European Journal of Operational Research 252 (2016) 829–836

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

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

Article history:
Received 5 February 2015
Accepted 20 January 2016
Available online 6 February 2016

Keywords:
Quality management
Conservative confidence region
Desirability function
Multi-response surface methodology
Seemingly unrelated regression model

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.

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).

http://dx.doi.org/10.1016/j.ejor.2016.01.037

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

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

where C is a $100(1-\alpha)\%$ confidence region for θ. The confidence region C is defined as $C = \{\theta : (\theta - \hat{\theta})'V^{-1}(\theta - \hat{\theta}) \le c_\alpha^2\}$, 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

$$y_i = f(x, \theta_i) + \varepsilon_i \tag{2}$$

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),

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

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

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

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),

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

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

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

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

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

where $[L, U] = \text{logit}(g(\hat{\theta}, r)) \pm z_{\alpha/2q}\hat{c}(r)$, 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 $\text{logit}(g(\hat{\theta}, r))$, 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.