Dual Response Surface Optimization

DENNIS K. J. LIN

University of Tennessee, Knoxville, TN 37996

WANZHU TU

University of South Carolina, Columbia, SC 29208

Vining and Myers adapted the dual response approach to achieve the goals of Taguchi's philosophy. This excellent approach contains some deficiencies that will be highlighted in this paper. A more satisfactory and substantially simpler optimization procedure on a dual response approach is proposed that allows more general response models to be entertained in reality. The new method is demonstrated using the example given in Vining and Myers in which it leads to a solution with 25% smaller mean square error.

Introduction

RESPONSE Surface Methodology, first introduced by Box and Wilson (1951), is designed to find the optimal settings for a set of input or design variables that maximize (or minimize) the response Y. Such a problem is typically focused on the mean value of Y. As a result, it works well when the variance of Y is relatively small and stable (i.e., variance is assumed at a known or unknown constant value).

Indeed, when the variance is not a constant, classical response surface methodology can be misleading. Vining and Myers (1990) made use of the dual-response approach (see Myers and Carter (1973)), and proposed an ingenious method to tackle such a problem. They first fit second-order models to both primary- and secondary-response surfaces and then applied the dual-response surface approach to optimize the primary response subject to an appropriate constraint on the value of the secondary response. (Henceforth, we will refer to Vining and Myers (1990) as VM.)

The basic idea behind VM's approach is excellent. However, we believe that their optimization procedure can be further improved. Specifically: (1) the optimization procedure via Lagrangian multipliers is somehow misleading. This may very well rule out better conditions due to the restriction that the "estimate" of secondary-response is forced to a fixed value. As we shall see later, a better procedure based on mean square error (MSE) criterion should be used; and (2) the restriction on the full second-order models for both primary- and secondary-response surfaces is unrealistic. Rather, the best subset models should be considered. In fact, the MSE procedure given here works well even for nonlinear (not polynomial type) response surfaces.

In the next section, we will briefly review the VM approach, and highlight its deficiencies. Following our critique, a more satisfactory formulation is described. We illustrate our method using the example from VM for a fair comparison.

Review of the Vining and Myers' Approach

Following VM notations, the primary and secondary (respectively) responses can be written as

$$\eta_p = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum\sum_{i<j}^{k} \beta_{ij} x_i x_j + \varepsilon_p \quad (1)$$

$$\eta_s = \gamma_0 + \sum_{i=1}^{k} \gamma_i x_i + \sum_{i=1}^{k} \gamma_{ii} x_i^2 + \sum\sum_{i<j}^{k} \gamma_{ij} x_i x_j + \varepsilon_s \quad (2)$$

and the resulting fitted response surfaces may be represented by

$$\widehat{\omega}_p = b_0 + \mathbf{x}'\mathbf{b} + \mathbf{x}'\mathbf{B}\mathbf{x} \quad (3)$$

$$\widehat{\omega}_s = c_0 + \mathbf{x}'\mathbf{c} + \mathbf{x}'\mathbf{C}\mathbf{x} \quad (4)$$

Dr. Lin is an Associate Professor in the Department of Statistics and an instructor in the Management Development Center. He is a Member of ASQC.

Mr. Tu is a graduate student in the Department of Statistics.

where $b_0 = \widehat{\beta}_0$, $c_0 = \widehat{\gamma}_0$, $\mathbf{b} = (\widehat{\beta}_1, \ldots, \widehat{\beta}_k)'$, $\mathbf{c} = (\widehat{\gamma}_1, \ldots, \widehat{\gamma}_k)'$,

$$\mathbf{B} = \frac{1}{2}\begin{bmatrix} 2\widehat{\beta}_{11} & \widehat{\beta}_{12} & \cdots & \widehat{\beta}_{1k} \\ \widehat{\beta}_{12} & 2\widehat{\beta}_{22} & \cdots & \widehat{\beta}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\beta}_{1k} & \widehat{\beta}_{2k} & \cdots & 2\widehat{\beta}_{kk} \end{bmatrix}$$

and

$$\mathbf{C} = \frac{1}{2}\begin{bmatrix} 2\widehat{\gamma}_{11} & \widehat{\gamma}_{12} & \cdots & \widehat{\gamma}_{1k} \\ \widehat{\gamma}_{12} & 2\widehat{\gamma}_{22} & \cdots & \widehat{\gamma}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{\gamma}_{1k} & \widehat{\gamma}_{2k} & \cdots & 2\widehat{\gamma}_{kk} \end{bmatrix}$$

where b_0, c_0, $\mathbf{b}$, $\mathbf{c}$, $\mathbf{B}$, and $\mathbf{C}$ are the appropriate vectors and matrices of the estimates for the coefficients.

Using Lagrangian multipliers, VM proposed a method of finding $\mathbf{x}$ that optimizes $\widehat{\omega}_p$ subject to $\widehat{\omega}_s = T$, where T is some desirable target value of the constraint response and a spherical region of interest is assumed. That is, find $\mathbf{x}$ that satisfies

$$\frac{\partial L}{\partial \mathbf{x}} = 0 \tag{5}$$

where

$$L = b_0 + \mathbf{b}'\mathbf{x} + \mathbf{x}'\mathbf{B}\mathbf{x} + \lambda(c_0 + \mathbf{c}'\mathbf{x} + \mathbf{x}'\mathbf{C}\mathbf{x} - T). \tag{6}$$

Denote $\widehat{\omega}_\mu$ as the fitted response surface for the mean and $\widehat{\omega}_\sigma$ as the fitted response surface for the standard deviation. The following three cases were discussed in VM:

Case 1: "Target value is best", which means keeping μ at a specified target value μ_0, while minimizing σ^2. Namely:

minimize $\widehat{\omega}_\sigma$
subject to $\widehat{\omega}_\mu = \mu_0$.

Case 2: "The larger, the better" which means making μ as large as possible, while controlling σ^2. Namely:

maximize $\widehat{\omega}_\mu$
subject to $\widehat{\omega}_\sigma = \sigma_0$.

Case 3: "The smaller, the better", which means making μ as small as possible, while controlling σ^2. Namely:

minimize $\widehat{\omega}_\mu$
subject to $\widehat{\omega}_\sigma = \sigma_0$.

We see that the determination of the primary and the secondary responses is determined by the goal of the experiment. One major concern here is the lack of realism of the equality constraints. Note that both $\widehat{\omega}_\mu$ and $\widehat{\omega}_\sigma$ are only approximations to the "true" responses (subject to certain random errors). Restricting the optimization to equality constraints will inevitably exclude globally preferred values. This is more obvious for Cases 2 and 3, where σ_0 is normally unknown (and in fact, the smaller the better).

Proposed Optimization Procedure

The proposed optimization procedure is best illustrated by the following example. We shall focus on the case "target value is best" to illustrate the basic idea and discuss other cases in a later section. Consider Figure 1 where the estimated mean response curve is denoted by $\widehat{\omega}_\mu$ and the estimated standard deviation response curve is denoted by $\widehat{\omega}_\sigma$. Our purpose is to find an optimal set of conditions such that $\widehat{\omega}_\mu$ will be close to the target value T, while the standard deviation $\widehat{\omega}_\sigma$ is kept small. Suppose the target for the mean is T as indicated. In this case, the VM approach will first restrict $\widehat{\omega}_\mu = T$. Four points (A, B, C, and D) satisfy this restriction. Among them, point A has the minimum variance and thus is the "optimal" setting.

Studying the behavior of $\widehat{\omega}_\mu$ and $\widehat{\omega}_\sigma$, it is easy to see that point E is a better choice than point A. By introducing a little bias, we may reduce the variance a great deal. In fact, Point E minimizes the mean square error (MSE $= (\widehat{\omega}_\mu - T)^2 + \widehat{\omega}_\sigma^2$).

The MSE criterion consists of two major terms: the bias and the variance. Since we must work with an estimate of both response functions, the MSE

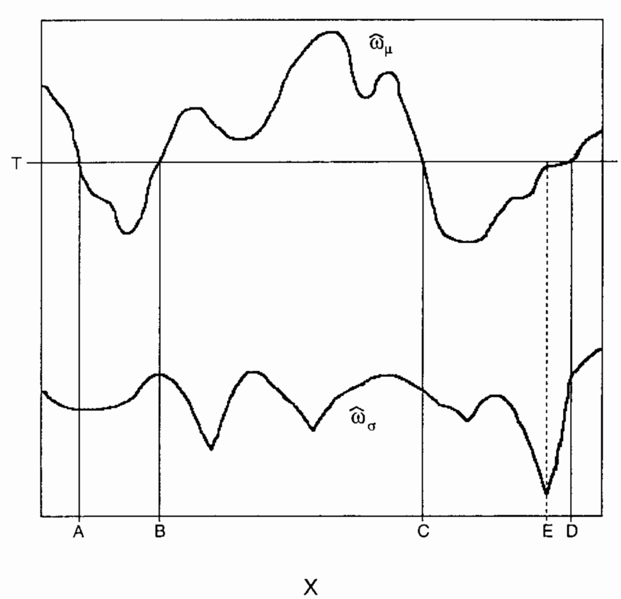

FIGURE 1. An Illustrative Example.

criterion allows some disparity around the target, meanwhile keeping the variance small. This is considered to be a better approach than the equality constraints in VM.

Thus, a more satisfactory formulation of the problem would be accomplished by the following steps:

1. Find a model for ω_μ and ω_σ (both are functions of $\mathbf{x}$).
2. Find $\mathbf{x}$ such that MSE $= (\widehat{\omega}_\mu - T)^2 + \widehat{\omega}_\sigma^2$ is minimized. Note that we have made no specific assumption about the models in Step 1.

Example

To provide a fair basis of comparison with the optimization in VM, we analyze the same data set they used, which is given in Table 1. The experiment, described in Box and Draper (1987), was conducted to determine the effect of the three variables x_1 (speed), x_2 (pressure), and x_3 (distance), on the quality of a printing process, that is, on the machine's ability to apply colored inks to package labels. The experiment is a 3^3 factorial design with 3 replicates at each point.

TABLE 1. The Printing Process Study Data

u	x_1	x_2	x_3	y_{u1}	y_{u2}	y_{u3}	$\bar{y}_u$	s_u
1	−1	−1	−1	34	10	28	24.0	12.49
2	0	−1	−1	115	116	130	120.3	8.39
3	1	−1	−1	192	186	263	213.7	42.80
4	−1	0	−1	82	88	88	86.0	3.46
5	0	0	−1	44	178	188	136.7	80.41
6	1	0	−1	322	350	350	340.7	16.17
7	−1	1	−1	141	110	86	112.3	27.57
8	0	1	−1	259	251	259	256.3	4.62
9	1	1	−1	290	280	245	271.7	23.63
10	−1	−1	0	81	81	81	81.0	0.00
11	0	−1	0	90	122	93	101.7	17.67
12	1	−1	0	319	376	376	357.0	32.91
13	−1	0	0	180	180	154	171.3	15.01
14	0	0	0	372	372	372	372.0	0.00
15	1	0	0	541	568	396	501.7	92.5
16	−1	1	0	288	192	312	264.0	63.50
17	0	1	0	432	336	513	427.0	88.61
18	1	1	0	713	725	754	730.7	21.08
19	−1	−1	1	364	99	199	220.7	133.80
20	0	−1	1	232	221	266	239.7	23.46
21	1	−1	1	408	415	443	422.0	18.52
22	−1	0	1	182	233	182	199.0	29.45
23	0	0	1	507	515	434	485.3	44.64
24	1	0	1	846	535	640	673.7	158.20
25	−1	1	1	236	126	168	176.7	55.51
26	0	1	1	660	440	403	501.0	138.90
27	1	1	1	878	991	1161	1010.0	142.50

Since the optimization method VM proposed is based on the quadratic form of the model, a full second-order model is required, regardless of the significance levels of the model fittings. Assuming the quadratic models were adequate (see next section), they fit a response surface for the mean of the characteristic of interest:

$$\begin{aligned}\widehat{\omega}_\mu &= 327.6 + 177.0x_1 + 109.4x_2 + 131.5x_3 \\ &\quad + 32.0x_1^2 - 22.4x_2^2 - 29.1x_3^2 \\ &\quad + 66.0x_1x_2 + 75.5x_1x_3 + 43.6x_2x_3 \qquad (7)\end{aligned}$$

and a response surface for the standard deviation:

$$\begin{aligned}\widehat{\omega}_\sigma &= 34.9 + 11.5x_1 + 15.3x_2 + 29.2x_3 \\ &\quad + 4.2x_1^2 - 1.3x_2^2 + 16.8x_3^2 \\ &\quad + 7.7x_1x_2 + 5.1x_1x_3 + 14.1x_2x_3. \qquad (8)\end{aligned}$$

Recall that VM sought (x_1, x_2, x_3) that would minimize $\widehat{\omega}_\sigma$, subject to $\widehat{\omega}_\mu = 500$, while our approach seeks (x_1, x_2, x_3) to minimize MSE $= (\widehat{\omega}_\mu - T)^2 + \widehat{\omega}_\sigma^2$.

Based on such models, Table 2 summarizes the results of the two approaches. We assume that a cuboidal region is of interest, that is, $-1 \le x_i \le 1$ for i = 1, 2 and 3. The VM approach based on a spherical region constraint leads to the setting (x_1, x_2, x_3) = (0.614, 0.228, 0.1) which results in an expected mean of 500 and a variance of 2679.698 (MSE = 2679.698). If we use the MSE criterion, the best setting is (x_1, x_2, x_3) = (1, 0.07, −0.25) with a slight bias on the expected mean but a much smaller variance (MSE = 2005.145, a 25.17% smaller MSE). Note that such a bias in the mean response is nonsignificant (at $\alpha = 0.05$) in the sense that one cannot reject the hypothesis $H_0\colon\ \mu = 500$, based on the variability associated with Equation (7).

We note the following:

1. When extrapolation beyond the original region of interest is inhibited, a solution for the VM approach may not exist in general. This is because the equation $\widehat{\omega}_\mu = \mu_0$ may not contain a solution in $-1 \le x_i \le 1$.
2. Since the VM solution, if it exists, is always on target, the expected variance is always equal to MSE.
3. Using the VM approach, we have found a "better" solution in a cuboidal region, $(x_1, x_2, x_3) = (1, 0.119, -0.26)$, which results in a mean of 500

TABLE 2. Comparison of the Optimal Settings for the Quadratic Model

Method	Optimal Setting	$\widehat{\omega}_\mu$	$\widehat{\omega}_\sigma^2$	MSE
Vining & Myers	(0.614, 0.228, 0.1)	500	2679.70	2679.70
MSE Method	(1.0, 0.07, −0.25)	494.44	1974.02	2005.14

and a variance of 2034.012, (MSE = 2034.012). Such an "optimal setting" is very close to our MSE-optimal setting.

The optimization algorithm used here is a standard subroutine in a nonlinear programming procedure. We used the subroutine BCPOL provided by IMSL MATH/LIBRARY (1987) on a VAX Cluster system. Such a subroutine is very flexible concerning the shape of the region. The result given in Table 2, for example, is based on a cuboidal region. Thus, a constraint $-1 \leq x_i \leq 1$ was added to the program. Likewise, if a spherical region is of interest, then a constraint $\mathbf{x}'\mathbf{x} \leq \rho$ can be added, for any specific value of ρ.

There are several packages available for large/small scale, linear/nonlinear optimization problems. In principle, any algorithm can be used here. For example, Del Castillo and Montgomery (1993) use a generalized reduced gradient (GRG) algorithm to optimize VM's problem with inequality constraints. Table 3 compares the results taken from Del Castillo and Montgomery (1993, Table 1). Our results are based upon the assumptions that extrapolation is allowed, and a spherical region is assumed. Recall that the dual response surfaces were built based upon the cuboidal region of (x_1, x_2, x_3). As shown in Table 3, by allowing a small bias on the estimated mean response, the variance can be reduced a great deal. In the case of $\rho = 3$, for example, a more than 25% reduction in MSE was found. In particular, the setting $(x_1, x_2, x_3) = (6.4347, -4.2938, 1.0371)$ for $\rho^2 = 7.88$, if acceptable under operating conditions, will have 500 as its expected value and a minimum possible variance (expected standard deviation = 0). Of course, the experimenter must remember that extrapolation is not always reliable. The GRG algorithm is a classical tool to solve a large scale optimization problem. Here, we also used a recent algorithm LANCELOT (Conn, Gould, and Toint (1992)).

Optimization on the Best Subset Model

As suggested by VM, the fitness or the prediction ability of the model is an extremely important consideration when optimizing a dual response problem.

TABLE 3. Comparison with Results in Del Castillo and Montgomery (1993)
(D&M = Del Castillo and Montgomery; L&T = Method Proposed Here)

$\mathbf{x}'\mathbf{x} \leq \rho$	Method	Optimal Setting	$\widehat{\omega}_\mu$	$\widehat{\omega}_\sigma^2$	MSE
1.0	D&M	(0.9839, 0.0265, −0.1760)	500	2053.75	2053.75
	L&T	(0.9831, 0.0036, −0.1829)	494.54	1992.73	2022.78
1.5	D&M	(1.1897, −0.2237, −0.1857)	500	1901.41	1901.41
	L&T	(1.1856, −0.2454, −0.1847)	495.20	1854.79	1877.84
2.0	D&M	(1.3395, −0.4261, −0.1544)	500	1802.41	1802.41
	L&T	(1.3347, −0.4421, −0.1547)	495.51	1761.06	1781.25
3.0	D&M	(0.9525, 1.2461, −0.7348)	500	2207.58	2207.58
	L&T	(1.5651, −.7373, 0.0883)	495.68	1615.88	1634.57

For models (7) and (8), the R^2 values are 0.8741 and 0.4542, respectively, which are not very satisfactory for this purpose. In addition, notice that many terms are nonsignificant in the full second-order model (see VM, Tables 2 and 3). However, the limitation of the MURSAC software (as described in VM) forces the use of the full second-order model.

To fit the best models for the mean and standard deviation, we start with the "full" cubic model (namely, with the terms $x_1x_2x_3$, $x_i^2x_j$'s, x_i^2's, x_ix_j's, x_i's, etc.). Several different model selection procedures (stepwise regression, all possible subset regression, C_p, PRESS, etc.) were employed to ensure the best subset models

$$\widehat{\omega}_\mu = 314.667 + 177.0x_1 + 109.426x_2 + 131.463x_3 + 66.028x_1x_2 + 75.472x_1x_3 + 43.583x_2x_3 + 82.792x_1x_2x_3 \quad (9)$$

and

$$\widehat{\omega}_\sigma = 47.994 + 11.527x_1 + 15.323x_2 + 29.190x_3 + 29.566x_1x_2x_3. \quad (10)$$

The analysis of variance tables for the mean and the standard deviation are given in Table 4 and Table 5, respectively. Here, the R^2 values are 0.9570 and 0.4839, respectively (as compared to 0.8741 and 0.4542 in the full second-order models). Consequently, the adjusted-R^2, R_a^2, are 0.9416 and 0.3901, respectively (as compared to 0.8074 and 0.1652).

Under these models and a cuboidal region, the optimal setting is found to be $(x_1, x_2, x_3) = (1, 1, -0.525)$ with the mean equal to 492.285 and a standard deviation of 44.01 (consequently, MSE = 1996.6). Such a setting is close to the MSE optimal setting for a full second-order model, but is far from VM's choice.

TABLE 4. Analysis of Variance for the Mean Response

Source	df	Sum of Squares	Mean Squares	F-Ratio	R^2
Model	7	1288838.2	184119.7	60.441	0.957
Error	19	57878.87	3042.26		

Variable	Partial t	p value
x_1	13.606	0.0001
x_2	8.411	0.0001
x_3	10.105	0.0001
x_1x_2	4.144	0.0006
x_1x_3	4.737	0.0001
x_2x_3	2.735	0.0131
$x_1x_2x_3$	4.243	0.0004

TABLE 5. Analysis of Variance for the Standard Deviation

Source	df	Sum of Squares	Mean Squares	F-Ratio	R^2
Model	4	28948.527	7237.1318	5.157	0.4839
Error	22	30871.972	1403.2715		

Variable	Partial t	p value
x_1	1.305	0.2052
x_2	1.735	0.0967
x_3	3.306	0.0032
$x_1x_2x_3$	2.232	0.0361

Concluding Remarks

The concept of a dual response approach is extremely important in industrial problems. Such an approach will become increasingly used in the future. However, we believe that the optimization procedure given by VM can be further improved. Specifically:

1. The MSE approach is not necessarily restricted in a full second-order model. In fact, it can handle more realistic models and much more complicated models than polynomial models; and
2. Forcing the estimated secondary-response to equal a specific value as a constraint can be misleading. One referee pointed out that certain computer packages can set inequality constraints on the mean response so that some disparity around the target can be taken into account. This is, in fact, our basic idea to implement dual response surface optimization.

Of course, there may be situations where tight adherence to the target is essential. The MSE approach can be modified to

$$\text{MSE}^* = \lambda_1(\widehat{\omega}_\mu - T)^2 + \lambda_2\widehat{\omega}_\sigma^2$$

where λ_1 and λ_2 are pre-specified positive constants. The user may, as in the case of ridge analysis, evaluate various operation settings based on values of $(\lambda_1,\ \lambda_2)$ that are of interest. Two special cases previously mentioned are:

1. $\lambda_1 = \lambda_2 = 1$, this is the objective function that was proposed in this paper; and

2. $\lambda_1 = \infty$ and $\lambda_2 = 1$, this is the objective function given in VM and Del Castillo and Montgomery (1993).

The optimization based upon the criteria of MSE gives a fairly general method to solve the dual response surface problem. It is a natural and appealing approach. This principle is not only applied to Case 1 (target is the best), but also to Case 3 (the smaller the better). Since the physical measurement of the response is normally nonnegative, it is naturally used to minimize MSE = $\widehat{\omega}_\mu^2 + \widehat{\omega}_\sigma^2$. For example, using the full second-order model, the setting $(x_1, x_2, x_3) = (-0.524, -1, -1)$ results in an expected mean of 68.99 and a standard deviation of 21.84. This is apparently better than VM's assumption that $\widehat{\omega}_\sigma = 60$, 75 and 90 (correspondingly, $\widehat{\omega}_\mu = 494.8, 585.0, 671.7$). While using the best subset model, the optimal setting is $(x_1, x_2, x_3) = (-1, -1, -0.3602)$ which has an expected mean of 60 and an expected standard deviation of 0. On the other hand, VM's approach will tend to minimize $\widehat{\omega}_\mu$ and fix $\widehat{\omega}_\sigma$ at a wild guess value. This is certainly not recommended.

The dual response approach will only work well when the responses are independent (i.e., the mean and variance are independent). For situations where such an assumption is violated, it is first necessary to understand their relationships. Consequently, a dual response approach seems inappropriate. This needs further investigation, however.

Acknowledgments

We thank two referees and Professor Robert Mee for their constructive comments on an earlier version of this paper. Dennis Lin was partially supported by The National Science Foundation via Grant DMS-9204007 and Professional Development Awards, The University of Tennessee. This paper was revised while the first author was visiting the IBM Watson Research Center. We are grateful to Dr. Andy R. Conn for introducting us to his LANCELOT package and performing the calculations for this paper.

References

BOX, G. E. P. and DRAPER, N. R. (1987). *Empirical Model Building and Response Surfaces.* John Wiley & Sons, New York, NY.

BOX, G. E. P. and WILSON, K. B. (1951). "On the Experimental Attainment of Optimum Conditions". *Journal of the Royal Statistical Society* B 13, pp. 1-45.

CONN, A. R.; GOULD, N. I. M.; and TOINT, P. L. (1992). *LANCELOT: A Fortran Package for Large-Scale Nonlinear Optimization (Release A).* Springer Verlag, Heidelberg, Germany.

DEL CASTILLO, E. and MONTGOMERY, D. C. (1993). "A Nonlinear Programming Solution to the Dual Response Problem". *Journal of Quality Technology* 25, pp. 199-204.

IMSL MATH/LIBRARY (1987). Version 1.0, IMSL Inc., Houston, TX.

MYERS, R. H. and CARTER, W. H. (1973). "Response Surface Techniques for Dual Response Systems". *Technometrics* 15, pp. 301-317.

VINING, G. G. and MYERS, R. H. (1990). "Combining Taguchi and Response Surface Philosophies: A Dual Response Approach". *Journal of Quality Technology* 22, pp. 38-45.

Key Words: *Constrained Optimization, Multiple Responses, Optimization, Response Surface Methodology.*