TILBURG UNIVERSITY

Response Surface Methodology's Steepest Ascent and Step Size Revisited

Kleijnen, J.P.C.; den Hertog, D.; Angun, M.E.

Publication date:
2002

Link to publication

Citation for published version (APA):
Kleijnen, J. P. C., den Hertog, D., & Angun, M. E. (2002). *Response Surface Methodology's Steepest Ascent and Step Size Revisited*. (CentER Discussion Paper; Vol. 2002-64). Operations research.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright, please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 08. mrt. 2021



No. 2002-64

RESPONSE SURFACE METHODOLOGY'S STEEPEST ASCENT AND STEP SIZE REVISITED

By Jack P.C. Kleijnen, Dick den Hertog, Ebru Angün

July 2002

ISSN 0924-7815

Discussion paper

D:\Data\WP\PAPERS\Angun\RsmCDP.wpd

Written: June 12, 2002

Printed: June 25, 2002 (12:50pm)

Response Surface Methodology's Steepest Ascent and Step Size Revisited

Jack P.C. Kleijnen [a], Dick den Hertog [b], and Ebru Angün [c]

[a] Corresponding author:

Department of Information Systems/Center for Economic Research (CentER)

Tilburg University (KUB), Postbox 90153, 5000 LE Tilburg, Netherlands

Phone: +3113-4662029; Fax: +3113-4663377; E-mail: kleijnen@kub.nl

Web: http://www.tilburguniversity.nl/faculties/few/im/staff/kleijnen/

[b] Department of Econometrics and Operations Research/CentER

Phone: +3113-4662122; Fax: +3113-4663280; E-mail: D.denHertog@kub.nl

[c] Department of Information Systems/CentER

Phone: +3113-4668202; Fax: +3113-4663377; E-mail: M.E.Angun@kub.nl

Abstract

Response Surface Methodology (RSM) searches for the input combination maximizing the output of a real system or its simulation. RSM is a heuristic that locally fits first-order polynomials, and estimates the corresponding steepest ascent (SA) paths. However, SA is scale-dependent; and its step size is selected intuitively. To tackle these two problems, this paper derives novel techniques combining mathematical statistics and mathematical programming. Technique 1 called 'adapted' SA (ASA) accounts for the covariances between the components of the estimated local gradient. ASA is scale-independent. The step-size problem is solved tentatively. Technique 2 does follow the SA direction, but with a step size inspired by ASA. Mathematical properties of the two techniques are derived and interpreted; numerical examples illustrate these properties. The search directions of the two techniques are explored in Monte Carlo experiments. These experiments show that - in general - ASA gives a better search direction than SA.

Keywords: Heuristics; Metaheuristics; RSM; Statistical analysis; Scale-dependence

JEL: C1

1. Introduction

Response Surface Methodology (RSM) is a stagewise heuristic that searches for the input combination maximizing the output (finding the minimum is equivalent to finding the maximum of 'minus' the output; the maximization problem is without explicit constraints or

side-conditions). Originally, Box and Wilson (1951) meant RSM for experiments with real (non-simulated) systems; see Box (1999), Khuri (1996), Khuri and Cornell (1996), Myers (1999), Myers and Montgomery (1995).

Later on, RSM was also applied to random simulation models; see Donohue, Houck, and Myers (1993, 1995), Hood and Welch (1993), Irizarry, Wilson, and Trevino (2001), Kleijnen (1998), Law and Kelton (2000, pp. 646-655), Neddermeijer et al. (2000), Safizadeh (2002), etc.

Note: (Hasty readers may skip paragraphs starting with '*Note*:'.)We assume that these random simulations are so complicated that methods such as stochastic approximation and sample path optimization, cannot be applied. RSM treats simulation models - either random or deterministic - and real systems as black boxes. Other black box methods are meta-heuristics, including tabu search, simulated annealing, and genetic algorithms. The black box approach is discussed by Jones et al. (1998). Various simulation optimization methods are presented by Fu (2002).

In this paper, we do not explain all stages of RSM, but refer to the literature cited above. We focus on the first stages, in which RSM fits a first-order polynomial in the inputs, per local area. This fitting uses Ordinary Least Squares (OLS), and estimates the steepest-ascent path, as follows. Let x_j denote the value of the original (non-standardized) input j with $j = 1, ..., k$. Hence k main or first-order effects (say) β_j are estimated. To enable this estimation, RSM uses a Resolution-3 design, which specifies which $n \approx k + 1$ input combinations are to be observed (simulated). These n input/output (I/O) combinations give the estimates $\hat{\beta}_j$. So the *SA path* implies $\Delta x_j / \Delta x_1 = \hat{\beta}_j / \hat{\beta}_1$; in other words, SA uses the local gradient, $\boldsymbol{\beta}' = (\hat{\beta}_1, ..., \hat{\beta}_k)'$.

Unfortunately, SA suffers from two well-known problems; see Myers and Montgomery (1995, pp. 192-194): (i) it is scale-dependent; (ii) the step size along its path is selected intuitively. In practice, analysts try an intuitively selected value for the step size. If that value yields a lower response, then they reduce the step size. Otherwise, they take one more step. An example is the case study in Kleijnen (1993), which uses a step size that doubles the most important input.

Note: Some disciplines interpret RSM completely differently: RSM becomes a *one-shot* approach that fits a single response surface - either a second-order polynomial or a Kriging model - to the I/O data of a random or deterministic simulation model, over the *whole* experimental area (instead of a series of local areas). Next, that single model is used to estimate the optimal input combination. See Sacks et al. (1989), and also Jones et al. (1999), Simpson et al. (2001), etc.

Our *research contribution* is the following. We derive *ASA*; that is, we adjust the estimated first-order factor effects through their estimated covariance matrix. We prove that ASA is scale independent. In most of our experiments, ASA gives a better search direction. We also propose and explore a solution for the step size.

The remainder of this article is organized as follows. §2 summarizes those parts of linear regression analysis that we need to formulate our two techniques. §3 derives these techniques, and their mathematical properties and interpretation. To get further insight into these properties, §4 applies the step-size selection of ASA to simple numerical problems. §5 compares the search directions of the two techniques, by means of Monte Carlo experiments. §6 gives conclusions and future research plans.

2. Linear-regression basics

We define the *estimated signal/noise ratio* (say) $\hat{\gamma}_j$ as

$$\hat{\gamma}_j = \frac{\hat{\beta}_j}{\sqrt{\hat{var}(\hat{\beta}_j)}} \quad (j = 1, ..., k) \tag{1}$$

where $\hat{\beta}_j$ denotes the estimate of β_j in the following local first-order polynomial approximation:

$$y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + e \tag{2}$$

where y denotes the *regression predictor* of the corresponding expected output; e denotes *white noise*; that is, e is normally, identically, and independently distributed (NIID) with zero mean μ_e and constant variance σ_e^2.

Note: Each of the k ratios in (1) equals Student's statistic t under the *null-hypothesis* of zero input effect, H_0: $\beta_j = 0$. However, this does not mean that we propose to test this H_0. Actually, if we use OLS to estimate $\beta = (\beta_0, \beta_1, ..., \beta_k)'$, then $\hat{\beta}$ is the Best Linear Unbiased Estimator (BLUE). Testing H_0: $\beta_j = 0$ makes sense only if we have good reasons to postulate such a hypothesis. But in our case, effects that are not significant in a certain stage, may still be practically important in that stage - or in later stages! And unimportant factors may be significant if the signal/noise ratio is high: in applications of RSM to random simulation the simulation may have small 'intrinsic' noise, or very many simulation runs may be executed if computer time per run is small; in deterministic simulation the intrinsic noise is zero, by definition.

The OLS estimator of the regression parameters β is

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{w} \tag{3}$$

with (in order of appearance)

$\hat{\boldsymbol{\beta}}$: vector with the q estimated effects in the regression model ($q = 1 + k$ in equation 2)

q: number of regression effects including the intercept β_0

$\boldsymbol{X}$: $N \times q$ matrix of explanatory (independent) regression variables including the 'dummy' variable x_0 with constant value 1; $\boldsymbol{X}$ is assumed to have linearly independent columns (so $\boldsymbol{X}$ has full column rank)

$N = \sum_{i=1}^{n} m_i$: number of observations (simulation runs)

m_i: number of replicates at input combination (point) i, with $m_i \in \mathbb{N} \wedge m_i > 0$

n: number of different observed combinations of the k inputs, with $n \in \mathbb{N} \wedge n \geq q$ (necessary condition for avoiding singularity in equation 3)

$\boldsymbol{w}$: vector with N outputs (real or simulated) corresponding to the N inputs.

Note: Obviously, the first m_1 rows of $\boldsymbol{X}$ are identical and equal to $(1, x_{1;1},\ldots, x_{1;k})$, ..., the last m_n rows of $\boldsymbol{X}$ are identical and equal to $(1, x_{n;1},\ldots, x_{n;k})$. A simulation example is a single-server simulation with $\boldsymbol{w}$ in (3) denoting the vector of average waiting times per replicate, and x_1 in (2) denoting the traffic rate.

Because of (1) through (3), we call $\hat{\beta}_j$ the estimated signal of input j. The *signal's noise* (see equation 1's denominator) is the square root of the corresponding element on the main diagonal of

$$\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\sigma_e^2 \ . \tag{4}$$

For example, (2) implies that $\boldsymbol{var}(\hat{\beta}_1)$ is the second element on the main diagonal of (4).

Equation (4) leads to the estimated noise: replace the unknown parameter σ_e^2 in (4) by the *mean squared residual* (MSR) estimator

$$\hat{\sigma}_e^2 = \frac{\sum_{i=1}^{n} \sum_{r=1}^{m_i} (w_{i;r} - \hat{y}_i)^2}{(N - q)} \tag{5}$$

where $w_{i;r}$ denotes the output for input combination *i* and replicate *r*, and $\hat{y}_i$ denotes the OLS regression predictor for the simulation's input combination *i* that follows from

$$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} \ . \tag{6}$$

For example, (6) combined with (2) gives $\hat{y}_1 = \hat{\beta}_0 + \sum_{j=1}^{k} \hat{\beta}_j \, x_{1;j}$.

The variance of this predictor is a function of $\boldsymbol{x}$ (input combination for which estimate is computed), since elementary regression analysis gives

$$var(\hat{y} | \ \boldsymbol{x}) = \boldsymbol{x}' \ cov(\hat{\boldsymbol{\beta}}) \ \boldsymbol{x} \ . \tag{7}$$

Notice that $\boldsymbol{x}$ in (7) may correspond with either one of the actually observed input combinations $\boldsymbol{X}$ - as in (6) - or a new point. A new point means *interpolation* or *extrapolation.*

To illustrate the implications of (7), suppose that $\boldsymbol{X}$ is *orthogonal*; that is, $\boldsymbol{X}'\boldsymbol{X} = N\boldsymbol{I}_{q \times q}$. Combining (4) and (7) then gives

$$var(\hat{y} | \ \boldsymbol{x}) = \boldsymbol{x}'\boldsymbol{x} \ \sigma_e^2/N \ . \tag{8}$$

Obviously, the regression predictor becomes less reliable, as the number of observations *N* decreases. Likewise, the predictor gets inaccurate, as the noise σ_e^2 increases (for example, a single-server simulation implies that a higher traffic rate not only increases the mean but also

the variance of waiting times so the intrinsic simulation noise increases and so does the white noise *e*). But what is the effect of $\boldsymbol{x}$, the point that we wish to predict?

In Appendix 1 we derive the design point that *minimizes* the variance of the regression predictor, (say) $\boldsymbol{d_o}$. Note that $\boldsymbol{x}' = (1, \boldsymbol{d}')$ where the element 1 corresponds with the intercept β_0. We find $\boldsymbol{d_o} = -\boldsymbol{C}^{-1}\boldsymbol{b}$ where $\sigma_e^2\boldsymbol{C}$ is the covariance matrix of $\hat{\boldsymbol{\beta}}_{-0}$ which equals $\hat{\boldsymbol{\beta}}$ excluding the intercept $\hat{\beta}_0$ (also see equation 4):

$$\boldsymbol{cov}(\hat{\boldsymbol{\beta}}) = \sigma_e^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma_e^2\begin{pmatrix} a & \boldsymbol{b}' \\ \boldsymbol{b} & \boldsymbol{C} \end{pmatrix}$$

where $\boldsymbol{a}$ is a scalar, $\boldsymbol{b}$ a *k*-dimensional vector, and $\boldsymbol{C}$ a $k \times k$ matrix. Hence, if $\boldsymbol{X}$ is orthogonal, then (8) is minimal at the center of the experimental area: $\boldsymbol{d_o} = \boldsymbol{0}$ (also see the 'funnel' shape of Figure 1, discussed below). Hence, extrapolation should be less *trusted* as the extrapolated point moves farther away into regions not yet observed. This property will guide our ASA. (The term 'trust region' is used in nonlinear optimization; see Conn, Gould, and Toint 2000.)

3. Two new search techniques

We consider a lower (one-sided) 1 - α confidence interval for the predictor based on (2), given $\boldsymbol{x}$. This interval ranges from infinity down to

$$\hat{y}_{\min}(\boldsymbol{x}) = \hat{y}(\boldsymbol{x}) - t_{N-q}^{\alpha}\,\hat{\sigma}(\hat{y}, \boldsymbol{x}) = \boldsymbol{x}'\hat{\boldsymbol{\beta}} - t_{N-q}^{\alpha}\,\hat{\sigma}_e\sqrt{\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}} \qquad (9)$$

where t_{N-q}^{α} denotes the 1 - α quantile of the distribution of *t* with *N* - *q* degrees of freedom, and $\hat{\sigma}(\hat{y}, \boldsymbol{x})$ follows from the basic linear-regression formulas in (4) through (7). The first term in (9) concerns the *signal*, whereas the second term concerns the *noise*.

Note: When we consider a *set* of $\boldsymbol{x}$ values, then the set of intervals following from (9) has a joint (simultaneous) probability lower than $1 - \alpha$. This complication is ignored in our two techniques.

Technique 1 (ASA) finds the $\boldsymbol{x}$ that *maximizes* the *minimum* output predicted through (9), (say) $\boldsymbol{x}^+$. This $\boldsymbol{x}^+$ gives both a *search direction* and a *step size*. First we prove in Appendix 2 that the objective function in (9) is concave in $\boldsymbol{d}$. Next in Appendix 3 we derive the following explicit solution for the optimal input values of the next observation:

$$\boldsymbol{d}^+ = -\boldsymbol{C}^{-1}\boldsymbol{b} + \lambda \boldsymbol{C}^{-1}\hat{\boldsymbol{\beta}}_{-0} \tag{10a}$$

where $\boldsymbol{C}^{-1}\hat{\boldsymbol{\beta}}_{-0}$ is the *ASA direction*, and λ the *step size* specified by

$$\lambda = \sqrt{\frac{a - \boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}}{(t^{\alpha}_{N-q}\hat{\sigma}_e)^2 - \hat{\boldsymbol{\beta}}'_{-0}\boldsymbol{C}^{-1}\hat{\boldsymbol{\beta}}_{-0}}} \tag{10b}$$

where $\hat{\boldsymbol{\beta}}_{-0}$ is the vector of estimated first-order effects (so it excludes the intercept $\hat{\boldsymbol{\beta}}_0$).

Technique 2 still maximizes $\hat{y}_{\min}(\boldsymbol{x})$, but the new point is restricted to the *SA* path; that is, the *search direction* is specified by the estimated local gradient, $\hat{\boldsymbol{\beta}}_{-0}$. In Appendix 4 we derive the *optimal step size* (say) ζ^+ along this path:

$$\zeta^+ = \sqrt{\frac{a - \boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}}{[\frac{\hat{\boldsymbol{\beta}}'_{-0}\boldsymbol{C}\hat{\boldsymbol{\beta}}_{-0}}{\hat{\boldsymbol{\beta}}'_{-0}\hat{\boldsymbol{\beta}}_{-0}} t^{\alpha}_{N-q}\hat{\sigma}_e]^2 - \hat{\boldsymbol{\beta}}'_{-0}\boldsymbol{C}\hat{\boldsymbol{\beta}}_{-0}}} \tag{11}$$

This step size is unique, because (9) is concave (see Appendix 2).

We derive the following mathematical properties and interpretations of these two techniques.

The first term in (10a) means that the path on which the next observation is placed, *starts* from the point with *minimal* predictor variance, namely $-\boldsymbol{C}^{-1}\boldsymbol{b}$ (also see end of §2). The second term means that this path is in the *ASA* direction; that is, the classic SA direction $\hat{\boldsymbol{\beta}}_{-0}$ (second term's last factor) is adjusted for the covariance matrix of $\hat{\boldsymbol{\beta}}_{-0}$, which is $\boldsymbol{C}$ (see §2, last paragraph). Finally, the step size λ is quantified in (10b).

For the *orthogonal* case ($\boldsymbol{X}'\boldsymbol{X} = N\boldsymbol{I}_{q\times q}$) it is easy to verify that $\boldsymbol{a} = 1/N$, $\boldsymbol{b} = \boldsymbol{0}$, and $\boldsymbol{C} = \boldsymbol{I}/N$, so (10) reduces to

$$\boldsymbol{d}^{+} = \frac{1}{\sqrt{(t^{\alpha}_{N-q})^2\hat{\sigma}^2_e/N - \hat{\boldsymbol{\beta}}'_{-0}\hat{\boldsymbol{\beta}}_{-0}}}\,\hat{\boldsymbol{\beta}}_{-0}\,. \tag{12}$$

This solution implies *identical search directions* for ASA and SA, in case of orthogonality. Moreover, for the orthogonal case we prove that the two techniques *coincide* (both the search direction and the step size are the same), provided SA starts from the design center; see Appendix 4.

In practice, however, designs are not orthogonal. The classic textbooks on Design Of Experiments (DOE) and RSM do present many orthogonal designs (for example, 2^{k-p} designs), but these designs use *standardized* inputs (say) z_j, that is, inputs ranging between -1 and +1, with an average value of zero. In practice, we apply the following linear transformation to obtain original inputs x_j that range between L_j and H_j:

$$x_j = a_j + b_j z_j \text{ with } a_j = \frac{L_j + H_j}{2};\ b_j = \frac{L_j - H_j}{2} . \quad (13)$$

Consequently, the first-order polynomial regression model (2) implies that β_j and (say) β_j - the main effects of the original and standardized inputs respectively - are related as follows: $\beta_j = \beta_j / b_j$. Hence, the steepest-ascent path directions for the original and the standardized inputs differ (unless $\forall j$: $b_j = 1$). (The interpretation of standardization is controversial in mathematical statistics; see the many references in Kleijnen 1987, pp. 221, 345.)

We prove that ASA is *scale independent*; see Appendix 5. So ASA is not affected by switching from (say) inches to centimeters when measuring inputs. Driessen et al. (2001) prove that ASA is also independent of *linear transformations* with $a_j \neq 0$ in (13).

In case of *large signal/noise ratios* (defined in equation 1), the denominator under the square root in (10b) is negative so this equation does not give a finite solution for d^+; that is, (9) can be driven to infinity (unbounded solution). Indeed, if the noise is negligible, we have a *deterministic* problem, which our technique is not meant to address (many other researchers - including Conn et al. (2000) - study optimization of deterministic simulation models.)

In case of a *small signal/noise ratio*, no step is taken. Actually, we distinguish two cases: (i) the signal is small, (ii) the noise is big. These two cases are discussed next.

In case (i), the signal may be small because the first-order polynomial approximation is bad. Then we should switch to an alternative metamodel using transformations of d_j such as $\log(d_j)$ and $1/d_j$ (inexpensive alternative), a second-order polynomial, which adds d_j^2 and $d_j d_{j'}$ with $j' > j$ (expensive because many more observations are required to estimate the corresponding effects), etc.; see the RSM literature (for example, Irizarry et al. 2001).

In case (ii), however, the first-order polynomial may fit, but the intrinsic noise may be

high (also see the comment below equation 8). To decrease this noise, we should increase the number of observations, *N*; see the denominator in (8). Hence, we should increase either *n* or m_i (see the definitions below equation 3). When our technique gives a value d^+ that is 'close' to one of the old points, then in practice we may increase m_i. Otherwise we observe a new combination: we increase *n*. So our technique suggests an approach to the old problem of how to choose between either using the next observation to increase the accuracy of the current local approximation, or trusting that approximation and moving into a new area! A different approach is discussed in Kleijnen (1975, p. 360). In the literature on maximizing the output of deterministic simulation, this is called the *geometry improvement* problem; see Conn et al. (2000). More research on this problem is needed.

If we specify a different α value in t^{α}_{N-q}, then (10) gives a different step size (in the same direction). Obviously, t^{α}_{N-q} increases to infinity, as α decreases to zero. So, a sufficiently small α always gives a finite solution. However, if we *increase* α, then we make a bigger step. And we prefer to take a *bigger step*, in order to get quicker to the top of the response surface! We feel that a reasonable maximum α value is 0.20 (so we are '80% sure'); however, more empirical research is needed.

Note: We assume that the *noise* (defined in equation 2) has *zero mean* when deriving the $1-\alpha$ confidence interval in (9), which leads to the techniques in (10) and (11). Actually, the locally fitted first-order polynomials may show lack of fit so the expected value of $\hat{\sigma}^2_e$ (defined in equation 5) exceeds σ^2_e; see the lack-of-fit tests in many RSM textbooks. Fortunately, this bias has the 'right' sign; that is, this bias increases $\hat{\sigma}_e$ in (10) and (11) so that it decreases the step size.

4. Numerical examples of step-size selection in ASA

To obtain a better understanding of ASA - especially its step size - we apply this technique to the following three numerical examples:

(i) single input, and orthogonal $\boldsymbol{X}$;

(ii) two inputs, and orthogonal $\boldsymbol{X}$;

(iii) two inputs, and one-at-a-time $\boldsymbol{X}$.

For each example, we study several *cases*; that is, different signal/noise ratios. We suppose that the regression estimates happen to equal the true values: $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$ and $\hat{\sigma}_e = \sigma_e$. Without loss of generality we take $\beta_0 = 0$ and $\sigma_e = 1$ (σ_e and $\boldsymbol{X}$ determine the noise of $\hat{\boldsymbol{\beta}}$; see equation 4). We start with example (iii), which is most relevant for practice; then we summarize results for the other two examples.

We use a non-orthogonal design, namely a one-factor-at-a-time design with $\boldsymbol{d}'_1 = (-1, -1)$, $\boldsymbol{d}'_2 = (1, -1)$, and $\boldsymbol{d}'_3 = (-1, 1)$ so $n = 3$ (= q). To estimate σ_e^2 through the MSR in (5), we duplicate combination 1: $m_1 = 2$ so $N = 4$. We consider two extreme signal/noise cases.

Case 1: high/low signal/noise, $\hat{\gamma}_1 = 10$ and $\hat{\gamma}_2 = 0.10$

The given $\boldsymbol{X}$ and $\hat{\sigma}_e$ result in $\hat{\beta}_1 = 6.124$ and $\hat{\beta}_2 = 0.061$. Equation (10b) does not give a *finite* step size for the traditional α values 0.20, 0.10, and 0.05.

Case 2: low signal/noise, $\hat{\gamma}_1 = 0.3$ and $\hat{\gamma}_2 = 0.5$

This case implies $\hat{\beta}_1 = 0.184$ and $\hat{\beta}_2 = 0.306$. Then $\alpha = 0.20$ gives $(x_1^+, x_2^+) = (-0.404, -0.212)$; $\alpha = 0.05$ gives (-0.4804, -0.4416): no move *outside* the original input area (the N outputs were obtained for $-1 \leq x \leq 1$).

The other two examples can be summarized as follows.

Example (i), single input and orthogonal $\boldsymbol{X}$

Obviously, we now have $q = 2$. Suppose $\boldsymbol{x}'_1 = (1, -1)$ and $\boldsymbol{x}'_2 = (1, 1)$, so $n = 2$. Suppose $m_i = 2$, so $N = 4$. Then, (4) gives $\hat{var}(\hat{\beta}_1) = \hat{\sigma}_e^2/N = 1/4$. Then (12) reduces to

$$d_1^+ = sign(\hat{\beta}_1)\sqrt{\frac{\hat{\gamma}_1^2}{(t_2^\alpha)^2 - \hat{\gamma}_1^2}}\ , \qquad (14)$$

which gives a finite solution if $t_2^\alpha > \hat{\gamma}_1$.

Consider a case with high signal/noise: $\hat{\gamma}_1 = 10$; that is, $\hat{\beta}_1 = 5$. A finite solution results only for $\alpha \leq 0.0049$; for example, $\alpha = 0.001$ gives such a solution in Figure 1 (where $\alpha = 0.50$ or $t_2^{0.50} = 0$ corresponds with $\hat{y}_{min}$ itself).

INSERT Figure 1: High signal/noise case: lower 1 - α confidence intervals for the regression predictor $\hat{y}_{min}$ for different α

Example (ii), two inputs and orthogonal ***X***

A 2^k design gives $n = 4$; $m_i = 1$ implies $N = 4$. So (12) becomes

$$d_j^+ = sign(\hat{\beta}_j)\sqrt{\frac{\hat{\gamma}_j^2}{(t_1^\alpha)^2 - (\hat{\gamma}_1^2 + \hat{\gamma}_2^2)}} \text{ with } j = 1, 2\ . \qquad (15)$$

We consider a case with high/low signal/noise: $\hat{\gamma}_1 = 10$, $\hat{\gamma}_2 = 0.10$; that is, $\hat{\beta}_1 = 5$ and $\hat{\beta}_2 = 0.05$. Then neither $\alpha = 0.20$ nor $\alpha = 0.10$ give a finite solution. So we might split the inputs into two parts: a high signal/noise input 1 treated as in deterministic simulation, and a low signal/noise input 2. Further, $\alpha = 0.025$ does give a finite solution, namely $(d_1^+, d_2^+) = (1.2759, 0.0128)$; so the input with the highest signal changes almost 30%.

5. Comparison of the ASA and SA search directions through Monte Carlo experiments

We perform Monte Carlo experiments to compare the search directions of the two techniques, ASA and SA. The *Monte Carlo* method is an efficient and effective way to estimate the behavior of search techniques applied to random simulations (such as discrete-event dynamic systems, including simulated queuing and inventory systems); see Appendix 6.

We experiment with two inputs: $k = 2$. Our Monte Carlo experiments generate output w (defined below equation 3) through second-order polynomials in two inputs with white noise:

$$w = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{1;\,1} x_1^2 + \beta_{2;\,2} x_2^2 + \beta_{1;\,2} x_1 x_2 + e \ . \qquad (16)$$

RSM fits first-order polynomials defined in (2) *locally*, and then estimates the SA (see §1). The *global* experimental area is the area over which the inputs of the real system can be varied, or the area over which the simulation model is assumed to be valid. We assume that this area is the unit square: $-1 \le x_1 \le 1$ and $-1 \le x_2 \le 1$. In the local area we use a specific design $\boldsymbol{D}$, namely a one-at-a-time design (as in §4). The specific local area is the upper corner of Figure 2, discussed below.

There are infinitely many polynomials that satisfy (16). To study the scale dependence problem, we apply a *linear transformation* to the *canonical* case, as follows. We define the canonical case as $\beta_0 = \beta_1 = \beta_2 = \beta_{1;\,2} = 0$; $\beta_{1;\,1} = \beta_{2;\,2} = -1$; see Figure 2. Our linear transformation replaces x_2 by $1000x_2 + 1$, so $\beta_0 = -1$, $\beta_1 = 0$, $\beta_2 = -2000$, $\beta_{1;\,2} = 0$; $\beta_{1;\,1} = -1$, $\beta_{2;\,2} = -10^6$; that is, the contours form ellipsoids parallel to the two main axes. In the canonical case, our local starting area is the upper-right corner (1, 1) with 0.2 input ranges: $\boldsymbol{d}_1 = (1, 1)$, $\boldsymbol{d}_2 = (1, 0.8)$, and $\boldsymbol{d}_3 = (0.8, 1)$. So, now the local starting area becomes $\boldsymbol{d}_1 = (1, 0)$, $\boldsymbol{d}_2 = (1, -0.0002)$, and $\boldsymbol{d}_3 = (0.8, 0)$. Obviously, the true optimal input combination (say) $\boldsymbol{x}^*$ is (0, -0.001).

INSERT Figure 2: Ellipsoid contours $E(w \mid d_1, d_2)$ with global and local experimental areas

For $\boldsymbol{d}'_1 = (1, 0)$ we observe the output *w* twice: $m_1 = 2$. After fitting the first-order polynomial, we start the search from $\boldsymbol{d}'_o = (0.95, -0.00005)$; see Appendix 1. In this Monte Carlo experiment we know that the *truly optimal search direction* is the vector (say) $\boldsymbol{e}$ that starts at $\boldsymbol{d}'_o$ and ends at the true optimum (0, -0.001); also see Figure 3 below. We compute the angle (say) $\hat{\theta}$ between this true search direction $\boldsymbol{e}$ and the estimated search direction $\boldsymbol{v}$:

$$\hat{\theta} = \left| \arccos\left(\frac{\boldsymbol{e}'\boldsymbol{v}}{\|\boldsymbol{e}\|_2 \|\boldsymbol{v}\|_2}\right) \right| . \qquad (17)$$

So, the smaller the resulting $\hat{\theta}$ is, the better the search technique performs.

We take 100 *macro-replications*. Each time we apply the two techniques to the *same* I/O data (w, x_1, x_2). Then we compute the 100 search directions $\boldsymbol{v}$ for ASA; see Figure 3. We characterize $\hat{\theta}$'s empirical distribution through its average, standard deviation, and specific quantiles. This gives Table 1 (left part), which demonstrates the superiority of ASA, unless we focus on the worst case for low variance or the 95% quantile for high variance. (Actually, we can prove that if $\sigma_e = 0$, then $\hat{\theta} \approx 0°$ and $\hat{\phi} \approx 90°$.)

INSERT Figure 3: ASA's search direction when $\sigma_e = 0.10$

Next we investigate the effects of σ_e (see *e* in equation 16): we increase σ_e from 0.10 to 0.25. We use the same random numbers as we used for the smaller noise. Now the

estimated search directions may be very wrong. ASA still performs better, unless we focus on outliers (see the 95% or 100% quantiles in the right-hand part of Table 1).

Finally, we consider *interaction* between the two inputs: we take $\beta_0 = \beta_1 = \beta_2 = 0$; $\beta_{1;1} = -2$, $\beta_{2;2} = -1$; $\beta_{1;2} = 2$ in (16). We again find that ASA is better; see Table 2.

6. Conclusions

In this paper we addressed the problem of searching for the input combination that gives the maximum output. RSM is a classic techniques for tackling this problem, but it has two well-known problems: (i) RSM uses steepest ascent (SA), which is *scale-dependent*; (ii) RSM i*ntuitively* selects the step size on the SA path.

To address these two problems, we devised two new techniques. In technique 1 - called adapted SA or ASA - we select both a step direction and a step size. In technique 2, we use classic SA but we select a step size inspired by ASA.

Our main conclusion is that - except for orthogonal cases - ASA gives a better search direction than SA, because ASA is scale-independent.

Appendix 1. Derivation of the minimum variance of the regression predictor

The variance of the regression predictor $\hat{y}$ at $\boldsymbol{x}' = (1, \boldsymbol{d}')$ follows from (4) and (7), where without loss of generality we take a unit variance, $\sigma_e^2 = 1$:

$$var(\hat{y} | \boldsymbol{x}) = \boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x} \ .$$

This can be rewritten as

$$var(\hat{y} \mid \boldsymbol{d}) = a + 2\boldsymbol{b}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{C}\boldsymbol{d}$$

where a, $\boldsymbol{b}$, and $\boldsymbol{C}$ are defined in §2.

Because $\boldsymbol{C}$ is positive definite, the necessary and sufficient condition for the point that gives minimal variance (say) $\boldsymbol{d}_o$ is

$$2\boldsymbol{b} + 2\boldsymbol{C}\boldsymbol{d}_o = 0$$

which gives

$$\boldsymbol{d}_o = -\boldsymbol{C}^{-1}\boldsymbol{b} \ .$$

If $\boldsymbol{X}$ is orthogonal, then $\boldsymbol{b} = \boldsymbol{0}$ so the variance is minimal at the design center: $\boldsymbol{d}_o = \boldsymbol{0}$.

Appendix 2. Proof of the concavity in *d* for the objective function (9)

In (9) the first term $\boldsymbol{x}'\hat{\boldsymbol{\beta}}$ is linear, and the second term has the factors t^{α}_{N-q} and $\hat{\sigma}_e$, which are positive. Hence, it suffices to show that

$$\sqrt{\boldsymbol{x}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{x}} \text{ with } \boldsymbol{x}' = (1, \boldsymbol{d}')$$

is convex in $\boldsymbol{d}$.

If in this expression the factor $(\boldsymbol{X}'\boldsymbol{X})$ is not orthogonal, then we orthogonalize by the well-known *Gramm-Schmidt QR* method. It can be proven that convexity is preserved by linear transformations, so it suffices to show convexity for the orthogonal case.

If $(\boldsymbol{X}'\boldsymbol{X})$ is orthogonal, then it suffices to show that

$$f(\boldsymbol{d}) = \sqrt{1 + \boldsymbol{d}'\boldsymbol{d}}$$

is convex in $\boldsymbol{d}$. Obviously,

$$\nabla f(\boldsymbol{d}) = \frac{\boldsymbol{d}}{\sqrt{1 + \boldsymbol{d}'\boldsymbol{d}}}$$

and

$$\nabla^2 f(\boldsymbol{d}) = (1 + \boldsymbol{d}'\boldsymbol{d})^{-3/2} \left[\boldsymbol{I} + (\boldsymbol{d}'\boldsymbol{d})\boldsymbol{I} - \boldsymbol{d}\boldsymbol{d}'\right].$$

In this expression, $\boldsymbol{d}'\boldsymbol{d}\boldsymbol{I} - \boldsymbol{d}\boldsymbol{d}'$ is positive semi-definite: for all (say) $\boldsymbol{v}$ we have

$$\begin{aligned} \boldsymbol{v}'(\boldsymbol{d}'\boldsymbol{d}\boldsymbol{I} - \boldsymbol{d}'\boldsymbol{d})\boldsymbol{v} &= \|\boldsymbol{d}\|^2\|\boldsymbol{v}\|^2 - (\boldsymbol{d}'\boldsymbol{v})^2 \\ &\geq \|\boldsymbol{d}\|^2\|\boldsymbol{v}\|^2 - \|\boldsymbol{d}\|^2\|\boldsymbol{v}\|^2 = 0. \end{aligned}$$

This means that $\nabla^2 f(\boldsymbol{d})$ is positive definite; hence, $f(\boldsymbol{d})$ is convex. Consequently, (9) is an 'easy' problem; that is, the local maximum is the global maximum.

Appendix 3. Maximization of the objective function (9)

We rewrite (9) as

$$\hat{\beta}_0 + \hat{\boldsymbol{\beta}}'_{-0}\boldsymbol{d} - t^{\alpha}_{N-q}\hat{\sigma}_e\sqrt{a + 2\,\boldsymbol{b}'\boldsymbol{d} + \boldsymbol{d}'\boldsymbol{C}\boldsymbol{d}}$$

where $\hat{\boldsymbol{\beta}}_{-0}$, $\boldsymbol{a}$, $\boldsymbol{b}$, and $\boldsymbol{C}$ are defined in §2. Since this function is concave (see Appendix 2), the necessary and sufficient first-order conditions for the maximizing point $\boldsymbol{d}^+$ are

$$\hat{\boldsymbol{\beta}}_{-0} - \frac{t^{\alpha}_{N-q}\hat{\sigma}_e}{\sqrt{a + 2\boldsymbol{b}'\boldsymbol{d}^+ + \boldsymbol{d}^{+\prime}\boldsymbol{C}\boldsymbol{d}^+}}(\boldsymbol{b} + \boldsymbol{C}\boldsymbol{d}^+) = \boldsymbol{0}.$$

Substituting

$$\boldsymbol{d}^{+} = -\boldsymbol{C}^{-1}\boldsymbol{b} + \lambda \boldsymbol{C}^{-1}\hat{\beta}_{-0}$$

in which λ is an unknown scalar, we get

$$\lambda = \sqrt{\frac{a - \boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b}}{(t^{\alpha}_{N-q})^2\hat{\sigma}^2_e - \hat{\beta}'_{-0}\boldsymbol{C}^{-1}\hat{\beta}_{-0}}} .$$

For an orthogonal design (implying $a = 1/N$, $\boldsymbol{b} = \boldsymbol{0}$, and $\boldsymbol{C} = \boldsymbol{I}/N$) this equation simplifies to

$$\lambda = \sqrt{\frac{1/N}{(t^{\alpha}_{N-q})^2\hat{\sigma}^2_e - N\hat{\beta}'_{-0}\hat{\beta}_{-0}}} .$$

Hence for an orthogonal design the new point is

$$\boldsymbol{d}^{+} = \frac{\hat{\beta}_{-0}}{\sqrt{(t^{\alpha}_{N-q})^2\hat{\sigma}^2_e/N - \hat{\beta}'_{-0}|}}$$

Appendix 4: Optimization of the step size in SA

We assume that the SA path starts from $\boldsymbol{d}_o = -\boldsymbol{C}^{-1}\boldsymbol{b}$, which is the point at which the predictor variance is minimal; if $\boldsymbol{X}$ is orthogonal, then $\boldsymbol{b} = \boldsymbol{0}$ so $\boldsymbol{d}_o = \boldsymbol{0}$ (see Appendix 1). In SA we make a step of size (say) ζ in the $\hat{\beta}_{-0}$ direction. This means

$$\boldsymbol{d}^{+} = -\boldsymbol{C}^{-1}\boldsymbol{b} + \zeta\hat{\beta}_{-0} .$$

Substitution of $\boldsymbol{x}^{+\prime} = (1, \boldsymbol{d}^{+\prime})$ into the regression predictor (9) gives

$$\hat{y}_{\min}(\boldsymbol{x}^{+}) = \hat{\beta}_0 - \boldsymbol{b}'\boldsymbol{C}^{-1}\hat{\beta}_{-0} + \zeta\hat{\beta}'_{-0}\hat{\beta}_{-0} - t^{\alpha}_{N-q}\hat{\sigma}^2_e\sqrt{\boldsymbol{a} - \boldsymbol{b}'\boldsymbol{C}^{-1}\boldsymbol{b} + \zeta^2\hat{\beta}'_{-0}\boldsymbol{C}}$$

Since this expression is concave in ζ, it is easy to verify that ζ^{+} defined in (11) indeed maximizes $\hat{y}_{\min}(\boldsymbol{x}^{+})$.

Comparison with Appendix 3 proves that in the *orthogonal* case the two techniques coincide - provided SA starts from the design center.

Appendix 5: Scale independence of ASA

Affine scaling means $\boldsymbol{A}\boldsymbol{z} = \boldsymbol{x}$ or $\boldsymbol{z} = \boldsymbol{A}^{-1}\boldsymbol{x}$ with non-singular $\boldsymbol{A}$. Then (9) expressed in $\boldsymbol{z}$ becomes

$$\boldsymbol{z}'\hat{\beta}_z - t^{\alpha}_{N-q}\hat{\sigma}_e\sqrt{\boldsymbol{z}'(\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{z}} \qquad (5.1)$$

where

$$\hat{\beta}_z = (\boldsymbol{Z}'\boldsymbol{Z})^{-1}\boldsymbol{y}$$

or

$$\begin{aligned}\beta_z &= [\boldsymbol{A}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{A}]\boldsymbol{A}^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{A}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{A}'\hat{\beta}_x .\end{aligned}$$

In (5.1) we write the square-root factor as

$$\sqrt{z'(Z'Z)^{-1}z} = \sqrt{z'[(XA^{-1})'(XA^{-1})]^{-1}z}$$

$$= \sqrt{z'A'(X'X)^{-1}Az} \ .$$

Hence (5.1) becomes

$$z'A'\hat{\beta}_x - t^{\alpha}_{N-q}\hat{\sigma}_e\sqrt{z'A'(X'X)^{-1}Az}$$

$$= (Az)'\hat{\beta}_x - t^{\alpha}_{N-q}\hat{\sigma}_e\sqrt{(Az)'(X'X)^{-1}Az}$$

$$= x'\hat{\beta}_x - t^{\alpha}_{N-q}\hat{\sigma}_e\sqrt{x'(X'X)^{-1}x} \ ,$$

which is indeed identical to the original expression in (9). This proves that ASA is scale independent.

Appendix 6: Black-box model of I/O behavior of dynamic stochastic simulation

As Law and Kelton (2000, p. 12) - the best selling textbook on simulation - states, a *single server* queuing system is quite representative of more complex, dynamic, stochastic simulation models. For further simplification, we suppose that the output of interest is the mean waiting time in the steady state, E(*W*). This output can be estimated through a simulation that uses the following non-linear stochastic difference equation:

$$W(i) = \max[W(i-1) + S(i) - A(i), 0] \text{ with } i = 1, 2, \ldots \qquad (6.1)$$

where $W(i)$ denotes the waiting time of customer i, $S(i)$ the service time of customer i, and $A(i)$ the interarrival time between customers i and i -1. It is standard to start the simulation

run in the empty (idle) state: $W(0) = 0$. For additional simplification, we assume that the arrival times form a Poisson process, and so do the service times. This gives the well-known *M/M/1* (which can actually be solved analytically; see equation 6.4 below). By definition, M/M/1 implies that $S(i)$ and $A(i)$ are identically and independently distributed (IID) so simulation is straightforward. The output E(W) is usually estimated through the simulation run's *average*

$$\bar{W} = \sum_{i=b}^{n} \frac{W(i)}{n - b + 1} \text{ with } 0 \le b < n \tag{6.2}$$

where b denotes the length of the initialization (start-up, transient) phase (which may be zero), and n the run length. (In M/M/1 analysis and simulation through renewal analysis, this initialization is no issue; in practical simulations, however, it is a major problem; see Law and Kelton (2000, pp 496-552).) In other words, the dynamic simulation model generates the time series (6.1), but this series is characterized through the single statistic (6.2).

Actually, simulation is done for sensitivity analysis (possibly followed by optimization). Such an analysis aims at estimating the *input/output (I/O) function* (say)

$$\boldsymbol{z} = f(\boldsymbol{x}) \tag{6.3}$$

where $\boldsymbol{z}$ denotes the (multiple) output and $\boldsymbol{x}$ the (multiple) input; in the M/M/1 example we have $\boldsymbol{z} = \bar{W}$, $x_1 = \lambda$, and $x_2 = \mu$ with arrival rate λ and service rate μ.

In general, $f(\boldsymbol{x})$ in (6.3) has unknown shape and parameters. However, when studying the performance of a specific simulation methodology (such as RSM), researchers often use the M/M/1 simulation model because the I/O function is then known - assuming that the methodology has selected an appropriate initialization length b in (6.2) (obviously,

knowledge of $f(x)$ may not be used by the methodology itself):

$$E(W) = \frac{\lambda}{\mu(\mu - \lambda)} \text{ with } \frac{\lambda}{\mu} < 1 \ . \tag{6.4}$$

Unfortunately, the latter assumption is very questionable: it is well-known that selecting an appropriate transient-phase length *b* and run length *n* in (6.2) is difficult (see Law and Kelton 2000).

Moreover, most methodologies (including RSM) assume that - in general - the simulation observations (say) *V* have additive *white noise* ϵ; that is,

$$V = f(x) + \epsilon \text{ with } \epsilon \sim NIID(0, \sigma) \tag{6.5}$$

where NIID(0, σ) stands for normally (Gaussian) IID with mean zero and standard deviation σ.

In the M/M/1 example, (6.5) gives (i) $V = \bar{W}$, (ii) $f(x) = \lambda/[\mu(1 - \lambda]$, (iii) normality holds if $\bar{W}$ in (6.2) uses a sample size ($n - b + 1$) such that an asymptotic central limit theorem holds, (iv) constant variances result if different simulated traffic rates use different and appropriate sample sizes, and (v) independence results if no common pseudorandom numbers are used. Altogether, (6.5) applies to the M/M/1 example only if a slew of assumptions hold!

Hence, it is much more efficient and effective to generate test data for the simulation methodology through sampling from the latter equation, (6.5), instead of (6.1) and (6.2). Indeed, using (6.5) requires less computer time, and guarantees that the white noise assumption holds, including the desired value for the variance of the white noise. The alternative -using (6.1) and (6.2) - would require very long runs, especially for high traffic rates $\lambda/\mu \uparrow 1$ this alternative requires $n \uparrow \infty$.

Finally, even if the researcher has solved the problems resulting from using (6.1) and

(6.2), the following additional problem arises. Often, sensitivity analysis and optimization *locally* approximate the I/O function $f(\boldsymbol{x})$ through *low*-order polynomials. But then the question is: for which values of the traffic rate λ/μ is a first-order polynomial adequate?

To solve the latter problem, we use a *second-order polynomial* for $f(\boldsymbol{x})$ in (6.5). After all, (6.4) is no more than a simple example of $f(\boldsymbol{x})$; in practice the simulation models are much more complex than M/M/1 so they have unknown $f(\boldsymbol{x})$.

In conclusion, to test a simulation methodology we generate data through a static, random Monte Carlo model such as (6.5); we do not use a dynamic stochastic simulation model such as (6.1) combined with (6.2). So, the Monte Carlo technique is both efficient and effective.

Acknowledgment

We thank our Tilburg colleague Gül Gürkan for comments on earlier drafts of this paper.

References

Box, G.E.P. (1999), Statistics as a catalyst to learning by scientific method, part II - a discussion. *Journal of Quality Technology*, 31, no. 1, pp 16-29

--- and K.B. Wilson (1951), On the experimental attainment of optimum conditions. *Journal Royal Statistical Society, Series B,* 13, no. 1, pp. 1-38

Conn, A.R., N. Gould, and Ph. L. Toint (2000), *Trust-region methods,* SIAM, Philadelphia

Donohue, J.M., E.C. Houck and R.H. Myers (1993), Simulation designs and correlation induction for reducing-order bias in first-order response surfaces. *Operations*

Research, 41, number 5, pp.880-902

--- (1995), Simulation designs for the estimation of response surface gradients in the presence of model misspecification. *Management Science,* 41, no. 2, pp. 244-262

Driessen, L., R. Brekelmans, H. Hamers, and D. den Hertog (2001), On D-optimality based trust regions for black-box optimization problems. CentER Discussion Paper No. 2001-69, Tilburg University

Fu, M.C. (2002), Optimization for simulation: theory vs. practice. *INFORMS Journal on Computing* (forthcoming)

Hood, S.J. and P.D. Welch (1993) Response surface methodology and its application in simulation. In *Proceedings of the 1993 Winter Simulation Conference,* edited by G.W. Evans, M. Mollaghasemi, E.C. Russell, and W.E. Biles, pp. 115-122

Irizarry, M., J.R. Wilson, and J. Trevino (2001), A flexible simulation tool for manufacturing-cell design, II: response surface analysis and case study. *IIE Transactions,* 33, pp. 837-846

Jones, D.R., M. Schonlau, W.J. Welch (1998), Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13, 455-492

Khuri, A.I. (1996), Multiresponse surface methodology. *Handbook of Statistics*, vol. 13, ed. S. Ghosh and C.R. Rao, Elsevier, Amsterdam

--- and J.A. Cornell (1996), *Response surfaces: designs and analyses, second edition.* Marcel Dekker, Inc., New York

Kleijnen, J.P.C. (1998), Experimental design for sensitivity analysis, optimization, and validation of simulation models. *Handbook of Simulation*, edited by J. Banks, Wiley, New York, 1998, pp. 173-223

--- (1993), Simulation and optimization in production planning: a case study. *Decision*

Support Systems, 9, pp. 269-280

--- (1987), *Statistical tools for simulation practitioners*. Marcel Dekker, Inc., New York

--- (1975), *Statistical techniques in simulation, Part II*. Marcel Dekker, Inc., New York

Koch, P.N., Simpson, T.W., J.K. Allen, and F. Mistree (1999), Statistical approximations for multidisciplinary design optimization: the problem of size. *Journal of Aircraft*, 36, no. 1, pp. 275-286

Myers, R.H. (1999), Response surface methodology - current status and future directions. (Including Discussion) *Journal of Quality Technology*, 31, no. 1, pp. 30-74

--- and D.C. Montgomery (1995), *Response surface methodology: process and product optimization using designed experiments*, Wiley, New York

Neddermeijer, H.G.,G.J. van Ootmarsum, N. Piersma, and R. Dekker (2000), A framework for response surface methodology for simulation optimization models. *Proceedings of the 2000 Winter Simulation Conference* (edited by J.A. Joines, R.R. Barton, K Kang, and P.A. Fishwick), pp. 129-136

Sacks, J., W.J. Welch, T.J. Mitchell and H.P. Wynn (1989), Design and analysis of computer experiments. (Includes comments and rejoinder.) *Statistical Science*, 4, no. 4, pp. 409-435

Safizadeh, M.H. (2002), Minimizing the bias and variance of the gradient estimate in RSM simulation studies. *European Journal of Operational Research*, 136, no. 1, pp. 121-135

Simpson, T.W., T.M. Mauery, J.J. Korte, and F. Mistree (2001), Kriging metamodels for global approximation in simulation-based multidisciplinary design optimization. *American Institute of Aeronautics and Astronautics Journal* (in press)

Table 1: Statistics for ASA and SA's estimated angle error $\hat{\theta}$ (defined in equation 17), for two noise values σ_e

Statistics	σ_e = 0.10		σ_e = 0.25	
	ASA	SA	ASA	SA
Average	1.83	89.87	19.83	89.84
Standard deviation	17.89	0.04	56.45	0.19
Median (50% quantile)	0.03	89.88	0.06	89.89
75% quantile	0.00	89.89	0.12	89.91
25% ,,	0.01	89.87	0.03	89.85
95% quantile	0.13	89.91	179.81	89.99
5% ,,	0.00	89.79	0.01	89.46
100% quantile	178.92	89.93	179.88	90.06
0% ,,	0.00	89.59	0.00	88.74

Table 2: Statistics in case of interaction, for ASA and SA's estimated angle error $\hat{\theta}$, for two noise values σ_e

Statistics	σ_e = 0.10		σ_e = 0.25	
	ASA	SA	ASA	SA
Average	9.72	16.01	10.14	17.33
Standard deviation	3.30	6.23	7.69	12.88
Median (50% quantile)	9.68	16.02	8.99	14.94
75% quantile	12.37	21.12	16.13	27.87
25% ,,	6.99	10.76	3.21	5.84
95% quantile	15.66	27.05	24.78	41.55
5% ,,	4.99	6.80	0.61	0.81
100% quantile	17.41	30.08	32.07	50.99
0% ,,	0.85	1.46	0.04	0.25

Figure 1

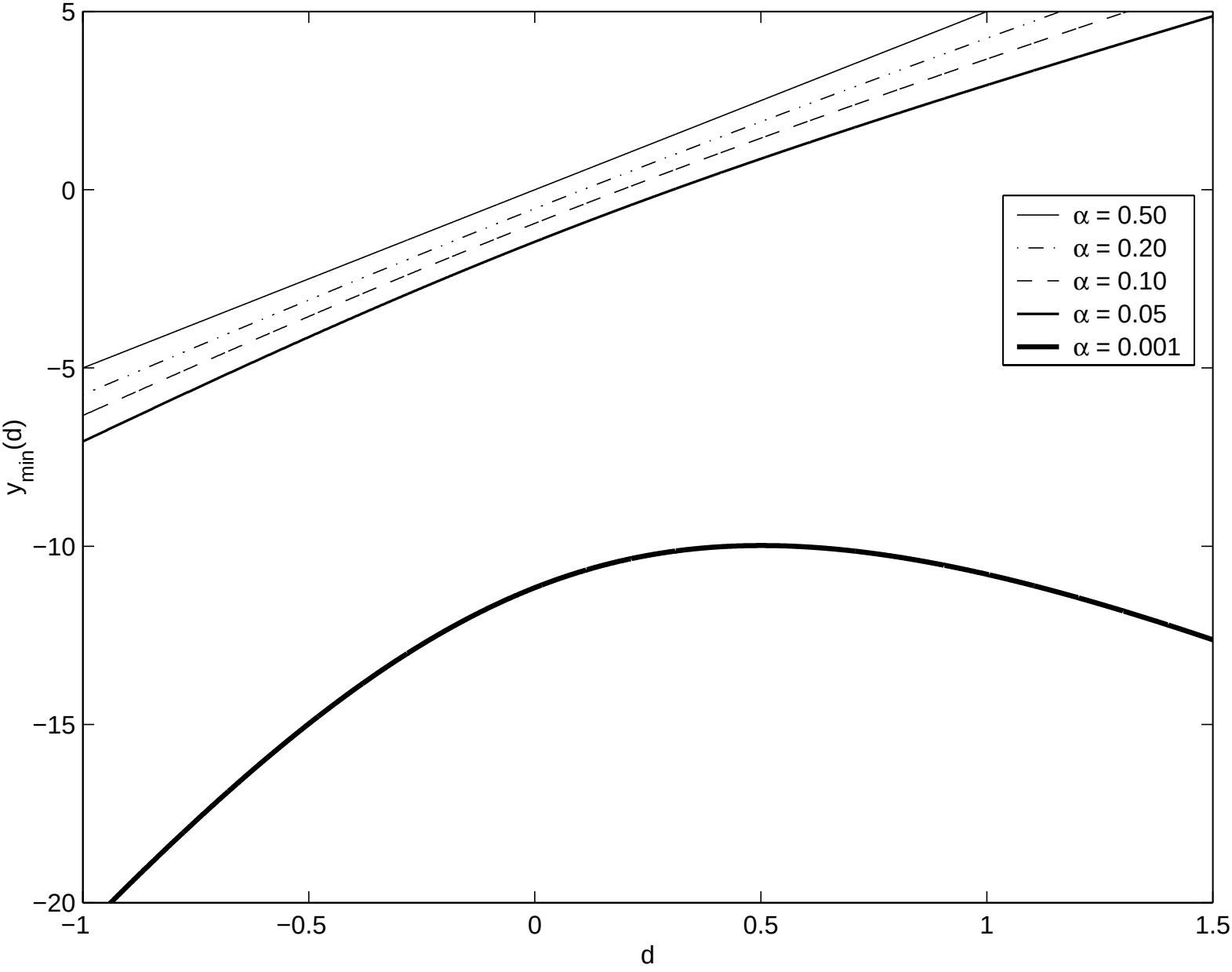


Figure 2

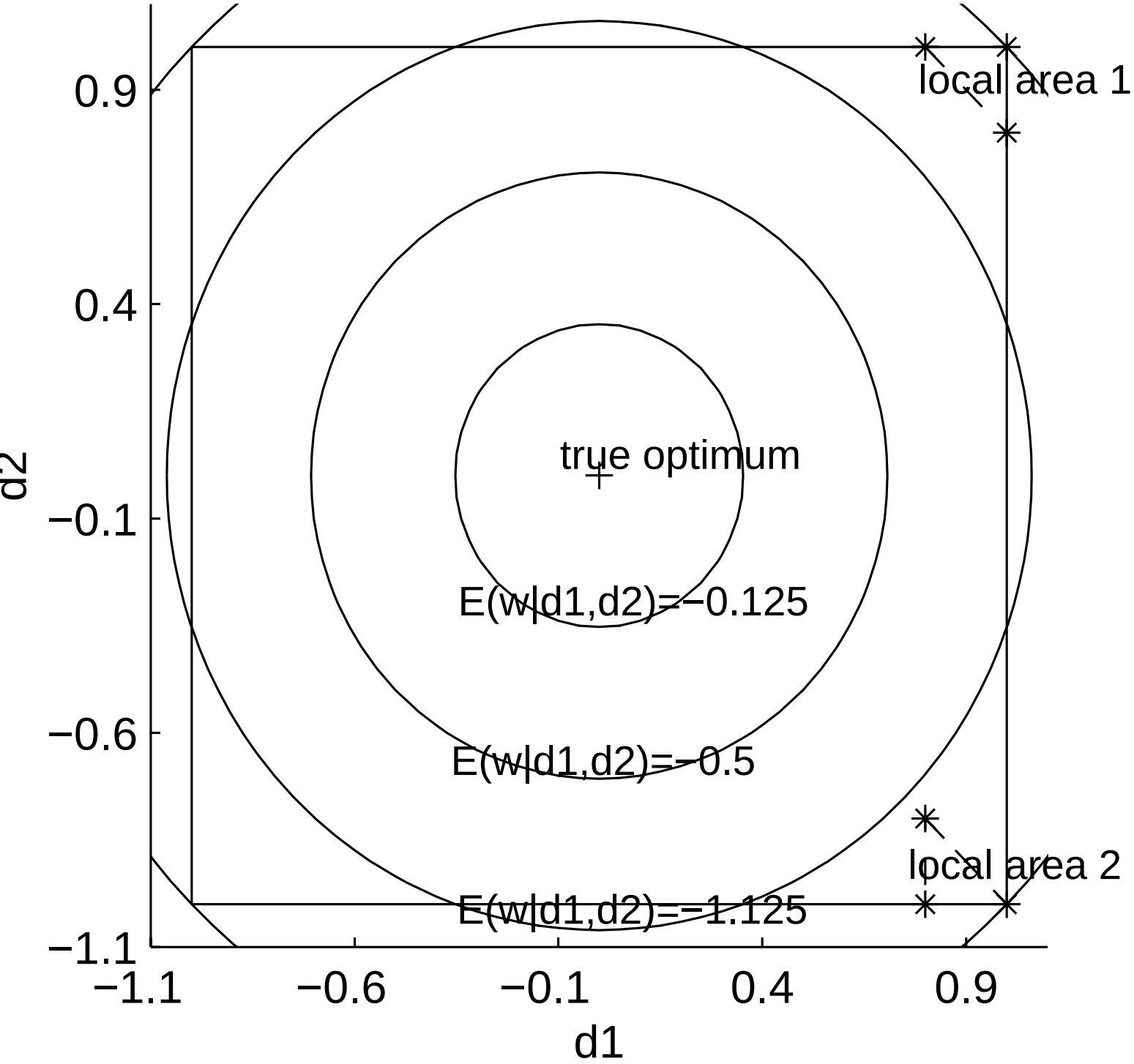

local area 1
0.9
0.4
true optimum
−0.1
E(w|d1,d2)=−0.125
−0.6
E(w|d1,d2)=−0.5
local area 2
E(w|d1,d2)=−1.125
−1.1
−1.1
−0.6
−0.1
0.4
0.9
d1
d2

Figure 3

