Journal of Korean
Data & Information Science Society
2003, Vol. 14, No.3 pp. 707～714

A New Approach for Selecting Fractional Factorial Designs[1]

DongKwon Park[2] · HyoungSoon Kim[3]

Abstract

Because of complex aliasing, nonregular designs have traditionally been used for screening only main effects. However, complex aliasing actually may allow some interactions entertained and estimated without making additional runs. According to hierarchical principle, the minimum aberration has been used as an important criterion for selecting regular fractional factorial designs. The criterion is not applicable to nonregular designs. In this paper, we give a criterion for choosing fractional factorial designs based on the fan theory. The criterion is focused on the partial ordering given by set inclusion on estimable sets which is called leaves.

Key words and Phrases: Maximal fan design, Minimum aberration, Locally maximal design, Nonregular design;

1. Introduction

Factorial experiments are conducted for simultaneously investigating a number of factors. If total runs consist of all possible combinations of the levels of the different factors, the experiment is called a complete factorial experiment. Often the run size may be too large to carry out a complete factorial experiment because of expensive cost and time limitation. For these reasons, we need to choose a fraction of the possible factorial combinations, which is called a fractional factorial (FF, for short) design.

Our primary concern is how to choose a good FF design from a complete factorial experiment. When we perform only a fraction of the complete factorial

---

1) This work was supported by grant No.2001-2-0226 from the Basic Research Program of the Korea Science & Engineering Foundation

2) Department of Statistics, Yonsei University, Wonju, 220-710, Korea.
3) Department of Mathematics, Yonsei University, Wonju, 220-710, Korea.

experiment, some factorial effects are aliased with some other factorial effects.

Most of the background theory of designs is related with regular FF designs. For regular designs, any two factorial effects can either be estimated independently of each other or fully aliased. A regular design is uniquely determined by independent defining words. Designs that do not possess this property are called nonregular designs, which include many mixed-level orthogonal arrays.

According to hierarchical principle, the minimum aberration (MA, for short) (see Fries and Hunter (1980) or Wu and Zhang (1993)) has been used as an important criterion for selecting regular FF designs. According to the MA criterion, to choose optimal FF designs we just sequentially minimize the wordlength. This criterion is basically same with sequentially minimizing the numbers of alias relations between factorial effects (Zhang and Park (2000)).

For reasons of run size economy or flexibility, nonregular designs may be used. For nonregular designs, some factorial effects may neither be uncorrelated nor fully aliased, that is, they have an absolute correlation strictly between 0 and 1. In these designs, the aliasing of effects may have a complex pattern, and are therefore referred to as designs with complex aliasing. Because of complex aliasing, nonregular designs have traditionally been used for screening only main effects. However, complex aliasing actually may allow some interactions entertained and estimated without making additional runs(see, for example, Wang and Wu (1995) for Plackett-Burman design). For nonregular designs, MA criterion cannot be applied. This motivates us to propose criteria for selecting FF designs.

In this paper, we propose a new approach for choosing FF designs based on the fan theory. The criterion is based on the partial ordering given by set inclusion on estimable sets which is called leaves.

2. Maximal Fan Design

In any particular problem we expect to find a design giving maximal estimable models(leaves in terms of fan theory). A maximal fan design is one for which every possible leaf can be estimated. It is based on the theory of aliasing to the study the fan of designs which is full range of estimable polynomial models for a particular design. Optimal fractions are chosen according to the number of leaves. Obviously, a maximal fan design, if exists, is the best with respect to estimability.

Consider for an example the 2^{3-1} regular fractional factorial design d_1 in three factors, x_1, x_2 and x_3, with defining relation $I = x_1x_2x_3$ consisting of the four points ($n=4$) with {(0,0,0), (0,1,1), (1,0,1), (1,1,0)}. The alias relationships are $x_1 = x_2x_3$, $x_2 = x_1x_3$ and $x_3 = x_1x_2$. It means that, for example, x_1 and x_2x_3 can not be estimated separately in a full model. A reduced (saturated) estimable model with estimable set $\{1, x_1, x_2, x_1x_2\}$ is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$$

One can check that it separately identifies the following three estimable sets of size $n=4$ $\{1, x_1, x_2, x_3\}$,$\{1, x_1, x_3, x_1x_3\}$ and $\{1, x_2, x_3, x_2x_3\}$. We will call this each estimable set as a leaf and the collection of leaves as a fan. Moreover, there are no other saturated estimable models by a 4-point designs. In fact, the collection of these four models is a maximal fan.

Let $\beta = (\beta_1, \ldots, \beta_m)$ denote a multi-exponent. An estimable set should satisfy divisibility condition(D) which is that if a term $x^{\beta} = x_1^{\beta_1} \cdots x_m^{\beta_m}$ is in the set, then every term which divides x^{β} is also in the set. Note that the size of a leaf is always equal to the sample size n of distinct points and leaves are saturated (see Kim, Park and Kim (2002)).

Definition 1. For fixed $\alpha = (\alpha_1, \ldots, \alpha_m)$, a finite set L of n monomials on variables $\{1, x_1, x_2, \ldots, x_m\}$ is called a *leaf* for fixed (n, α) if it satisfies D-condition and the power of each variable x_i of monomials is less than α_i.

Let $\Lambda(n, \alpha)$ denote the set of all leaves for (n, α). Note that the number of all possible leaves for given n and $\alpha = (\alpha_1, \ldots, \alpha_m)$ is same as the number $p(I^{\alpha}, n)$ of partitions of positive integer n in the multi-dimensional integer grid $I^{\alpha} = \{0, 1, \ldots, \alpha_1 - 1\} \times \ldots \times \{0, 1, \ldots, \alpha_m - 1\}$.

We denote by $D(n, \alpha)$ the set of all possible n-point designs in the grid I^{α} and denote the design matrix for a leaf L at d in $D(n, \alpha)$ by $X(L, d)$ and the determinant of a matrix A by $det[A]$.

Definition 2. A leaf L is called *estimable* by a design d if its design matrix $X(L, d)$ at d is invertible or equivalently $det[X(L, d)] \neq 0$.

Let $E(d)$ be the collection of estimable leaves of d and we call this a fan of d.

Definition 3. A design d in $D(n, \alpha)$ is called a *maximal fan* design if $E(d)$ is the same set of all possible leaves, that is, $E(d) = \Lambda(n, \alpha)$.

Existence of a maximal fan design depends on the number n of design points and the size α of multi-dimensional integer grid I^{α}. We give an example as an illustration.

Example 1. Consider *FF* designs with 4 design points from mixed level experiments $D(4,(2,3,2))$. There are 6 leaves in $\Lambda(4,(2,3,2))$ as follows :

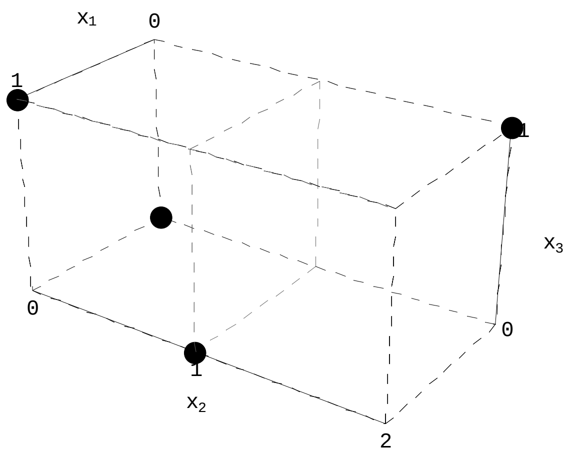
$L_1 = \{1, x_1, x_2, x_1x_2\}$ $L_2 = \{1, x_1, x_2, x_2^2\}$ $L_3 = \{1, x_1, x_2, x_3\}$

$L_4 = \{1, x_1, x_3, x_1x_3\}$ $L_5 = \{1, x_2, x_2^2, x_3\}$ $L_6 = \{1, x_2, x_3, x_2x_3\}$

By using an exhaustive method, we can verify only eight maximal fan designs out of 495 four-point designs in the grid $I^{\alpha} = \{0,1\}\times\{0,1,2\}\times\{0,1\}$ exist. They are all isomorphic in sense that a design can be obtained by switching the levels from another design. A design as shown Figure 1 is

$$d = \{\{0,\ 0,\ 0\},\ \{1,\ 1,\ 0\},\ \{1,\ 0,\ 1\},\ \{0,\ 2,\ 1\}\}$$

The corresponding $\{det[X(L_i, d)]|\ i=1,2,...,6\} = \{2,\ -2,\ -3,\ -1,\ -2,\ 2\}$ and hence all not zero, so the design is a maximal fan design and best.



[Figure 1 : A maximal fan design for 2×3×2 mixed level design]

3. Locally maximal designs

If there exists a maximal fan design like Example 1, it would be the best choice of *FF* design. Actually, a maximal fan design does not always exist and the existence problem of maximal fan design is still open. Caboara, Pistone, Riccomago and Wynn (1997) proved that a maximal fan design with n distinct points in m dimensions always exist. Then they conjectured that a maximal fan design on the integer grid $\{0,1,2,\ldots,n-1\}^m$ exists for any n and m. It was proved partially in Kim, Park and Kim (2002) that maximal fan design for the case α is

$n^m = (n, n, \ldots, n)$ must be a latin hypercube.

In case of nonexistence of a maximal fan design, it is very natural to think that a design which has more estimable leaves is better and hence we give the following definition.

Definition 4.

(i) Suppose that d_1 and d_2 are two *FF* designs in $D(n, \alpha)$. Define an equivalence relation on the set of designs by $d_1 \sim d_2$ if $E(d_1)$ is equal to $E(d_2)$ and say that d_1 has the *same power* of d_2.

(ii) Define an order relation on the set of equivalence classes $D(n, \alpha)/\sim$ by $[d_1] \ll [d_2]$ if $E(d_2)$ contains $E(d_1)$ and say that d_1 has *more power* than d_2.

The order relation defined above obviously gives a partial ordering in the set $D(n, \alpha)/\sim$. We say a design d is *locally maximal* if there is no design which has more power than d.

Example 2. Consider *FF* designs with 8 design points from symmetric experiments $D(8, (2, 2, 2, 2))$. There are 24 leaves in $\Lambda(8, (2, 2, 2, 2))$ as follows :

$L_1 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_1x_4\}$ $\quad$ $L_2 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_2x_4\}$

$L_3 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_3x_4\}$ $\quad$ $L_4 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_1x_4\}$

$L_5 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_2x_4\}$ $\quad$ $L_6 = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_3x_4\}$

$L_7 = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_2x_3, x_1x_4\}$ $\quad$ $L_8 = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_2x_3, x_2x_4\}$

$L_9 = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_2x_3, x_3x_4\}$ $\quad$ $L_{10} = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_4, x_2x_4\}$

$L_{11} = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_4, x_3x_4\}$ $\quad$ $L_{12} = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_2x_4, x_3x_4\}$

$L_{13} = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_1x_4, x_2x_4\}$ $\quad$ $L_{14} = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_1x_4, x_3x_4\}$

$L_{15} = \{1, x_1, x_2, x_3, x_4, x_1x_3, x_2x_4, x_3x_4\}$ $\quad$ $L_{16} = \{1, x_1, x_2, x_3, x_4, x_2x_3, x_1x_4, x_2x_4\}$

$L_{17} = \{1, x_1, x_2, x_3, x_4, x_2x_3, x_1x_4, x_3x_4\}$ $\quad$ $L_{18} = \{1, x_1, x_2, x_3, x_4, x_2x_3, x_2x_4, x_3x_4\}$

$L_{19} = \{1, x_1, x_2, x_3, x_4, x_1x_2, x_1x_3, x_2x_3\}$ $\quad$ $L_{20} = \{1, x_1, x_2, x_3, x_4, x_1x_4, x_2x_4, x_3x_4\}$

$$L_{21} = \{1, x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}$$
$$L_{22} = \{1, x_1, x_2, x_4, x_1x_2, x_1x_4, x_2x_4, x_1x_2x_4\}$$
$$L_{23} = \{1, x_1, x_3, x_4, x_1x_3, x_1x_4, x_3x_4, x_1x_3x_4\}$$
$$L_{24} = \{1, x_2, x_3, x_4, x_2x_3, x_2x_4, x_3x_4, x_2x_3x_4\}$$

By using an exhaustive method, we found out that no maximal fan design out of 12870 eight-point designs in the grid $I^{\alpha} = \{0, 1\} \times \{0, 1\} \times \{0, 1\} \times \{0, 1\}$.

We consider the following two designs:

d_1:{(0, 0, 0, 0),(1, 1, 0, 0),(1, 0, 1, 0),(0, 1, 1, 0),(1, 0, 0, 1),(0, 1, 0, 1),(0, 0, 1, 1),(1, 1, 1, 1)}
d_2:{(0, 0, 0, 0),(1, 0, 0, 0),(1, 0, 1, 0),(0, 1, 1, 0),(1, 0, 0, 1),(0, 1, 0, 1),(0, 0, 1, 1),(1, 1, 1, 1)}

Design d_2 has the same design points as d_1 except one point as shown in Figure 2. A work on computer shows that two designs are locally maximal and the corresponding determinants of design matrices and estimable leaves are as follows:

$$\{det[X(L_i, d_1)]\}_{i=1}^{24} = \{-4, 0, 0, 0, 4, 0, 0, 0, 4, -4, 0, 0, 0, -4, 0, 0, 0, 4, 4, -4, 1, 1, 1, 1\}$$

$$\{\det[X(L_i, d_2)]\}_{i=1}^{24} = \{-2, 1, 1, 1, 0, -1, -1, 1, 2, -1, 1, -1, -1, -3, 1, 1, 1, 1, 1, -2, 0, 0, 1, 0\}$$

$$E(d_1) = \{L_1, L_5, L_9, L_{10}, L_{14}, L_{18}, L_{19}, L_{20}, L_{21}, L_{22}, L_{23}, L_{24}\}$$
$$E(d_2) = \Lambda(8, (2,2,2,2)) - \{L_5, L_{21}, L_{22}, L_{24}\}$$

Locally maximal designs d_1 and d_2 are the best if we wish to estimate estimable leaves $E(d_1)$ and $E(d_2)$ or less respectively. However, the comparison directly among locally maximal designs is meaningless because $E(d_i)'$s do not contain each other. If an ordering is given as the importance such as the hierarchical principle on factorial effects or leaves, then we could find locally maximal designs which are most effective based on the ordering.

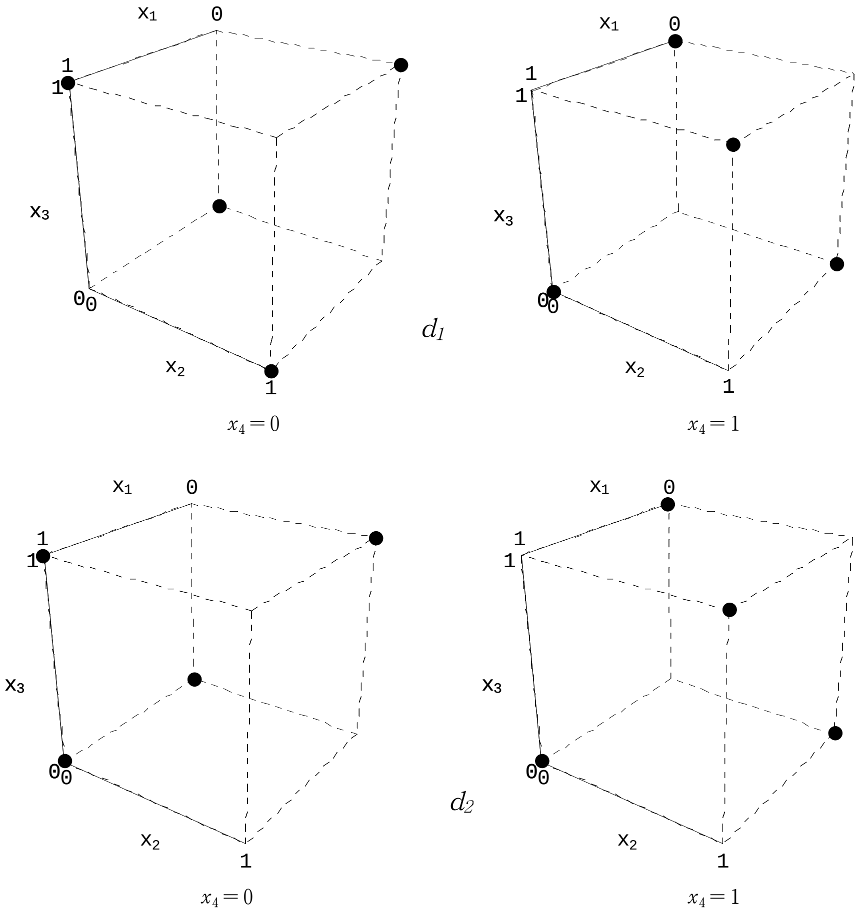
Usually, we have assumed that the following hierarchical principle.

(i) Lower-order interactions are more likely to be important than higher-order interactions.

(ii) Interactions of the same order are equally likely to be important.

Example 3. Consider the two locally maximal designs d_1 and d_2 in Example 2. Under the usual hierarchical principle, we say that $L1 \sim L20$ is more important than $L21 \sim L24$ which are not able to estimate all main effects, and leaves in $L1 \sim L20$ are equally good. Thus, under the hierarchical principle, we say that a design which has more estimable leaves in $L1 \sim L20$ than the other design is better. In this respect, d_2 is much better than d_1.

Actually, d_1 is a regular design and is formed by the defining relation $I = x_1x_2x_3x_4$; the other hand d_2 is a nonregular design. From algebraic relation, x_1x_2 and x_3x_4, for example, can not be estimated simultaneously in d_1. Even though d_1 is a minimum aberration design and best among regular designs, the design is not allowed to have enough estimable leaves from the beginning because of very strong algebraic restrictions. That is a good reason why we have to look at nonregular designs in respect to estimability.



[Figure 2 : Two *FF* designs for 2^4-experiment]

4. Concluding Remarks

Traditionally, nonregular designs were not advocated because of their complex aliasing structure. However, in last decade, they have received increasing attention in the literature. The problem is how to assess, compare and rank nonregular designs in a systematic fashion. In this paper, we propose maximal and locally maximal criteria for this purpose, paralleling the minimum aberration criterion used for assessing regular designs. In principle, it is not hard to apply the criterion given here. However, as n or α goes to larger, difficulties of calculating arise more rapidly. We have lots of work to do to make easier comparisons.

References

1. Caboara, M., Pistone, G., Riccomagno. E. and Wynn, H.P. (1997). The fan of an experimental design, Unpublished Manuscipt.
2. Fries, A. and Hunter, W.G. (1980). Minimum aberration 2^{n-p} designs, *Technometrics*, 26, 225-232.
3. Kim, H.S., Park,D.K. and Kim, K.H. (2002). On the existence of maximal fan design, *Korean Communications in Statististics*, 9, 347-354.
4. Wang, J. C. and Wu, C.F.J. (1995). A hidden projection property of Plackett-Burman and related designs, *Statistica Sinica*, 5, 235-250.
5. Wu, C.F.J. and Zhang, R.C. (1993). Minimum aberration designs with two-level and four-level factors, *Biometrika*, 80, 203-209.
6. Zhang, R.C. and Park, D.K. (2000). Optimal blocking of two-level fractional factorial designs, *Journal of Statistival Planning and Inference*, 91(1), 107-121.

[received date : May. 2003, accepted date : Aug. 2003]