Stochastics and Statistics

A confidence region for the ridge path in multiple response surface optimization

Liangxing Shi [a,*], Dennis K. J. Lin [b], John J. Peterson [c]

[a] *College of Management and Economics, Tianjin University, 92 Weijin Rd., Nankai District, Tianjin 300072, PR China*
[b] *Department of Statistics, The Pennsylvania State University, 317 Thomas Building, University Park, PA 16802, USA*
[c] *Quantitative Sciences Department, GlaxoSmithKline Pharmaceuticals 1250 So. Collegeville Road (UP-4315), Collegeville, PA 19426, USA*

ARTICLE INFO

ABSTRACT

Ridge analysis allows the analyst to explore the optimal operating conditions of the experimental factors. A confidence region is desirable for the estimated ridge path. Most literature concentrates on the univariate response situation. Little is known for the confidence region of the ridge path for the multivariate response; only a large-sample confidence interval for the ridge path is available. The simultaneous coverage rate for the existing interval is typically too conservative in practice, especially for small sample sizes. In this paper, the ridge path (via desirability function) is estimated based on the seemingly unrelated regression (SUR) model as well as standard multivariate regression (SMR) model, and a conservative confidence interval suitable for small sample sizes is proposed. It is shown that the proposed method outperforms the existing methods. Real-life examples and simulative study are given for illustration.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Ridge analysis, first introduced by Hoerl (1959), is used to explore the optimal setting of the experimental variables. Consider the response surface model $y = f(x, \theta) + \varepsilon$, where y is the response variable, x is the vector of input variables, θ is the vector of model parameters, and ε is the error. Without loss of generality, suppose that maximization of the response is desirable. Let $g(\theta, r) = \max_{x'x=r^2} f(x, \theta)$ represent the constrained optimal mean response value, where r is the distance from the center of the experiment region. A ridge path is the locus of the $g(\theta, r)$ on different radii (r) of the surface. The typical output of a ridge analysis is presented as two two-dimensional plots: a plot of $g(\theta, r)$ vs. r and an overlay plot of x_{ir} vs. r ($i = 1, \ldots, l$), where l is the number of input variables. These are typically used to locate the optimal operating conditions.

The true value of the model parameter θ is unknown in practice, and the estimated value $\hat{\theta}$ is used. Thus, the plot of $g(\hat{\theta}, r)$ vs. r is only a statistical point estimate of the true ridge path. To construct the confidence region of the ridge path is obviously important since it can measure the accuracy of the estimation. Carter, Chinchilli, Myers, and Campbell (1986) proposed the use of simultaneous confidence bounds for a ridge path. Peterson

(1993) gave a general approach to ridge analysis with confidence intervals. Both of them are limited to univariate responses. When multiple responses are involved in experiments, the common approach converts the multiple responses into a univariate index. Such a conversion is intentionally biased, however. Thus, it is desirable to investigate the standard error of the fitted parameters and their effects on optimization indices (See Hunter, 1999). Furthermore, the ridge path is well defined in univariate cases, but as by Lin (1999) it is hard to extend those ideas from univariate to multivariate cases straightforwardly. How to appropriately apply ridge analysis to multivariate cases deserves further study.

Ding, Lin, and Peterson (2005) applied the standard multivariate regression (SMR) model to fit the response surface model (RSM) and developed a large-sample simultaneous confidence interval for a multi-response ridge path based on the desirability function. However, their method may not be appropriate when the sample size is small. When the SMR model is used, it likely leads to overfitting for some responses because the design matrix is identical for each response in SMR but the significant terms for each response may be different. Here, a new approach to construct confidence intervals with multiple response surfaces is proposed. The seemingly unrelated regressions (SUR) model (Zellner, 1962) is employed in our method. The SUR model could fit the model with different experimental factors for each response, meanwhile it estimates the correlations among all responses. Compared with the existing methods, the proposed method using the SUR model, results in a smoother and more reliable confidence interval when the

* Corresponding author. Tel.: +86 22 87401783.
  E-mail addresses: shi@tju.edu.cn (L. Shi), DKL5@psu.edu (D.K.J. Lin), john.peterson@gsk.com (J.J. Peterson).

sample size is small. This will help experimenters to locate the optimal setting in an efficient manner.

The paper is organized as follows. In Section 2, a brief review of ridge analysis is presented. Estimating a ridge path based on SUR model and a conservative confidence approach are then proposed in Section 3. Section 3 also provides the algorithm of the proposed method and its general properties. In Section 4, the tire tread example (with small sample size) is used for illustration with comparisons to previous works, as well as a large sample case. A further simulation study is also provided. The conclusion is given in Section 5.

2. Statistical inference for the ridge path

2.1. Confidence intervals for a single response ridge path

Peterson (1993) took $f(x, \theta)$ as $z(x)'\theta$, where $z(x)$ is a $p \times 1$ vector-valued function of a $k \times 1$ vector of factors. Thus, the response surface model $y = f(x, \theta) + \varepsilon$ can be represented as $y = z(x)'\theta + \varepsilon$. Then, the ridge path $g(\theta, r)$ becomes $g(\theta, r) = \max_{x'x=r^2} z(x)'\theta$. Carter et al. (1986) proposed the simultaneous confidence bounds of the optimal responses for various r, and the form of the confidence bounds can be written as

$$\left[\min_{\theta \in C} \{ \max_{x'x=r^2} z(x)'\theta \}, \max_{\theta \in C} \{ \max_{x'x=r^2} z(x)'\theta \} \right], \quad (1)$$

where C is a $100(1 - \alpha)\%$ confidence region for θ. The confidence region C is defined as $C = \{ \theta : (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}) \leq c_\alpha^2 \}$, where $\hat{\theta}$ is an estimate of θ, V is an estimate of $var(\hat{\theta})$, and $c_\alpha^2 = pF(1 - \alpha, p, n - p)$, with n being the sample size and $F(1 - \alpha, p, n - p)$ is the $100(1 - \alpha)$th percentile of the F distribution with p and $(n - p)$ degrees of freedom. Peterson (1993) proposed an alternative confidence bound as $\left[\max_{x'x=r^2} \{ \min_{\theta \in C} z(x)'\theta \}, \max_{x'x=r^2} \{ \max_{\theta \in C} z(x)'\theta \} \right]$. Because $z(x)'\theta$ is linear in θ, the confidence interval can be written as $\max_{x'x=r^2} \{ z(x)'\hat{\theta} \pm c_\alpha (z(x)'Vz(x))^{\frac{1}{2}} \}$. For a rotatable design, this can be further simplified as $\max_{x'x=r^2} \{ z(x)'\hat{\theta} \pm c_\alpha \hat{\sigma} v(r)^{\frac{1}{2}} \}$, where $\hat{\sigma}^2$ is the sample-error mean square and $v(r) = z(x)'(Z'Z)^{-1}z(x)$, with $x'x = r^2$. Z is the regression model matrix.

Note that the Carter et al. (1986) approach requires a nonlinear optimization solver for $\max z(x)'\theta$ subject to $x'x = r^2$ to obtain x^*, such that it maximize $z(x^*)'\hat{\theta}$. One then applies another nonlinear solver for min or $\max \{ z(x^*)'\theta \}$ subject to $\theta \in C$. Peterson (1993) argued that solving $\max_{x'x=r^2} \{ z(x)'\hat{\theta} \pm C_\alpha (z(x)'Vz(x))^{\frac{1}{2}} \}$ in reality is much more manageable than solving Eq. (1). However, his approach utilizes the property that $z(x)'\theta$ is linear in θ which is an unrealistic assumption for multivariate response problems in many situations. It is usually highly nonlinear in both θ and x in the desirability function.

2.2. Confidence intervals for a multi-response ridge path

A general multi-response problem can be written as

$$y_i = f(x, \theta_i) + \varepsilon_i \quad (2)$$

for $i = 1, 2, \ldots, p$, where y_i is the response vector, $x = (x_1, x_2, \ldots, x_k)$ is input variable vector, θ_i is the vector of model parameters, and the ε_i is random error term, typically assumed to be $N(0, \sigma^2)$. The model function $f(x, \theta_i)$ represents the functional relation between the ith response and the input variables.

The parameters θ are usually estimated by fitting multivariate linear regression models in the matrix form (see, e.g., Arnold, 1981, p. 349),

$$Y \sim N_{n,p}(X\Theta, \Sigma), \quad (3)$$

where n is the number of independent experiment runs, and m is the number of response variables in each run, with a fixed covariance matrix Σ. The matrices Y, X and Θ are the response matrix $(n \times p)$, design matrix $(n \times m)$ and parameter matrix $(m \times p)$, respectively.

The optimization for a multi-response issue is to find a set of operating conditions x^* that optimizes all responses in the given ranges. Many methods have been proposed for optimization of multiple responses (Bera & Mukherjee, 2015; Kim & Lin, 2006). See, for examples, the generalized distance measure (Khuri & Conlon, 1981), and the squared error loss approach (Ames, Mattucci, Macdonald, Szonyi, & Hawkins, 1997; Pignatiello, 1993; Vining, 1998). The most popular approach is probably the desirability function. The desirability function (Derringer & Suich, 1980; Harrington, 1965; He, Zhu, & Park, 2012; Jeong & Kim, 2009) transforms an estimated response y_i to a scale free value $d_i(\cdot) \in [0, 1]$, called a desirability. The overall desirability function is then defined as the geometric mean

$$D(x, \theta) = \left(\prod_{i=1}^{m} d_i(\hat{y}_i) \right)^{1/m}. \quad (4)$$

Kim and Lin (2000) used an exponential form of the desirability function and illustrated its application to the simultaneous optimization of mechanical properties of steel. This approach also considered the predictive of every individual response surface model. In general, any reasonable desirability function can be used here, as long as it is continuous and differentiable. Following Ding et al. (2005), we adapt the desirability functions of Gibb, Carter, and Myers (2001),

$$d_i(\hat{y}_i) = \begin{cases} \left[1 + e^{-\frac{E(y_i)-a_i}{b_i}} \right]^{-1} & \text{if } y_i \text{ is LTB;} \\ e^{-0.5 \left(\frac{E(y_i)-a_i}{b_i} \right)^2} & \text{if } y_i \text{ is NTB;} \\ \left[1 + e^{\frac{E(y_i)-a_i}{b_i}} \right]^{-1} & \text{if } y_i \text{ is STB.} \end{cases} \quad (5)$$

For the nominal-the-best (NTB) case, a_i is the target value of response, and $b_i = \frac{\delta_i}{\sqrt{-2ln(\gamma_i)}}$ is to control the spread of the function, where $\gamma_i \in (0, 1)$. For the larger-the-better (LTB) or the smaller-the-better (STB) case, $a_i = \frac{y_i^{max} - y_i^{min}}{2}$ and $b_i = \frac{y_i^{max} - y_i^{min}}{2ln(\frac{1-\gamma_i}{\gamma_i})}$, where $y_i^{max} > y_i^{min}$, and $\gamma_i \in (0, 1)$. The values of δ_i and γ_i can be determined via the guideline given by Gibb et al. (2001).

Ding et al. (2005) developed a large-sample simultaneous confidence interval for a multi-response ridge path based on the desirability function. They defined the multi-response ridge path as the plot of $g(\theta, r)$ vs. radius r, where

$$g(\theta, r) = \max_{x'x=r^2} D(\theta, x), \quad (6)$$

Assuming that $x_0 = x_0(\theta, r) = arg\max_{x'x=r^2} D(x, \theta)$ is unique for each r, Ding et al. (2005) construct $100(1 - \alpha)\%$ asymptotic simultaneous confidence intervals for $g(\theta, r)$ which have the form of

$$\left[\frac{e^L}{1 + e^L}, \frac{e^U}{1 + e^U} \right], \quad (7)$$

where $[L, U] = logit(g(\hat{\theta}, r)) \pm z_{\alpha/2q}\hat{c}(r)$, where $z_{\alpha/2q}$ is the upper $\alpha/2q$ critical value of standard normal distribution, q is the number of radii. $\hat{c}(r)$ is the estimated standard error of $logit(g(\hat{\theta}, r))$, and $\hat{c}(r)^2 = \frac{D_\theta(\hat{x}_0, \hat{\theta})'(\hat{\Sigma} \otimes (X'X)^{-1})D_\theta(\hat{x}_0, \hat{\theta})}{(D(\hat{x}_0, \hat{\theta})(1 - D(\hat{x}_0, \hat{\theta}))^2)}$, in which $\hat{x}_0 = x_0(\hat{\theta}, r)$. Logistic regression is popularly used in many areas, especially in bioscience (see Hosmer and Lemeshow (2005), for example). This results a Bonferroni's z type confidence band, since the critical value is based on Bonferroni's inequality.

One immediate question, however, is how to construct a confidence band of a path ridge of multiple responses for the small sample case. Here, we propose a new method to construct a confidence region for ridge path of multiple responses, and it is shown to be reliable when the sample size is small.

3. A conservative confidence region for multi-response ridge path

3.1. Seemingly unrelated regression (SUR) model

Since SMR model is likely to overfit for one or more of the responses, a SUR model proposed by Zellner (1962) is believed to be much more appropriate here. Suppose X_j (for $j = 1, \ldots, p$) is an $n \times q_j$ design matrix which includes all the selected significant terms. These significant terms can be identified from fitting the univariate model for the jth response. If we also assume that the X_j have full column rank $q_j (j = 1, \ldots, p)$, a SUR model can be written as

$$y = G\theta + \varepsilon \qquad (8)$$

where $y = vec(Y)$, $cov(y) = \Sigma \otimes I_n$, $\theta = (\theta_1, \ldots, \theta_m)'$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_p)'$, G is the block-diagonal design matrix which can be written as

$$G_{np \times q} = \begin{bmatrix} X_1 & 0 & \ldots & 0 \\ 0 & X_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & X_p \end{bmatrix},$$

and $q = q_1 + q_2 + \cdots + q_p$. The estimation of θ in Eq. (8) can be obtained by Maximum Likelihood (ML) estimate as,

$$\hat{\theta} = (G'(\Omega)^{-1}G)^{-1}G'\Omega^{-1}y, \qquad (9)$$

where $\Omega = \Sigma \otimes I_n$. However, the true covariance matrix Σ is generally unknown. The solution for this problem is a feasible generalized least squares (FGLS) estimation. Zellner and Huang (1962) estimated Σ via $\hat{\Sigma} = S = [s_{ij}]$, where $s_{ij} = \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_j}{n - q_i - q_j + q_{ij}}$ and $q_{ij} = trace[(X_i'X_i)^{-1}X_i'X_j(X_j'X_j)^{-1}X_j'X_i]$. Substituting $\hat{\Sigma}$ for Σ in Eq. (9), the FGLS estimator of θ is $\hat{\hat{\theta}} = (G'(\hat{\Sigma} \otimes I_n)^{-1}G)^{-1}G'(\hat{\Sigma} \otimes I_n)^{-1}y$. Therefore, a two-stage process is applied to estimate θ: Σ is firstly estimated with $\hat{\Sigma}$ such that $\hat{\Sigma} \xrightarrow{p} \Sigma$, and then $\hat{\hat{\theta}}$ is obtained. The ML estimate is obtained by iterating the process. The $\Omega = \Sigma \otimes I_n$ can be re-estimated with $\hat{\hat{\theta}}$ as $\hat{\Omega}_2 = (y - G\hat{\hat{\theta}}_1)(y - G\hat{\hat{\theta}}_1)'$, where $\hat{\hat{\theta}}_1$ is the two stage estimate. Then a revised estimate of θ is $\hat{\hat{\theta}}_2 = (G'(\hat{\Omega}_2)^{-1}G)^{-1}G'\hat{\Omega}_2^{-1}y$. Continuing in this manner, the ith iteration is $\hat{\Omega}_i = (y - G\hat{\hat{\theta}}_{i-1})(y - G\hat{\hat{\theta}}_{i-1})'$, and $\hat{\hat{\theta}}_i = (G'(\hat{\Omega}_{i-1})^{-1}G)^{-1}G'\hat{\Omega}_{i-1}^{-1}y$. The process stops when $\hat{\hat{\theta}}_i$ converges such that $\| \hat{\hat{\theta}}_i - \hat{\hat{\theta}}_{i-1} \|^2 < \delta$ for some $\delta > 0$. Such a δ is typically pre-specified.

3.2. The confidence intervals of the multi-response ridge path

We define the ridge path for the multiple response model based on the desirability function. A multi-response ridge path $g(\theta, r)$ can be defined on desirability function D as in Eq. (6). It should be noticed that the desirability function $D(x, \theta)$ is usually highly nonlinear in both θ and x: thus Peterson's (1993) approach may not be appropriate. Following Carter et al. (1986), a conservative $100(1 - \alpha)\%$ confidence band for $\max_{x'x=r^2} D(x, \theta)$ can be defined as

$$\left[\min_{\theta \in C} \{ \max_{x'x=r^2} D(x, \theta) \}, \max_{\theta \in C} \{ \max_{x'x=r^2} D(x, \theta) \} \right], \qquad (10)$$

where C is a $100(1 - \alpha)\%$ confidence region for θ.

The important next step is to find a closed form of C. From Timm (2002), a test of the hypothesis $H_0 : \theta = \theta^*$ for SUR model (Eq. 8) can be conducted using the approximate F-test statistic of $F^* = (W/v_h)/MSe$, where $MSe = \frac{(y - G\hat{\hat{\theta}})'\hat{\Omega}^{-1}(y - G\hat{\hat{\theta}})}{v_e}$, and $W = (\hat{\hat{\theta}} - \theta^*)'[G'(\hat{\Sigma} \otimes I_n)^{-1}G](\hat{\hat{\theta}} - \theta^*)$. Henningsen and Hamann (2007) recommend using the degrees of freedom $v_e = np - q$ and $v_h = q$. Here, $v_h = 2$ is suggested instead of q. The rationale for using $v_h = 2$ in ridge analysis was given by Peterson (1993) as follows. Define Z as the factor variable space. As the radius value r goes from 0, the optimal vector x_0 produces a ridge trace in Z extending from the origin. Suppose this ridge trace is a ray with only moderate curvature, a $100(1 - \alpha)\%$ confidence band about a mean-response line corresponding to a straight line in Z uses a critical value of $2F(2, v_e)$. This is because this one-dimensional straight-line ray is associated with a two-dimensional linear subspace of the Z-space (Miller, 1981). Simulation studies by Peterson (1993) and Peterson and Kuhn (2005) found that the $2F(2, v_e)$ critical value worked well for all their ridge analysis examples. The close form of C can be created by F^* as $C = \{\theta^* : F^* \le F_{1-\alpha}(v_h, v_e)\}$. Namely,

$$C = \left\{ \theta^* : \frac{(\hat{\hat{\theta}} - \theta^*)'[G'(\hat{\Sigma} \otimes I_n)^{-1}G](\hat{\hat{\theta}} - \theta^*)}{MSe} \le v_h F_{1-\alpha}(v_h, v_e) \right\}. \qquad (11)$$

3.3. Algorithm for confidence interval

The ridge path and its confidence bands can be constructed step-by-step. Basically it consists of three steps: (1) Model fitting; (2) Ridge path estimation; and (3) Confidence interval estimation.

Step 1. Fitting the model via SUR. There may exist some insignificant terms. The R_i^2 and R_*^2 (McElroy, 1977) criteria are recommended to select the final SUR model. R_i^2 is to measure the goodness of fit for each $y_i = f(x, \theta_i) + \varepsilon_i$, while R_*^2 is to measure the goodness of fit for the entire SUR model. They are defined as follows:

$$R_i^2 = 1 - \frac{\hat{\varepsilon}_i'\hat{\varepsilon}_i}{(y_i - \bar{y}_i)'(y_i - \bar{y}_i)}, \qquad (12)$$

and

$$R_*^2 = 1 - \frac{\hat{\varepsilon}'\hat{\Omega}^{-1}\hat{\varepsilon}}{y'[\hat{\Sigma}^{-1} \otimes (I_n - \frac{\mathbf{11}'}{n})]y}, \qquad (13)$$

where $\bar{y}_i$ is the mean value of y_i. $\mathbf{1}$ is a column vector of n ones. Note that R_*^2 takes into account the correlation structure for the SUR model.

Step 2. Estimating the ridge path. Choose the desirability function as the optimization measure for multiple responses. Estimate the ridge path by solving Eq. (6). The set $x : x'x = r^2$ can be converted into a set of angles by using a polar-coordinate transformation $A = \{a : -\pi/2 < a_i \le \pi/2 (i = 1, \ldots, k - 2), -\pi < a_{k-1} \le \pi\}$, such that we can obtain in polar coordinates $x = rt(a)$, where a is a $(k - 1) \times 1$ vector of A. The form of $t(a)$ given by Peterson (1993) is:

$$\begin{cases} x_1 = r \sin a_1; \\ x_i = r\{\prod_{j=1}^{i-1} \cos a_j\} \sin a_i, \quad i = 2, \ldots, k - 1; \\ x_k = r \prod_{j=1}^{k-1} \cos a_j. \end{cases} \qquad (14)$$

Then, the optimal desirability function with the constraints $\max_{a \in A} D(rt(a), \theta)$ can be obtained by grid search once the desirability function is specified. As a result, two ridge path plots, i.e., $\max_{a \in A} D(rt(a), \theta)$ vs. r and x vs. r, can be drawn. Due to many appealing theoretical properties, we choose Gibb's desirability function as the optimization measure Eq. (5).

Step 3. Estimating the confidence interval. Through the ridge path, find $x_0 = x_0(\hat{\theta}, r) = arg\max_{a \in A} D(rt(a), \theta)$ for each r. A closed form of region C can be estimated by Eq. (11), then the confidence bands $\min_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$ and $\max_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$ can be obtained. Note that both $\min_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$ and $\max_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$ are optimized under nonlinear constraints. The sequential quadratic programming (SQP) approach is proposed because it is known to be the most effective method for the nonlinearly constrained optimization (Nocedal & Wright, 2006). SQP generates steps by solving quadratic subproblems, which are the optimization problems with a quadratic objective function and linear constraints. Boggs and Tolle (1995) describe the most popular manifestations of SQP, discuss their theoretical properties and comment on their practical implementations. The algorithm of SQP for nonlinear programming can refer to Nocedal and Wright (2006, page 546).

4. Case studies and simulations

4.1. The tire-tread example

Derringer and Suich (1980) described the following tire-tread example. The objective of this study is to determine the best combination of three chemical components in a tire such that all four responses are optimized. The three chemical components are: hydrated silica level (x_1), silane coupling agent level (x_2) and sulfur level (x_3), denoted by $x = \{x_1, x_2, x_3\}$. The four responses taken into account are: PICO abrasion index (y_1), 200 modulus (y_2), elongation at break (y_3) and hardness (y_4), denoted by $y = \{y_1, y_2, y_3, y_4\}$. The directions of optimization of those four responses are not the same, however. Among them, y_1 and y_2 are the larger-the-better (LTB), with their target values 170 and 1300, respectively, while y_3 and y_4 are the nominal-the-best (NTB), and the target values are 500 and 67.5.

A central composite design (CCD) with six center runs was conducted, and the design matrix and the result of this experiment are displayed in Table 1. Following Derringer and Suich (1980), we assume that the true response functions are quadratic in x,

$$E(y_i) = \theta_{i,0} + \sum_{l=1}^{3} \theta_{i,l}x_l + \sum_{\substack{l, m = 1 \\ l \neq m}}^{3} \theta_{i,lm}x_l x_m + \sum_{l=1}^{3} \theta_{i,ll}x_l^2.$$

where $i = 1, 2, 3, 4$. To estimate the ridge path, we follow the steps mentioned in algorithm section.

Table 1
Data set of the experiments.

No.	x_1	x_2	x_3	y_1	y_2	y_3	y_4
1	−1	−1	1	102	900	470	67.5
2	1	−1	−1	120	860	410	65
3	−1	1	−1	117	800	570	77.5
4	1	1	1	198	2294	240	74.5
5	−1	−1	−1	103	490	640	62.5
6	1	-1	1	132	1289	270	67
7	−1	1	1	132	1270	410	78
8	1	1	−1	139	1090	380	70
9	−1.633	0	0	102	770	590	76
10	1.633	0	0	154	1690	260	70
11	0	−1.633	0	96	700	520	63
12	0	1.633	0	163	1540	380	75
13	0	0	−1.633	116	2184	520	65
14	0	0	1.633	153	1784	290	71
15	0	0	0	133	1300	380	70
16	0	0	0	133	1300	380	68.5
17	0	0	0	140	1145	430	68
18	0	0	0	142	1090	430	68
19	0	0	0	145	1260	390	69
20	0	0	0	142	1344	390	70

Step 1. Apply SUR to fit the whole model. Each of equations initially evaluated by p-values from SUR model fits, and all significant terms are kept in the model. The resulting model is obtained by Matlab with $R_1^2 = 96.91\%$, $R_2^2 = 67.33\%$, $R_3^2 = 97.37\%$, $R_4^2 = 95.28\%$ and $R_*^2 = 95.48\%$. Note that the goodness of fit for y_2 is relatively low.

$$\hat{y}_1 = 137.9 + 16.5x_1 + 17.9x_2 + 10.9x_3 + 5.2x_1x_2 + 7.0x_1x_3$$
$$+ 8.2x_2x_3 - 3.8x_1^2 - 3.4x_2^2,$$
$$\hat{y}_2 = 1195.2 + 268.2x_1 + 246.5x_2 + 139.5x_3 - 119.7x_2^2 + 209.3x_3^2,$$
$$\hat{y}_3 = 406.3 - -99.7x_1 - 31.4x_2 - 73.9x_3 + 16.8x_2^2, \quad and$$
$$\hat{y}_4 = 68.7 - -1.4x_1 + 4.3x_2 + 1.6x_3 - 1.6x_1x_2 + 1.6x_1^2.$$

The estimated covariance matrix under SUR model can be obtained as

$$\hat{\Sigma} = \begin{pmatrix} 31.69 & 49.04 & -4.48 & 1.70 \\ 49.04 & 97814.22 & -930.89 & 21.17 \\ -4.48 & -930.89 & 399.43 & -1.10 \\ 1.70 & 21.17 & -1.10 & 1.29 \end{pmatrix}.$$

Step 2. Choosing a desirability function of Gibb et al. (2001) with $\gamma_i = 0.025, i = 1, 2, 3, 4$, we can obtain the maximal desirability value for each r, $\max_{a \in A} D(rt(a), \theta)$. The ridge path and the constrained optimal coordinates are shown in Figs. 1 and 2, respectively. From Fig. 1, the ridge path shows that the overall desirability value is maximized when r is around 1.0. The corresponding setting is found to be $x = [-0.045, 0.318, -0.947]$ as shown in Fig. 2.

Step 3. This step constructs the confidence bounds for the ridge path. The 95% confidence region of $\hat{\theta}$, C, can be obtained through Eq. (11), where $v_h = 2$ and $v_h = 10$. SQP is employed to solve $\min_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$ and $\max_{\theta \in C}\{\max_{a \in A} D(rt(a), \theta)\}$. The resulting ridge path is shown on Fig. 1, with a 95% conservative confidence band.

This is a typical multiple response surface methodology problem to locate the optimal treatment combinations. We use the Gibb et al. (2001) desirability function to measure overall optimal response, then follow the proposed algorithm to draw a desirability-based multi-response ridge path step by step. From the ridge path, we can see that the overall optimum desirability value will be achieved when r is around 1.0.

4.2. The SOVRING example

The SOVRING example by Eriksson, Johansson, Kettaneh-Wold, and Wold (2001) is discussed below to show a typical performance of the proposed method when the sample size is large. The term "SOVRING" was used by Tano (1996)–it is a Swiss word, meaning "screening". The data originated from a mineral sorting plant. The purpose of the experiment is to test the effect of three input variables, raw material input Tonin (x_1), magnetic grinder speed variables HS1 (x_2) and HS2 (x_3). Two response variables of product output in a mine are PAR (y_1) and FAR (y_2). Both responses are larger-the-better. This data have been investigated by Ding et al. (2005) and their results will be served as a benchmark for comparison.

The models for both responses fitted by SUR are:

$$\hat{y}_1 = 283.2 + 115.2x_1 + 6.2x_2 + 22.1x_3 + 13.6x_1x_2 + 30.7x_1x_3$$
$$+ 18.4x_2x_3 + 24.7x_1^2 - 7.4x_2^2 + 10.1x_3^2,$$
$$\hat{y}_2 = 689.8 + 244.1x_1 + 29.8x_2 - 46.5x_3 + 1.1x_1x_2 - 61.5x_1x_3$$
$$- 55.8x_2x_3 + 6.8x_1^2 - 9.5x_2^2 - 35.4x_3^2.$$

The estimated covariance matrix under SUR model is

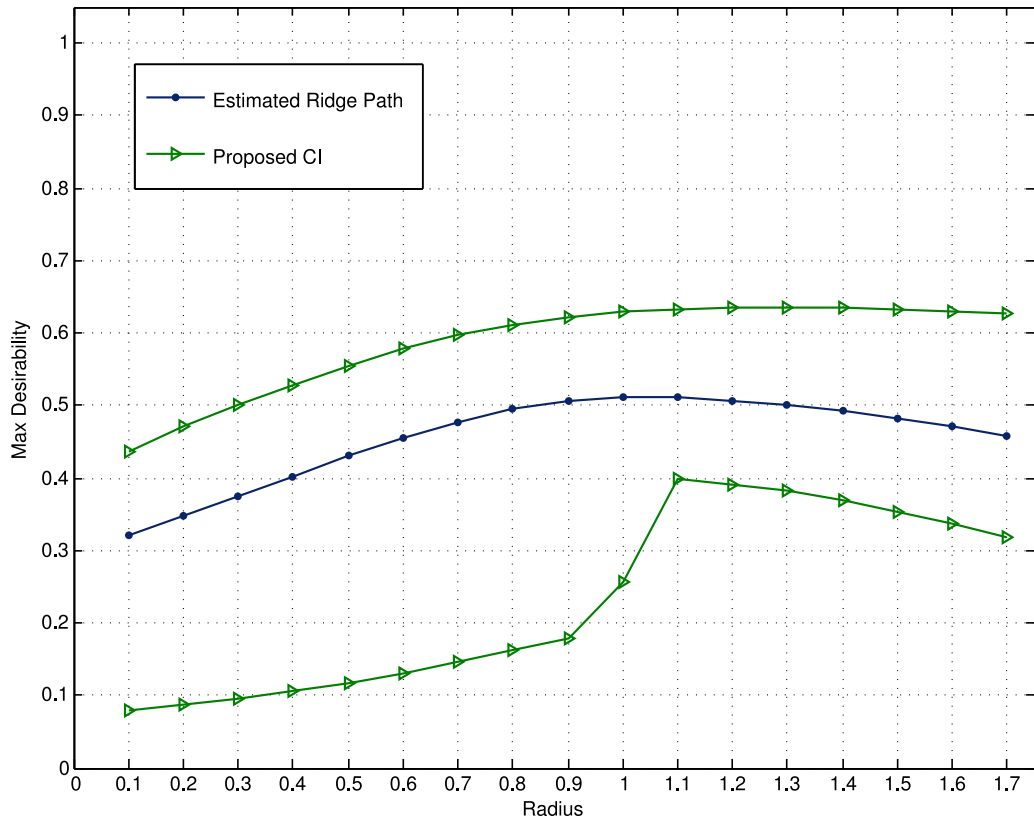$$\hat{\Sigma} = \begin{pmatrix} 195.13 & 101.91 \\ 101.91 & 1063.35 \end{pmatrix}.$$
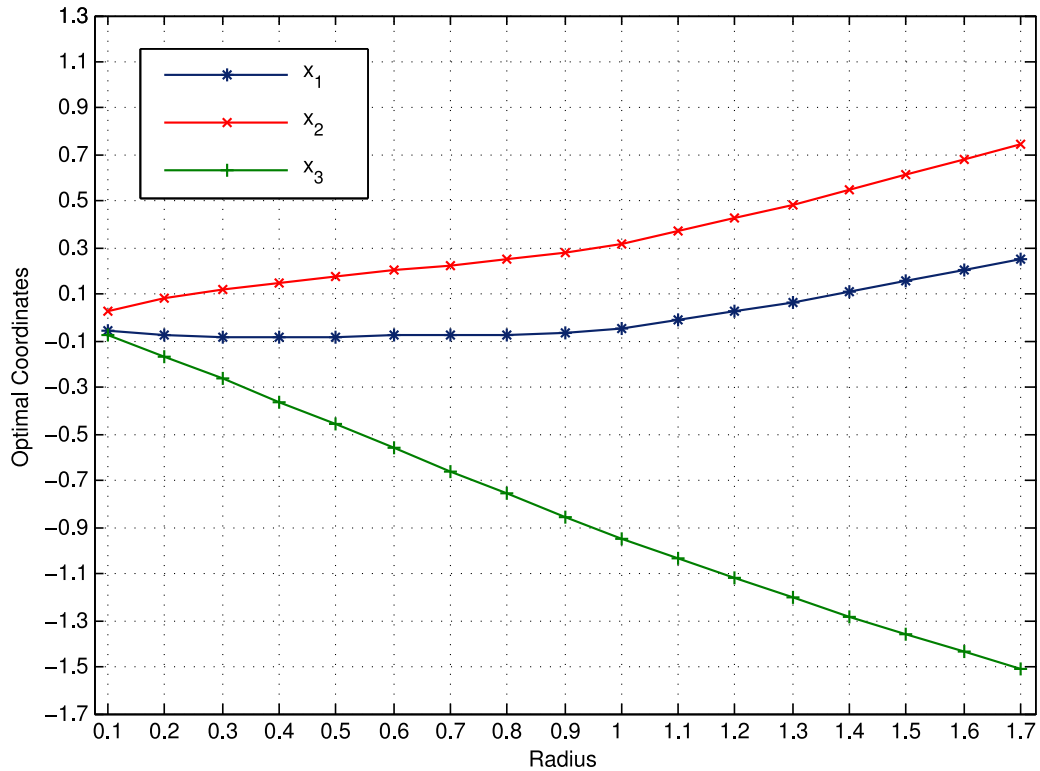
Fig. 1. Ridge path of the tire-tread responses.

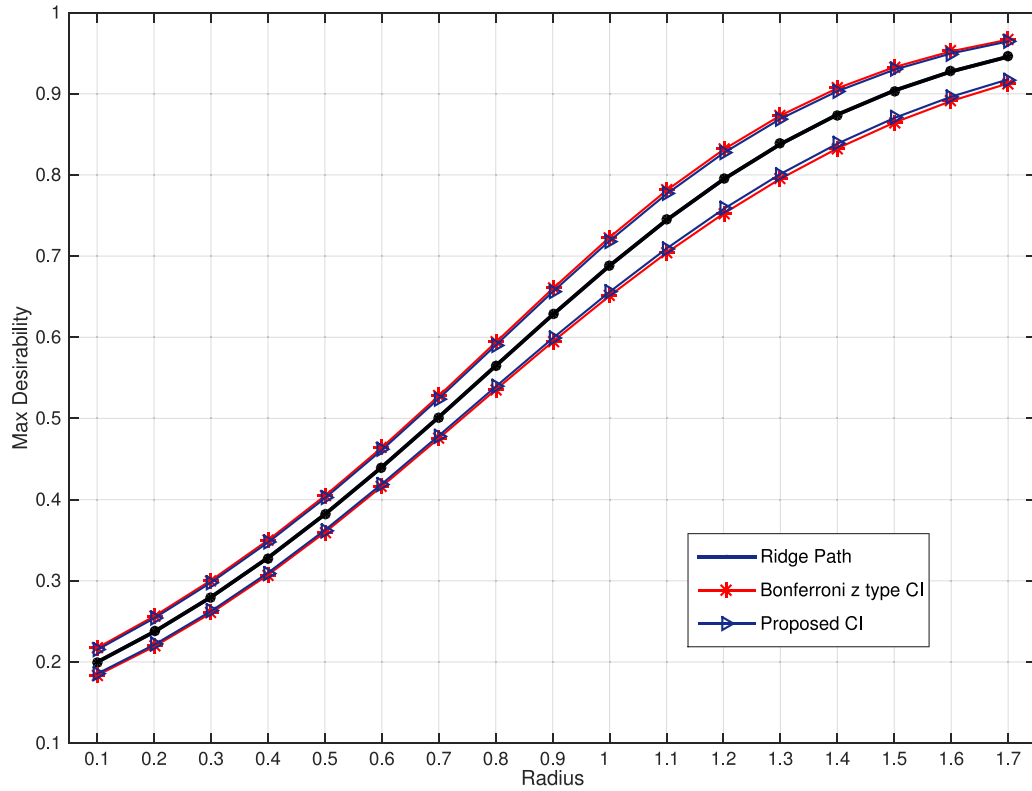**Fig. 2.** Constrained optimal coordinate plot.

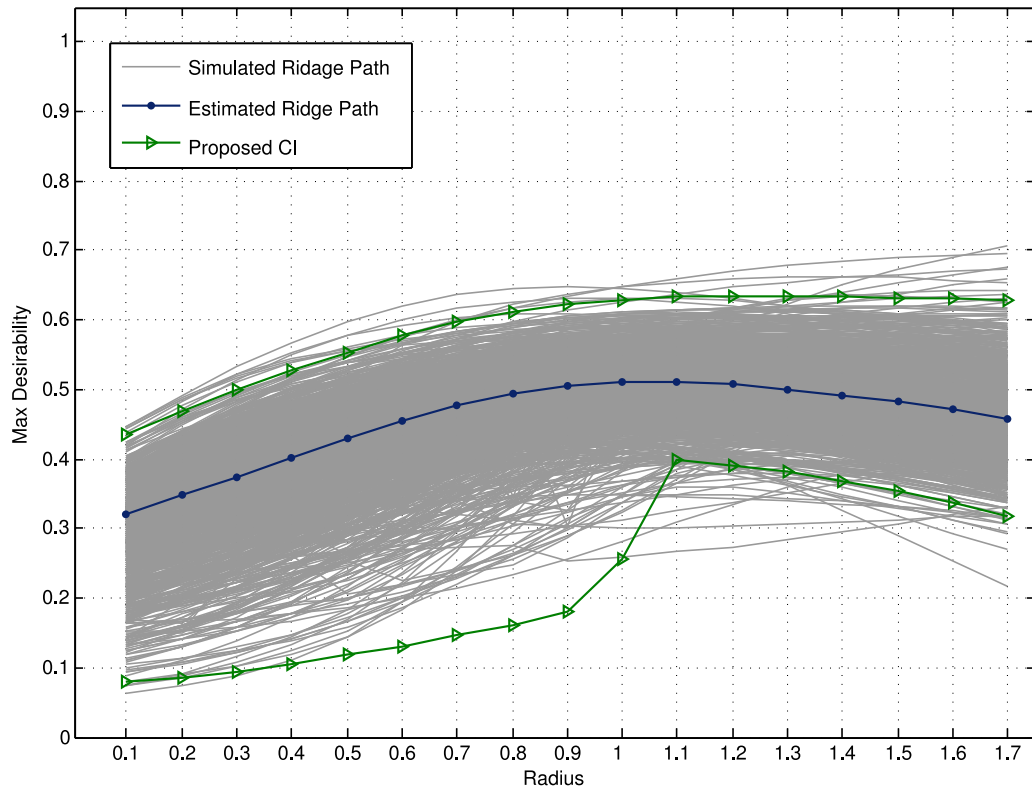**Fig. 3.** The ridge path and confidence intervals of SOVRING example.

Fig. 4. The simulated ridge path and simultaneous coverage of proposed confidence interval.
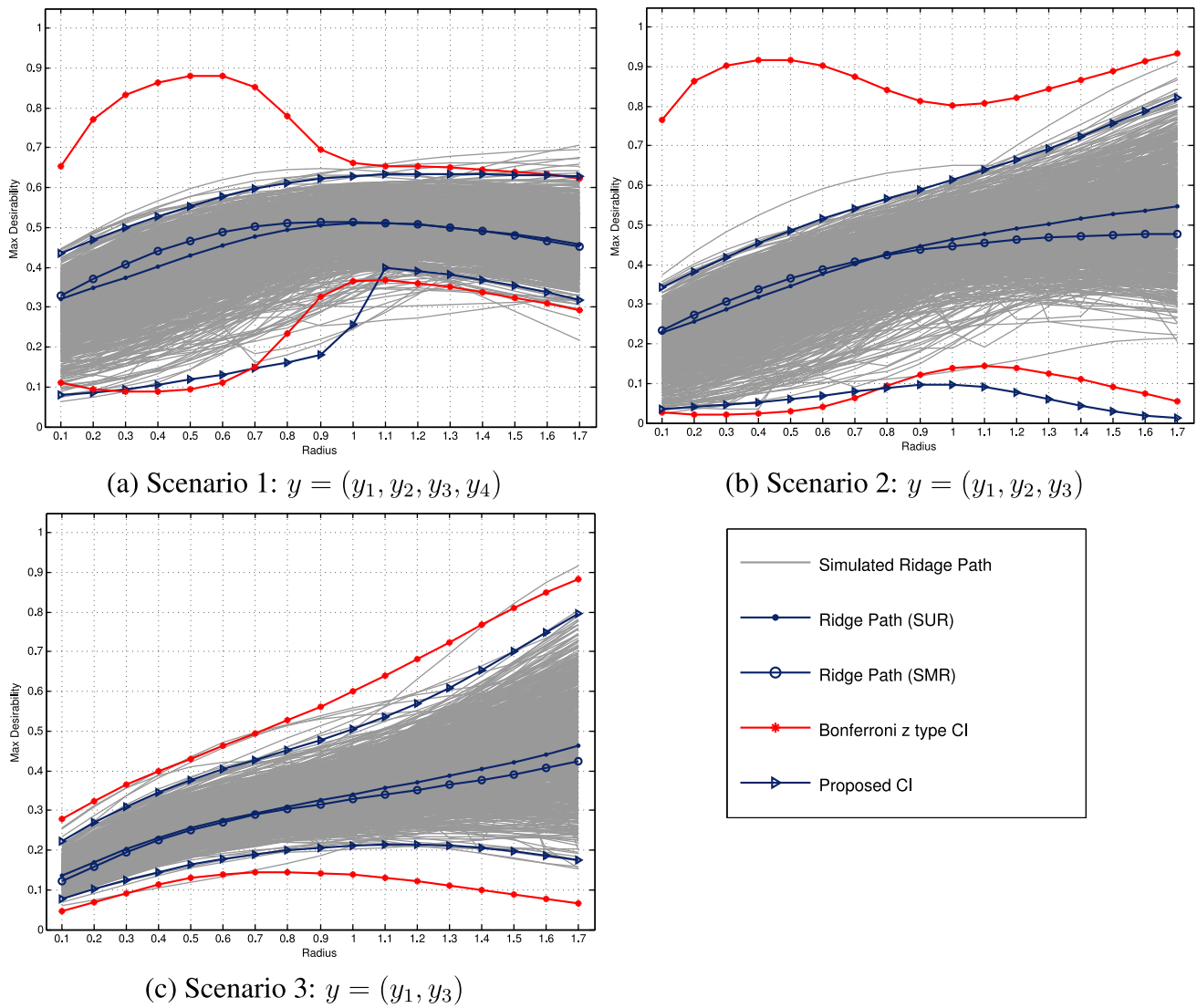
(a) Scenario 1: $y = (y_1, y_2, y_3, y_4)$

(b) Scenario 2: $y = (y_1, y_2, y_3)$

(c) Scenario 3: $y = (y_1, y_3)$

Fig. 5. Simulation study: comparing the performance of different CI in different scenarios.

Fig. 3 shows the ridge path, the confidence band via Bonferroni (Ding et al., 2005) as well as confidence band via the proposed method. It is shown that the confidence band of both approaches are nearly identical with the Bonferroni type confidence being a little bit wider (especially as r increases). This indicates that the proposed method also performs well when sample size is large.

4.3. Simulation study and comparison

To asses the performance of the proposed method in general, a thorough simulation study is employed. The general observations from the simulation are: (1) the proposed confidence region provides approximately 95% coverage; (2) the performance of the proposed method is more appropriate than the large-sample method for small sample situations; (3) the estimators which include confidence interval construction of the multiple response ridge path are instructive and meaningful for locating the optimal setting of experiments.

Assuming the fitted model and the related covariance in the tire-tread example as the true model, 1,000 simulated samples are generated in each setting where the error follows a multivariate normal distribution with the same mean and covariance structure. The ridge path and its confidence region can be easily obtained by

following the proposed procedure. The simulation result is shown in Fig. 4.

In Fig. 4, the ridge path and its 95% confidence bands are plotted and they are used as the benchmark. Each simulated ridge path was estimated for each simulation sample and shown in the figure. The simultaneous coverage proportion of proposed confidence region is about 93.4%.

We compare our method with the large-sample method under three different scenarios: Scenario 1 includes all of the four responses, i.e. $y = (y_1, y_2, y_3, y_4)$; Scenario 2 has three responses, i.e. $y = (y_1, y_2, y_3)$; while Scenario 3 has two responses of y_1 and y_3. Note that y_1 and y_2 are the LTB type responses, and y_3 as well as y_4 are NTB type responses. The simulation results of these scenarios are shown in Fig. 5. Fig. 5(a)–(c) are for Scenario 1, 2 and 3, respectively.

For each sub-figure in Fig. 5, the ridge path and its 95% confidence interval were obtained with the relative data set in Table 1. Both the Bonferroni type confidence and the proposed type confidence are displayed; 1,000 simulated ridge paths for each scenario are shown in the figure. Note that both of the Large-sample method and the proposed approach are based on the SUR model, while the ridge path is estimated based on both the SUR and the SMR models. It is shown that the Bonferroni z-type confidence are much more conservative than the proposed

ones, and the proposed method provides much smoother confidence bands. Specifically, the coverage probability of the Bonferroni *z*-type intervals are 94.2%, 99.8% and 99.6%, respectively for Scenarios 1, 2 and 3, while the corresponding coverage probabilities of conservative confidence bands are 93.4%, 96.8% and 94.9%. Though the coverage probability of both approaches are almost same in Scenario 1, it is clearly shown in Fig. 5(a) that the confidence region from the large-sample approach is wider than that of the proposed method especially when radius ranges from 0.1 to 0.8. It also can be observed that the two types of estimated ridge paths are different in all scenarios. This indicates that including the non-significant terms in the model will result in a different prediction. For example, the maximal desirability value exists on the ridge path based on SUR model at $r = 1.4$ in Scenario 2 (Fig. 5b), while it appears around at $r = 1.6$ on the ridge path based on SMR model. Although the estimated ridge paths are different in each scenario, they are not very different from each other. That is, using the typical multivariate regression model versus SUR does not seem to give drastically differing results.

Through the estimate of the ridge path and its confidence interval, it is desirable that experimenters can locate the optimal combination of treatments within the experimental space or decide to run more experiments along certain directions. In Fig. 5(a), the optimal treatment combination takes place at $r = 1.0$ in Scenario 1. Fig. 5(b) shows that the maximal desirability value is obtained at about $r = 1.4$ in Scenario 2. For a ridge path in Fig. 5(c), it is monotone increasing along the radius and the confidence region gets wider, it can be expected that the optimal treatment combination may exist at a certain farther radius, which is beyond the current experimental space. To seek the potential optimal location, experimenters may conduct some experiments along the ridge path direction. Furthermore, it is shown that the optimal location is different in each scenario. This brings up another important issue that the experimenter must be cautious in choosing the number of responses, because different numbers of responses can produce very different ridge paths. The effects of this variation depend on the process of interest. This is mainly in the knowledge domain and should be explored by the practitioners.

5. Conclusions

A new approach for construction of a confidence region for multiple responses with small sample size is proposed. The proposed approach employs the SUR (seemingly unrelated regressions) model to estimate parameters and covariance, as opposed to the SMR (standard multivariate regression) model used in the large-sample case. The SMR model usually over-fits for some responses. The coverage rate is too conservative for small samples. The Gibb's desirability function (to represent the overall desirability for all responses) uses a quadratic region for $\hat{\theta}$ by the Lawley–Hotelling test, and conducts a three-step algorithm for constructing the desirability function-based ridge path confidence region. Gibb's desirability function is merely one of these many to be applied. In fact, any continuous and differentiable desirability function can be applied to the proposed algorithm directly. Through real-life examples in the literature (both small and large sample cases) as well as decent simulations, the proposed approach performs significantly better than the existing method which is limited to large samples.

A confidence region for the ridge path had been studied in the literature, but mainly for univariate response problems. Little had been known about confidence regions for the ridge path for multiple response cases. The proposed method given here allows us to construct a (relatively tight) confidence region for the ridge path for multiple responses under small sample sizes. This will help experimenters to locate the optimal setting in an efficient manner.

References

Ames, A. E., Mattucci, N., Macdonald, S., Szonyi, G., & Hawkins, D. (1997). Quality loss functions for optimization across multiple response surfaces. *Journal of Quality Technology, 29*, 339–346.
Arnold, S. F. (1981). *The theory of linear models and multivariate analysis*. New York, NY: Wiley.
Bera, S., & Mukherjee, I. (2015). A multistage and multiple response optimization approach for serial manufacturing system. *European Journal of Operational Research, 248*, 444–452.
Boggs, P. T., & Tolle, J. W. (1995). Sequential quadratic programming. *Acta Numerica, 4*, 1–51. doi:10.1017/S0962492900002518.
Carter, W. H., Chinchilli, V. M., Myers, R. H., & Campbell, E. D. (1986). Confidence intervals and an improved ridge analysis of response surfaces. *Technometrics, 28*, 339–346.
Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variable. *Journal of Quality Technology, 12*, 214–219.
Ding, R., Lin, D. K. J., & Peterson, J. J. (2005). A large-sample confidence band for a multi-response ridge path. *Quality And Reliability Engineering International, 21*, 669–675.
Eriksson, I., Johansson, E., Kettaneh-Wold, N., & Wold, S. (2001). Multi- and megavariate data analysis: Part I-Basic principles and applications. *Umetrics Academy*, Umeå, Sweden. ISBN 91-973730-1-X.
Gibb, R. D., Carter, W. H., & Myers, R. H. (2001). Incorporating experimental variability in the determination of desirable factor levels. *Unpublished Manuscript*.
Harrington, E. C. J. (1965). The desirability function. *Industrial Quality Control, 21*, 494–498.
He, Z., Zhu, P.-F., & Park, S.-H. (2012). A robust desirability function method for multi-response surface optimization considering model uncertainty. *European Journal of Operational Research, 221*, 241–247.
Henningsen, A., & Hamann, J. (2007). A package for estimating systems of simultaneous equations in r. *Journal of Statistical Software, 23(4)*, 1–40.
Hoerl, A. E. (1959). Optimum solution of many variables equations. *Chemical Engineering Progress, 55*, 69–78.
Hosmer, D. W., & Lemeshow, S. (2005). *Applied Logistic Regression* (2nd). New York, NY: Wiley.
Hunter, J. S. (1999). Discussion on "response surface methodology–current status and future directions" by Myers, Raymond H. *Journal of Quality Technology, 31(1)*, 54–57.
Jeong, I.-J., & Kim, K.-J. (2009). An interactive desirability function method to multiresponse optimization. *European Journal of Operational Research, 195(2)*, 412–426.
Khuri, A. I., & Conlon, M. (1981). Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics, 23(4)*, 363–375.
Kim, K.-J., & Lin, D. K. J. (2000). Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. *Applied Statistics, 49(3)*, 311–325.
Kim, K.-J., & Lin, D. K. J. (2006). Optimization of multiple responses considering both location and dispersion effects. *European Journal of Operational Research, 169(1)*, 133–145.
Lin, D. K. J. (1999). Discussion on "response surface methodology–current status and future directions". *Journal of Quality Technology, 31*, 61–66.
McElroy, M. (1977). Goodness of fit for seemingly unrelated regressions. *Journal of Econometrics, 6*, 381–387.
Miller, R. (1981). *Simultaneous Statistical Inference* (2nd). Springer: New York.
Nocedal, J., & Wright, S. J. (2006). Numerical optimization. *Springer Series in Operations Research* (2nd). Springer Verlag.
Peterson, J. (1993). A general approach to ridge analysis with confidence intervals. *Technometrics, 35*(2), 204–214.
Peterson, J. J., & Kuhn, A. M. (2005). Ridge analysis with noise variables. *Technometrics, 47(3)*, 274–283.
Pignatiello, J. J. (1993). Strategies for robust multi-response quality engineering. *IIE Transaction, 25(3)*, 5–15.
Tano, K. (1996). Multivariate modelling and monitoring of mineral processes using partial least square regression. Licentiate Thesis No. 1996:36L, *Luleå University of Technology*, ISSN: 0280-8242.
Timm, N. H. (2002). *Applied multivariate analysis*. Springer-Verlag.
Vining, G. G. (1998). A compromise approach to multiresponse optimization. *Journal of Quality Technology, 30(4)*, 309–313.
Zellner, A. (1962). An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. *Journal of the American Statistical Association, 57*, 348–368.
Zellner, A., & Huang, D. (1962). Further properties of efficient estimators for seemingly unrelated regression equations. *International Economic Review, 3*, 300–313.