain rule, the Hessian matrix B takes the form

$$B = (\beta_1, \ldots, \beta_k) M (\beta_1, \ldots, \beta_k)',$$

where M is the k by k Hessian matrix of $h(\cdot)$. Since the least squares estimate is unbiased $E\hat{B} = B$, from (2.5) we see that

$$v_{qi} = (\lambda_{qi})^{-1}(\beta_1, \ldots, \beta_k)[M(\beta_1, \ldots, \beta_k)'\Sigma_x v_{qi}]. \tag{3.1}$$

When $\lambda_{qi} \neq 0$, the right side of (3.1) expresses v_{qi} as a linear combination of β vectors. This establishes the Fisher consistency.

Theorem 3.1 *If the response function* (1.2) *is quadratic, then the q-based pHd method is consistent in finding e.d.r. directions.*

We now turn to the discussion of affine invariance. This is a property of the pHd procedure itself, independent of whatever the true regression surface is. The reason we bring it up in this section is to draw an interesting connection with the canonical quadratic surface analysis as explored in Box and Draper (1987).

First, if $\boldsymbol{x}$ is transformed to have identity covariance by $\tilde{\boldsymbol{x}} = \Sigma_x^{-1/2}\boldsymbol{x}$, then the dimension reduction assumption (1.2) becomes $h((\Sigma_x^{1/2}\beta_1)'\tilde{\boldsymbol{x}}, \ldots, (\Sigma_x^{1/2}\beta_k)'\tilde{\boldsymbol{x}}, \epsilon)$ and the e.d.r. directions become those in the space spanned by $\Sigma_x^{1/2}\beta_i$. Thus if the q-based pHd is applied to the transformed $\tilde{\boldsymbol{x}}$, it would be preferable to have the invariance property that the new pHd directions $\tilde{v}_i$ are related to the original pHd directions $\hat{v}_i$ via the same transformation $\Sigma_x^{1/2}$. This is indeed the case because the new pHd directions are just the eigenvectors for the matrix $\Sigma_x^{1/2}\hat{B}\Sigma_x^{1/2}$, the Hessian matrix of the new fitted quadratic function. Multiplying $\Sigma_x^{1/2}$ on both sides of (2.3) reveals that $\tilde{v}_i = \Sigma_x^{1/2}\hat{v}_i$, as desired.

Consider the new coordinates given by $\boldsymbol{u} = U'\boldsymbol{x} = (u_1, \ldots, u_p)'$ where U is the matrix of eigenvectors of B : $BU = UD$ where D is the diagonal matrix of eigenvalues λ_i. The quadratic model (2.2) now takes a canonical form: $g(\boldsymbol{x}) = \mu + (U'\alpha)'\boldsymbol{u} + \frac{1}{2}\sum_{i=1}^{k}\lambda_i u_i^2$. Thus, we see that projection of $\boldsymbol{x}$ along the first few q-based pHd directions amounts to finding the most important canonical variates u_i.

For locating the maximum, minimum, stationary ridge, etc. of quadratic response surfaces, the eigenvalue decomposition of $\hat{B}$ is suggested in Box and Draper (1987). This is the same procedure as the q-based pHd if the covariance Σ_x of $\boldsymbol{x}$ is proportional to the identity matrix. In fact, the nonzero eigenvectors of B span the same space as the one spanned by the nonzero eigenvectors of $B\Sigma_x$. But the adjustment by Σ_x is necessary in order to achieve affine invariance, a property desirable for the purpose of data visualization.

Another procedure for exploring quadratic surfaces based on maximum likelihood is also mentioned in Box and Draper (1987). A discussion on its connection with pHd is given in Appendix B.

3.1. A response surface example

Consider the data set from page 370 in Box and Draper (1987). It has five variables and 32 runs. The covariance matrix for the design points is not proportional to the identity. We applied the q-based pHd; see Table 3.1. The first two directions are used to construct a 3-D scatterplot for y against $\hat{\beta}'_{q1}\boldsymbol{x}$ and $\hat{\beta}'_{q2}\boldsymbol{x}$; see Figures 3.1(a)-(d). Spinning the plot helps us to learn the 3D geometric shape of the data points. We discover a cluster of points, as highlighted there, taking the shape of a ring. The entire data cloud appears like a cone with the ring at the bottom. We can also use the pHd directions to examine the residuals obtained by fitting y with linear functions of $\boldsymbol{x}$ and with quadratic models. The cone shape is again visible for the residuals of the linear fit, but no visible pattern in the residuals is present for the quadratic fit.

These figures suggest several directions for further study. For example, the trajectory of the ring can be tracked more closely by focusing on the highlighted points. We find that they come from the outer boundary of the design region when projected along the first two pHd directions; Figure 3.1(e). It also appears that the rest of the data points, particularly those near the top of the cone, have a larger response variation than the points on the ring. For example, the two points near the center in Figure 3.1(e), labeled as $*$ (case number 30) and $+$ (case number 27), have a difference in y about five times as large as the estimated standard deviation from the quadratic model. This indicates that more caution should be taken when exploring the center region for a maximum response (the goal of the experiment). The fitted quadratic model may not be adequate there and the response may be unstable too.

3.2. Application in 2^p factorial designs

Consider the data for efficiency study of chemical reactor from Box, Hunter and Hunter (1978, page 377). A 2^5 complete factorial design was conducted on five factors, feed rate (liters/min), catalyst (%), agitation rate (rpm), temperature (C), concentration (%). Three main effects and two two-factor interactions were found active from the normal probability plot (Figure 3.2). In addition to their relative sizes as shown in the plot, what else can one learn from these five numbers? We apply pHd to this data set; consult Table 3.2 for the output. The rotation plot for y against the first two pHd directions (Figures 3.3 (a), (b)) reveals two clusters of points, each with a linear pattern: the highlighted dots are with $x_4 = 1$, and the others are with $x_4 = -1$.

Table 3.1. First two q-based pHd directions and eigenvalues for the response surface example of Section 3.1.

$\hat{\beta}_{1q}$	$-.32$	$-.15$	.77	.59	.02
$\hat{\beta}_{2q}$	$-.61$	$-.47$	$-.01$	$-.47$	$-.17$
$\lvert\hat{\lambda}_{iq}\rvert$	8.31	5.99	3.28	1.32	.08

Figure 3.1 (a)-(e). 3-D plot of the response variable against the first two projections found by q-based pHd. The response surface example. By rotating about the y-axis, (a)-(d), the highlighted points are seen to form a ring, which is the basis of the cone-shaped data cloud. (e) is the scatterplot for the two projections.

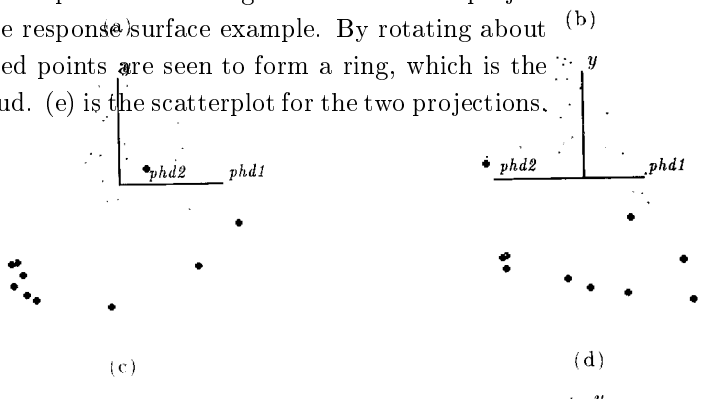


Table 3.2. First two pHd directions for the Chemical Reactor example.

$\hat{\beta}_{1y}$	.07	$-.54$	$-.08$	.67	.46
$\hat{\beta}_{2y}$	.01	.53	.10	.71	$-.41$
$\lvert\hat{\lambda}_{iy}\rvert$	9.22	8.21	1.49	.34	.15

Figure 3.2. Normal plot for the Chemical reactor example.

Figures 3.3 (a), (b). Two views of the 3-D plot of the response variable against the first two projections of x found by pHd. Chemical reactor example. Two clusters are revealed, each showing a linear trend, when we rotate the plot about the y-axis.

Geometrically, our pHd plot suggests that approximately, the five dimensional response surface $E(y|\boldsymbol{x})$ contains two hyper-planes, one at $x_4 = 1$ and another at $x_4 = -1$. But the big gap between the two indicates that no further information is available regarding how the response surface may behave at intermediate temperature levels, say $x_4 = 0$. This message highlights the place of the weakness in the prediction equation obtained from the standard ANOVA. For better interpolation, more levels have to be assigned to the temperature factor if more data are to be collected in the future.

The pHd plot also suggests a split of the data into two groups according to x_4 (temperature) as a follow-up analysis. We fit each group linearly in $\boldsymbol{x}$:

$$\begin{aligned} y &= 70.875 - 1.125x_1 + 16.375x_2 + .75x_3 - 8.625x_5, \quad \text{for } x_4 = 1, \\ y &= 60.125 - .25x_1 + 3.125x_2 - 1.375x_3 + 2.375x_5, \quad \text{for } x_4 = -1. \end{aligned}$$

Respectively, the residual standard deviations are 2.9 and 3.5; the standard errors for slope coefficients are .73, and .87; the R-squared values are .98 and .67. Scatterplots of y against the fitted values are given in Figures 3.4(a), (b). The tighter linear pattern in Figure 3.4(a) reflects the higher R-squared value for the group with $x_4 = 1$; but be aware of the different scales. We also apply the Yates algorithm to each group. The half-normal probability plots of estimated effects are given by Figures 3.5(a), (b). The main effects due to factor x_2 (catalyst), and factor x_5 (concentration) stand out clearly in Figure 3.5(a). This explains the four clusters found in Figure 3.4(a); they correspond to the level combinations of x_2 and x_5. The detection of significant effects from Figure 3.5(b) is less clear-cut. Putting Figure 3.2 side by side with Figures 3.5(a), (b), we may also appreciate the net gain of our analysis in the reduction of pattern-complexity.

Box, Hunter and Hunter also analyzed the above reactor data using only a half-fraction of the complete factorial, with the defining relation $\boldsymbol{I} = \mathbf{12345}$. We apply pHd to this subsample and find the results similar to the full sample case; details are omitted.

4. General Nonlinear Surfaces

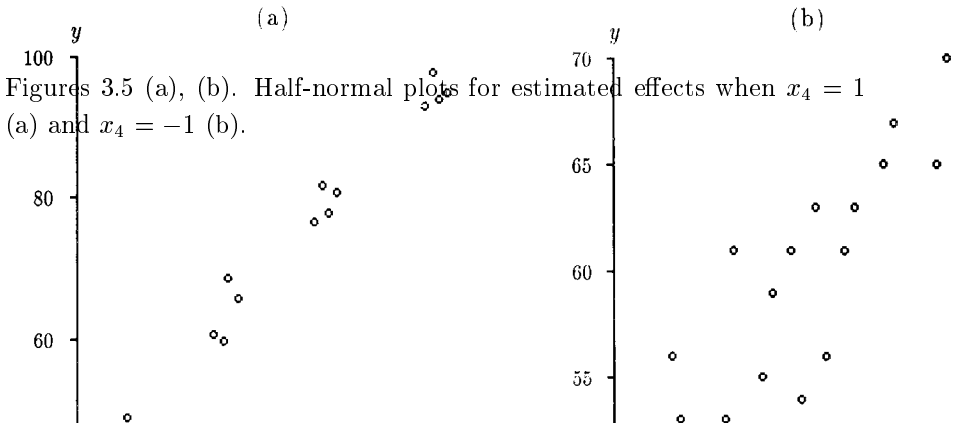
We study the performance of pHd when the function h in (1.2) deviates from quadratics. A general discussion is given in Section 4.1 first. Then we shall derive more results for two-level factorial designs in Section 4.2.

4.1. General designs

To proceed, let $\boldsymbol{u} = (\beta_1'\boldsymbol{x}, \ldots, \beta_k'\boldsymbol{x})$ and consider the quadratic fit of (1.2):

$$\min E(h(\boldsymbol{u}) - Q_p(\boldsymbol{x}))^2, \tag{4.1}$$

Figures 3.4 (a), (b). Plots of the response variable against the fitted values for the main effect model. (a) applies to the split sample with high temperature level, $x_4 = 1$; (b) applies to $x_4 = -1$.



Figures 3.5 (a), (b). Half-normal plots for estimated effects when $x_4 = 1$ (a) and $x_4 = -1$ (b).

where the minimum is taken over all quadratic functions $Q_p(\cdot)$ with p arguments, and the expectation notation denotes the average over all design points as before. This fitting can be equivalently carried out in two steps. The first step is to fit $h(\boldsymbol{u})$ with quadratic polynomials of $\boldsymbol{u}$:

$$\min E(h(\boldsymbol{u}) - Q_k(\boldsymbol{u}))^2. \tag{4.2}$$

Denote the solution by $Q_k^0(\boldsymbol{u})$ and let r_q be the residual:

$$h(\boldsymbol{u}) = Q_k^0(\boldsymbol{u}) + r_q. \tag{4.3}$$

The second step is to fit r_q further with quadratic polynomials of $\boldsymbol{x}$:

$$\min E(r_q - Q_p(\boldsymbol{x}))^2, \tag{4.4}$$

and denote the solution by $Q_p^*(\boldsymbol{x})$. Since the fitting (4.2) is more restrictive than (4.1), the sum $\bar{Q}_p(\boldsymbol{x}) = Q^0(\boldsymbol{u}) + Q_p^*(\boldsymbol{x})$ must be the best fit for (4.1).

We claim that if $Q_p^*(\boldsymbol{x})$ is linear, then the q-based pHd is consistent. To see this, let $\bar{B}$, B_0 be the Hessian matrices of $\bar{Q}_p(\boldsymbol{x})$, $Q^0(\beta_1'\boldsymbol{x}, \ldots, \beta_k'\boldsymbol{x})$, respectively. By the chain rule, B_0 is seen to take the form of $(\beta_1, \ldots, \beta_k)M(\beta_1, \ldots, \beta_k)'$ for some k by k matrix M. Since $E\hat{B} = \bar{B} = B_0$, (2.5) leads to an expression with the same form as (3.1), which implies the consistency result.

Theorem 4.1. *For a general regression function, if the residual r_q as defined in* (4.3) *is orthogonal to any quadratic function of $\boldsymbol{x}$ after eliminating its correlation with linear functions of $\boldsymbol{x}$, then the q-based pHd is consistent.*

Another way of explaining this result is to examine (4.4) more closely. For this purpose, let us introduce the variable $\boldsymbol{v}$ defined by

$$\boldsymbol{v} = (\gamma_1'\boldsymbol{x}, \ldots, \gamma_{p-k}'\boldsymbol{x})', \tag{4.5}$$

where the direction γ_j, $j = 1, \ldots, (p-k)$, are orthogonal to the e.d.r. space with respect to the design measure; i.e., $\gamma_j'\Sigma_x\beta_i = 0$. We refer to $\boldsymbol{v}$ as model-redundant complement of $\boldsymbol{u}$ because it is redundant for modeling the relationship between y and $\boldsymbol{x}$ once we find out the e.d.r. variates.

Now let us decompose any quadratic polynomial in $\boldsymbol{x}$ into four orthogonal parts: (i) a quadratic function of $\boldsymbol{u}$; (ii) a linear function of $\boldsymbol{v}$; (iii) a pure quadratic function of $\boldsymbol{v}$ after eliminating the linear part (ii); and (iv) the product of a linear function of $\boldsymbol{u}$ and a linear function of $\boldsymbol{v}$. The residual r_q by definition is already uncorrelated with part (i). The correlation between r_q and part (ii) is allowed, because it only leads to the linear term in Q_p^*. But r_q should be uncorrelated with parts (iii) and (iv) in order to have the consistency result of Theorem 4.1.

The source of possible bias for pHd is now clear. It is mainly due to the correlation between the higher order non-quadratic terms from the e.d.r. variates $\boldsymbol{u}$ and quadratic polynomials of its model-redundant complement $\boldsymbol{v}$. To give a more detailed account on the size of bias, further study on this confounding issue is called for. On the other hand, this correlation is likely to be smaller for designs with higher degrees of orthogonality.

Projection of data along the directions found by pHd can be very useful in visualizing the nonlinearity of the function $h(\boldsymbol{u})$. A simulation is reported here for illustration.

Example 4.1. We consider a 2^5 complete factorial design. Generate the data according to the model:

$$y = x_1(x_2 + .25x_3 + x_2x_3) + .5e, \quad e \sim N(0,1).$$

This model involves a three-factor interaction, but can be viewed as a two-component model with $\beta_1 = (1,0,0,0,0)'$, $\beta_2 = (0,1,.25,0,0)'$, for example, by putting $h(\boldsymbol{u}) = u_1(u_2 + 2u_2^2 - 2.125)$. The normal probability plot is given in Figure 4.1. The rotation plot found by pHd contains two clusters, each hanging around a different parabolic curve; one of them is highlighted in Figure 4.2(b). The variable x_1 is identified from the projection direction in Figure 4.2(a). The output of pHd is given in Table 4.1. Follow-up analysis can be done by further conditioning on x_1 and applying pHd again. Theorem 4.1 applies to this example because the residual r_q, equal to $2u_1u_2^2$, is uncorrelated with any two-factor interactions.

It is interesting to observe that from the normal probability plot, the roles of factor 1 and factor 2 appear to be exchangeable. This wrong conclusion is due to the failure of detecting the small interaction between factors 1 and 3. However, pHd analysis has bypassed this difficulty through the recognition of factor 1 as an e.d.r. variate.

4.2. Two-level factorial designs

For two-level factorial designs, due to the orthogonality among main effects and interactions, the result in Section 4.1 can be phrased in simpler terms. For example, consider a resolution **IV** design, where any two-factor interaction is unconfounded with any main effect. Theorem 4.1 implies that pHd is consistent if no two-factor interactions are present in the ANOVA decomposition for the reduced-quadratic-fit residual r_q. In general, this condition depends on the unknown function h. However, there are still important cases where pHd is always consistent, regardless of the form of h. This is what we want to characterize next for two-level factorial designs. We shall assume that the e.d.r. space has only one dimension:

$$E(y|\boldsymbol{x}) = g(\boldsymbol{x}) = h(\beta'\boldsymbol{x}). \tag{4.6}$$

Table 4.1 First two pHd directions for Example 4.1.

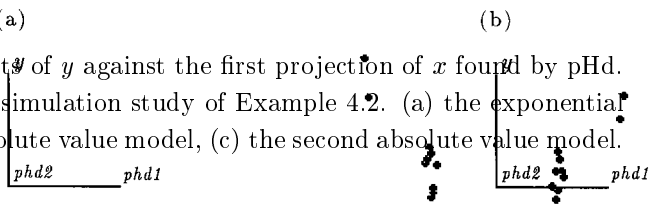
$\hat{\beta}_{1y}$	$-.69$	.66	.22	$-.02$	.01
$\hat{\beta}_{2y}$	$-.70$	$-.67$	$-.17$	.06	$-.06$
$\lvert\hat{\lambda}_{iy}\rvert$	1.2	1.19	.22	.12	.08

Figure 4.1. Normal plot for estimated effects. The simulation example in Section 4.1.

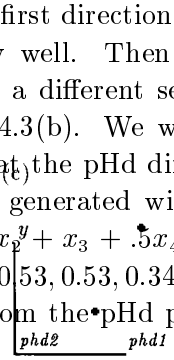
Figure 4.2 (a)-(b). Two views of the pHd plot for the simulation example in Section 4.1. Two clusters are found by rotating the 3-D plot about the y-axis.

4.2.a. Complete factorial designs

To motivate the discussion, we begin with a simulation example.



Figures 4.3 (a)-(c). Plots of y against the first projection of x found by pHd. The data are from the simulation study of Example 4.2. (a) the exponential model, (b) the first absolute value model, (c) the second absolute value model.



Example 4.2. Consider the 2^5 complete factorial design again. First, generate data from the model $y = .5\exp(x_1 + x_2 + x_3 + x_4) + \epsilon$ with the standard normal error. Figure 4.3(a) shows the first direction of pHd, which reveals the exponential shape of the function very well. Then change the model to $y = 2|x_1 + x_2 + x_3 + x_4| + \epsilon$ and generate a different set of data. Again the pHd is doing well in this case; see Figure 4.3(b). We will show later on that for these two cases, pHd is consistent so that the pHd direction will be exactly proportional to $(1, 1, 1, 1, 0)$ if the data are generated without ϵ. Finally, generate data from another model $y = |x_1 + x_2 + x_3 + .5x_4|$. Then we find that pHd has produced a biased direction, $(0.53, 0.53, 0.53, 0.34, 0)'$. But the shape of the absolute value function is still visible from the pHd plot; see Figure 4.3(c).

If desired, re-estimation of the regression coefficients can be pursued using the transformation suggested by the pHd plot. The random error in this last model is omitted in order to illustrate the bias effect on the pHd plot.

The following theorem explains what happens in the above example.

Theorem 4.2. *Consider the* 2^p *complete factorial design. Regardless of the form of* h *in* (4.6)*, for* β *in the set* $\mathcal{B}$ *defined by*

$$\mathcal{B} = \textit{the set of } p\textit{-dimensional vectors with coordinate values being either } 0, \textit{ or } 1, \textit{ or } -1 \tag{4.7}$$

the first eigenvector for the q*-based* pHd *method is consistent.*

This theorem follows from Theorem 4.3 of the next subsection. Note that since the diagonal elements of the matrix of two-term interactions $\hat{B}$ are identically zero, it is impossible for $E\hat{B}$ to have only one non-zero eigenvalue. However, the other eigenvectors with nonzero eigenvalues are redundant for exploring the shape of h.

Theorem 4.2 characterizes a special set of directions for which pHd works without any knowledge on h. But for other β vectors whose coordinates are not as simple as specified in Theorem 4.2, some bias may be present. An important reason why consistency cannot occur in more directions is due to a direction identifiability problem caused by the finiteness of the design support: the response function $g(\boldsymbol{x})$ on the design points can be represented by two rather different β directions using different h functions. For example, we can represent $y = \exp(x_1 + .5x_2)$ as $h(x_1 + .8x_2)$ (or many other β of the form $(b_1, b_2, 0, \ldots, 0)'$, $b_1 \neq b_2, -b_2$) for some function h. But for directions falling on the set $\mathcal{B}$, such direction identifiability problems disappear and the pHd method is shown to be consistent. As exhibited in Example 4.2, bias might not seriously affect the main features in the pHd plots if the true e.d.r. direction is still close to $\mathcal{B}$.

4.2.b. 2^{p-m} fractional factorial designs

Aliasing is an important issue in fractional factorial designs. It has some impact on consistency for pHd.

Theorem 4.3. *Consider a* 2^{p-m} *design with* $2^m - 1$ *defining contrasts:* $I = \cdots = \cdots = \cdots$*. Regardless of the form of* h *in* (4.6)*, for* β *in the set* $\mathcal{B}$ *defined by* (4.7)*, the first eigenvector for the* q*-based pHd method is consistent, if each of the defining contrasts contains at least three letters with the corresponding coordinates of* β *being zero.*

Factors appearing in $\beta'\boldsymbol{x}$ (namely those with the corresponding β coordinates being 1 or -1) are called the active factors. Others are called the inactive factors.

With this terminology, we can rephrase the condition on the defining contrast in this theorem:

The two-term interactions between the active factors and the inactive factors, and the two-term interactions within the inactive factors, cannot be aliased with any interactions of any order within the active factors.

As a simple example for applying this theorem, consider a 2^{6-1} design with defining relation $I = ABCDEF$. The consistency result holds if the number of active factors is no more than three.

Remark 4.1. In Theorem 4.3, suppose that each of the defining contrasts contains at least two letters with the corresponding coordinates of β being zero. Then we can only show that one of the eigenvectors found by pHd is consistent, but we cannot guarantee that it is the first one. For the aforementioned half replicate of the two-level factorial design involving six variables, this means that the number of active factors could be as many as four.

5. Conclusion

In spite of the growing interest in developing new methods for analyzing high dimensional data, traditional parametric methods still dominate the analysis in designed experiments. While we might apply these new tools mechanically to any data sets for exploratory analysis, how much can be learned from such exercises has never been carefully discussed. Consequently, the gap between the analysis of designed experiments and the analysis of general regression data is widening.

How to close this gap is one of the motivations behind this article. We focus on the performance of the pHd method under the regression dimension reduction framework of Li (1991). The q-based pHd method is shown to be consistent when the response surface is quadratic. For general surfaces, the Fisher consistency property requires that the nonquadratic part of the response function, which is orthogonal to any quadratic polynomials of the e.d.r. variates, has to be also orthogonal to quadratic terms from its model-redundant complement. This result suggests that smaller bias may be anticipated for designs with higher degrees of orthogonality.

We discuss the two-level factorial designs in greater detail. We characterize a special set of directions for which the pHd method is always consistent without any assumption about the response function.

We have illustrated how the plots of y against the pHd variates may help reveal the shape of the unknown regression function. This information can be used in building up a better model for further study. The pHd directions can also be used to examine residuals. This offers information different from the standard

residual plot of residuals against the fitted values.

In general, consistency of pHd in designed experiments does require more stringent assumptions than in general regression settings where the regressors tend to follow continuous distributions. Thus more caution needs to be taken in confirmatory analysis. On the other hand, our main interest is often to reveal nonlinear data structure by projecting along the e.d.r. directions. The pHd plots are still informative unless the bias is so large that the projection angle of pHd deviates substantially from the e.d.r. space. Roughly speaking, for angles less than 30 degrees, global features like trend or clustering can still be recognizable. This greatly releases the seemingly stringent conditions in the theorems for consistency.

Are there other dimension reduction methods that may work better than pHd in designed experiments? The answer is certainly worth pursuing. But as a reversal of our common perception, this study suggests that the more balanced data collected by designed experiments may turn out much harder to analyze than the unbalanced data from general regression studies. A fundamental obstacle is the issue of direction identifiability as raised in section 4.2.a, due to the lack of space coverage by design points along the e.d.r. directions. For most commonly used designs, the number of levels for each factor is usually small. This may prohibit us from getting more accurate information about the true shape of the response function, even if we are able to focus on the crucial factors that do affect the response variable.

The best way to alleviate this difficulty is to increase the number of levels for each factor in the design. This may be conducted in such a way that the total number of runs can still be small. Latin-hyper-cube designs (McKay, Conover and Beckman (1979)), or the space-filling designs (Wang and Fang (1981), Conway and Sloane (1987), Johnson, Moore and Ylvisaker (1990)) are often used in studying computer models (Sacks, Shiller and Welch (1989), and the references given there). We would like to add that the property of spherical symmetry is desirable, as shown in the general theory for pHd and SIR. Brillinger (1983) has discussed an interesting procedure to find designs that have nearly Gaussian distributions. Another easy-to-implement design is to apply a Latin-hyper-cube design on the radius r and angels θ_i in the polar coordinate transformation, $x_1 = r\cos\theta_1$, $x_2 = r\sin\theta_1\cos\theta_2, \ldots$.

For many experiments, however, level switching can be costly and the number of levels for each factor must be kept small. As we have seen, for such data, pHd can still squeeze out some useful information which might otherwise be hard to extract out by traditional ANOVA. The concerted use pHd with ANOVA is expected to lead to more fruitful analysis of designed experiments.

Acknowledgement

This work was sponsored in part by NSF grants. The second author was also funded by a Guggenheim Fellowship during the revision.

Appendix A: Proof of Theorem 2.1

To prove this result, first observe that the variable $x_j x_{j'}$ is uncorrelated with all other terms. This implies that $\hat{b}_{jj'}$, the jj' element of $\hat{B}$ can be obtained by fitting y against this variable directly, yielding

$$\hat{b}_{jj'} = \frac{1}{2c} n^{-1} \sum_{i=1}^{n} y_i x_{ij} x_{ij'} = \frac{1}{2c} \cdot jj'th \text{ element of } \hat{\Sigma}_y.$$

On the other hand, the diagonal elements of $\hat{B}$ can be obtained directly by fitting the regression of y against $x_1^2 - c, \ldots, x_p^2 - c$ without the intercept term. The information matrix for this regression takes the form of $2cI + (c - a^2)J$, where J is the matrix of *ones*. Hence the vector $\mathrm{Diag}(\hat{B})$ of the diagonal elements in $\hat{B}$ turns out to be

$$(2cI + (c - a^2)J)^{-1} \mathrm{Diag}(\hat{\Sigma}_y) = \left(\frac{1}{2c} I + \frac{a^2 - c}{2c(2c + p(c - a^2))} J \right) \mathrm{Diag}(\hat{\Sigma}_y).$$

The rest of the proof is straightforward.

Appendix B: Maximum Likelihood Estimate

The maximum likelihood estimate $\hat{B}_{mle}$ as suggested in Box and Draper (1987) is to fit the quadratic model (2.2) subject to the constraint that the rank of B is at most k. This becomes a nonlinear least squares problem and requires an adequate iterative algorithm, details of which are yet to be developed. If this is to be pursued further, the pHd directions can be used as the initial values to start the iteration. The following theorem shows a close relationship between the pHd and the m.l.e. method.

Theorem B.1. *If the design is second-order rotatable, then the eigenvectors (associated with nonzero eigenvalues) of the m.l.e. $\hat{B}_{mle}$ are also eigenvectors of pHd.*

The proof of this theorem is to be given later. Note that the eigenvalue sequences may not be in the same order. After a further elaboration (details omitted here), the set of nonzero eigenvectors for the m.l.e. $\hat{B}_{mle}$ can be chosen from the set of the eigenvectors for the r-based pHd, using the following rule:

$$\max_s \max \left\{ \frac{\sum_{i \in s} (\hat{\lambda}_{ri} - \bar{\lambda}_{rs})^2}{2c}, \frac{k \bar{\lambda}_{rs}^2}{k(c - a^2) + 2c} \right\},$$

where the maximum is over any subset $\boldsymbol{s}$ of k integers between 1 and p; $\bar{\lambda}_{rs} = k^{-1}\sum_{i\in s}\hat{\lambda}_{ri}$; $\hat{\lambda}_{ri}$ are eigenvalues for r-based pHd; a, c are moment constants defined in (2.4).

Another advantage for rotatable designs is the simplicity in finding the distribution of eigenvalues. If y is normally distributed with a common variance σ^2, then the distribution of $\hat{B}$ is also normal. The distribution of the sum of the $p-k$ smallest squared eigenvalues for the q-based pHd is approximately equal to the distribution of the random variable $\Lambda = \text{trace}\,(P_2\sum_x^{1/2}\hat{B}\sum_x^{1/2}P_2)^2$, where P_2 is the projection matrix on the null space of $\sum_x^{1/2}B\sum_x^{1/2}$. Arrange the elements of the upper triangle of $\hat{B}$ in a vector form, and let the covariance matrix be denoted by COV_1. Then the distribution of Λ can be represented as the distribution of the squared length of a normal random vector with mean zero and covariance matrix determined by COV_1 and P_2. However, for general designs, no closed form expression is available.

When the design is second-order rotatable, the distribution of Λ takes a simpler form. The off-diagonal elements are uncorrelated, with a common variance $(2cn)^{-1}\sigma^2$. They are also uncorrelated with the diagonal elements. The diagonal elements are correlated, however, with the covariance matrix

$$n^{-1}\sigma^2\left(\frac{1}{2c}I + \frac{a^2-c}{2c(2c+p(c-a^2))}J\right),$$

where J denotes the matrix of ones. From these, the distribution of Λ can be expressed as the sum of two independent rescaled chi-squares:

$$\frac{a^2\sigma^2}{2cn}\chi^2_{(p-k+2)(p-k-1)/2} + \frac{a^2\sigma^2}{n}\,\frac{2c+(c-a^2)k}{2c(2c+p(c-a^2))}\chi^2_1,$$

where the subscripts give the degrees of freedom of the chi-squared distributions. Instead of considering the sum of the smallest $p-k$ squared eigenvalues, $\sum_{i=k+1}^p\hat{\lambda}^2_{iq}$, we may break it into two parts: $(p-k)(\bar{\lambda}_{(k+1)q})^2$ and $\sum_{i=k+1}^p(\hat{\lambda}_{iq}-\bar{\lambda}_{(k+1)q})^2$, where $\bar{\lambda}_{(k+1)q} = (p-k)^{-1}\sum_{i=k+1}^p\hat{\lambda}_{iq}$. Then each part accounts for one rescaled chi-squared distribution in the above expression. We can easily readjust the weights between the two parts in order to get a single chi-squared distribution with $(p-k+1)(p-k)/2$ degrees of freedom.

Proof of Theorem B.1

Consider the linear residual r_{li}. The m.l.e. method can be reformulated as the maximization:

$$\max_{M_k}(\text{corr}\,(r_l, \boldsymbol{x}'M_k\boldsymbol{x}))^2, \tag{B.1}$$

where corr denotes the sample correlation coefficient, and the maximum is taken over any p by p matrix M_k that has rank at most k. Now (B.1) is equivalent to

the maximization of

$$\frac{(Er_l \boldsymbol{x}' M_k \boldsymbol{x})^2}{\mathrm{Var}(\boldsymbol{x}' M_k \boldsymbol{x})}. \tag{B.2}$$

Here we have used E and Var to denote the sample mean and sample variance. The numerator of (B.2) can be rewritten as $(\text{trace } \hat{\Sigma}_r M_k)^2$ and with the rotatability condition (2.4), the denominator is reduced to $2c \,\text{trace}\, M_k^2 + (c - a^2)(\text{trace } M_k)^2$. Now using Lagrange's multiplier, the maximization of (B.1) can be solved by

$$\max_{M_k} \text{trace}[\hat{\Sigma}_r M_k - d_1 M_k^2 - d_2 M_k]$$

for some constants d_1, d_2. By completing the square, this is equivalent to

$$\max_{M_k} \text{trace}(M_k - \frac{1}{2d_1}(\hat{\Sigma}_r - d_2 I))^2$$

whose solution is simply the matrix formed by the largest k components in the eigenvalue decomposition of the matrix $(2d_1)^{-1}(\hat{\Sigma}_r - d_2 I)$. The eigenvectors for the latter matrix are the same as the eigenvectors of $\hat{\Sigma}_r$, although the eigenvalues may be different. The proof of Theorem B.1 is complete.

Appendix C: Proof of Theorem 4.3

By permutation and sign change, without loss of generality, we can take $\beta = (1, \ldots, 1, 0, \ldots, 0)'$, where there are k ones. We also assume that the mean effect is zero, $E\bar{y} = 0$. By conditioning, we have $E\hat{\Sigma}_y = E(h(t)E(\boldsymbol{x}\boldsymbol{x}'|x_1 + \cdots + x_k = t))$. Consider the partition

$$E(\boldsymbol{x}\boldsymbol{x}'|x_1 + \cdots + x_k = t) = \begin{pmatrix} A_{11} & A_{12} \\ A_{12}' & A_{22} \end{pmatrix},$$

where A_{11} has the dimension k by k and A_{22} is $p - k$ by $p - k$. We can evaluate each matrix as follows.

Franklin and Bailey (1977, page 324) described an algorithm for using the defining relationship to choose a set of basic factors, those factors which form a complete factorial and whose interactions (called basic effects) can be used to define the other factors. It is clear from their algorithm that if each defining contrast contains at least one factor other than $x_1, \ldots, x_k$, then $x_1, \ldots, x_k$ can be chosen as part of a set of $p - m$ basic factors. The permutation invariance within $x_1, \ldots, x_k$ implies that A_{11} is permutation invariant. Thus the off-diagonal elements of A_{11} are identical. From this observation, we see that the first k by k submatrix of $E\hat{\Sigma}_y$ (corresponding to the A_{11} part) takes a very simple form: the diagonal elements are zero (because of the definition of the matrix $\hat{\Sigma}_y$) and the off-diagonal elements take the same value. This submatrix has $(1, \ldots, 1)$ as the

first eigenvector. To complete the proof, we need only to show that A_{12} and A_{22} are identically zero.

Consider an element in A_{12},

$$E(x_i x_j | x_1 + \cdots + x_k = t), \tag{C.1}$$

$i \leq k$, $j > k$. By Franklin and Bailey's algorithm, if each defining contrast contains at least two factors other than $x_1, \ldots, x_k$, then x_j can also be chosen as one of the basic factors, and therefore is independent of $x_1, \ldots, x_k$. It follows that (C.1) is zero. Finally, by a similar argument and the assumption that each defining contrast contains at least three factors other than $x_1, \ldots, x_k$, we can see that all elements of A_{22} are zero and Theorem 4.3 is proved.

If each defining contrast contains at least two factors other than $x_1, \ldots, x_k$, then A_{22} may not be zero. In this case, we can only conclude that $(1, \ldots, 1, 0, \ldots, 0)'$ is one of the eigenvectors of $E\hat{\Sigma}_y$.

References

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations (with discussion). J. Roy. Statist. Soc. Ser.B **26**, 211-252.

Box, G. E. P. and Draper, N. R. (1987). Empirical Model Building and Response Surfaces. John Wiley, New York.

Box, G. E. P. and Hunter, J. S. (1957). Multi-factor experimental designs for exploring response surfaces. Ann. Math. Statist. **28**, 195-241.

Box, G. E. P., Hunter, W. G. and Hunter, J. S. (1978). Statistics for Experimenters. John Wiley, New York.

Breiman, L. and Friedman, J. H. (1985). Estimating optimal transformations for multiple regression and correlation, J. Amer. Statist. Assoc. **80**, 580-598.

Breiman, L., Friedman, J., Olshen, R. and Stone, C. (1984). Classification and Regression Trees. Wadsworth.

Brillinger, D. R. (1983). A generalized linear model with "Gaussian" regressor variables. In A Festschrift for Erich L. Lehmann, 97-114, Wadsworth.

Chen, H. (1991). Estimation of a projection-pursuit type regression model. Ann. Statist. **19**, 142-157.

Chaudhuri, P., Huang, M.-C., Loh, W.-Y. and Yao, R. (1994). Piecewise-polynomial regression trees. Statistica Sinica **4**, 143-167.

Conway, J. and Sloane, N. (1987). Sphere-Packings, Lattices and Groups. Grundlehren der Wissenschaften Vol. 290, Springer-Verlag, New York.

Draper, N. R. and Pukelsheim, F. (1990). Another look at rotatability. Technometrics **32**, 195-202.

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

Fedorov, V. V. (1972). Theory of Optimal Experiments, translated and edited by W. J. Studden and E. M. Klimko. Academic Press, New York.

Franklin, M. F. and Bailey, R. A. (1977). Selection of defining contrasts and confounded effects in two-level experiments. Appl. Statist. **26**, 321-326.

Friedman, J. H. (1991). Multivariate adaptive regression splines (with discussion). Ann. Statist. **19**, 1-141.

Friedman, J. H. and Stuetzle, W. (1981). Projection pursuit regression. J. Amer. Statist. Assoc. **76**, 817-823.

Hall, P. (1989). On projection pursuit regression. Ann. Statist. **17**, 573-588.

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

Hsing, T. and Carroll, R. J. (1992). An asymptotic theory for sliced inverse regression. Ann. Statist. **20**, 1040-1061.

Huber, P. J. (1985). Projection pursuit, with discussion. Ann. Statist. **13**, 435-525.

Kiefer, J. (1959). Optimum experimental designs. J. Roy. Statist. Soc. Ser.B **21**, 272-319 (with discussion).

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). Ann. Statist. **2**, 849-879.

Johnson, M. E., Moore, L. M. and Ylvisaker, D. (1990). Minimax and maxmin distance designs. J. Statist. Plann. Inference **26**, 131-148.

Li, K. C. (1991). Sliced inverse regression for dimension reduction, with discussion. J. Amer. Statist. Assoc. **86**, 316-342.

Li, K. C. (1992). On principal Hessian directions for data visualization and dimension reduction: Another application of Stein's lemma. J. Ameri. Statist. Assoc. **87**, 1025-1039.

McKay, M. D., Conover, W. J. and Beckman, R. 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.

Sacks, J. Schiller, S. B. and Welch, W. J. (1989). Designs for computer experiments. Technometrics **31**, 41-47.

Steinberg, D. M. and Hunter, W. G. (1984). Experimental design: Review and comment. Technometrics **26**, 71-97.

Wang, Y. and Fang, K. T. (1981). A note on uniform distribution and experimental design. Kexue Tongbao **26**, 485-489.

Department of Statistics, University of California, Berkeley, CA 94720, U.S.A.

Department of Mathematics, University of California, Los Angeles, CA 90024, U.S.A.

(Received October 1993; accepted February 1995)