© Grace Scientific Publishing

Journal of
Statistical Theory and Practice
Volume 5, No. 1, March 2011

A Note on Small Composite Designs for Sequential Experimentation

Nam-Ky Nguyen, *International School & Centre for High Performance Computing, Vietnam National University, Hanoi, Vietnam.*
Email: namnk@isvnu.vn

Dennis K. J. Lin, *Department of Statistics, Pennsylvania State University, University Park, USA &*
School of Statistics, Renmin University of China, Beijing, China.
Email: DKL5@psu.ed

Received: May 4, 2010 *Revised:* May 31, 2010; July 12, 2010

Abstract

The recommended approach to experiments using the response surface methodology is sequential, i.e., experiments should be conducted iteratively. At the first stage, a first-order design, usually an orthogonal two-level design (with a few center points) is used to find out whether the current region is appropriate and to allow the estimation of main effects (and possibly some interactions). The design at the first stage is then augmented with more runs in the second stage. The combined design allows the estimation of the remaining interaction and quadratic effects. Some well-known classes of designs which allow such a sequential experimentation are the central composite designs, the small composite designs and the augmented-pair designs. This paper reviews these designs and introduces a new algorithm which is able to augment any first order design with additional design points to form a good design for a second-order model.

AMS Subject Classification: ?

Key-words: Augmented-pair design; Box-Behnken design; Composite design; Orthogonal quadratic effect property; Response surface design; Two-level design.

1. Introduction

Response surface methodology (RSM) considers the situation in which a *response* y depends on k factors, $x_1, x_2, \cdots, x_k$. The true response function is unknown, and we shall approximate it over a limited experimental region by a polynomial representation. This is

* 1559-8608/11-1/$5 + $1pp - see inside front cover

done by fitting a local response surface from a typically small set of observations. One of the main purposes of RSM is to determine which level combinations of the k input factors will optimize the response, y. Under certain smooth conditions, this response function may be approximated well by lower-order polynomial models over a limited experimental region, $\mathscr{X}$. Usually the first-order polynomial model is employed at the initial stage, i.e.

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + \varepsilon,$$

where ε is a white noise. If it suffers from lack of fit arising from the existence of surface curvature, then the first-order polynomial model would be modified by adding higher-order terms into the model. Therefore, we might fit a second-order polynomial model of

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_{ii} x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \varepsilon. \tag{1.1}$$

Such a sequential feature is proved to be efficient in practice.

Some (second-order) response surface designs (RSDs) for sequential experimentations are the central composite designs (CCDs), the small composite designs (SCDs) and the augmented-pair designs (APDs). The CCDs developed by Box and Wilson (1951) and the SCDs by Draper and Lin (1990) have been popularly used among experimenters and discussed in most textbooks and papers, see for example, Myers, Montgomery and Anderson-Cook (2009) and Draper and Lin (1996).

A composite design consists of (i) a fractional factorial portion called *cube* portion, (ii) a set of $2k$ *axial* points at a distance α from the origin, plus (iii) n_0 center points. The cube portion for CCD is a fractional factorial 2^{k-p} of resolution V or higher; while for SCD, this is reduced to a proper size of Plackett and Burman designs (Plackett and Burman, 1946).

The augmented-pair designs (APDs) were proposed by Morris (2000). It consists of a first-order two-level orthogonal design with n_1 runs and n_0 center points in the first stage. This design is then augmented by $n_2 = \binom{n_1}{2}$ runs. For each pair of runs x_u and x_v in n_1, a run in n_2 is generated as $x_{uv} = -0.5(x_u + x_v)$. There are two reasons why APDs deserve special attention: (i) unlike the CCDs and SCDs, the run size of the APD design in the first stage is minimal; and (ii) the quadratic effects of APDs are always orthogonal to all main-effects and interaction effects. We consider such an orthogonal quadratic effect (OQE) property, Property (ii), an important property, as the quadratic effects which could not be estimated in the first stage should be estimated with the maximum precision in the second stage.

Let the u-th row of $X_{n\times p}$, the expanded model matrix of a design for k factors in n runs, be written as $(1, x_{u1}^2, x_{u2}^2, \ldots, x_{u1}, x_{u2}, \ldots, x_{u1}x_{u2}, x_{u1}x_{u3}, \ldots)$. Here, $p = \frac{1}{2}(k+1)(k+2)$ is the number of parameters in (1.1). The $X'X$ matrix of designs with the *OQE* property will have the form

$$\left(\begin{array}{c|c} A & \mathbf{0} \\ \hline \mathbf{0} & B \end{array}\right), \tag{1.2}$$

where the square matrix A has $k+1$ columns, and the square matrix B has $k(k+1)/2$ columns. The $(X'X)^{-1}$ matrix will be of the form

$$\left(\begin{array}{c|c} A^{-1} & \mathbf{0} \\ \hline \mathbf{0} & B^{-1} \end{array}\right). \tag{1.3}$$

Since not all SCDs have the *OQE* property, in this paper we denote an SCD with this property as SCD* (and an RSD with this property as RSD*). Note the $X'X$ and $(X'X)^{-1}$ matrices of the CCDs and the Box-Behnken designs (BBDs) (see Box and Behnken, 1960 and Nguyen and Borkowski, 2008) are also of the form in (1.2) and (1.3). In addition, B and B^{-1} in (1.2) and (1.3) are diagonal matrices.

In this paper, an algorithm to construct RSD*'s for sequential experimentation (including SCD* as a special case) is proposed. In Section 2, a general algorithm which can be used to augment designs in the first stage with additional design points is proposed. Section 3 will show how to adapt this algorithm to construct RSD*'s and SCD*'s. Section 4 will catalogue some existing designs and some new designs generated by the proposed algorithm. Section 5 provides the concluding remarks.

2. An SOD Algorithm

Without loss of generality, a three-level design factor ($x_i,\ i = 1,\ldots,k$) can be coded as $-1,0,1$. Let $D_{k\times n}$ be a three-level RSD for k factors in n runs with each factor having the same number of $+1$'s and -1's. We then have $\sum x_i = 0$, $\sum x_i^3 = 0$ and $\sum x_i^2 = \sum x_i^4 = b_i$, where b_i is the number of ± 1 of factor i. Now, impose the following conditions on D:

(i) $\sum x_i^2 x_j = 0\ (i < j)$;

(ii) $\sum x_i^2 x_j x_k = 0\ (i < j < k)$;

(iii) $\sum x_i x_j = 0\ (i < j)$;

(iv) $\sum x_i x_j x_k = 0\ (i < j < k)$;

(v) $\sum x_i x_j x_k x_l = 0\ (i < j < k < l)$; and

(vi) $\sum x_i^2 x_j^2 - b_i b_j / n = 0\ (i < j)$;

where the summations are taken over the n design points. There are, respectively, $q_1 = k(k-1)$ summations in (i), $q_2 = k\binom{k-1}{2}$ summations in (ii), $q_3 = \binom{k}{2}$ summations in (iii), $q_4 = \binom{k}{3}$ summations in (iv), $q_5 = \binom{k}{4}$ summations in (v), and $q_6 = \binom{k}{2}$ summations in (vi). It can be seen that these conditions are the conditions for D to be orthogonal (see Section 10.2 of John, 1971). A 3^k full factorial, an orthogonal design will satisfy all six conditions. The CCDs and BBDs will satisfy the first five conditions (i)–(v), while the first three conditions (i)–(iii) will imply the *OQE* property.

The u-th row of D can be used to construct a vector J_u of length $q = \sum q_i$. Define the first q_1 elements of J_u as $x_{u1}^2 x_{u2}, x_{u1}^2 x_{u3}, \ldots$, the next q_2 elements of J_u as $x_{u1}^2 x_{u2} x_{u3}, x_{u1}^2 x_{u2} x_{u4}, \ldots$, and the last q_6 elements of J_u as $x_{u1}^2 x_{u2}^2 - b_1 b_2/n, x_{u1}^2 x_{u3}^2 - b_1 b_3/n, \ldots$. Let $J_{1\times q} = \sum J_u$. Further define f as the sum of squares of the first $q - q_6$ elements of J, and g as the sum of squares of the last q_6 elements of J. If the value of any element in row u^* of D changes, say from -1 to $+1$ or from 0 to -1, to recalculate J (and consequently f and g), we only have to recalculate just J_{u^*} instead of the entire J_u's. This observation motivates us to propose the following SOD (second-order RSD) algorithm:

1. Start with a random design $D_{k\times n}$. Each column of D has a pre-specified number of 0's and an equal number of $+1$'s and -1's (If the number of 0's in some or all columns of D is 0, these columns will become two-level columns.)

2. Randomly permute the positions of 0's and $+1$'s and -1's in each column.
3. Calculate $J_u,\ u = 1,\ldots,n$ and $J = \sum J_u$. Then evaluate f and g.
4. Sequentially minimize f and g by swapping the positions of $-1,\ +1$ and 0 in each column of D. The algorithm stops when (i) both f and g become 0, i.e. D becomes orthogonal; or (ii) only f becomes 0 (only the first five orthogonality conditions are satisfied) and each $\sum x_i^2 x_j^2 = c$, a constant, i.e. D becomes a slope-rotatable design (see Park, 1987); or (iii) there is no further improvement of f in the swapping.

The above steps correspond to one *try* of SOD. Several tries are recommended to ensure a good resulting design.

Remarks:

- Let D_1 and D_2 be two designs with objective functions f_1, g_1 and f_2, g_2. Design D_1 is preferred over D_2 if $f_1 < f_2$; or $f_1 = f_2$ and $g_1 < g_2$; or $f_1 = f_2$ and $g_1 = g_2$ and d-value for D_1 is higher than d-value for D_2, where d-value= $|X'X|^{1/p}/n$. This d-value, known as "information per point", is a popular measure of goodness of a design.
- SOD can augment additional factors to a *base* design D_b. This makes it possible for easy-to-change factors to be added to a design containing hard-to-change factors (see Parker, Kowalski and Vining (2006).
- SOD can also augment additional runs to a base design D_b. This feature is very handy for sequential experimentation and to construct the SCD*'s in the next Section.

Our SOD algorithm has no difficulty in generating standard RSDs such as CCDs for $k \le 8$ and BBD-type designs for $k \le 7$. Our BBD-type design for $k = 6$ actually improves the the corresponding BBD in terms of rotatability and D- and G-optimality (see `http://designcomputing.net/gendex/sod/`).

As an illustrated example, consider the following eight-point design used at the first stage for an investigation of $k = 5$ factors (excluding the center points):

1	1	1	-1	1
-1	1	1	1	-1
-1	-1	1	1	1
1	-1	-1	1	1
-1	1	-1	-1	1
1	-1	1	-1	-1
1	1	-1	1	-1
-1	-1	-1	-1	-1

These points made up an orthogonal two-level design which was also used in the first stage of the 5-factor APD shown in Table 1 of Morris (2000). Our base design D_b in this case will consist of these eight points. We next add, say 20 runs to this D_b to form a second order design. Given below are 20 cube points from 3^5 augmented to this D_b found by SOD (with the number of 0's in each column set to be eight).

1	-1	-1	0	0
-1	0	0	1	-1
1	0	0	-1	-1
0	-1	0	1	-1
-1	-1	1	0	0
0	-1	-1	0	-1
-1	0	-1	1	0
-1	1	-1	0	0
0	1	0	-1	-1
0	1	1	0	-1
1	0	1	1	0
1	0	-1	-1	0
0	-1	0	-1	1
-1	0	1	-1	0
0	-1	1	0	1
0	1	0	1	1
0	1	-1	0	1
1	1	1	0	0
1	0	0	1	1
-1	0	0	-1	1

It can be shown that the combined design is a 5-factor RSD* (i.e. an RSD with the *OQE* property) in 28 runs. This design has eight runs less than the 5-factor APD of Morris (2000), and in fact, has a higher d-value.

3. Using SOD to Construct SCDs with OQE Property

In this section, the proposed SOD algorithm is applied to SCD*'s, i.e. SCDs with *OQE* property. When it may not be possible to enforce all orthogonality conditions (as in the case of 3^k full factorials) or the first five orthogonality conditions (as in the case of CCDs and BBDs), it is more sensible to enforce just the first three orthogonality conditions. Recall that Conditions (i)–(iii) implies the *OQE* property. To enforce these three conditions, we redefine f as the sum of squares of the first $q_1 + q_2 + q_3$ elements of J, and g as the sum of squares of the next $q_4 + q_5$ elements of J (Since the $\sum x_i^2 x_j^2$ values of an SCD will be n_c, i.e. the size of the SCD's cube portion, it is not necessary to include the last q_6 elements of J in the objective functions.) The basic idea in constructing an SCD* is to construct a *good* augmented design, given a base design D_b. Two types of SCD* are reported here. For Type-I SCD*, the $2k$ axial points are fixed at the second stage, SOD is used to search the *best* first-order design which should be used at the first stage. Here D_b is the design with all $2k$ axial points. For Type-II SCD*, a *small* first-order design is used at the first stage. The $2k$ axial points are anticipated at the second stage. The SOD algorithm is then used to search for the *best* additional design points. Here, D_b is the design with the initial design plus all $2k$ axial points.

As an illustrative example, we show how to construct a Type I SCD* for five factors. Here, the base design D_b consists of a set of 10 axial runs. SOD was used to augment this D_b with the 12 cube points from 2^5. The 12 points obtained below by SOD will be used in the first stage and the set of 10 axial points will be used in the second stage. This is, of course, five columns from a 12-run Plackett and Burman design as shown in Draper and Lin (1990).

-1	-1	-1	-1	-1
-1	1	1	-1	1
-1	1	1	1	1
1	-1	1	-1	1
1	1	-1	-1	1
-1	-1	-1	1	1
1	-1	-1	1	1
1	1	1	1	-1
-1	-1	1	1	-1
-1	1	-1	-1	-1
1	-1	1	-1	-1
1	1	-1	1	-1

Next, we show how to construct a Type-II SCD* for five factors. Suppose an eight-point design, as given in previous section, is used at the first stage. The base design D_b now consists of these eight points plus a set of 10 axial runs. SOD is then used to augment this D_b with additional eight design points from 2^5. Thus, the 18 runs to be conducted at the second stage consists of the set of 10 axial points plus the eight points below, obtained via SOD.

-1	-1	1	-1	-1
1	1	1	-1	-1
-1	1	-1	1	-1
1	1	-1	-1	1
1	-1	-1	1	-1
1	-1	1	1	1
-1	-1	-1	-1	1
-1	1	1	1	1

Type-II SCD*'s, like APDs (but unlike Type I SCD*'s), could have a minimal number of points at the first stage. It can be seen that the number of cube points in an SCD* must be a multiple of four, regardless of whether it is an SCD* of Type I or Type II.

4. Designs for Sequential Experimentation

Table 4.1 displays the d-value $\times 10^3$ (and run sizes n) of selected designs with no center points for sequential experimentations. For SCDs and CCDs, α is set to 1. Unlike BBDs which can only be used non-sequentially, these designs can be used either sequentially or non-sequentially. The first two columns of Table 4.1 are the number of factors k and parameters p. The 3rd column is associated with the SCDs of Draper and Lin (1990). The columns of these SCDs were selected from the appropriate Plackett and Burman designs (Plackett and Burman, 1946). Only SCDs for $k = 3$, 4 and 6 are SCD*'s, as none of the

Table 4.1. Comparison of d-value $\times 10^3$ (and run sizes) of SCDs and CCDs ($\alpha = 1$ and $n_0 = 0$) and APDs ($n_0 = 0$).

k	p	SCDs of Draper & Lin (1990)	SCD* Type I	SCD* Type II	CCDs	APDs
3	10	303 (10)	303 (10)	303 (10)	463 (14)	303 (10)
4	15	308 (16)	308 (16)	308 (16)	457 (24)	373 (36)
5	21	241 (21)	259 (22)	355 (26)	440 (26)	308 (36)
6	28	263 (28)	263 (28)	368 (36)	456 (44)	298 (36)
7	36	196†(36)	262 (38)	226 (38)	465 (78)	269 (36)
8	45	221†(46)	280 (48)	252 (48)	474 (80)	272 (78)
9	55	200†(56)	246 (58)	231 (58)	480 (146)	253 (78)
10	66	165†(66)	224 (68)	207 (68)	493 (148)	238 (78)

†We have improved the d-value $\times 10^3$ of these designs for $k = 7$, 8, 9, 10 to 234, 243, 232 and 219 respectively.

rows of the selected columns of the Plackett and Burman designs for these SCDs is deleted. The remaining columns of Table 4.1 are associated with SCD*'s of Types I and II, the CCDs and APDs. For SCDs, Type I SCD*'s and CCDs, the numbers of runs in the first and second stages are $n - 2k + n_0$ and $2k$ respectively. For APDs and Type II SCD*'s, the number of runs in the first and second stages and are $4 + n_0$ and $n - 4$ (for $k = 3$), $8 + n_0$ and $n - 8$ (for $k = 4$, 5, 6 and 7), and $12 + n_0$ and $n - 12$ (for $k = 8$, 9, and 10) respectively. The cube points for the 5-factor SCD*'s Types I and II are shown in the previous Section.

It is interesting to note that all SCD, SCD* and APD for $k = 3$ have a similar structure: a saturated orthogonal two-level design in four runs and six axial runs. It can be seen that no class of design in Table 4.1 is a clear winner. If the experimenters wish to conduct their experiments sequentially and do not wish to spend a lot of resources initially, Type II SCD*'s and APDs are attractive alternatives. All Type II SCD*'s for $k \leq 6$ and the APD for $k = 7$ have high d-value.

When the experiment is conducted in a single stage, the SCDs of Draper and Lin (1990) and Type I SCD*'s should be considered if the runs are expensive or when an independent estimate of error is available while CCDs are highly recommended if resources are readily available and a high degree of the precision of parameter estimates is expected. Note that for $k = 7$, 8, 9 and 10, with just two additional runs, Type I SCD*'s increase the d-value of SCDs of Draper and Lin (1990) substantially.

SCD*'s and APDs can be viewed as good substitutes to BBDs for two reasons: (i) these designs have far fewer runs than BBDs; and (ii) the percentage of the 0-level of each factor (the level of least interest to the experimenters) of these designs is more acceptable than those of BBDs. The number of runs including the recommended number of center points of BBDs for 3-7 factors are 15, 27, 46, 54, and 62 runs respectively. The percentages of the 0-level for each factor of BBDs are 47, 56, 65, 56 and 61% respectively.

Designs in Table 4.1 are available at `http://designcomputing.net/SCD/`. The SCDs for $k = 7$, 8, and 9 of Draper and Lin (1990) have also been improved by Angelopoulos, Evangelaras and Koukouvinos (2009) using complete search. Complete search seems only feasible for seven or less factors. For more than seven factors, we have to resort to heuristic methods. Angelopoulos, Evangelaras and Koukouvinos (2009) have discussed the maximization of the rotatability index Q^* (see Draper and Pukelsheim, 1990) by varying the α values in the axial runs.

5. Concluding Remarks

An algorithm for construction of SCD*'s is proposed. These are SCDs with the *OQE* property, i.e. the property that the quadratic effects are orthogonal to all main-effects and interaction effects. These designs are not only more efficient but also more flexible than those of Draper and Lin (1990). The purpose of this paper is, however, not just to provide a catalogue of designs for sequential experimentation but to introduce an algorithm to construct this type of design (`http://designcomputing.net/gendex/sod/`). An experimenter looking for a 6-factor RSD can use this algorithm to construct a Type I SCD* for six factors in 32 runs with d-value=0.322 instead of using any 6-factor design in Table 4.1. This SCD* requires four less runs than the corresponding APD and at the same time has a higher value of d-value than the latter. As mentioned, RSM is an iterative process. Consider an experiment for eight factors using an orthogonal two-level design for 12 runs in the first stage. In the second stage, the experimenter might decide to drop the two non-significant factors and augment these 12 runs with additional runs so that the resulting design is a good second order one. The APD algorithm requires 66 additional runs. The resulting design is an APD with 78 runs and d-value=0.317. Our algorithm requires only 24 additional runs (12 axial runs plus 12 runs from a 2^6). The resulting design is an SCD* with 36 runs and d-value=0.359.

References

Angelopoulos, P., Evangelaras, H., Koukouvinos, C., 2009. Small, balanced, efficient and near-rotatable central composite designs. *J. of Statist. Planning and Inference,* 139, 2010–2013.

Box, G.E.P., Behnken, D.W., 1960. Some new three level designs for the study of quantitative variables. *Technometrics,* 2, 455–477.

Box, G.E.P., Wilson, K.B., 1951. On the experimental attainment of optimum conditions. *J. Roy. Statist. Soc. Ser. B,* 13, 1–45.

Draper, N.R., Lin, D.K.J., 1990. Small response surface designs. *Technometrics,* 32, 187–194.

Draper, N.R., Lin, D.K.J., 1996. Response Surface Designs. In *Handbook of Statistics,* Vol. 13, Ghosh, S. and Rao, C.R., (Editors), Elsevier Science, 343–375.

Draper, N.R., Pukelsheim, F., 1990. Another look at rotatability. *Technometrics,* 32, 195–202.

John, P.M.W., 1971. *Statistical Design and Analysis of Experiments.* McMillan.

Morris, M.D., 2000. A class of three-level experimental designs for response surface. *Technometrics,* 42, 111–121.

Myers, R.H., Montgomery D.C., Anderson-Cook C.M., 2009. *Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 3rd ed.* Wiley.

Nguyen, N-K., Borkowski, J.J., 2008. New three-level response surface designs constructed from incomplete block designs. *J. of Statist. Planning and Inference,* 138, 294–305.

Park, S.H., 1987. A class of multifactor designs for estimating the slope of response surfaces. *Technometrics,* 29, 449–453.

Parker, A.P., Kowalski, S.M., Vining, G.G., 2006. Classes of split-plot response surface designs for equivalent estimation. *Quality and Reliability Engineering International,* 22, 291–305.

Plackett, R.L., Burman, J.P., 1946. The designs of optimum multi-factorial experiments. *Biometrika,* 33, 305–325.