Journal of the Korean Statistical Society (2001), **30**: *4, pp 585–595*

Selection of Canonical Factors in Second Order Response Surface Models†

Sung H. Park and Seong K. Han[1]

Abstract

A second-order response surface model is often used to approximate the relationship between a response factor and a set of explanatory factors. In this article, we deal with canonical analysis in response surface models. For the interpretation of the geometry of second-order response surface model, standard errors and confidence intervals for the eigenvalues of the second-order coefficient matrix play an important role. If the confidence interval for some eigenvalue includes 0 or the estimate of some eigenvalue is very small (near to 0) with respect to other eigenvalues, then we are able to delete the corresponding canonical factor.

We propose a formulation of criterion which can be used to select canonical factors. This criterion is based on the IMSE(=Integrated Mean Squared Error). As a result of this method , we may approximately write the canonical factors as a set of some important explanatory factors.

Keywords: Canonical analysis, Canonical factor, Response surface model, Eigenvalue, Integrated mean squared error.

1. Introduction

Suppose that an experimenter is concerned with a system involving some response η which depends on several independent factors $\xi_1, \xi_2, \xi_3, \cdots, \xi_k$. In general, the functional relationship between the independent factors and the mean response can be written as

$$\eta = f(\xi_1, \xi_2, \xi_3, \cdots, \xi_k) \tag{1}$$

†This research was supported in part by KOSEF through Statistical Research Center for Complex Systems at Seoul National University.

[1]Department of Statistics, Seoul National University, San 56-1, Shinrim-dong, Kwanak-ku, Seoul, 151-742

where the explicit form of f is unknown or extremely complicated. Response surface methodology(RSM) often involves the approximation of f by a low order polynomial in some region of the independent factors. Usually, the original factors($\xi's$) are coded to the design factors ($x's$) in order to locate the origin to the center of region, with the latter normally being simple linear functions of the former. In recent years, interests in RSM have been increased and several books on this subject have been published by authors such as Myers(1976), Box and Draper(1987), Khuri and Cornell(1987), Myers and Montgomery(1995) and so on.

Let us consider the second order response surface model.

$$\eta(\underline{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i<j} \beta_{ij} x_i x_j \tag{2}$$

which may be written in matrix form as

$$\begin{aligned} \eta(\underline{x}) &= \beta_0 + \underline{x}'\underline{\beta} + \underline{x}'B\underline{x} \\ &= \underline{x}'_f \underline{\beta}_f, \end{aligned} \tag{3}$$

where,

$$\begin{aligned} \underline{x} &= (x_1, x_2, \cdots, x_k)', \\ \underline{\beta} &= (\beta_1, \beta_2, \cdots, \beta_k)', \\ \underline{x}_f &= (x_1, x_2, \cdots, x_k, x_1^2, x_2^2, \cdots, x_k^2, x_1x_2, \cdots, x_{k-1}x_k)', \\ \underline{\beta}_f &= (\beta_0, \beta_1, \beta_2, \cdots, \beta_k, \beta_1^2, \beta_2^2, \cdots, \beta_k^2, \beta_1\beta_2, \cdots, \beta_{k-1}\beta_k)' \end{aligned}$$

The coefficients in the second order models are estimated, by the method of

$$B = \begin{pmatrix} \beta_{11} & \frac{1}{2}\beta_{12} & \cdots & \frac{1}{2}\beta_{1k} \\ \frac{1}{2}\beta_{12} & \beta_{22} & \cdots & \frac{1}{2}\beta_{2k} \\ & & \ddots & \\ \frac{1}{2}\beta_{1k} & \frac{1}{2}\beta_{2k} & \cdots & \beta_{kk} \end{pmatrix}$$

least squares from N observations on the response factor,

$$y_u = \eta(\underline{x}_u) + \varepsilon_u \ , \qquad u = 1, 2, \cdots, N \tag{4}$$

where $\varepsilon_u's$ are assumed to be uncorrelated and have zero means and constant variance, σ^2 . The $\underline{\beta}_f$ is then estimated by the method of least squares as follows.

$$\hat{\underline{\beta}}_f = (X'X)^{-1}X'\underline{y} \tag{5}$$

where X is the $N \times m$ matrix whose ith row consists of $1 \times m$ vector

$$(\underline{x}'_f)_i = (1, x_{i1}, x_{i2}, \cdots, x_{ik}, x_{i1}^2, \cdots, x_{ik}^2, x_{i1}x_{i2}, \cdots, x_{ik-1}x_{ik})$$

and $\underline{y}$ is a vector of N observations.

It is important to determine the nature of the local response surface. Canonical analysis is a method of rewriting a fitted second order model in a form in which it can be more readily understood. This is achieved by a rotation of axes which removes all cross-product terms.

Consider a fitted second order model

$$\hat{y} = \hat{\beta}_0 + \underline{x}'\underline{\hat{\beta}} + \underline{x}'\hat{B}\underline{x}. \tag{6}$$

Let $\hat{\lambda_1}, \hat{\lambda_2}, \cdots, \hat{\lambda_k}$ be the eigenvalues of the symmetric matrix $\hat{B}$, and $\underline{m_1}, \underline{m_2} \cdots$, $\underline{m_k}$ the corresponding eigenvectors so that, by definition,

$$\hat{B}\underline{m_i} = \hat{\lambda}_i\underline{m_i} \qquad i = 1, 2, \cdots, k \tag{7}$$

If we standardize each eigenvector so that $\underline{m_i'}\ \underline{m_i'} = 1$ and if the $k \times k$ matrix M has $\underline{m_i}$ for its ith column, then M is an orthogonal matrix and the k equations (7) may be written simultaneously as

$$\hat{B}M = M\hat{\Lambda} \tag{8}$$

where $\hat{\Lambda}$ is a diagonal matrix. i.e.

$$\hat{\Lambda} = diag(\hat{\lambda_1}, \hat{\lambda_2}, \cdots, \hat{\lambda_k}).$$

Premultiplying by $M'(= M^-1)$ gives $M'\hat{B}M = \hat{\Lambda}$. By making use of the fact that $M'M = MM' = I_k$, we can write the equation (6) as

$$\hat{y} = \hat{\beta}_0 + (\underline{x}'M)(M'\underline{\hat{\beta}}) + (\underline{x}'M)(M'\hat{B}M)(M'\underline{x}) \tag{9}$$

By adopting the notation

$$\underline{w} = M'(\underline{x} - \underline{x_s}) \tag{10}$$

where $\underline{x_s}$ *is the stationary point*

we obtain the fitted second order model

$$\begin{aligned} \hat{y} &= \hat{\beta}_0 + \underline{x}'\underline{\hat{\beta}} + \underline{x}'\hat{B}\underline{x} \\ &= \hat{y}_s + \underline{w}\hat{\Lambda}\underline{w} \\ &= \hat{y}_s + \sum_{i=1}^{k} \hat{\lambda}_i w_i^2 \end{aligned} \tag{11}$$

where $\hat{y_s}$ is the fitted response at the stationary point. i.e

$$\hat{y_s} = \hat{\beta}_0 + \frac{1}{2}\underline{x_s}\,\hat{\underline{\beta}}.$$

We call this simplification the canonical form. The factors $w_1, w_2, \cdots, w_k$ are called canonical factors. The canonical form nicely describes the nature of the stationary point and the nature of the system around the stationary point.

2. Selection of canonical factors

2.1. Consequences of elimination of factors

Now a question is raised in using the canonical form. If the experimenter decides the number of canonical factors to be used, does he or she have to include every term in the canonical form to fit the response surface? It is possible to obtain a "better" response in the sense of precision in $y(\hat{\underline{x}})(= y(\hat{\underline{w}}))$, the least squares estimator of $\eta(\underline{x})$, by deleting some terms in the canonical form. We have k canonical factors $w_1, w_2, \cdots, w_k$ and a dependent factor y. A linear model that represents y in terms of k factors is

$$y_j = \lambda_0 + \lambda_1 w_{1j}^2 + \cdots \ \lambda_k w_{kj}^2 + \varepsilon_j \tag{12}$$

where $\varepsilon_j \sim N(0, \sigma^2)$. In matrix form, $\underline{y} = W\underline{\lambda} + \underline{\epsilon}$
where $\underline{\epsilon} \sim N(\underline{0}, \sigma^2 I_n)$. and

$$W = \begin{pmatrix} 1 & w_{11}^2 & w_{21}^2 & \cdots & w_{k1}^2 \\ 1 & w_{12}^2 & w_{22}^2 & \cdots & w_{k2}^2 \\ & & & \ddots & \\ 1 & w_{1n}^2 & w_{2n}^2 & \cdots & w_{kn}^2 \end{pmatrix}$$

Note that each w_i is a function of $x_1, x_2, \cdots, x_k$ through the equation (10), and the equation (12) is in fact the quadratic response surface model in $x_1, x_2, \cdots, x_k$. Also as shown in (11), the fitted value of y in (12) is the same as the original original quadratic response surface model in (6).

When k is large enough, the equation (12) is lengthy and it is practically not useful to use the full model. Instead of dealing with the full set of factors, we might want to delete a number of factors in order to use a small number of important factors for real application in practice. Then we want to construct an equation with a subset of factors. Let us denote the set of factors retained by $w_1, w_2, \cdots, w_p$ and those deleted by $w_{p+1}, w_{p+2}, \cdots, w_k$.

Instead of fitting (12), we fit the subset model

$$y_j = \lambda_0 + \lambda_1 w_{1j}^2 + \cdots \lambda_p w_{pj}^2 + \varepsilon_j^* \tag{13}$$

In matrix form,

$$\underline{y_p} = W_p \underline{\lambda_p} + \underline{\epsilon^*}$$

where

$$W_p = \begin{pmatrix} 1 & w_{11}^2 & w_{21}^2 & \cdots & w_{p1}^2 \\ 1 & w_{12}^2 & w_{22}^2 & \cdots & w_{p2}^2 \\ & & & \ddots & \\ 1 & w_{1n}^2 & w_{2n}^2 & \cdots & w_{pn}^2 \end{pmatrix}$$

We recall the following consequences of deletion of factors. The following lemma is explained in detail in Hocking(1976) and Park(1977, 1998).

Lamma

If the matrix $Var(\underline{\lambda_r^*}) - \underline{\lambda_r}\underline{\lambda_r'}$ *is positive semi-definite, then*

1. $MSE(\lambda_p^*) - MSE(\hat{\underline{\lambda_p}})$ *is positive semi-definite.*

2. $MSE(\hat{y}) \geq MSE(\hat{y}_p)$

Remark

i) $\hat{\lambda} = \begin{bmatrix} \underline{\lambda^*}_p \\ \underline{\lambda^*}_r \end{bmatrix} = (W'W)^{-1}W'Y$ (full model)

$\hat{\underline{\lambda}}_p = (W_p'W_p)^{-1}W_p'\underline{Y}$ (reduced model)

where $W = [W_p \ \vdots \ W_r]$

ii) $\hat{y} = \underline{\omega'}\ \hat{\underline{\lambda}}$ (full model)

$\hat{y}_p = \underline{\omega'}_p \hat{\underline{\lambda}}_p$ (reduced model)

where $\omega = \begin{bmatrix} \underline{\omega}_p \\ \underline{\omega}_r \end{bmatrix}$

where $\underline{\omega}_p = (1, \omega_1^2, \omega_2^2, \cdots, \omega_p^2)'$, $\underline{\omega}_r = (\omega_p^2 + 1, \omega_p^2 + 2, \cdots, \omega_k^2)'$

The motivation for elimination of canonical factors is provided by this lemma. That is, even if $\underline{\lambda_r} \neq \underline{0}$, $\underline{\lambda_p}$ or $\eta = E(y)$ may be estimated with smaller variance using the subset model.

However, the penalty is in bias. Lemma describes a condition under which the gain in precision is not offset by bias, i.e. , the gain is still favorable to the subset model.

Note that$Var(\hat{y}) = MSE(\hat{y})$ and $Var(\underline{\lambda^*}_p) = MSE(\underline{\lambda^*}_p)$ since $\hat{y}$ and $\underline{\lambda^*}_p$ are unbiased estimators.

2.2. Formulation of $\hat{Q}$ criterion

It is important to estimate response $\eta = E(y)$ with smallest possible MSE at any point of interest in the region R, in particular, at the stationary points(e.g. maximum or minimum response point), since the estimation of the stationary points with smaller MSE's is often pursued in response surface experiment.

It was observed that if the matrix $Var(\underline{\lambda_r^*}) - \underline{\lambda_r}\underline{\lambda_r'}$ is positive semi-definite, then it is possible to estimate parameters and responses with a smaller MSE by use of the subset model. If we write $Var(\underline{\lambda_r^*}) = \sigma^2 H$ in which is the appropriate submatrix of $(W'W)^{-1}$, the condition is that the matrix $\sigma^2 H - \underline{\lambda_r}\ \underline{\lambda_r}'$ is positive semi-definite.

The parameters, σ^2 and $\underline{\lambda_r}$ are are unknown, but we assume that they may be estimated from the current data using the full model. It may be argued that the condition is too restrictive for response estimation since it applies for any point over the region of interest. Therefore, as long as an average precision of estimation is acceptable, the requirement that $MSE(\hat{y}) - MSE(\hat{y_p})$, when integrated over the region of interest, be positive is a reasonable choice as discussed in Park(1977).

Consequently, the proposed criterion for selection of factors is "select the p factors" which maximize the quantity

$$Q = \int_R [Var(\hat{y}) - MSE(\hat{y_p})]dW(\underline{w}) \tag{14}$$

where $W(\underline{w})$ is a weighting function that can be treated as a probability distribution function on R. It can be readily shown that

$$\begin{aligned} Q &= \int_R [Var(\hat{y}) - Var(\hat{y_p})]dW(\underline{w}) - \int_R [Bias(\hat{y_p})]^2 dW(\underline{w}) \\ &= \sigma^2\Big[Tr[(W'W)^{-1}M] - Tr[(W_p'W_p)^{-1}M_{pp}]\Big] \\ &\quad - \underline{\lambda}'_r[A'M_{pp}A - 2AM_{pr} + M_{rr}]\underline{\lambda}_r \end{aligned}$$

where

$$\begin{aligned} M &= \int_R \underline{w}\ \underline{w}'dW(\underline{w})\ ,\ M_{ij} = \int_R \underline{w_i}\ \underline{w_j}'dW(\underline{w}), \\ A &= (W_p'W_p)^{-1}W_p'W_r \ \ \textit{and Tr denotes trace.} \end{aligned}$$

After replacement of the parameters σ^2 and $\underline{\lambda_r}$ by their estimates, the quantity to be maximized is

$$\begin{aligned} \hat{Q} &= \hat{\sigma^2}\Big[Tr[(W'W)^{-1}M] - Tr[(W_p'W_p)^{-1}M_{pp}]\Big] \\ &\quad - \hat{\underline{\lambda}'_r}[A'M_{pp}A - 2AM_{pr} + M_{rr}]\hat{\underline{\lambda}_r} \end{aligned} \tag{15}$$

Therefore, in essence, we are looking for a subset of polynomial terms whose gain in precision is not offset by the squared bias over the whole region of interest, R.

3. An example

3.1. Canonical analysis

In this example (Box and Draper, 1987), the chemical process under study had two stages, and five factors considered were stage one temperature (x_1), stage one reaction time (x_2), stage one concentration (x_3), stage two temperature (x_4), and stage two reaction time (x_5). The data set is listed in Table 1.

The objective of the experiment is to maximize the yield of chemical reaction. A preliminary application of the steepest ascent procedure had brought the experimenter close to a near-stationary region, and a second-order model was now to be fitted and examined.

At first, we deal with canonical analysis in second-order response surface models. From canonical analysis, if we find very small or non-significant eigenvalues, then we delete the corresponding canonical factors and refit the reduced model.

The second-order response surface model is fitted as follows.

$$\begin{aligned}\hat{y} \;=\;& 68.72 + 3.26x_1 + 1.58x_2 + 1.16x_3 + 3.47x_4 + 1.49x_5 - 1.61x_1^2 \\ & -1.35x_2^2 - 2.58x_3^2 - 2.34x_4^2 - 1.42x_5^2 - 1.90x_1x_2 + 2.10x_1x_3 \\ & -0.35x_1x_4 - 0.04x_1x_5 + 0.60x_2x_3 - 0.17x_2x_4 - 1.10x_2x_5 - 3.55x_3x_4 \\ & -0.75x_3x_5 + 0.40x_4x_5\end{aligned}$$

From the canonical analysis result as shown in Table 2, the second-order response surface model can be represented as canonical factors as follows.

$$\hat{y} = 72.51 - 4.46w_1^2 - 2.62w_2^2 - 1.78w_3^2 - 0.40w_4^2 - 0.04w_5^2$$

where,

$$\begin{aligned}w_1 \;=\;& -0.28(x_1 - 2.50) - 0.14(x_2 + 1.09) + 0.74(x_3 - 1.24) \\ & + 0.59(x_4 + 0.30) + 0.03(x_5 - 0.54) \\ w_2 \;=\;& 0.64(x_1 - 2.50) + 0.60(x_2 + 1.09) + 0.003(x_3 - 1.24) \\ & + 0.43(x_4 + 0.30) + 0.21(x_5 - 0.54) \qquad (16)\end{aligned}$$

Table1: Data of a chemical process experiment

Run number	x_1	x_2	x_3	x_4	x_5	Response, y
1	-1	-1	-1	-1	1	49.8
2	1	-1	-1	-1	-1	51.2
3	-1	1	-1	-1	-1	50.4
4	1	1	-1	-1	1	52.4
5	-1	-1	1	-1	-1	49.2
6	1	-1	1	-1	1	67.1
7	-1	1	1	-1	1	59.6
8	1	1	1	-1	-1	67.9
9	-1	-1	-1	1	-1	59.3
10	1	-1	-1	1	1	70.4
11	-1	1	-1	1	1	69.6
12	1	1	-1	1	-1	64.0
13	-1	-1	1	1	1	53.1
14	1	-1	1	1	-1	63.2
15	-1	1	1	1	-1	58.4
16	1	1	1	1	1	64.3
17	3	-1	-1	1	1	63.0
18	1	-3	-1	1	1	63.8
19	1	-1	-3	1	1	53.5
20	1	-1	-1	3	1	66.8
21	1	-1	-1	1	3	67.4
22	1.23	-0.56	-0.03	0.69	0.70	72.3
23	0.77	-0.82	1.48	1.88	0.77	57.1
24	1.69	-0.30	-1.55	-0.50	0.62	53.4
25	2.53	0.64	-0.10	1.51	1.12	62.3
26	-0.08	-1.75	0.04	-0.13	0.27	61.3
27	0.78	-0.06	0.47	-0.12	2.32	64.8
28	1.68	-1.06	-0.54	1.50	-0.93	63.4
29	2.08	-2.05	-0.32	1.00	1.63	72.5
30	0.38	0.93	0.25	0.38	-0.24	72.0
31	0.15	-0.38	-1.20	1.76	1.24	70.4
32	2.30	-0.74	1.13	-0.38	0.15	71.8

$$
\begin{aligned}
w_3 &= -0.25(x_1 - 2.50) + 0.25(x_2 + 1.09) + 0.23(x_3 - 1.24) \\
&\quad - 0.38(x_4 + 0.30) + 0.82(x_5 - 0.54) \\
w_4 &= 0.37(x_1 - 2.50) - 0.73(x_2 + 1.09) - 0.19(x_3 - 1.24) \\
&\quad + 0.22(x_4 + 0.30) + 0.50(x_5 - 0.54) \\
w_5 &= 0.56(x_1 - 2.50) - 0.16(x_2 + 1.09) + 0.60(x_3 - 1.24) \\
&\quad - 0.52(x_4 + 0.30) - 0.19(x_5 - 0.54)
\end{aligned}
$$

Table 2: Canonical analysis result

Canonical Analysis of Response Surface

Factor	Stationary point
X1	2.495548
X2	-1.093360
X3	1.243882
X4	-0.304204
X5	0.535206

Predicted value at stationary point 72.509519

Eigenvectors

Eigenvalues	X1	X2	X3	X4	X5
-0.040525	0.558012	-0.156681	0.601001	-0.518112	-0.185556
-0.397526	0.368751	-0.730077	-0.188156	0.221910	0.496351
-1.782351	-0.254125	0.254818	0.225386	-0.383098	0.820320
-2.624728	0.638495	0.598723	0.003226	0.433868	0.213550
-4.460949	-0.283532	-0.137977	0.743361	0.589330	0.026008

3.2. Factor selection and its results

For the case of including the canonical factors $w_1^2, w_2^2, w_3^2, w_4^2 (p = 4)$, we obtain the reduced second-order response surface model in canonical factors as follows.

$$\begin{aligned} \hat{y} &= 72.21 - 4.48w_1^2 - 2.61w_2^2 - 1.79w_3^2 - 0.39w_4^2 \\ \hat{Q} &= 7.34 \end{aligned}$$

For the case of including the canonical factors $w_1^2, w_2^2, w_3^2 (p = 3)$ we can similarly obtain the reduced second-order response surface model in canonical factors as follows.

$$\begin{aligned} \hat{y} &= 72.19 - 4.43w_1^2 - 2.65w_2^2 - 1.72w_3^2 \\ \hat{Q} &= -123.27 \end{aligned}$$

Similarly we can obtain the results for the cases of $p = 2$ and $p = 1$, and they are summarized in Table 3.

Table 3: Criteria of factor selection

Number of selected factor	criterion	
	$\hat{Q}$	R_p^2
p=5($w_1^2 \sim w_5^2$)	0	0.986
p=4($w_1^2 \sim w_4^2$)	7.34	0.985
p=3($w_1^2 \sim w_3^2$)	-123.37	0.958
p=2($w_1^2 \sim w_2^2$)	-1238.93	0.884
p=1(w_1^2)	-4198.13	0.463

Now we want to choose p factors to maximize $\hat{Q}$ in (15). Table 3 shows that only $p = 4$ gives a positive value for Q, which means that $p = 4$ is the most favorable number of selected factors for the model. Consequently, we can delete the fifth canonical factor, which means that, we are able to select canonical factors w_1, w_2, w_3, w_4.

From the equation (16), we might approximately write the canonical factors as

$$\begin{aligned} w_1 &= 0.74(x_3 - 1.24) + 0.59(x_4 + 0.30) \\ w_2 &= 0.64(x_1 - 2.50) + 0.60(x_2 + 1.09) \\ w_3 &= 0.82(x_5 - 0.54) \\ w_4 &= -0.73(x_2 + 1.09) \\ w_5 &= 0.56(x_1 - 2.50) + 0.60(x_3 - 1.24) - 0.52(x_4 + 0.30) \end{aligned}$$

Note that the factor w_1 influences $\hat{y}$ most, which means that in order to maximize $\hat{y}$, we need to increase factors (x_2, x_3), in the direction of $0.74(x_3 - 1.24) + 0.59(x_4 + 0.30)$. The factor w_2 followed by w_1 influences $\hat{y}$ next, which means that in order to maximize $\hat{y}$, we need to increase the factors (x_1, x_2), in the direction of $0.64(x_1 - 2.50) + 0.60(x_2 + 1.09)$. We can similarly interpret the results for the remaining w_3, w_4, and w_5. These results can be usefully applied in industry.

ACKNOWLEDGEMENTS

The authors are really thankful for the referees' useful and thoughtful comments which improved the quality of this paper.

REFERENCES

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

Hocking, R. R. (1976), "The analysis and selection of variables in linear regression", Biometrics, 32, pp. 1-49.

Khuri, A. I., and Cornell, J. A. (1987), Response Surface: Designs and Analyses, New York: Marcel Dekker.

Myers, R. H. (1976), Response Surface Methodology, Boston: Allyn and Bacon.

Myers, R. H. and Montgomery, D. C. (1995), Response Surface Methodology: Process and Product Optimization Using Designed Experiments: Wiley-Interscience, New York.

Park, S. H. (1977), "Selecting of polynomial terms for response surface experiment," Biometrics, 33, 225-229.

Park, S. H. (1998), Regression Analysis(in Korean): Third edition, Minyoung-sa, Seoul.