Journal of Statistical Planning and Inference 141 (2011) 2449–2462

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

# Optimal selection and ordering of columns in supersaturated designs

N. Niki [a], M. Iwata [b], H. Hashiguchi [c,*], S. Yamada [d]

[a] Tokyo University of Science, Department of Management Science, Kagurazaka, Shinjuku-ku, Tokyo 162-8601, Japan
[b] DENSO Corporation, Showa-cho, Kariya, Aichi 448-8661, Japan
[c] Saitama University, Graduate School of Science and Engineering, Shimo-Okubo, Sakura, Saitama 338-8570, Japan
[d] University of Tsukuba, Graduate School of Business Sciences, Otsuka, Bunkyo-ku, Tokyo 112-0012, Japan

ARTICLE INFO

*Article history:*
Received 22 October 2009
Received in revised form
27 January 2011
Accepted 8 February 2011
Available online 16 February 2011

*Keywords:*
Corresponding analysis
Exhaustive enumeration
Factor assignment
Non-orthogonality
Optimal design

ABSTRACT

Two methods to select columns for assigning factors to work on supersaturated designs are proposed. The focus of interest is the degree of non-orthogonality between the selected columns. One method is the exhaustive enumeration of selections of $p$ columns from all $k$ columns to find the exact optimality, while the other is intended to find an approximate solution by applying techniques used in the corresponding analysis, aiming for ease of use as well as a reduction in the large computing time required for large $k$ with the first method. Numerical illustrations for several typical design matrices reveal that the resulting "approximately" optimal assignments of factors to their columns are exactly optimal for any $p$. Ordering the columns in $E(s^2)$-optimal designs results in promising new findings including a large number of $E(s^2)$-optimal designs.


## 1. Introduction

Supersaturated designs are, essentially, fractional factorial designs of which can be assigned more factors than the standard fractional factorial design. They are helpful for screening many factors to find active factors from many candidates such as the primary stage of a scientific investigation, technology development and product innovation. The designs are utilized under the assumption of the effect sparsity, *i.e.*, a small number of factors are active while the remaining factors are not. A distinctive aspect of supersaturated designs is that the degree of non-orthogonality between two columns, *e.g.*, the inner products of the two columns in two-level designs, does not always vanish to zero. Following pioneering studies by Satterthwaite (1959) and Booth and Cox (1962), Lin (1993) has introduced a method using half fractions of Hadamard matrix to construct supersaturated designs. This simple and practical method has inspired a number of active researchers in the field of experimental design.

The first aim to construct supersaturated designs is to attempt to minimize the degree of non-orthogonality under given numbers of runs and columns. One promising approach to find a solution to this requirement is to construct a design for any two of the columns which have a lower measure of non-orthogonality at least on an average. In fact, many researchers have proposed designs with optimality in $E(s^2)$, which is an average of the squared inner products between all paired columns. Most of the previous studies including Wu (1993), Lin (1995), Nguyen (1996), Cheng (1997), Li and Wu (1997), Tang and Wu (1997), Yamada and Lin (1997), Butler et al. (2001), Lejeune (2003), Bulutoglu and Cheng (2004), and Nguyen and Cheng (2008) have been devoted to the construction methodologies for two-level supersaturated designs with better orthogonality in term of $E(s^2)$. For other types of supersaturated designs, Yamada and Lin (1999) and Yamada

* Corresponding author.
*E-mail address:* hiro@mail.saitama-u.ac.jp (H. Hashiguchi).

0378-3758/$ - see front matter 
doi:10.1016/j.jspi.2011.02.007

et al. (1999) have introduced a class of three-level supersaturated designs, then, which has been extended to mixed-level by Yamada and Matsui (2002) and Yamada and Lin (2002). Some construction methods for optimal mixed-level supersaturated designs have been discussed in Chen and Liu (2008) and Nguyen and Liu (2008) as a recent work.

In many practical cases, the technical knowledge in the experimented field implies that the possibilities of active factor are not uniform over the candidate factors. One typical case is that some of the candidate factors are assumed to be more important than the remaining factors. In order to utilize the technical knowledge, the factor with high possibility of activeness should be assigned the columns whose degree of non-orthogonality with the other columns are low, while the remaining factors are assigned the remaining columns whose degree of non-orthogonality with the other columns are not low. Since the previous studies have attempted to minimize the average of degree of non-orthogonality over all paired columns, it is not possible to assign factors to columns with considering the possibility of activeness.

In order to solve this problem, this paper discusses the optimality for ordering of columns in a supersaturated design, where this ordering is to arrange columns in the left hand side and dependent columns in the right hand side of a supersaturated design. The ordered designs have a property that the degree of non-orthogonality is low by selecting the columns from the left hand side to any column. This property assures that the degree of non-orthogonality is low for the important factors by assigning factors from the left column based on the possibility.

Section 2 discusses two approaches to optimal selection of the columns from an existing design. One is the exhaustive enumeration of selections of $p$ columns from $k$ columns to obtain the exact optimality, while the other is intended to find an approximate solution by ordinating the columns of a design matrix, as in corresponding analysis, for ease of use as well as to avoid a combinatoric explosion in computation.

Numerical illustrations are provided in Section 3 to demonstrate the tactics required in ordination and additional discussion concerning the optimal designs including all $8\times k$ E($s^2$)-optimal designs for $k\leq 35$ aligned in only one design matrix with ordinated columns. This example shows that an arrangement of 35 columns realize an overall optimal design, namely, from the 18th to 35th column leads to overall optimal and that, by reverse arrangement, from the 4th to 17th column does so as well.

## 2. Selection of columns

### 2.1. Definitions and preliminaries

Let $D$ be the $n\times k$ design matrix of a $t$-level design with $n$ runs and $k$ columns, *i.e.*,

$$D=(d_{ij})=(\boldsymbol{d}_1\ldots\boldsymbol{d}_k),\quad d_{ij}\in T=\{1,2\ldots,t\},\ i\in N=\{1,2\ldots,n\},\ j\in K=\{1,2\ldots,k\}. \tag{1}$$

The design is called supersaturated design when $n-1<k(t-1)$. We treat a balanced design with equal occurrence of $\{1, 2, \ldots,t\}$ in $d_{i,j}$.

Then the $\chi^2$ value is defined as

$$\chi^2_{ij}=\sum_{a=1}^{t}\sum_{b=1}^{t}\frac{(\#\{r\in N|d_{ri}=a,d_{rj}=b\}-n/t^2)^2}{n/t^2}, \tag{2}$$

which is proportional to the measure proposed by Booth and Cox (1962) in the case of $t$=2. When $t$=2, *i.e.*, for two-level designs, the measures

$$s^2_{ij}=((2\boldsymbol{d}_i-3\mathbf{1}_n)^\top(2\boldsymbol{d}_j-3\mathbf{1}_n))^2=n\chi^2_{ij} \tag{3}$$

and

$$\mathrm{E}(s^2)=\binom{k}{2}^{-1}\sum_{1\leq i<j\leq k}s^2_{ij}=n\chi^2(K) \tag{4}$$

are commonly used for representing non-orthogonality between the two and all the $k$ columns, respectively, where $\mathbf{1}_n$ is the $n$-dimensional column vector of all ones. Hereafter, for convenience, $U$=($u_{ij}$) denotes the $k\times k$ matrix of $\chi^2$ values of which all the diagonal elements are zeroes provided as

$$U=(\chi^2_{ij})-\mathrm{diag}(\chi^2_{11},\ldots,\chi^2_{kk}). \tag{5}$$

It is clear that $U^\top=U$, where the superscript $^\top$ denotes transposition.

Several papers including Yamada and Lin (1999) use the criterion

$$\chi^2(K)=\binom{k}{2}^{-1}\sum_{1\leq i<j\leq k}\chi^2_{ij}=\frac{1}{k(k-1)}\mathbf{1}_k^\top U\mathbf{1}_k \tag{6}$$

to evaluate total non-orthogonality among columns of $D$, where $K$ is defined in (1).

For any $S=\{s_1,s_2,\ldots,s_p\}\in\mathcal{K}(p)$, where

$$\mathcal{K}(p)=\{S\in 2^K|\#S=p\leq k\}, \tag{7}$$