The Korean Communications
in Statistics Vol. 9, No. 2, 2002
pp. 347-354

On the Existence of Maximal Fan Design[1)]

Hyoungsoon Kim[2)], Dongkwon Park[3)], KyungHee Kim[4)]

Abstract

An n-point design is maximal fan if all the models with n-terms satisfying the divisibility condition are estimable. Such designs tend to be space filling and look very similar to the "Latin-hypercube" designs used in computer experiments. Caboara, Pistone, Riccomago and Wynn (1997) conjectured that a maximal fan design on an integer grid exists for any n and m, where m is the number of factors. In this paper we examine the relationship between maximal fan design and latin-hypercube to give a partial solution for the conjecture.

Keywords : Algebraic geometry; Gröbner bases; Ideal; Leaf; Maximal fan 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 cannot 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 (hereafter, *FF*) design.

Our primary concern is how to choose a *FF* design from full factorial experiments. When we perform only a fraction of the complete factorial experiment, some factorial effects are aliased(or confounded) with some other factorial effects. In other words, in order to estimate a factorial effect, we have to sacrifice other factorial effects. Most of the background theory of designs is related with regular *FF* designs that have relatively simple aliased structure.

Regular designs form a subgroup or coset according to defining relation. The aliased structure for a regular design may be easily determined by using a defining relation. Criteria

---

1) Kim's work was supported by Yonsei University Research Fund of 1998. Park's work was supported by Yonsei University Research Fund of 1999.

2) Associate Professor, Department of Mathematics, Yonsei University, Wonju, 220-710, Korea
E-mail : kimh@dragon.yonsei.ac.kr

3) Professor, Department of Statistics, Yonsei University, Wonju, 220-710, Korea
E-mail : statpdk@dragon.yonsei.ac.kr

4) Associate Professor, Department of Mathematics, Yonsei University, Wonju, 220-710, Korea
E-mail : khlkim@dragon.yonsei.ac.kr

for selecting of an efficient FF design such as resolution (Box and Hunter, 1961) or minimum aberration (Fries and Hunter, 1980) are based on the number of aliased factors.

Consider a 2^{3-1} two-level FF design, which has three factors(x_1, x_2 and x_3) with two levels 0 and 1, and 2^2 runs and is determined by defining relation $I = x_1x_2x_3$. The design consists of the four points with {(0,0,0), (0,1,1), (1,0,1), (1,1,0)}. The aliased structure is $x_1 = x_2x_3$, $x_2 = x_1x_3$ 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 is

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

with estimable set $\{1, x_1, x_2, x_1x_2\}$. One can check that it separately identifies three more estimable sets $\{1, x_1, x_2, x_3\}$,$\{1, x_1, x_3, x_1x_3\}$ and $\{1, x_2, x_3, x_2x_3\}$.

On the other hand, non-regular designs have traditionally been used for screening only main effects because of complex aliasing. Piston and Wynn (1996) introduced algebraic geometry ideas into the design of experiments and showed how the theory of Gröbner bases(G-bases) can be used to find a saturated estimable factorial effect(linear polynomial model) for a given design. This allows the study of "generalized aliasing" in non-regular designs. In section 2, we introduce the basic fan theory using the algebraic geometry (see Cox, O'Shea(1992) for details).

Our main results are included in the section 3. In any particular problem we expect to find a design giving a maximal estimable model. An n-point design is *maximal fan* if all the models with n-terms satisfying the divisibility condition are estimable. Such designs tend to be space filling and look very similar to the "latin-hypercube" designs used in computer experiments. Caboara, Pistone, Riccomago and Wynn (1997) conjectured that a maximal fan design exists on an integer grid for any n and m. In this paper we examine the relationship between maximal fan design and latin-hypercube to give a partial solution for the conjecture. If an n-point maximal fan design exists on the integer grid $\{0, 1, 2, \ldots, n-1\}^m$, then the design should be a latin-hypercube. However, not every latin-hypercube design is a maximal fan design.

2. Fan Theory

The starting point is to consider a design point as an algebraic variety. Then one may express the design points as the solution given by the intersection of polynomial ideals. Given a design $d = \{x^{(1)}, \ldots, x^{(n)}\}$, if $x = (x_1, \ldots, x_m)$ are the independent factors and $p(x)$ is a polynomial model and $g_1(x), \ldots, g_r(x)$ are the polynomials with respect to $x = (x_1, \ldots, x_m)$ forming a G-basis then

$$p(x) = \sum_{j=1}^{r} s_j(x) g_j(x) + r(x)$$

where $r(x)$ is unique and of low order than $g_j(x)$. The equation forces $p(x) = r(x)$ for all x in the design d.

We say that two models(factorial effects) $p_1(x)$ and $p_2(x)$ are aliased relative to a design d and a G-basis if they have the same remainder $r(x)$ with respect to the G-basis or if they are equal to at all x in the design d.

Let Q represent the rational numbers and $I(d)$ (or I) be the design ideal for a given ordering τ. The set of all remainders is in one-to-one correspondence with the quotient ring $Q[x_1, \ldots, x_m]/I$. In addition the dimension of the vector space $Q[x_1, \ldots, x_m]/I$ is exactly equal to the number of distinct experimental design points. Pistone and Wynn (1996) stated the fact, and we'll give a full proof in Theorem 2.1.

Theorem 2.1 The dimension of the vector space $Q[x_1, \ldots, x_m]/I$ is exactly equal to the number of distinct design point n those used to define the ideal I.

Proof : Note that the ideals defined by each design point has co-dimension 1, that is, $\dim(Q[x_1, \ldots, x_m]/I(\{p_i\})) = 1$. Therefore it is clear that the dimension of the vector space $Q[x_1, \ldots, x_m]/(I_1 \cap \cdots \cap I_n)$ is less than or equal to n where $I_j = I(\{p_j\})$.

We will show that the dimension is equal to n using induction. For the case $n=1$, it is obvious. Now, we assume that our assertion holds when $n=q$. Then the dimension of $Q[x_1, \ldots, x_m]/(I_1 \cap \cdots \cap I_q)$ is equal to q. Let I_{q+1} be the ideal generated by another distinct design point $p_{q+1} = (b_1, \ldots, b_m) \in Q^m$ and $I = I_1 \cap \cdots \cap I_{q+1}$.

Since the design point p_{q+1} is distinct from $p_1, \ldots, p_q$, there exist distinct coordinates, say $a_{1j_1}, \ldots, a_{qj_q}$ such that a_{ij_i} is a coordinate of p_i $(1 = 1, \ldots, q)$ and $a_{ij_i} \neq b_{j_i}$. Then the polynomial $f(x) = (x_{j_1} - a_{1j_1})(x_{j_2} - a_{2j_2}) \cdots (x_{j_q} - a_{qj_q})$. Clearly belongs to $I_1 \cap \cdots \cap I_q$ but it does not to I since $f(p_{q+1}) \neq 0$. (Note that $f(x) \in I(p) \Leftrightarrow f(p) = 0$). Therefore the dimension of $I_1 \cap \cdots \cap I_q / I$ is at least one and hence the dimension $Q[x_1, \ldots, x_m]/I$ is at least $q+1$. According to the previous assertion, we have $\dim(Q[x_1, \ldots, x_m]/I_1 \cap \cdots \cap I_{q+1}) = q+1$ and this completes the proof.

A vector-space basis of the remainder set $Q[x_1, \ldots, x_m]/I$ is calculated as the set of terms not divisible by any leading term in the G-basis. It follows that $Q[x_1, \ldots, x_m]/I$ is the set of all models identifiable by the design with respect to the ordering τ. In particular the elements of a vector space basis of $Q[x_1, \ldots, x_m]/I$ gives the terms of a saturated model identifiable using d. This is the set $E_{d,\tau}$, called a *leaf* of d, and the remainder is a linear combination of elements of $E_{d,\tau}$. The leaf $E_{d,\tau}$ depends on ordering τ. The collection of $E_{d,\tau}$ for all τ is called the *fan* of d. A leaf should satisfy divisibility condition(D) which is that if a term $x^{\alpha}=x_1^{\alpha_1}\cdots x_m^{\alpha_m}$ is in $E_{d,\tau}$ then every term which divides x^{α} is also in $E_{d,\tau}$. As we mentioned before, the size of a leaf is always equal to the sample size n of distinct points and leaves are saturated.

Let $F(d)$ be the fan of the design d and $E(d)$ be the set of models satisfying the D -condition and with n terms and such that their design matrices at d are invertible. We say that the elements of $F(d)$ are algebraically identifiable and elements of $E(d)$ are identifiable in a statistical sense. Caboara, Pistone, Riccomago and Wynn (1997) stated $F(d) \subseteq E(d)$, we give a full proof in Theorem 2.2.

Theorem 2.2 Let $E_{d,\tau} = \{ p_1(x), \ldots, p_n(x) \}$ be a particular leaf of the fan of d. Then the design matrix

$$X(E_{d,\tau}) = \{p_j(x^{(i)})\}_{i,j=1}^{n}$$

is invertible.

Proof : Suppose that $X(E_{d,\tau})$ is not invertible. Then the column vectors are not independent, that is, there exists $(\alpha_1, \ldots, \alpha_n)$ which are not all zero such that

$$\alpha_1 p_1(x^{(i)}) + \ldots + \alpha_n p_n(x^{(i)}) = 0, \text{ for all } i=1,2,\ldots,n.$$

This implies that the polynomial $\alpha_1 p_1(x) + \ldots + \alpha_n p_n(x)$ is an element of the ideal I generated by the design d. Therefore we obtain the equation $\alpha_1 p_1(x) + \ldots + \alpha_n p_n(x) = 0$ in the vector space $Q[x]/I$. This contradicts the linear independency of $p_1, \ldots, p_n$ and the proof is completed.

3. Maximal Fan Design

Caboara, Pistone, Riccomago and Wynn (1997) proved that a maximal fan design with n distinct points in m dimensions always exist when the range of levels of the factors is not restricted. However, because the situation is not realistic, they conjectured that a maximal fan design on the integer grid $\{0,1,2,\ldots,n-1\}^m$ exists for any n and m. Such designs tend to be space filling and look very similar to the "latin-hypercube" designs used in computer experiments. From some examples the conjecture looks very affirmative.

In this paper we examine the relationship between maximal fan design and latin-hypercube to give a partial solution for the conjecture. If a maximal fan design exists, then the design should be a latin-hypercube. However, not every latin-hypercube design is a maximal fan design. Also, the following theorem shows that the size of the integer grid cannot be smaller than n in order to make d a maximal fan design.

Theorem 3.1 Let d be a maximal fan design with n distinct points on the integer grid $\{0,1,2,\ldots,n-1\}^m$. Then it is a latin-hypercube.

Proof : If d is not a latin-hypercube, then the same number appears in a certain component, say x_k, of the points in d. This implies that there are the same rows in the design matrix $X(E)=\{p_j(x^{(i)})\}_{i,j=1}^n$ where $E=\{1,x_k,x_k^2,\ldots,x_k^{n-1}\}$, and hence d cannot be a maximal fan design. By the contrapositive statement, the theorem is proven.

Examples of $n(=3,\ldots,8)$-point maximal fan designs on the integer grid $\{0,1,2,\ldots,n-1\}^2$ are shown in Figure 1. Those designs are found by an exhaustive search with computer. We easily verify that the designs are latin-hypercubes by looking and the assertion in Theorem 3.1 is justified.

The following proposition eliminates the special latin-hypercubes which are not maximal fan designs.

Proposition 3.1 Let $m=2$, $n\geq 6$, and d be a design on the grid $\{0,1,2,\ldots,n-1\}^2$ consists of n distinct points which lie on any two straight lines. Then d is not a maximal fan design.

Proof : Suppose that every point of d lies on either $a_1x_1+b_1x_2=c_1$ or $a_2x_1+b_2x_2=c_2$, then it satisfies the equation

$$\begin{aligned}0&=(a_1x_1+b_1x_2-c_1)\ (a_2x_1+b_2x_2-c_2)\\&=c_1c_2-(a_1c_2+a_2c_1)x_1-(b_1c_2+b_2c_1)x_2+a_1a_2x_1^2+b_1b_2x_2^2+(a_1b_2+a_2b_1)x_1x_2\end{aligned}$$

This implies that the design matrix $X(E)$ where $E=\{1,x_1,x_1^2,x_2,x_1x_2,x_2^2\}$ is not invertible. Therefore d cannot be a maximal fan design.

4. Conclusive Remark

For the case m=2 and n=6, we may separate 84 cases out of 136 non-maximal fan designs of latin hypercube type by simply checking if the design points are on any two straight lines.

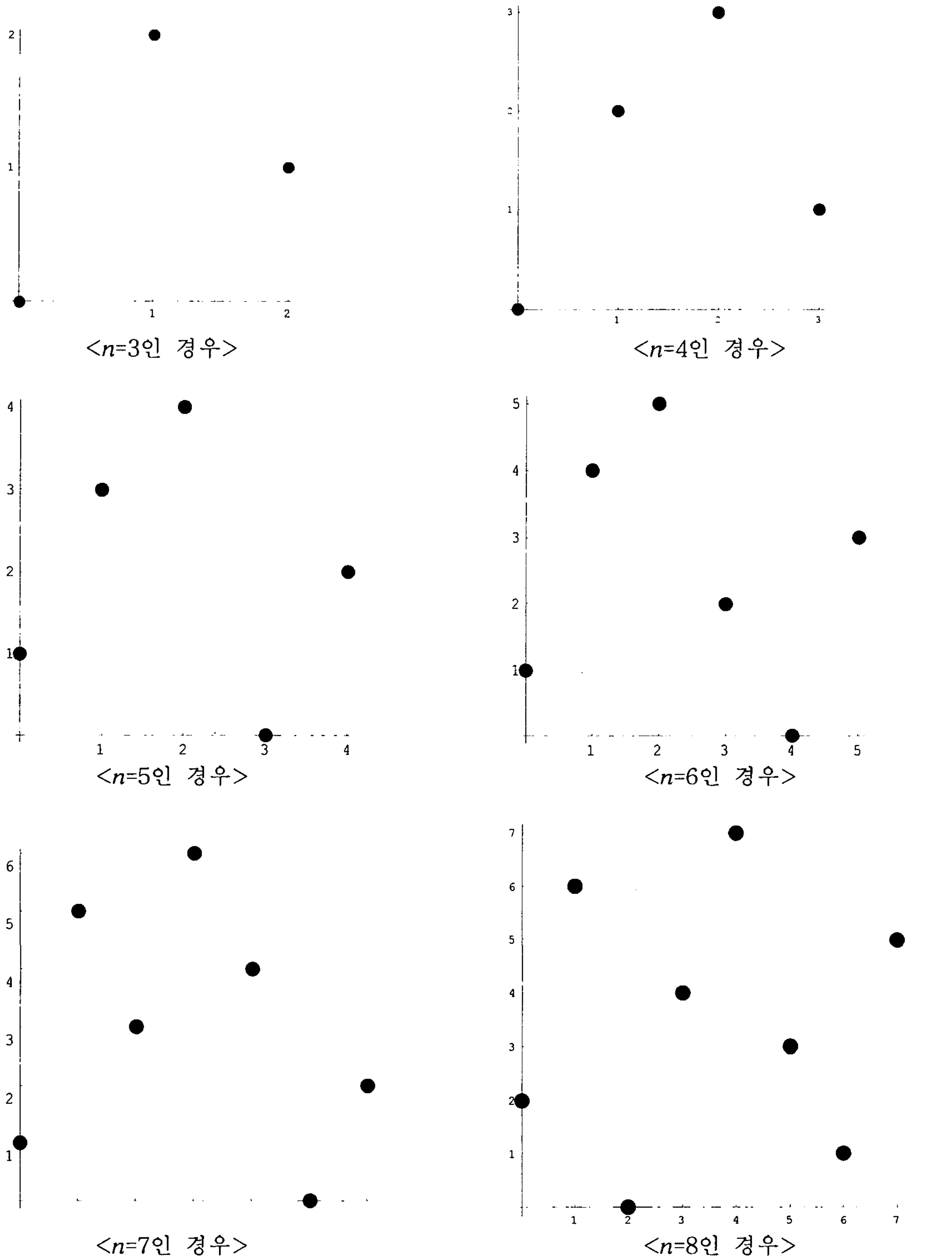

Figure 1. Examples of n-point maximal fan designs on the integer grid $\{0,1,2,\dots,n-1\}^2$

For the case n=5, we could not apply the proposition directly. However, a similar result can be drawn. If d is a design on $\{0,1,2,3,4\}^2$ consists of the points on any two lines with slopes a and $-a$, then it cannot be a maximal fan design. There are 52 non-maximal fan designs among 120 latin-hypercubes on the grid $\{0,1,2,3,4\}^2$ for the case m=2, n=5. And 31 out of 52 non maximal cases can be separated by the fact mentioned above.

For the case of higher dimension ($m \geq 3$) we have similar result with larger n. Fixing n=6, the proposition can be generalized for higher dimensional case by substituting 'two lines' to 'two hyper planes whose projections on a coordinate plane are lines'.

In fact, many different maximal fan designs would exist. They can be divided by different types. If one can be obtained by rotation or reflection from another then we call them *of the same type*. Figure 2 represents the different types of maximal fan designs when n=4. There are only one type of two dimensional 3-point maximal fan designs and five types for 4-point case ; but the numbers glow rapidly as n grows.

In our mind, we easily guess that a latin-hypercube type maximal fan design exists because the number of latin-hypercubes increases rapidly as n increases and we prove that a maximal fan design should be a latin-hypercube. However, the proof for the conjecture is still open for us. A main difficulty to prove the conjecture arises from the irregularity of the number of models with n terms which satisfies the D-condition, say $p(n)$. Note that $p(n)$ is just the number of partitions of positive integer n. Its generating function is $f(q)=\Pi_{n\in N}(1-q^n)^{-1}$, and we can compute $p(n)$ for fixed n, but we don't have any general formula for $p(n)$ yet. We refer to Andrews(1976) for details on the theory of partitions.

So far we dealt with designs on a square grid $\{0,1,2,\ldots,n-1\}^m$. But in most of the practical experiments the grid for a design may not be a square. So it would be a nice project to generalize the concept of maximal fan to the case of non-square type grid like $\{0,1,2,3,4\}\times\{0,1,2\}$.

n	number of latin hypercubes	number of maximal fan designs	number of different types of latin hypercubes	number of different types of maximal fan designs	$p(n)$, number of partitions
3	6	4	2	1	3
4	24	20	7	5	5
5	120	68	23	11	7
6	720	584	115	81	11
7	5040	3820	694	495	15
8	40320	37716	5282	4800	22

Table 1. The number of n-point maximal fan designs on the integer grid $\{0,1,2,\ldots,n-1\}^2$

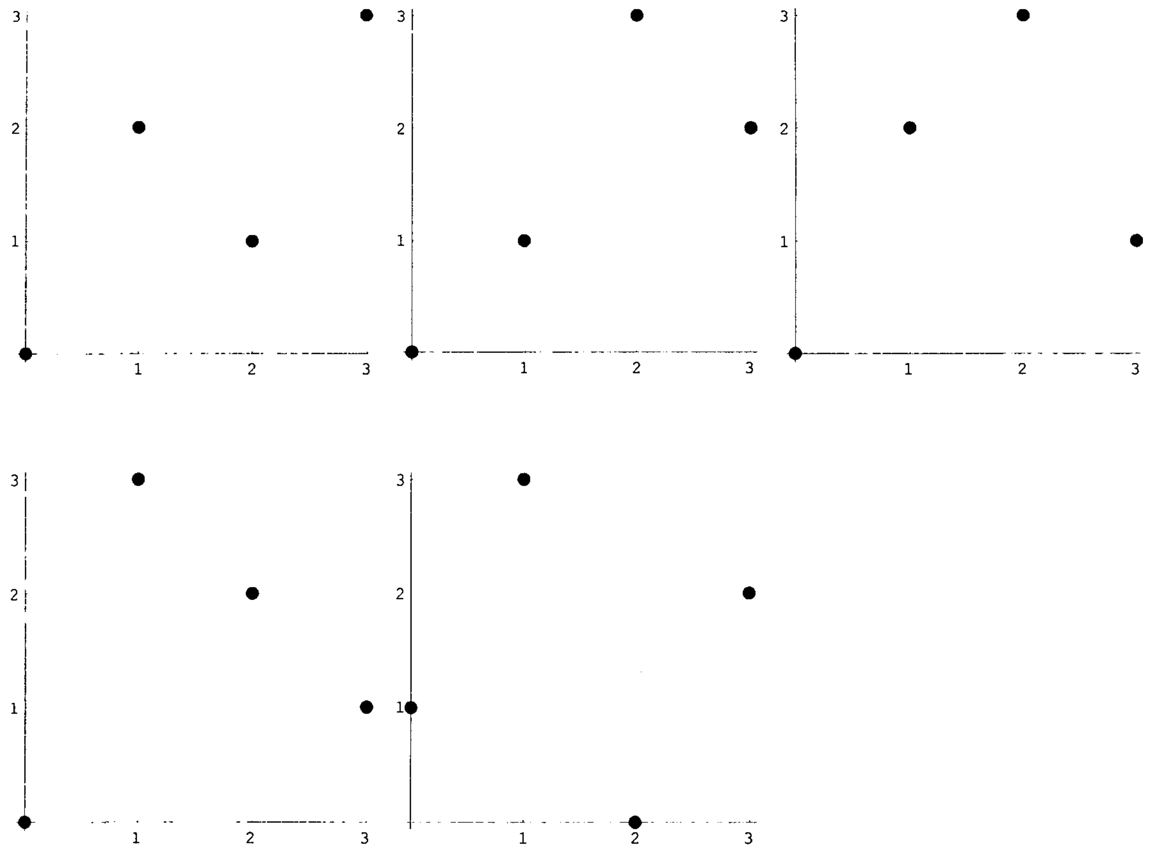

Figure 2. Five different types of 4-point maximal fan designs on the integer grid $\{0,1,2,3\}^2$

References

[1] Andrews, G. E. (1976). *The Theory of Partitions,* Addison-Wesley.

[2] Box, G.E.P. and Hunter, J.S. (1961). The 2^{k-p} fractional factorial designs. *Technometrics*, 3, 311-351, 449-458.

[3] Caboara, M., Pistone,G., Riccomagno, E. and H.P. Wynn (1997). "The fan of an experimental design," Unpublished Manuscript.

[4] Cox, D., Little, J., and O'Shea, D. (1992). *Ideal, Varieties, and Algorithms,* Spring-Verlag, New York.

[5] Fries, A. and Hunter, W.G. (1980). Minimum aberration 2^{n-p} designs, *Technometrics*, 26, 225-232,

[6] Pistone, G. and Wynn, H. P. (1996). "Generalised confounding with Gröbner bases," *Biometrika*, **83**, 653-666.

[2002년 1월 접수, 2002년 7월 채택]