Available online at www.sciencedirect.com
SCIENCE DIRECT®
COMPUTATIONAL STATISTICS & DATA ANALYSIS
Computational Statistics & Data Analysis 50 (2006) 254 – 265
www.elsevier.com/locate/csda

A general construction method for mixed-level supersaturated design

Shu Yamada[a,*], Michiyo Matsui[b], Tomomi Matsui[c], Dennis K.J. Lin[d], Takenori Takahashi[b]

[a] *Graduate School of Business Sciences, University of Tsukuba, Otsuka 3-29-1, Tokyo 112-0012, Japan*
[b] *Department of Management Science, Tokyo University of Science, Kagurazaka, Tokyo 162-8601, Japan*
[c] *Department of Mathematical Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Hongo 7-3-1, Tokyo 113-8656, Japan*
[d] *Department of Supply Chain and Information Systems, The Pennsylvania State University, University Park, PA 16802-3005, USA*

Available online 18 August 2004

Abstract

When the number of the experimental variables is large, the first and most critical step is to identify the (few) active factors among those (many) candidate factors. Supersaturated design is shown to be helpful for such a critical first step. A general construction method for mixed-level supersaturated design is proposed. The newly constructed design has several advantages, including the flexibility for the number of runs and the assurance of upper bound of the (pairwise) dependency among all design columns. Specific applications to the construction of two-level and three-level mixed-level designs are discussed in detail.

Keywords: Algorithmic approach; χ^2 value; Non-orthogonality

1. Introduction

Quality Engineers are constantly faced with distinguishing between the effects that are caused by particular factors and those that are due to random noise. In variation reduction

* Corresponding author. Tel.: +81-3-3942-6871, Fax:+81-3-3942-6829.
E-mail address: shu@gssm.otsuka.tsukuba.ac.jp (S. Yamada).

0167-9473/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2004.07.018

of quality characteristics, for example, the conventional approaches is to reduce variation for the factors that affect the quality characteristics. In such experimental studies, the first step aims to find active factors from many candidate factors in order to conduct additional experiments based on these selected factors. Supersaturated design works well for finding active factors from many candidate factors by only a small number of experimental runs.

Supersaturated design was originated by Satterthwaite (1959) and has received a great deal of attention in the recent literature. This includes construction methods in Lin (1991, 1993, 1995), Wu (1993), Iida (1994), Nguyen (1996) and Li and Wu (1997); theoretical justifications in Deng et al. (1994), Cheng (1997), Tang and Wu (1997), and Yamada and Lin (1997); data analysis methods in Westfall et al. (1998), Abraham et al. (1999), Beattie et al. (2002), and Li and Lin (2002, 2003). Recently, some extensions to multi-level and mixed-level supersaturated designs have been proposed. See, for example, Yamada and Lin (1999), Yamada et al. (1999), Fang et al. (2000), and Yamada and Matsui (2002).

In this paper, we show a general construction method for mixed-level supersaturated design which allows flexibility of the number of runs, as opposed to the previous methods which generate designs with only some particular numbers of runs. A special feature of the proposed method is that the maximum dependency among all pairs of design columns can be specified. An example of construction with two- and three-level designs is given in details for illustration. For the simplicity of presentation, all proofs are given in the Appendix.

2. Preliminaries

Let $\mathscr{C}_l^n$ be the set of n-dimensional column vectors such that each vector $\boldsymbol{c} \in \mathscr{C}_l^n$ consists of an equal number of 1s, 2s, …, ls; and $\mathscr{D}_m^n$ be the set of n-dimensional column vectors $\boldsymbol{d} \in \mathscr{D}^n$ consists of an equal numbers of 1s, 2s, …, ms. An $n \times (p+q)$ matrix $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$, with $\boldsymbol{c}_i \in \mathscr{C}_l^n$ and $\boldsymbol{d}_j \in \mathscr{D}_m^n$, is called a mixed-level design consisting of l-level and m-level columns. An n-dimensional design matrix $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$ is called supersaturated when a degree of saturation defined by

$$v = \frac{(l-1)p + (m-1)q}{n-1} \tag{1}$$

is greater than 1 (Yamada and Matsui, 2002). We are interested in constructing a method of mixed-level supersaturated designs with $n = lmt$ runs, where t is a positive integer.

Let $n^{ab}(\boldsymbol{c}, \boldsymbol{d})$ be the number rows in the $n \times 2$ matrix $(\boldsymbol{c}, \boldsymbol{d})$, $\boldsymbol{c} \in \mathscr{C}_l^n$ and $\boldsymbol{d} \in \mathscr{D}_m^n$, whose values are (a, b). It is clear that

$$\sum_{a \in \{1,\ldots,l\}} \sum_{b \in \{1,\ldots,m\}} n^{ab}(\boldsymbol{c}, \boldsymbol{d}) = n.$$

A design is called *orthogonal* if $n^{ab}(\boldsymbol{c}, \boldsymbol{d})$ is a constant for all $(a, b) \in \{1, \ldots, l\} \times \{1, \ldots, m\}$. For a two-level design, the orthogonality implies that the inner product of any two design columns is zero. In other words, a uniform frequency on $n^{ab}(\boldsymbol{c}, \boldsymbol{d})$ is preferable in terms of estimation of factor effects, namely $n^{ab}(\boldsymbol{c}, \boldsymbol{d}) = n/(lm)$ for all combinations of $(a, b) \in \{1, \ldots, l\} \times \{1, \ldots, m\}$. Some measures for non-orthogonality (dependency) based on $n^{ab}(\boldsymbol{c}, \boldsymbol{d})$ have been proposed; for example, Yamada and Lin (1999) proposed the χ^2 statistic criterion

$$\chi^2(\boldsymbol{c}, \boldsymbol{d}) = \sum_{a \in \{1,\ldots,l\}} \sum_{b \in \{1,\ldots,m\}} \frac{(n^{ab}(\boldsymbol{c}, \boldsymbol{d}) - n/(lm))^2}{n/(lm)}; \tag{2}$$

Fang et al. (2000) defines

$$\sum_{a \in \{1,\ldots,l\}} \sum_{b \in \{1,\ldots,m\}} |n^{ab}(\boldsymbol{c}, \boldsymbol{d}) - n/(lm)|. \tag{3}$$

These measures are defined based on the differences between the actual frequency from the expected frequency, namely a function of the quantity: $n^{ab}(\boldsymbol{c}, \boldsymbol{d}) - n/(lm)$. Note that Eq. (3) is equal to zero if and only if χ^2 value in Eq. (2) is equal to 0.

Motivated by the popular criterion of $E(s^2)$ proposed by Booth and Cox (1962), we use the average of χ^2 values as the criterion for the overall design optimality. Consider a mixed-level supersaturated design $(\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$, where $\boldsymbol{c}_i \in \mathscr{C}_l^n$ and $\boldsymbol{d}_j \in \mathscr{D}_m^n$. The average χ^2 values over all pairs of two l-level columns, all pairs of one l-level and one m-level column and all pairs of two m-level columns are defined by

$$\text{ave}\,\chi^2_{l,l} = \sum_{1 \leqslant i < j \leqslant p} \chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j) \Big/ \binom{p}{2}, \tag{4}$$

$$\text{ave}\chi^2_{l,m} = \sum_{1 \leqslant i \leqslant p} \sum_{1 \leqslant j \leqslant q} \chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j)/(pq), \tag{5}$$

$$\text{ave}\chi^2_{m,m} = \sum_{1 \leqslant i < j \leqslant q} \chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j) \Big/ \binom{q}{2}, \tag{6}$$

respectively. A lower bound of the sum of χ^2 values for any level of mixed-level supersaturated design obtained by Yamada and Matsui (2002), can be shown to be

$$\tfrac{1}{2} v(v-1)n(n-1). \tag{7}$$

Furthermore, the measurement

$$\chi^2_{\text{eff}} = \frac{v(v-1)n(n-1)/2}{\chi^2_{\text{sum}}} \tag{8}$$

is utilized to evaluate the efficiency for an attainment of χ^2-optimality, where χ^2_{sum} is the sum of all the χ^2 values in Eqs. (4)–(6).

Another design optimality criterion under consideration is the maximum χ^2 value over all pairs, defined by

$$\max \chi^2_{l,l} = \max\{\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j) \mid \boldsymbol{c}_i, \boldsymbol{c}_j \in \mathscr{C}_l^n, i \neq j\}, \tag{9}$$

$$\max \chi^2_{l,m} = \max\{\chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j) \mid \boldsymbol{d}_i \in \mathscr{C}_l^n, \boldsymbol{d}_j \in \mathscr{D}_m^n\}, \tag{10}$$

$$\max \chi^2_{m,m} = \max\{\chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j) \mid \boldsymbol{d}_i, \boldsymbol{d}_j \in \mathscr{D}^n_m, i \neq j\}. \tag{11}$$

We will deal with a constructing problem of mixed-level supersaturated design consisting of l-level and m-level columns, that is a generation of the columns from the sets $\mathscr{C}^l_n$ and $\mathscr{D}^m_n$ while maintaining low dependency, i.e., keeping a small value of $\max \chi^2_{l,l}$, $\max \chi^2_{l,m}$, $\max \chi^2_{m,m}$, $\mathrm{ave}\chi^2_{l,l}$, $\mathrm{ave}\chi^2_{l,m}$ and $\mathrm{ave}\chi^2_{m,m}$.

3. The construction of supersaturated design

An initial design matrix with $n = lm$ rows is constructed, say $(\boldsymbol{C}_l, \boldsymbol{D}_m) = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p, \boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$, where $\boldsymbol{c}_i \in \mathscr{C}^{lm}_l$ and $\boldsymbol{d}_j \in \mathscr{D}^{lm}_m$. Next, let $\boldsymbol{T}^t_l$ and $\boldsymbol{T}^t_m$ be $(t \times r_l)$ and $(t \times r_m)$ matrices whose elements are $0, 1, \ldots, l-1$ (l levels) and $0, 1, \ldots, m-1$ (m levels), respectively. An n-run design can be generated by

$$\boldsymbol{C} = \boldsymbol{T}^t_l \oplus \boldsymbol{C}_l, \tag{12}$$

$$\boldsymbol{D} = \boldsymbol{T}^t_m \oplus \boldsymbol{D}_m. \tag{13}$$

where, the operator $\oplus$ determines the $((i-1)lm + u,\ (j-1)p + v)$ element in the matrix $\boldsymbol{C}$ by $\mathrm{mod}(t_{ij} + c_{uv} - 1, l) + 1$, and t_{ij} and c_{uv} denote the (i, j) and (u, v) elements in $\boldsymbol{T}^t_l$ and $\boldsymbol{C}_l$, respectively. The location of elements is determined based on the original elements in the same way as the "Kronecker Product," while the calculation at each element is different.

Eq. (12) generates an $lmt \times pr_l$ design matrix $\boldsymbol{C}$ from an $lm \times p$ initial matrix $\boldsymbol{C}_l$ and a $t \times r_l$ generating matrix $\boldsymbol{T}^t_l$. In the same manner, Eq. (13) generates an $lmt \times qr_m$ design matrix $\boldsymbol{D}$ from an $lm \times q$ initial matrix $\boldsymbol{D}_m$ and a $t \times r_m$ generating matrix $\boldsymbol{T}^t_m$. The resulting design is then $(\boldsymbol{C}, \boldsymbol{D})$. Theorem 1 below gives a theoretical justification of this constructing method. We first provide an example.

As an illustrative example, consider the following matrices:

$$\boldsymbol{T}^2_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad \boldsymbol{T}^2_3 = \begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}, \quad \boldsymbol{C}_2 = \begin{pmatrix} 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 \\ 2 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 \end{pmatrix} \quad \text{and}$$

$$\boldsymbol{D}_3 = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 3 & 3 & 3 \\ 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{pmatrix}.$$

The (9, 5) element at the generated matrix $\boldsymbol{T}^2_2 \oplus \boldsymbol{C}_2$, for example, is determined by the (2, 2) element at $\boldsymbol{T}^2_2$ and (3, 1) at $\boldsymbol{C}_2$ by $\mathrm{mod}(t_{22} + c_{31} - 1, l) + 1 = 2$. In the same manner, all elements in $\boldsymbol{T}_2^2 \oplus \boldsymbol{C}_2$ and $\boldsymbol{T}_3^2 \oplus \boldsymbol{D}_3$ are generated as follows:

$$\boldsymbol{C} = \boldsymbol{T}_2^2 \oplus \boldsymbol{C}_2 = \begin{pmatrix} 1 & 1 & 1 & 2 & 1 & 1 & 1 & 2 \\ 1 & 1 & 2 & 1 & 1 & 1 & 2 & 1 \\ 1 & 2 & 1 & 1 & 1 & 2 & 1 & 1 \\ 2 & 2 & 2 & 1 & 2 & 2 & 2 & 1 \\ 2 & 2 & 1 & 2 & 2 & 2 & 1 & 2 \\ 2 & 1 & 2 & 2 & 2 & 1 & 2 & 2 \\ 1 & 1 & 1 & 2 & 2 & 2 & 2 & 1 \\ 1 & 1 & 2 & 1 & 2 & 2 & 1 & 2 \\ 1 & 2 & 1 & 1 & 2 & 1 & 2 & 2 \\ 2 & 2 & 2 & 1 & 1 & 1 & 1 & 2 \\ 2 & 2 & 1 & 2 & 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 2 & 1 & 2 & 1 & 1 \end{pmatrix},$$

$$\boldsymbol{D} = \boldsymbol{T}_3^2 \oplus \boldsymbol{D}_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 & 3 & 3 \\ 1 & 2 & 3 & 1 & 2 & 3 \\ 2 & 3 & 1 & 2 & 3 & 1 \\ 3 & 1 & 2 & 3 & 1 & 2 \\ 2 & 2 & 2 & 3 & 3 & 3 \\ 3 & 3 & 3 & 1 & 1 & 1 \\ 1 & 1 & 1 & 2 & 2 & 2 \\ 2 & 3 & 1 & 3 & 1 & 2 \\ 3 & 1 & 2 & 1 & 2 & 3 \\ 1 & 2 & 3 & 2 & 3 & 1 \end{pmatrix}. \tag{14}$$

Theorem 1. *Consider the supersaturated design* $(\boldsymbol{C}, \boldsymbol{D})$ *generated by initial matrices* $\boldsymbol{C}_l = (\boldsymbol{c}_1, \ldots, \boldsymbol{c}_p)$ *and* $\boldsymbol{D}_m = (\boldsymbol{d}_1, \ldots, \boldsymbol{d}_q)$ *and generating matrices* $\boldsymbol{T}_l^t$ *and* $\boldsymbol{T}_m^t$*, as in Eqs.* (12) *and* (13). *For all* $1 \leqslant k_1, k_2 \leqslant t$*, we have the following*:

$$\begin{array}{lll} (a) & \chi^2(\boldsymbol{c}_{i^*}^*, \boldsymbol{c}_{j^*}^*) \leqslant t\,\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j), & i^* = (k_1-1)p+i,\ \ j^* = (k_2-1)p+j, \\ & & 1 \leqslant i \leqslant j \leqslant p,\ \ i^* \neq j^*, \\ (b) & \chi^2(\boldsymbol{c}_{i^*}^*, \boldsymbol{d}_{j^*}^*) \leqslant t\,\chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j) & i^* = (k_1-1)p+i,\ \ j^* = (k_2-1)q+j, \\ & & 1 \leqslant i \leqslant p,\ \ 1 \leqslant j \leqslant q, \\ (c) & \chi^2(\boldsymbol{d}_{i^*}^*, \boldsymbol{d}_{j^*}^*) \leqslant t\,\chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j), & i^* = (k_1-1)q+i,\ \ j^* = (k_2-1)q+j, \\ & & 1 \leqslant i \leqslant j \leqslant q,\ \ i^* \neq j^*. \end{array} \tag{15}$$

Theorem 1 provides the property of assuring the maximum χ^2 values in the resulting designs. This is convenient for design construction. Specifically, the maximum $\chi^2(\boldsymbol{c}_{i^*}^*, \boldsymbol{c}_{j^*}^*)$ satisfies

$$\max\{\chi^2(\boldsymbol{c}_{i^*}^*, \boldsymbol{c}_{j^*}^*)\} \leqslant \begin{cases} t\max\{\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j)\}, & (i \neq j), \\ tml(l-1), & (i = j), \end{cases} \tag{16}$$

where $i^* = (k_1-1)p+i,\ j^* = (k_2-1)p+j$, and $1 \leqslant k_1 \leqslant k_2 \leqslant t$. This suggests the importance of selecting initial matrices as well as generating matrices $\boldsymbol{T}_l^t$ and $\boldsymbol{T}_m^t$.

Inequality (16) implies that a small value of the maximum χ^2 value in the initial design matrix implies a small value of the maximum χ^2 value in the resulting design matrix. It also implies the possibility to generate a design matrix whose maximum χ^2 value is lower than the level specified by the right-hand side of Inequality (16). For example, for $l=2$, $m=3$, $t=2$, $\boldsymbol{c}=(1,1,1,2,2,2)$, and $\boldsymbol{T}_2^2=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$, the χ^2 value between the two constructed vectors by $\boldsymbol{T}_2^2 \oplus \boldsymbol{c}$ is 0; while the upper bound specified by Inequality (16) is 12. Although the above is a case of two l-level (or m-level) columns, similar properties hold on the maximum χ^2 value between one l-level and one m-level columns.

4. Two-level and three-level designs

We first consider the case of $t=1$. For any $\boldsymbol{c} \in \mathscr{C}_2^6$ and $\boldsymbol{d} \in \mathscr{D}_3^6$, χ^2 value varies as follows: $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j) \in \{\frac{2}{3}, 6\}$, $\chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j) \in \{0, 4\}$ and $\chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j) \in \{3, 6, 12\}$. It is desirable to explore designs which maintain low level of χ^2 values, such as $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j)=\frac{2}{3}$ and $\chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j)=3$. Mixed two- and three- level supersaturated design can thus be constructed as follows:

$$\boldsymbol{C}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 2 & 2 & 2 & 2 & 2 & 2 \\ 1 & 2 & 2 & 2 & 1 & 1 & 1 & 2 & 2 & 2 \\ 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 2 \\ 2 & 2 & 1 & 2 & 2 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 2 & 1 & 2 & 2 & 1 & 2 & 1 & 1 \end{pmatrix}, \quad \boldsymbol{D}_3 = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 3 & 3 \\ 3 & 2 & 3 & 3 & 1 \\ 1 & 2 & 2 & 2 & 2 \\ 2 & 3 & 3 & 1 & 2 \\ 3 & 3 & 1 & 2 & 3 \end{pmatrix} \tag{17}$$

The designs $\boldsymbol{C}_2$, $\boldsymbol{D}_3$ were constructed by a lexicographical enumeration and a computer search, respectively. These designs are justified by the following properties:

- A mixed-level supersaturated design $(\boldsymbol{C}_2, \boldsymbol{D}_3)$ satisfies $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j)=\frac{2}{3}$ and $\chi^2(\boldsymbol{d}_i, \boldsymbol{d}_j)=3$.
- A two-level supersaturated design $\boldsymbol{C}_2$ and a three-level supersaturated design $\boldsymbol{D}_3$ are optimal in terms of χ^2 efficiency, respectively. Thus the mixed-level supersaturated design $(\boldsymbol{C}_2, \boldsymbol{D}_3)$ is also optimal in terms of χ^2 efficiency.

The following Lemma implies the difficulty to construct supersaturated designs with the property of $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j)=\frac{2}{3}$ and $\chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j)=0$. Specifically, χ^2 values between a two-level column and any three-level column in the set $\mathscr{D}^6$ must consist of four 0s and six 4s. This implies the impossibility of keeping $\chi^2(\boldsymbol{c}_i, \boldsymbol{d}_j)=0$ with more than four two-level columns.

Lemma 1. *Let* $S=\{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{10}\}$ *be a set of two-level columns from the set* $\mathscr{C}_2^6$ *satisfying* $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j)=\frac{2}{3}$ $(i \neq j)$. *For any pair of three-level columns* $\boldsymbol{d}, \boldsymbol{d}' \in \mathscr{D}_3^6$, *we have* $\sum_{i=1}^{10} \chi^2(\boldsymbol{c}_i, \boldsymbol{d})=24$ for each vector $\boldsymbol{d} \in \mathscr{D}_3^6$.

This Lemma can be regarded as an extension from the results in two-level and mixed-level supersaturated design including orthogonal designs (Yamada and Lin, 1997; Yamada and Matsui, 2002). The proof is given in the Appendix.

For the case of $t \geqslant 2$, we apply a computer search to obtain the generating matrices $\boldsymbol{T}_2^t$ and $\boldsymbol{T}_3^t$, while $\boldsymbol{C}_2$ and $\boldsymbol{D}_3$ in Eq. (17) are treated as initial design matrices. Furthermore, we only consider designs for $r_m = r_l = t$. The outline of searching algorithm is as follows:

- *Step* 1: A pair of generating matrices $\boldsymbol{T}_2^t$ and $\boldsymbol{T}_3^t$ are constructed by using uniform random numbers. When the matrices have at least one pair of equivalent columns, another pair of generating matrices is constructed. Without loss of generality, the first row in the matrices are all zeros. The pair of generating matrices is dealt with a candidate mixed-level design.
- *Step* 2: We update a mixed-level supersaturated design using the candidate design in the previous step.
    - If the value of max $\chi^2_{2,2}$ in the candidate mixed-level design is lower than the value in the tentative design, then the tentative design is replaced by the candidate design.
    - If the value of max $\chi^2_{2,2}$ in the candidate mixed-level design is higher than the value in the tentative design, then the examination of the candidate design is terminated. Return to Step 1.
    - If the value of max $\chi^2_{2,2}$ in the candidate mixed-level design is equal to the value in the tentative design, the candidate design is compared to tentative design matrix in terms of max $\chi^2_{2,3}$.
- *Step* 3: The update procedure compares the candidate design to the tentative design in terms of max $\chi^2_{2,2}$, max $\chi^2_{2,3}$, max $\chi^2_{3,3}$, ave$\chi^2_{2,2}$, ave$\chi^2_{2,3}$ and ave$\chi^2_{3,3}$, sequentially. We have investigated other orders of sequence, the resulting designs are rather similar, and thus are not reported here.
- The above algorithm is iterated 100,000 times.

The resulting (generating) matrices are shown in Table 1. These designs are then evaluated in terms of the number of columns, maximum and average χ^2 values and χ^2eff. The summary of evaluation is displayed in Table 2.

For a mixed two- and three-level design, the run size must be a multiple of six. As seen in Table 2, the proposed construction method is able to generate all these mixed-level supersaturated designs, whereas, all previous studies can only generate design with some specific run sizes, such as $n = 12, 24$. As such, the proposed method has an advantage of flexibility of number of columns. Furthermore, a straightforward comparison with the previous studies is difficult because the previous study only involves specific number of runs/columns and the numbers of columns are different from the resulting designs in this paper. In terms of the maximum χ^2 value, however, the resulting designs are better than or equal to the previous designs.

The proposed construction method works well for relatively small numbers of runs. This can be seen from Table 2, along with Lemma 1. For larger numbers of runs, the decreasing χ^2_{eff} along with the number of runs is an evidence for potential improvement. Moreover, some of the maximum values in the resulting designs for $t \geqslant 2$ are rather small (as compared to the bound specified in Theorem 1). This fact indicates that the random search for generating matrices $\boldsymbol{T}_2^t$ and $\boldsymbol{T}_3^t$ works well.

Table 1
The generating matrices ($t = 2, 3, \ldots, 8$)

t	$\boldsymbol{T}_2^t$	$\boldsymbol{T}_3^t$
2	$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 \\ 1 & 2 \end{pmatrix}$
3	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{pmatrix}$
4	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 2 \\ 2 & 0 & 0 & 2 \\ 2 & 1 & 2 & 0 \end{pmatrix}$
5	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 2 \\ 2 & 1 & 0 & 0 & 1 \\ 2 & 0 & 1 & 2 & 2 \end{pmatrix}$
6	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 & 2 & 2 \\ 1 & 0 & 1 & 0 & 2 & 2 \\ 1 & 2 & 1 & 0 & 1 & 2 \\ 1 & 2 & 2 & 1 & 1 & 0 \\ 2 & 2 & 1 & 1 & 1 & 1 \end{pmatrix}$
7	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 2 & 0 & 1 & 1 & 1 & 2 \\ 1 & 2 & 1 & 1 & 1 & 2 & 0 \\ 0 & 2 & 2 & 0 & 1 & 2 & 0 \\ 2 & 1 & 0 & 0 & 2 & 1 & 2 \\ 1 & 0 & 1 & 1 & 0 & 1 & 0 \end{pmatrix}$
8	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 2 & 0 & 2 & 1 & 1 & 2 & 2 \\ 0 & 2 & 1 & 1 & 0 & 2 & 0 & 1 \\ 2 & 2 & 1 & 2 & 1 & 2 & 0 & 1 \\ 0 & 0 & 2 & 2 & 2 & 2 & 1 & 0 \\ 1 & 1 & 0 & 1 & 2 & 2 & 0 & 2 \\ 0 & 1 & 0 & 1 & 1 & 1 & 2 & 0 \end{pmatrix}$

5. Concluding remarks

In this paper, we proposed a general construction method for mixed-level supersaturated design. The proposed method has three major advantages: (1) this procedure does not require iterative computation to generate mixed-level design; (2) the maximum χ^2 values are under control, and (3) this procedure allows various levels combination. We also apply the method to construct two-level and three-level supersaturated designs. The essences of such an application are the selection of initial design matrices and the generating matrices. The theoretical justification of the design optimality (on χ^2 efficiency) with the optimal

Table 2
The properties of the resulting designs

		χ^2_{22}					χ^2_{33}					χ^2_{23}		Full design	
t	n	p	D.s.	Max	Ave	χ^2_{eff}	Q	D.s.	Max	Ave	χ^2_{eff}	Max	Ave	D.s.	χ^2_{eff}
1	6	10	2.00	0.67	0.67	1.00	5	2.00	3.00	3.00	1.00	4.00	2.40	4.00	1.00
2	12	20	1.82	1.33	0.63	0.82	10	1.82	6.00	3.60	0.61	6.00	2.40	3.64	0.83
3	18	30	1.76	2.00	0.90	0.53	15	1.76	9.00	2.57	0.77	5.33	2.40	3.53	0.79
4	24	40	1.74	2.67	0.62	0.74	20	1.74	12.00	3.57	0.52	4.00	2.40	3.48	0.77
5	30	50	1.72	3.33	0.81	0.55	25	1.72	31.20	4.57	0.40	5.60	2.23	3.45	0.71
6	36	60	1.71	4.00	1.03	0.42	30	1.71	42.00	6.21	0.29	14.00	2.29	3.43	0.61
7	42	70	1.71	7.71	1.09	0.40	35	1.71	53.14	6.97	0.25	16.00	2.52	3.41	0.55
8	48	80	1.70	12.00	1.09	0.39	40	1.70	96.00	6.78	0.26	18.50	2.57	3.40	0.55

d.s. = degree of saturation

initial designs for $n = 6$ is given (see Lemma 1). For other cases, the random search for generating matrices works well, although improvement in the efficiencies seem possible.

Acknowledgements

Dennis Lin's research was partially supported by *National Security Agency*, via Grant MDA904-02-1-0054 and by Smeal College of Business Administration, Penn State University.

Appendix. Proofs

Proof of Theorem 1. First, consider Case (i). Obviously, $\chi^2(\boldsymbol{c}_{i^*}^*, \boldsymbol{c}_{j^*}^*)$ has a maximum χ^2 value among possible paired columns when $\boldsymbol{c}_{i^*}^* = (\boldsymbol{c}_i^\top, \ldots, \boldsymbol{c}_i^\top)^\top$ and $\boldsymbol{c}_{j^*}^* = (\boldsymbol{c}_j^\top, \ldots, \boldsymbol{c}_j^\top)^\top$. Under this condition, χ^2 value between $\boldsymbol{c}_{i^*}^*$ and $\boldsymbol{c}_{j^*}^*$ is

$$\sum_{(a,b)\in\mathscr{P}^2} \frac{(tn^{ab}(\boldsymbol{c}_{i^*}^*, \boldsymbol{c}_{j^*}^*) - tlm/l^2)^2}{tlm/l^2} = t \sum_{(a,b)\in\mathscr{P}^2} \frac{(n^{ab}(\boldsymbol{c}_i, \boldsymbol{c}_j) - lm/l^2)^2}{lm/l^2} = t\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j). \tag{18}$$

The above equation implies the inequality given in the theorem. In the same manner, Case (ii) one l-level column and one m-level column and Case (iii) two m-level columns are obtained. $\square$

Proof of Lemma 1. Let σ be a permutation defined on the set of numbers $\{1, 2, \ldots, l\}$. For any column $\boldsymbol{c} \in \mathscr{C}_l^n$, $\sigma(\boldsymbol{c})$ denotes the vector in $\mathscr{C}_l^n$ satisfying that jth element of $\sigma(\boldsymbol{c})$ is equivalent to $\sigma(c_j)$, where c_j is the jth element of $\boldsymbol{c}$. Then it is easy to see that $\forall(\boldsymbol{c}, \boldsymbol{d}) \in \mathscr{C}_l^n \times \mathscr{D}_m^n$, $\chi^2(\boldsymbol{c}, \boldsymbol{d}) = \chi^2(\sigma(\boldsymbol{c}), \boldsymbol{d})$. We introduce a relation $\sim$ on the set $\mathscr{C}_l^n$ satisfying that $\boldsymbol{c} \sim \boldsymbol{c}'$ if and only if there exists a permutation σ satisfying $\sigma(\boldsymbol{c}) = \boldsymbol{c}'$. It is clear that the relation $\sim$ is an equivalence relation.

When $l = 2$, each equivalence class of the system $(\mathscr{C}_2^n, \sim)$ consists of 2 columns. Additionaly, for $n=6$, $\boldsymbol{c} \sim \boldsymbol{c}'$ if and only if $\chi^2(\boldsymbol{c}, \boldsymbol{c}')=6$ and thus $\boldsymbol{c} \not\sim \boldsymbol{c}'$ if and only if $\chi^2(\boldsymbol{c}, \boldsymbol{c}')=\frac{2}{3}$. Since the set $\mathscr{C}_2^6$ contains 20 columns, the system $(\mathscr{C}_2^6, \sim)$ has 10 equivalence classes. From the assumption, the set $S = \{\boldsymbol{c}_1, \ldots, \boldsymbol{c}_{10}\}$ satisfies that $i \neq j$ implies $\chi^2(\boldsymbol{c}_i, \boldsymbol{c}_j) = \frac{2}{3}$ i.e., $\boldsymbol{c}_i \not\sim \boldsymbol{c}_j$. Thus, the set S of 10 columns is a set of class representatives.

Next, we consider a permutation π defined on the set of indices $\{1, 2, \ldots, n\}$. For any column vector $\boldsymbol{x} \in \mathscr{C}_l^n \cup \mathscr{D}_m^n$, $\pi(\boldsymbol{x})$ denotes the column vector obtained by changing the indices of elements of $\boldsymbol{x}$ by the permutation π. Clearly, $\boldsymbol{c} \sim \boldsymbol{c}'$ if and only if $\pi(\boldsymbol{c}) \sim \pi(\boldsymbol{c}')$ for any pair of vectors $\boldsymbol{c}, \boldsymbol{c}' \in \mathscr{C}_l^n$. Since S is a set of class representatives of the system $(\mathscr{C}_2^6, \sim)$, for any permutation π on $\{1, 2, \ldots, 6\}$, there exists a bijection $f : S \to S$ satisfying that $\boldsymbol{c} \sim f(\pi(\boldsymbol{c}))$ for each $\boldsymbol{c} \in S$.

Lastly, we show the desired equality. Let $\boldsymbol{d}$ and $\boldsymbol{d}'$ be a pair of vectors in $\mathscr{D}_m^6$. Then there exists a permutation π on $\{1, 2, \ldots, 6\}$, satisfying that $\pi(\boldsymbol{d}) = \boldsymbol{d}'$. From the previous discussion, there exists a bijection $f : S \to S$ satisfying $\forall \boldsymbol{c} \in S,\ \boldsymbol{c} \sim f(\pi(\boldsymbol{c}))$. Thus, we have the following result:

$$\begin{aligned}\sum_{i=1}^{10} \chi^2(\boldsymbol{c}_i, \boldsymbol{d}) &= \sum_{i=1}^{10} \chi^2(\pi(\boldsymbol{c}_i), \pi(\boldsymbol{d})) = \sum_{i=1}^{10} \chi^2(f(\pi(\boldsymbol{c}_i)), \pi(\boldsymbol{d})) \\ &= \sum_{i=1}^{10} \chi^2(\boldsymbol{c}_i, \pi(\boldsymbol{d})) = \sum_{i=1}^{10} \chi^2(\boldsymbol{c}_i, \boldsymbol{d}').\end{aligned}$$

The desired result can be shown for $m = 3$ by a straightforward calculation. □

References

Abraham, B., Chipman, H., Vijayan, K., 1999. Some risks in the construction and analysis of supersaturated designs. Technometrics 41 (2), 135–141.

Beattie, S.D., Fong, D.K.H., Lin, D.K.L., 2002. A two-stage Bayesian model selection strategy for supersaturated designs. Technometrics 44 (1), 55–63.

Booth, K.H.V., Cox, D.R., 1962. Some systematic supersaturated designs. Technometrics 4 (4), 489–495.

Cheng, C.S., 1997. E(s^2)-optimal supersaturated designs. Statist. Sinica 7 (4), 929–939.

Deng, L.Y., Lin, D.K.J., Wang, J.N., 1994. Supersaturated design using Hadamard matrix. IBM Research Report, RC19470, IBM Watson Research Center.

Deng, L.Y., Lin, D.K.J., Wang, J.N., 1999. A resolution rank criterion for supersaturated designs. Statist. Sinica 9 (2), 605–610.

Fang, K.T., Lin, D.K.J., Ma, C.X., 2000. On the construction of multi-level supersaturated designs. J. Statist. Plann. Inference 86, 239–252.

Iida, T., 1994. A construction method of two-level supersaturated design derived from L12. Jpn. J. Appl. Statist. 23 (3), 147–153 (in Japanese).

Li, R., Lin, D.K.L., 2002. Data analysis in supersaturated designs. Statist. Probab. Lett. 59, 135–144.

Li, R., Lin, D.K.L., 2003. Analysis methods for supersaturated design: some comparisons. J. Data Sci. 1, 103–121.

Li, W.W., Wu, C.F.J., 1997. Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. Technometrics 39 (2), 171–179.

Lin, D.K.J., 1991. Systematic supersaturated designs. Technical Report, No. 264, College of Business Administration, University of Tennessee.

Lin, D.K.J., 1993. A new class of supersaturated designs. Technometrics 35 (1), 28–31.

Lin, D.K.J., 1995. Generating systematic supersaturated designs. Technometrics 37 (2), 213–225.

Nguyen, N.K., 1996. An algorithmic approach to constructing supersaturated designs. Technometrics 38 (1), 69–73.

Satterthwaite, F.E., 1959. Random balance experimentation (with discussion). Technometrics 1, 111–137.

Tang, B., Wu, C.F.J., 1997. A method for constructing supersaturated designs and its E(s^2) optimality. Canadian J. Statist. 25 (2), 191–201.

Wu, C.F.J., 1993. Construction of supersaturated designs through partially aliased interactions. Biometrika 80 (3), 661–669.

Westfall, P.H., Young, S.S., Lin, D.K.L., 1998. Forward selection error control in the analysis of supersaturated design. Statist. Sinica 8 (1), 101–117.

Yamada, S., Lin, D.K.J., 1997. Supersaturated design including an orthogonal base. Canadian J. Statist. 25 (2), 203–213.
Yamada, S., Lin, D.K.J., 1999. Three-level supersaturated design. Statist. Prob. Lett. 45, 31–39.
Yamada, S., Matsui, T., 2002. Optimality of mixed-level supersaturated designs. J. Statist. Plann. Inference 104, 459–469.
Yamada, S., Ikebe, Y., Hashiguchi, H., Niki, N., 1999. Construction of three-level supersaturated design. J. Statist. Plann. Inference 81, 183–193.