Thailand Statistician
July 2017; 15(2): 167-183
http://statassoc.or.th
Contributed paper

An Application of Nelder-Mead Algorithm in Response Surface Methodology: CCD

Chantha Wongoutong, Jirawan Jitthavech* and Vichit Lorchirachoonkul
Graduate School of Applied Statistics, National Institute of Development Administration, Bangkok 10240, Thailand.
*Corresponding author; e-mail: jirawan@as.nida.ac.th

Received: 14 October 2016
Accepted: 26 December 2016

Abstract

A novel method using the Nelder-Mead algorithm is proposed to be used instead of the first order model in moving the experiment in the response surface methodology toward the neighbor of the optimum. A second order model similar to the second order model in the CCD is constructed to estimate the optimum design point and the optimum response. From the simulation using five published test functions and five different normal generators, it can be concluded that the proposed method outperforms the traditional CCD in terms of the number of experiments, the MAPEs of the estimated optimum design point and estimated optimum response and the coverage probabilities.

Keywords: Response surface methodology, central composite design, Nelder-Mead algorithm.

1. Introduction

Response surface methodology (RSM) is useful for experiments with n quantitative factors that are under taken so as to determine the combination of levels of factors at which each of the quantitative factors must be set in order to optimize the response in some sense. The traditional response surface method originally proposed by Box and Wilson (1951) is based on initially conducting steepest ascent or descent searches until a significant curvature is detected. Box and Wilson (1951) use this method of interest to maximize the response based on experiments conducted on the direction defined by the gradient of an estimated main effects model. The observed responses along the steepest ascent direction are used to locate the neighborhood of the maximum. This method can theoretically locate the maximum through numerous iterations as long as it exists.

However, if it is used on a badly scaled system, the rate of convergence may become too slow and the method is impractically to be used. Normally, the step size is estimated by using the coefficient of regression in the first order model based on the results from the experiments. In the other word, the effectiveness of the traditional response surface method depends on the step sizes given by the first order models.

This study proposes an alternative to solve the step size problem in the traditional RSM by using only the values in the experiments without using the steepest direction in searching the optimum of the response surface. The best-known methods in the direct search class include genetic algorithms (Holland 1992), simulated annealing (Brooks and Morgan 1995), Tabu search (Glover and Laguna 1997), neural networks (Sexton et al. 1998) and Nelder-Mead (NM) algorithm (Nelder and Mead 1965). The NM algorithm is quite simple to understand and very easy to use (Gavin 2016). This leads to its wide applications in many fields of science and technology, especially in chemistry and medicine.

The Nelder-Mead (NM) method is a derivative free method for searching a local optimum of a function. In this optimization process, the initial simplex adapts itself iteratively to the local surface landscape by varying its size and orientation. The NM algorithm is especially suitable for exploring the "unwieldy" terrains and has been widely accepted as the most robust and efficient of current sequential techniques for unconstrained optimization (Lagarias et al. 1998). Numerous software packages include the NM algorithm as an optimization solver such as Mathematica, MultiSimplex, PROC IML in SAS, etc. The idea behind the NM algorithm is to "crawl" to the optimum in the selected direction by moving one vertex of the simplex at each iteration. The vertices are moved by performing four basic operations: Reflection, Expansion, Contraction, and Multiple Contraction (shrink).

Aimed at having better convergence, several variants of the simplex method have been proposed (Torczon 1989; Dennis and Torczon 1991; Torczon 1997; McKinnon 1998; Byatt 2000; Kelley 2000; Tseng 2000; Price et al. 2002). The NM algorithm generally performs well for solving small dimensional real life problems and continuously remains as one of the most popular direct search methods (Wright 1996; Lagarias et al. 1998; Kolda et al. 2003; Han and Neumann 2006; Gavin 2016). It has been observed by many researchers that the NM algorithm may become inefficient for large dimensional problems (Parkinson and Hutchinson 1972, Byatt 2000; Torczon 1989). However, for the usual problems arising from the RSM practice, the number of influential process factors included in the final model is rarely larger than half a dozen (Olsson and Nelson 1975, and Myers and Montgomery 2002). Typically, a "pre-experiment" via fractional factorial is carried out in the earlier phase to eliminate the irrelevant factors, leaving only a small number of relevant factors.

The NM algorithm has demonstrated its wide versatility, accuracy and ease of use for solving different types of optimization problems in the noise-free environments in the area of applied statistics (Olsson 1974; Olsson and Nelson 1975; Copeland and Nelson 1996; Khuri and Cornell 1996). But the application of the NM algorithm in the response surface optimization in the presence of errors has been scarcely reported in the RSM literature.

This study proposes to use the NM algorithm searching for the optimum region instead of the steepest direction in the traditional RMS. The second order model similar to the traditional RMS is still used to estimate in the optimum. This paper is structured as follows. The next section summarizes the 2^k factorial design in the traditional RMS. The NM algorithm is proposed to be used in the 2^k factorial design in Section 3. Section 4 is the simulation study of experimental designs by the traditional RMS and the proposed NM algorithm using the five published functions with the random noises and the comparisons of simulation results are also presented. The final section is conclusions.

2. Response Surface Methodology (RSM)

The response surface methodology (RSM) is a collection of mathematical and statistical techniques for the analysis of problems in many fields such as engineering, manufacturing, agricultural, etc. in which a response is influenced by several independent variables. The objective is to determine the optimum design point in an experiment (Montgomery 2005). In general, such a relationship between a response of interest and independent variables, which are the factors in an experiment, is unknown and is approximated by a low-degree polynomial model (Box and Draper 1987) of the form

$$y = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta} + \varepsilon \tag{1}$$

where $\mathbf{x}^T = (x_1, x_2, ..., x_k)$, $\mathbf{f}(\mathbf{x})$ is a vector function of p elements that consists of a d^{th} order polynomial of $x_1, x_2, ..., x_k$, $d \geq 1$, $\boldsymbol{\beta}$ is a vector of p unknown constant parameters, and ε is a random experimental error distributed as $N(0, \sigma^2)$. The expected response is $E(y) = \mathbf{f}^T(\mathbf{x})\boldsymbol{\beta}$ but the vector function $\mathbf{f}$ and the parameter vector $\boldsymbol{\beta}$ are unknown and are to be estimated. Frequently, in the first experiment of a 2^k factorial design where k is the number of factors, the initial 2^k vertices might be far from the optimum region. Additional experiments of a 2^k factorial design without replication at a group of center points of the simplex are performed to construct the first order model, $d = 1$, consisting of only k first order terms of the independent variables and can be written as

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \varepsilon, \tag{2}$$

in order to estimate the error, to test the hypothesis of the existence of the cross product terms in the model and the quadratic effect which indicate the existence of the curvature. The random error, ε, is assumed to be normally distributed with mean 0 and variance σ^2.

If no curvature is detected, the experiment will move to the points sequentially in the steepest direction to optimize the change in the response. The steepest direction method is a procedure for moving sequentially along the path of steepest direction toward the neighborhood of the optimum. Consider the fitted first-order response surface model,

$$\hat{y} = \hat{\beta}_0 + \sum_{i=1}^{k} \hat{\beta}_i x_i, \tag{3}$$

where $\hat{y}$ is the predicted response, x_i represents the coded variable of the i^{th} independent factor, $\hat{\beta}_0$ is the estimated intercept, and the individual $\hat{\beta}_i$'s are the estimated coefficients of the i^{th} coded variable. The steepest direction method seeks for a point determined by a set of independent variables that produce the estimated optimum response over all points that are a fixed distance r from the center of the design which is the point $\mathbf{x} = \mathbf{0}$. As a result; the optimization problem involves the use of a Lagrange multiplier (λ) subject to the restraint given by $\sum_{i=1}^{k} x_i^2 = r^2$. Taking partial derivatives the Lagrangian function

$$L = \hat{\beta}_0 + \hat{\beta}_1 x_1 + ... + \hat{\beta}_k x_k - \lambda\left(\sum_{i=1}^{k} x_i^2 - r^2\right),$$

with respect to x_j and setting it equal to zero yield

$$x_j = \frac{\hat{\beta}_j}{2\lambda}, \; j = 1, 2, \ldots, k. \tag{4}$$

The expression $\frac{1}{2\lambda}$ may be viewed as a constant of proportionality ρ and (4) becomes

$$x_j = \rho\hat{\beta}_j, \; j = 1, 2, \ldots, k, \tag{5}$$

where

$$\rho = \frac{\pm r}{\left(\sum_{j=1}^{k} \hat{\beta}_j^2\right)^{1/2}}. \tag{6}$$

The positive (negative) value of ρ in (6) is used when searching for the minimum (maximum) on the response surface. The origin of independent variables is moved to the point given by (5) which is the point where the largest absolute change of (4) occurs on the hyper-sphere of radius r. The value of x_j in (5) can be considered as an increment of x_j, Δx_j, moving away from the origin. Let $|\beta_i| = \max_{1 \le j \le k} |\beta_j|$. The increment of x_j, Δx_j, in term of the largest absolute increment Δx_i can be written as

$$\Delta x_j = \frac{\hat{\beta}_j}{\hat{\beta}_i / \Delta x_i} \quad j = 1, 2, ..., k, \; i \ne j. \tag{7}$$

The incremental Δx_j in (7) is converted back to the increment of the corresponding natural variable before running the next experiment. After the experiment moves toward the neighborhood of the optimum along the steepest direction, step by step given by (7), until a curvature is detected by direct comparison of the successive responses, the 2^k vertices of the current simplex with a number of center points are performed as in the initial experiment to test the hypothesis of the existence of a curvature. If the test cannot reject the null hypothesis, a number of axial points by using the central composite design (CCD) are added in the experiment to construct the second order model, $d = 2$, for estimating the optimum design point. The second order model consists of k first order terms, k second order terms and $k(k-1)/2$ interaction terms and can be written as

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j>i}^{k} \beta_{ij} x_i x_j + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \varepsilon, \tag{8}$$

which is used to approximate the response surface in the neighborhood of the optimum. After the parameters of the second order model are estimated, the optimum design point in terms of the coded variables can be written as (Anderson, et al. 2009)

$$\mathbf{x}_{opt} = -\frac{1}{2}\mathbf{B}^{-1}\mathbf{b}, \tag{9}$$

where

$$\mathbf{b} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} \quad \text{and} \quad \mathbf{B} = \begin{bmatrix} \hat{\beta}_{11} & \hat{\beta}_{12}/2 & \cdots & \hat{\beta}_{1k}/2 \\ \hat{\beta}_{12}/2 & \hat{\beta}_{22} & \cdots & \hat{\beta}_{2k}/2 \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\beta}_{1k}/2 & \hat{\beta}_{2k}/2 & \cdots & \hat{\beta}_{kk} \end{bmatrix},$$

and the predicted response at the optimum design point is given by

$$\hat{y}_{opt} = \hat{\beta}_0 + \frac{1}{2}\mathbf{x}'_{opt}\mathbf{b}, \tag{10}$$

which is to be converted back in terms of the natural variables.

3. The Proposed Method

The proposed method consists of two phases to search for the optimum design point in an experiment of a 2^k factorial design. The first phase is to move the experiment from the initial points toward the neighborhood of the optimum in the direction given by the NM algorithm instead of the steepest direction as in the first order model in the traditional RSM. After the experiment approaches the vicinity of the optimum by moving in the directions given by the NM algorithm, the first phase is terminated and the second phase of the proposed method is to construct a second order model without using CCD as in the traditional RSM such that the variances of the estimated parameters can be estimated.

The NM algorithm is a heuristic, iterative and derivative free procedure for multidimensional unconstrained optimization problems. The movement is determined by comparing the responses of the experiments at the vertices of the simplex. The simplex adapts itself to the local landscape and moves on to the final optimum.

A simplex is described in a geometric figure with k dimensions which is a convex hull of $k+1$ vertices, i.e. a simplex with vertices of the natural variables, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k+1}$ denoted by Δ. The method iteratively generates a sequence of simplexes to approximate an optimal design point of the response in the experiment. In minimization problem, the simplexes are ordered in an iteration according to the responses with $y(\mathbf{x}_1) \le y(\mathbf{x}_2) \le \ldots \le y(\mathbf{x}_{k+1})$. In maximization problem, the simplexes are ordered reversely with $y(\mathbf{x}_1) \ge y(\mathbf{x}_2) \ge \ldots \ge y(\mathbf{x}_{k+1})$. Let $\mathbf{x}_1$ be the best vertex and $\mathbf{x}_{k+1}$ be the worst vertex.

Four possible operations are determined in the algorithm: reflection, expansion, contraction, and shrink, each being associated with a scalar parameter: α (reflection), β (expansion), γ (contraction), and δ (shrink) (see Figure 1). These parameters should satisfy $\alpha > 0$, $\beta > 1$, $\beta > \alpha$, $0 < \gamma < 1$, and $0 < \delta < 1$ (Lagarias et al. 1998). In the standard implementation of the NM algorithm, the parameters are chosen to be $\{\alpha, \beta, \gamma, \delta\}$ = {1, 2, 1/2, 1/2}. Let $\bar{\mathbf{x}}$ be the centroid of the first k best vertices, excluding the worst vertex, $\bar{\mathbf{x}} = \frac{1}{k}\sum_{i=1}^{k}\mathbf{x}_i$.

Each iteration in the NM algorithm starts with a simplex specified by $k+1$ vertices in a k dimensional space and can be described, for a minimization problem, as follows.

1. Sort. Perform the experiment at the $k+1$ vertices of Δ and sort the vertices such that $y(\mathbf{x}_1) \le y(\mathbf{x}_2) \le \ldots \le y(\mathbf{x}_{k+1})$. If $|\, y(\mathbf{x}_1) - y(\mathbf{x}_{k+1}) \,| < e$ where e is the specified tolerance, then stop the algorithm; otherwise, go to Step 2.

2. Reflection. Compute the reflection point $\mathbf{x}_r = \bar{\mathbf{x}} + \alpha(\bar{\mathbf{x}} - \mathbf{x}_{k+1})$ and perform the experiment at $\mathbf{x}_r$. If $y(\mathbf{x}_1) \le y(\mathbf{x}_r) \le y(\mathbf{x}_k)$, then replace $\mathbf{x}_{k+1}$ with $\mathbf{x}_r$ and terminate the iteration.

3. Expansion. If $y(\mathbf{x}_r) < y(\mathbf{x}_1)$, then compute the expansion point $\mathbf{x}_e = \bar{\mathbf{x}} + \beta(\mathbf{x}_r - \bar{\mathbf{x}})$ and perform the experiment at $\mathbf{x}_e$. If $y(\mathbf{x}_e) < y(\mathbf{x}_r)$, then replace $\mathbf{x}_{k+1}$ with $\mathbf{x}_e$; otherwise replace $\mathbf{x}_{k+1}$ with $\mathbf{x}_r$. Terminate the iteration.

4. Outside Contraction. If $y(\mathbf{x}_k) \le y(\mathbf{x}_r) < y(\mathbf{x}_{k+1})$, compute the outside contraction point $\mathbf{x}_{oc} = \bar{\mathbf{x}} + \gamma(\mathbf{x}_r - \bar{\mathbf{x}})$ and perform the experiment at $\mathbf{x}_{oc}$. If $y(\mathbf{x}_{oc}) \le y(\mathbf{x}_r)$, then replace $\mathbf{x}_{k+1}$ with $\mathbf{x}_{oc}$ and terminate the iteration; otherwise go to step 6.

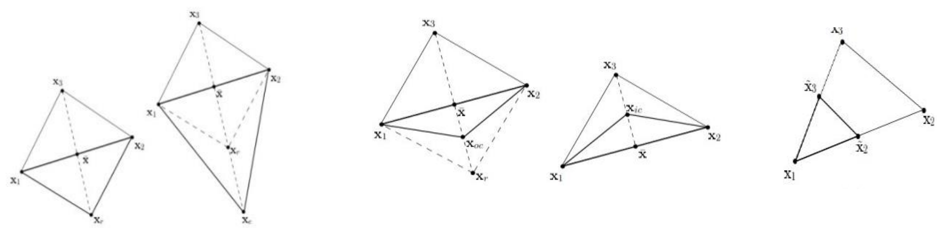
5. Inside Contraction. If $y(\mathbf{x}_r) \ge y(\mathbf{x}_{k+1})$, compute the inside contraction point

$$\mathbf{x}_{ic} = \bar{\mathbf{x}} - \gamma(\mathbf{x}_r - \bar{\mathbf{x}}),$$

and perform the experiment at $\mathbf{x}_{ic}$. If $y(\mathbf{x}_{ic}) \le y(\mathbf{x}_{k+1})$, then replace $\mathbf{x}_{k+1}$ with $\mathbf{x}_{ic}$ and terminate the iteration; otherwise go to step 6.

6. Shrink. Compute the shrinkage point $\mathbf{v}_i = \mathbf{x}_1 + \delta(\mathbf{x}_i - \mathbf{x}_1)$, $2 \le i \le k+1$ and perform the experiment at $\mathbf{v}_i$, $2 \le i \le k+1$. The new vertices of the simplex at the next iteration are $\mathbf{x}_1, \mathbf{v}_2, \mathbf{v}_3, \ldots, \mathbf{v}_{k+1}$.

The possible operations in a two-dimensional space are shown in Figure 1.



Reflection Expansion Outside Contraction Inside Contraction Shrink

Figure 1 Possible operations performed on a simplex in a two-dimensional space

In a 2^k factorial design, the initial experiment is performed at 2^k vertices and a group of center points. But in the first phase of the proposed method, the initial experiment is performed at only $k+1$ points which may be randomly selected from the 2^k vertices in the 2^k factorial design. For example, in a 2^2 factorial design, any three of the four starting points as shown in Figure 2. In the next experiment, only one additional point given by one of the five operations: reflection, expansion, outside contraction, inside contraction and shrink is required to be performed. Let $y(\mathbf{x}_i^l)$ be the response of the experiment at the i^{th} vertex in the l^{th} iteration. The first phase is terminated at the l^{th} iteration if $|\, y(\mathbf{x}_1^l) - y(\mathbf{x}_{k+1}^l) \,| < e$, the acceptable tolerance.

In an iteration in the NM algorithm, the number of points required to perform the experiment depends on the operation in the iteration as shown in Table 1. The numbers of points in Table 1 are in accordance with the NM algorithm as described above. A shrink operation may lead to an increase in every response, $y(\mathbf{x}_i)$, except $y(\mathbf{x}_1)$. However, the shrink operation very rarely happens in practice (Torczon 1989). In the non-shrink iteration l, it can be shown that

$$y(\mathbf{x}_i^{l+1}) \le y(\mathbf{x}_i^l), \quad \text{for } 1 \le i \le k+1,$$

with strict inequality for at least one value of *i* if the response is bounded below (Lagarias et al. 1998). In other words, in a minimization problem in the experimental design and no shrink operation occurs, the difference $y(\mathbf{x}_{k+1}^l) - y(\mathbf{x}_1^l)$ approaches to 0 as *l* becomes large since, in practice, the response is bounded below. Therefore, under the condition that no shrink operation occurs, it can be concluded that when $y(\mathbf{x}_{k+1}^l) - y(\mathbf{x}_1^l) < e,$ the experiment has moved to a neighborhood of the minimum. No statistical hypothesis testing is required in the proposed method.

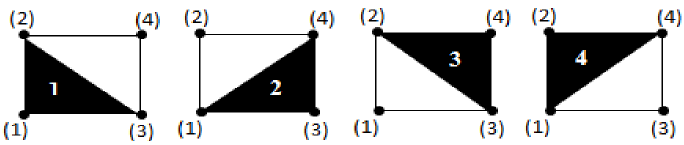


Figure 2 Four possible starting points in the first phase of the proposed method

Table 1 Number of points required performing the experiment under different operations in the NM algorithm

Operation	Reflection	Expansion	Outside contraction	Inside contraction	Shrink
No. of points	1	2	2	2	k

The second phase of the proposed method constructs the second order model in which the total number of parameters including the intercept term is equal to $\frac{k(k+3)}{2}+1$. The responses at the last $\frac{k(k+3)}{2}+2$ points in the first phase are used to estimate the parameters in the second order model. The optimum design point is given by (9) and the predicted response at the optimum design point is also given by (10) and are in terms of the natural variables. However, if the estimates of the parameters are also of interests, the number of points used in the parameter estimation should be increased to the same as in the case of CCD second order model.

4. Simulation Study

Five published test functions are used in the simulation to compare the efficiency between the classical RSM with CCD and the proposed method. The first four test functions,

$f_i(x_1, x_2),\ i = 1, 2, 3, 4$ are published in Fan and Zahara (2004) and the last one, $f_5(x_1, x_2)$, in Myers and Montgomery (2002). The responses and the contours of the five test functions are illustrated in Figures 3-7 respectively with the corresponding the minimum points and the minimum response shown in Table 2.

$$f_1(x_1, x_2) = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1) + 2(2 - x_2)^2 + 7\sin\left(\frac{x_1}{2}\right)\sin\left(\frac{7x_1x_2}{10}\right); x_1, x_2 \in [1, 4]$$

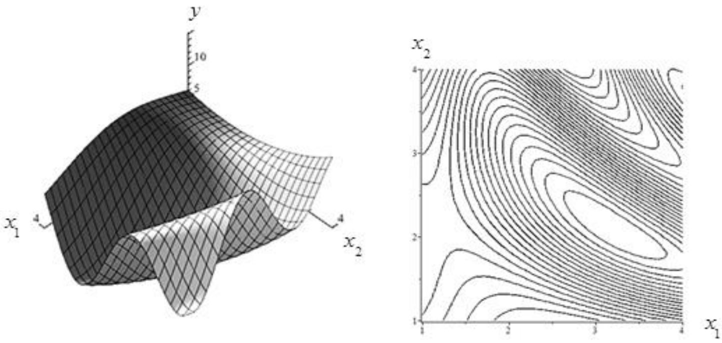


Figure 3 The response surface and the contour of f_1

$$f_2(x_1, x_2) = -(x_1^2 + x_2 - 11)^2 - (x_1 + x_2^2 - 7)^2; x_1, x_2 \in [-2, 2]$$

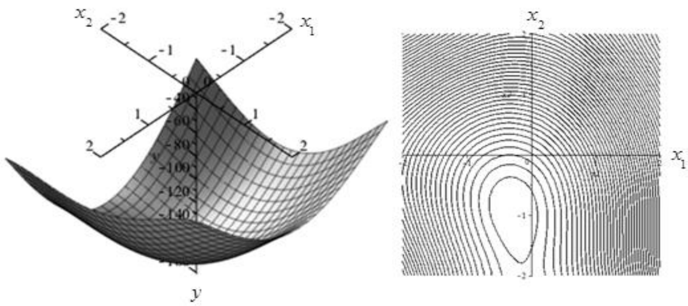


Figure 4 The response surface and the contour of f_2

$$f_3(x_1, x_2) = 4x_1^2 - 2.1x_1^4 + \frac{1}{3}x_1^6 + x_1x_2 - 4x_2^2 + 4x_2^4; x_1 \in [-1, 0.5], x_2 \in [0, 1]$$

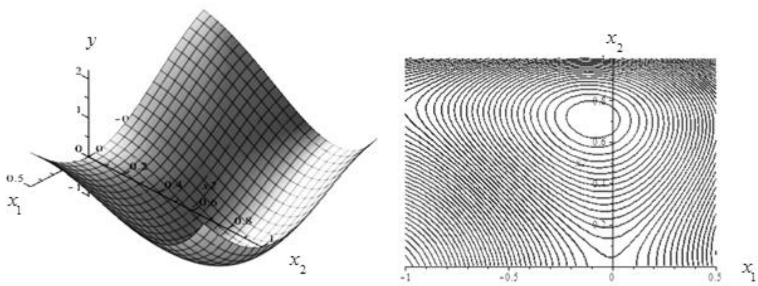


Figure 5 The response surface and the contour of f_3

$$f_4(x_1, x_2) = x_1\sin(4x_1) + 1.1x_2\sin(2x_2); x_1, x_2 \in [1.5, 3.5]$$

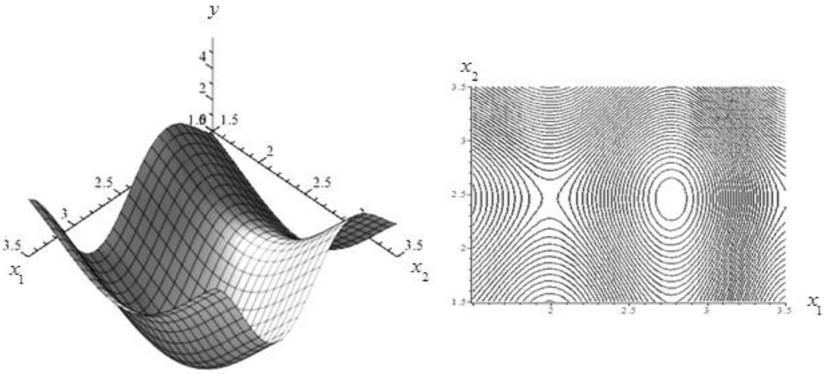


Figure 6 The response surface and the contour of f_4

$$f_5(x_1,x_2)=1431-7.81x_1-13.3x_2+0.0551x_1^2+0.0401x_2^2-0.01x_1x_2;\ x_1\in[50,120],\ x_2\in[150,200]$$

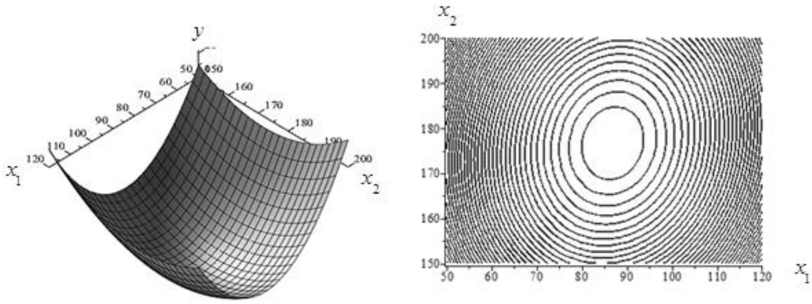


Figure 7 The response surface and the contour of f_5

Table 2 The minima and the minimum responses of the test functions

Test function	Minimum points		Minimum response
	$x_{\min_1}$	$x_{\min_2}$	$y_{\min}$
f_1	3.200	2.100	-6.510
f_2	-0.270	-0.920	-181.600
f_3	-0.092	0.713	-1.032
f_4	2.770	2.460	-5.410
f_5	86.900	176.670	-83.220

In simulation, the response is evaluated at a specific point by using a stochastic function which is the sum of one of the five test functions and a normal random generator. The random generators associated with the corresponding test functions are $\varepsilon_1 \sim N(0,0.01^2)$, $\varepsilon_2 \sim N(0,1)$, $\varepsilon_3 \sim N(0,0.01^2)$, $\varepsilon_4 \sim N(0,0.01^2)$ and $\varepsilon_5 \sim N(0,1)$. The variances of the random generators are specified by considering the corresponding ranges of the independent variables of the function.

4.1. RSM with CCD simulation

For each response given by the test function and the associated random generator, one hundred simulation runs are conducted, each of which consists of 50 replications. The natural variable, x_i, is generated by the uniform generator $U(a_i, b_i)$ where a_i and b_i are the lower bound and upper bound of x_i. Let L_{ij} be the low level of x_i in the j^{th} replication of a simulation run generated by the uniform generator $U[a_i, (a_i + b_i)/2]$ for $i = 1, 2;$ $j = 1, 2, \ldots, 50.$ The high level of x_i in the j^{th} replication H_{ij} is given by $L_{ij} + \delta_{ij}$ where δ_{ij} is generated by the uniform generator $U(0, b_i - (a_i + b_i)/2),$ $i = 1, 2;$ $j = 1, 2, \ldots, 50.$ The low level and high level of the natural variables are converted to the coded variables $L(-)$ and $H(+)$ respectively for analysis in the factorial design in the CCD.

The center point in a 2^2 factorial design, C_{ij} is given by $(L_{ij} + H_{ij})/2$ for $i = 1, 2,$ $j = 1, 2, \ldots, 50.$ Initially, the experiment is performed at four axial points and five center points for statistical testing of the curvature existence. The simulation follows the procedure as described in Section 2 until the curvature is detected. The experiment is performed at five center points and four axial points and four additional axial points for constructing the second order model. The axial point in the simulation is determined by using $\alpha = (\text{number of factorial runs})^{1/4} = 1.414$ for a 2^2 factorial design. After the parameters of the second order model have been estimated by the ordinary least squares (OLS) method, the minimum design point in terms of coded variables is calculated by (9) and the minimum response is given by (10). Then, the coded minimum design point and the coded minimum response are converted back to in terms of the natural variables. After 50 replications have been completed, the 95% confidence intervals (CI) for the mean of minimum design point and the mean of the minimum response can be estimated.

For each of the five stochastic test functions, the simulation runs 100 times for statistical analysis of the simulation results.

4.2. Simulation by the proposed method

Generally, the proposed method may start with any $k+1$ vertices in a 2^k factorial design selected from the vertices in the initial experiment in the traditional RSM with CCD for the sake of performance comparisons. However, in this simulation of a 2^2 factorial design, all four possible set of three vertices: NM(1), NM(2), NM(3) and NM(4) as shown in Figure 2 are used as the starting points of the simulation by the proposed method for analysis of the performances of the proposed method. The stochastic test functions, the number of replications in a run and the number of runs are the same as in the RSM simulation. The acceptable tolerance e in the stopping criterion of the NM algorithm is set to be a reasonable small value of 0.10 since the NM algorithm converges to a minimum provided that no shrink operation occurs during the search. Furthermore, the objective of NM search is to locate the neighborhood of the minimum for constructing the second order model in the second phase.

The same statistics of the minimum design point and the minimum response are calculated as in the CCD simulation.

4.3. Analysis of the simulation results

In each run of both simulations under one of the five test functions, the means of the absolute percentage errors (MAPE) of the natural variables at the estimated minimum design point and of the estimated minimum response are given by

$$\text{MAPE}(\hat{x}_{\min_1}) = \frac{1}{50}\sum_{j=1}^{50}\left|\frac{\hat{x}_{\min_{1j}} - x_{\min_1}}{x_{\min_1}}\right| \times 100; \ \text{MAPE}(\hat{x}_{\min_2}) = \frac{1}{50}\sum_{j=1}^{50}\left|\frac{\hat{x}_{\min_{2j}} - x_{\min_2}}{x_{\min_2}}\right| \times 100;$$

$$\text{MAPE}(\hat{y}_{\min}) = \frac{1}{50}\sum_{j=1}^{50}\left|\frac{\hat{y}_{\min} - y_{\min}}{y_{\min}}\right| \times 100.$$

The means and standard deviations of the estimated minimum design points and the estimated minimum responses from the 100 runs, each of which has 50 replications in both simulations under each of the five test functions are summarized in Table 3 and the averages of the MAPEs of the estimated minimum design point and the estimated minimum response are shown in Table 4 and 5 respectively.

Table 3 The true value, mean and standard deviation of the estimated values of the minimum design points and the minimum responses in the simulation by the CCD and the proposed method

Test function	Variables	True value	Methods				
			CCD	NM(1)	NM(2)	NM(3)	NM(4)
f_1	x_1	3.200	3.085	3.210	3.198	3.194	3.195
			(0.049)	(0.036)	(0.019)	(0.018)	(0.029)
	x_2	2.100	2.208	2.113	2.103	2.120	2.119
			(0.034)	(0.023)	(0.016)	(0.026)	(0.031)
	y	-6.510	-4.089	-6.396	-6.504	-6.493	-6.519
			(0.352)	(0.168)	(0.044)	(0.052)	(0.017)
f_2	x_1	-0.270	-0.383	-0.284	-0.290	-0.282	-0.280
			(0.070)	(0.009)	(0.005)	(0.008)	(0.010)
	x_2	-0.920	-0.993	-0.955	-0.945	-0.940	-0.945
			(0.099)	(0.034)	(0.039)	(0.050)	(0.045)
	y	-181.600	-178.537	-181.603	-181.243	-181.255	-181.645
			(1.140)	(0.424)	(0.657)	(0.453)	(0.447)
f_3	x_1	-0.092	-0.032	-0.096	-0.097	-0.095	-0.096
			(0.015)	(0.005)	(0.006)	(0.003)	(0.006)
	x_2	0.713	0.334	0.716	0.710	0.709	0.708
			(0.094)	(0.021)	(0.012)	(0.013)	(0.019)
	y	-1.032	-0.514	-1.038	-1.038	-1.017	-1.017
			(0.102)	(0.050)	(0.051)	(0.022)	(0.016)

Table 3 (Continued)

Test function	Variables	True value	Methods				
			CCD	NM(1)	NM(2)	NM(3)	NM(4)
f_4	x_1	2.770	2.747	2.746	2.776	2.758	2.767
			(0.030)	(0.025)	(0.018)	(0.024)	(0.018)
	x_2	2.460	2.376	2.454	2.455	2.455	2.456
			(0.045)	(0.025)	(0.029)	(0.024)	(0.009)
	y	-5.410	-4.828	-5.272	-5.442	-5.413	-5.414
			(0.171)	(0.117)	(0.034)	(0.006)	(0.008)
f_5	x_1	86.900	86.816	86.572	86.818	86.650	86.801
			(0.044)	(0.477)	(0.436)	(0.483)	(0.463)
	x_2	176.670	176.655	176.628	176.624	176.470	176.613
			(0.055)	(0.666)	(0.588)	(0.815)	(0.582)
	y	-83.220	-82.353	-83.131	-83.152	-83.119	-83.234
			(0.112)	(0.184)	(0.196)	(0.218)	(0.030)

Note: The number in the parenthesis is the standard deviation of the estimated mean.

From Table 3, it is obvious that the estimated minimum responses and the minimum points of the five test functions given by the proposed NM method from all four possible starting points are closer to the true values than the ones given by the CCD.

It can be seen from the averages of the MAPEs of the estimated minimum design point, $\hat{\mathbf{x}}_{\min}$, and the estimated minimum response, $\hat{y}_{\min}$, in Tables 4 and 5 that the proposed method with any one of the four possible starting vertices clearly outperforms the CCD when the stochastic test functions are f_i, $i = 1, 2, 3$ and 4 which are not quadratic functions. But in the case of a quadratic function f_5 the averages of the MAPES of $\hat{x}_{\min_1}$ and $\hat{x}_{\min_2}$ in the proposed method are 0.415-0.514% and 0.253-0.340% while those in the CCD are 0.097% and 0.025% respectively. However, in the case of quadratic function f_5, the averages of the MAPEs of the estimated minimum response in the proposed method are 0.026-0.211% less than 1.042% in the CCD. This leads to the conclusion that the proposed method using any one of the possible four vertices as the starting vertices yields the better estimates than the traditional RSM with CCD since most of the responses in practice are not quadratic.

Table 4 The average MAPEs of the estimated minimum design point in the simulation by the CCD and the proposed method

Test function	Average MAPE of $\hat{x}_{\min_1}$, %					Average MAPE of $\hat{x}_{\min_2}$, %				
	CCD	NM(1)	NM(2)	NM(3)	NM(4)	CCD	NM(1)	NM(2)	NM(3)	NM(4)
f_1	3.697	0.866	0.451	0.481	0.687	5.361	0.936	0.488	1.267	1.423
f_2	40.779	5.188	7.386	4.643	4.325	10.237	4.312	4.155	4.900	4.581
f_3	64.970	5.946	6.541	4.139	6.237	53.166	2.512	1.522	1.643	2.413
f_4	1.281	0.903	0.579	0.763	0.557	3.296	0.832	0.338	0.808	0.331
f_5	0.097	0.514	0.415	0.485	0.440	0.025	0.281	0.257	0.340	0.253

Table 5 The average MAPEs of the estimated minimum response in the simulation by the CCD and the proposed method

Test function	Average MAPE of $\hat{y}_{\min}$, %				
	CCD	NM(1)	NM(2)	NM(3)	NM(4)
f_1	36.822	1.879	0.382	0.693	0.239
f_2	1.805	0.198	0.345	0.256	0.211
f_3	50.119	4.209	4.414	1.957	1.624
f_4	12.409	0.909	0.717	0.104	0.134
f_5	1.042	0.198	0.189	0.211	0.026

The average number of points performed before estimating the minimum design points in the CCD and the proposed method from the 100 runs of simulation are shown in Table 6. The number of points in the proposed method is 80.85-92.77%, 55.67-60.62%, 68.25-75.52%, and 66.01-71.07%, of the corresponding one in the CCD for of the non-quadratic test functions, f_i, $i = 1, 2, 3, 4,$ respectively. But in the case of quadratic test function f_5, the number of required experiments in the proposed method is 6.85-12.38% more than the CCD.

Table 6 The average number of required points in the CCD and the proposed method

Test function	Average number of required points in				
	CCD	NM(1)	NM(2)	NM(3)	NM(4)
f_1	13.00	12.06(92.77)	11.95(91.92)	10.97(84.38)	10.51(80.85)
f_2	19.83	11.04(55.67)	12.02(60.62)	11.12(56.08)	11.51(58.04)
f_3	13.89	9.48(68.25)	10.49(75.52)	10.41(74.95)	10.36(74.59)
f_4	14.83	9.79(66.01)	10.54(71.07)	9.91(66.82)	10.46(70.53)
f_5	13.00	14.19(109.15)	14.61(112.38)	14.23(109.46)	13.89(106.85)

Note: The number in the parenthesis is the percentage of the required points relative to the one in the CCD.

Let define the coverage probability be the probability that the true value is in the 95% confidence interval of the estimate. The coverage probabilities in the CCD and the proposed method in the cases of the stochastic functions, $f_i,\ i=1,2,\ldots,5$ are summarized in Table 7. It can be seen that the coverage probabilities of the 95% CI of the estimates in the proposed method are higher than in the traditional RSM except the coverage probability of $\hat{x}_{\min_2}$ in the case of quadratic function f_5 which equals 0.94 in the CCD higher than 0.77-0.84 in the proposed method, and the coverage probability of $\hat{x}_{\min_1}$ in the case of f_4 which equals 0.88 in the CCD higher than 0.82 and 0.76 in the proposed method with the starting vertices NM(2) and NM(4) respectively. However, from the coverage probability of the estimated minimum design point, it can be concluded that the proposed method performs much better than the CCD.

5. Conclusions

In this study, a novel method using the NM algorithm with the starting $k+1$ vertices in the k dimensional space to move the experiment toward the neighborhood of the optimum is proposed to be used instead of the first order model in the traditional RMS. A second order model is constructed based on the last $\frac{k(k+3)}{2}+2$ points before the termination of the NM algorithm. The response in the simulation is given by a stochastic test function which is a combination of the published test function and a normal generator. From the simulation results, the average MAPE of $\hat{y}_{\min}$ in the proposed method is less than the corresponding values in the CCD for five stochastic test functions as shown in Tables 5. Also, the coverage probabilities in the proposed method are higher than the corresponding values in the CCD except the case x_2 in the quadratic stochastic test function f_5. The average MAPEs of $\hat{x}_{\min_1}$ and $\hat{x}_{\min_2}$ in the proposed method and the number of required experiments are clearly less than the ones in the CCD except in the case of the quadratic stochastic test function f_5 for which the CCD slightly outperforms the proposed method as shown in Tables 4 and 6. It can be concluded from the simulation that the proposed method may randomly select $k+1$ vertices from 2^k vertices in a 2^k factorial design for the initial experiment and outperforms the CCD for a non-quadratic response from the points of view of the minimum design point within the smaller neighborhood of the true minimum, the smaller number of points required performing the experiment and the higher coverage probabilities of the estimates. However, in the case of a quadratic response, the CCD slightly outperforms the proposed method in terms of the number of points required performing the experiment and the MAPE of the estimate of the minimum design point. One should be noted that the less the number of points required performing the experiment, the less the experiment cost and the time taken.

Table 7 The coverage probabilities in the CCD and the proposed method under the stochastic functions, $f_i,\ i = 1, 2, \ldots, 5$

Test function	Method	Coverage probability of			
		x_1	x_2	Both $x_1,\ x_2$	Response(y)
f_1	CCD	0.27	0.01	0.00	0.00
	NM(1)	0.96	0.98	0.95	0.93
	NM(2)	0.99	0.99	0.98	0.85
	NM(3)	0.98	0.96	0.94	0.84
	NM(4)	0.88	0.93	0.87	0.98
f_2	CCD	0.60	0.65	0.28	0.16
	NM(1)	0.96	0.91	0.89	0.84
	NM(2)	0.89	0.86	0.84	0.78
	NM(3)	0.98	0.96	0.94	0.80
	NM(4)	0.95	0.91	0.89	0.82
f_3	CCD	0.08	0.08	0.04	0.01
	NM(1)	0.94	0.89	0.83	0.89
	NM(2)	0.97	0.87	0.85	0.88
	NM(3)	0.97	0.95	0.95	0.95
	NM(4)	0.96	0.94	0.91	0.97
f_4	CCD	0.88	0.46	0.37	0.04
	NM(1)	0.99	0.84	0.83	0.93
	NM(2)	0.82	0.79	0.73	0.90
	NM(3)	0.88	0.83	0.74	0.88
	NM(4)	0.76	0.84	0.72	0.82
f_5	CCD	0.55	0.94	0.53	0.01
	NM(1)	0.95	0.83	0.78	0.94
	NM(2)	0.93	0.84	0.78	0.95
	NM(3)	0.92	0.77	0.69	0.91
	NM(4)	0.97	0.81	0.79	0.89

From the simulation results, it can be observed that in the case of non-quadratic response, the first order model may not move the experiment sufficiently close to the minimum even though a curvature is statistically detected. Consequently, the second order model in the CCD cannot satisfactorily fit the curvature of the non-quadratic response. But, in the case of quadratic response, the second order model in the CCD will give a better result. Since the objective of this study is to locate the minimum design point and the minimum response the degree of freedom of the error in the proposed method in the simulation is only 1. If the estimates of parameters in the second order model is of interests the number of points used in the construction of the second order model in the proposed method should be increased to in the same order in the CCD.

For the future research, other direct search, such as Tabu search, pattern search, Powell's method, etc., should be explored to search the optimum design point and different designs in the RSM also should be investigated.

References

Anderson C, Borror CM, Montgomery DC. Response surface design evaluation and comparison. J. Stat. Plann. Infer. 2009; 139: 629-674.

Box GEP, Wilson KB. On the experimental attainment of optimum conditions. J. Roy. Stat. Soc. B Stat. Meth.1951; 13: 1-45.

Box GEP, Draper NR. Empirical model-building and response surfaces. New York: Wiley; 1987.

Brooks SP, Morgan BJT. Optimization using simulated annealing. Statistician. 1995; 44: 241-257.

Byatt D. Convergent variants of the Nelder-Mead algorithm. MS [dissertation]. Christchurch: New Zealand University of Canterbury; 2000.

Copeland KAF, Nelson PR. Dual response optimization via direct function minimization. J. Qual. Tech. 1996; 28: 331-336.

Dennis JE, Torczon V. Direct search methods on parallel machines. SIAM J. Optim. 1991; 1: 448-474.

Fan SKS, Zahara E. Simulation optimization using an enhanced Nelder-Mead simplex search algorithm. Proceedings of the Fifth Asia Pacific Industrial Engineering and Management Systems Conference, 2004.

Gavin HP. The Nelder-Mead algorithm in two dimensions. CEE 201L. Duke University. 2016.

Glover F, Laguna M. Tabu search. Boston: Kluwer Academic Publishers; 1997.

Han L, Neumann M. Effect of dimensionality on the Nelder-Mead simplex method. Optim. Methods Software. 2006; 21: 1-16.

Holland JH. Genetic algorithms. Sci. Am. 1992; 267: 66-72.

Kelley CT. Detection and remediation of stagnation in the Nelder-Mead algorithm using a sufficient decrease condition. SIAM J. Optim. 2000; 10: 43-55.

Khuri AI, Cornell JA. Response surfaces: Design and analysis. New York: Marcel Dekker; 1996.

Kolda TG, Lewis RM, Torczon V. Optimization by direct search: New perspectives on some classical and modern methods. SIAM J. Rev. 2003; 45: 385-482.

Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence properties of the Nelder-Mead simplex method in low dimensions. SIAM J. Optim. 1998; 9: 112-147.

McKinnon KIM. Convergence of the Nelder-Mead simplex method to a nonstationary point. SIAM J. Optim. 1998; 9: 148-158.

Montgomery DC. Introduction to statistical quality control. New York: Wiley; 2005.

Myers RH, Montgomery DC. Response surface methodology: Process and product optimization using designed experiments. New York: Wiley; 2002.

Nelder JA, Mead R. A simplex method for function minimization. Comput. J. 1965; **7**: 308-313.

Olsson DM. A sequential simplex program for solving minimization problems. J. Qual. Tech. 1974; 6: 53-57.

Olsson DM, Nelson LS. The Nelder-Mead simplex procedure for function minimization. Technometrics. 1975; 17: 45-51.

Parkinson JM, Hutchinson D. An investigation into the efficiency of the variants on the simplex method. In: F.A. Lootsma (Ed.) Numerical methods for non-linear optimization. London and New York: Academic Press; 1972.

Price CJ, Coope ID, Byatt D. A convergent variant of the Nelder-Mead algorithm. J. Optim. Theory Appl. 2002; 113: 5-19.

Sexton RS, Alidaee B, Dorsey RE, Johnson JD. Global optimization for artificial neural networks: A Tabu search application. Eur. J. Oper. Res. 1998; 106: 570-584.

Torczon V. Multi-directional search: A direct search algorithm for parallel machines. PhD [dissertation]. Texas: Rice University; 1989.

Torczon V. On the convergence of pattern search methods. SIAM J. Optim. 1997; 7: 1-25.

Tseng P. Fortified-descent simplicial search method: A general approach. SIAM J. Optim. 2000; 10: 269-288.

Wright MH. Direct search methods: Once scorned, now respectable. In: D.F. Griffiths and G.A. Watson (Eds.) Numerical Analysis 1995: Proceedings of the 1995 Dundee Biennial Conference in Numerical Analysis (Harlow, UK: Addison Wesley Longman. 1996; 191-208).