# Optimality of Mixed-level Supersaturated Designs

**Shu Yamada and Tomomi Matsui**

METR 98-05 April 1998

**Shu Yamada**
Department of Management Science
Science University of Tokyo
Kagurazaka 1-3, Shinjuku, Tokyo 162-8601, Japan
Tel +81 3 3260 4271, ext 3336
Fax +81 3 3235 6479
e-mail: shu@ms.kagu.sut.ac.jp

**Tomomi Matsui**
Department of Mathematical Engineering and Information Physics
Faculty of Engineering
University of Tokyo
Hongo 7-3-1, Bunkyo, Tokyo 113-8656, Japan
Tel +81 3 3812 2111 ext.6921
Fax + 81 3 5689 7358
e-mail: tomomi@misojiro.t.u-tokyo.ac.jp

# Optimality of mixed-level supersaturated designs

**Abstract**

Supersaturated design is an important aspect of experimental design. Several properties of supersaturated designs have been obtained that enable supersaturated design to be constructed while maintaining low dependency over all pairs of columns. This paper presents generalized theorems on the optimality of supersaturated designs. Mixed-level supersaturated designs are generated using a construction method based on these theorems. An index is proposed for measuring the efficiency of supersaturated design and is applied to the constructed mixed-level supersaturated designs.

*Key words and phrases:* $\chi^2$ statistic, dependency between two columns, saturated and orthogonal design, lower bound of dependency

*AMS 1991 subject classifications:* Primary 62K15, Secondary 05B20

*Abbreviated title:* Mixed-level supersaturated design

## 1 Introduction

When an experiment is expensive and the number of factors is large, a useful first step is to identify the active factors so that the number of factors can be limited. Supersaturated design is helpful doing this, because it is a kind of fractional factorial design in which the number of columns is greater than or equal to the number of rows. Supersaturated design was originated by Satterthwaite (1959) as a random balance design and formulated by Booth and Cox (1962) in a systematic manner. The main aim of construction of supersaturated design is to generate two-level columns while maintaining low dependency over all pairs of columns. The dependency has been measured using the squared inner product between two columns because the dependency between the two estimates of the effects of the assigned factors can be represented by a function of the inner product. Lin (1993), Wu (1993), Iida (1994) and Deng, Lin and Wang (1994) have constructed supersaturated designs from existing designs such as the half fraction of the Plackett and Burman design proposed by Lin (1993). Lin (1995) and Cheng (1997) have gained insight into the properties of the design criteria. The construction methods proposed by Nguyen (1996) and Li and Wu (1997) are algorithmic approaches that maintain low dependency. Tang and Wu (1997) and Yamada and Lin (1997) have constructed supersaturated designs by applying the properties of an orthogonal base. For example, in Tang and Wu's method, the design matrices generated by permutation of rows in an initial matrix are sequentially added, where the initial matrix is an orthogonal base.

The three-level supersaturated design defined by Yamada, Lin and Yasunari (1997) is a natural extension of the two-level supersaturated design. In it, the measure for dependency between two columns is defined by $\chi^2$ statistic which is applied to the hypothesis test in a two-way contingency table. They have constructed three-level supersaturated designs from two-level supersaturated designs while maintaining low dependency. Yamada, Ikebe, Hashiguchi and Niki (1997) have shown a method for constructing a three-level supersaturated design based on the fundamental theorem of multi-level supersaturated design proven by Niki, Hashiguchi, Yamada and Ikebe (1997), in which the dependency is measured by $\chi^2$ statistic.

In this paper, we first present generalized theorems on the optimality of mixed-level supersaturated designs, where the dependency between two columns is defined by the $\chi^2$ statistic, because it can also be applied to mixed-level supersaturated design. Next we describe mixed-level supersaturated designs generated using a construction method based on our theorems. Finally, we propose an index to represent the efficiency of supersaturated designs and apply it to the constructed mixed-level designs.

## 2 The construction problem

Given a pair of positive integers, $n$ and $p$, we define the following families:

$$\begin{aligned}
\mathcal{D}_p^n &\equiv \{\boldsymbol{d} = (d_1, d_2, \ldots, d_n)^t \in \{0,1\}^n \mid d_1 + \cdots + d_n = n/p\}, \\
\mathcal{M}_p^n &\equiv \{\{\boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_p\} \mid \boldsymbol{d}_1, \boldsymbol{d}_2, \ldots, \boldsymbol{d}_p \in \mathcal{D}_p^n,\ \boldsymbol{d}_1 + \boldsymbol{d}_2 + \cdots + \boldsymbol{d}_p = \mathbf{1}\}, \\
\mathcal{M}^n &\equiv \mathcal{M}_1^n \cup \mathcal{M}_2^n \cup \cdots \cup \mathcal{M}_n^n.
\end{aligned}$$

Any element $M$ in $\mathcal{M}^n$ is called a column. For any column $M$, $p(M)$ denotes the integer $p$ such that $M \in \mathcal{M}_p^n$, so $M$ is called a $p(M)$-level column. Any multiset $\mathcal{F}$ of columns is called a design. A design $\mathcal{F} = \{M^1, M^2, \ldots, M^q\}$ corresponds to an $(n \times q)$ design matrix $[\boldsymbol{f}^1, \boldsymbol{f}^2, \ldots, \boldsymbol{f}^q]$ such that $r$-th column vector $\boldsymbol{f}^r$ is defined by $1\boldsymbol{d}_1^r + 2\boldsymbol{d}_2^r + \cdots + p_r\boldsymbol{d}_{p_r}^r$ where $p_r = p(M^r)$ and $M^r = \{\boldsymbol{d}_1^r, \boldsymbol{d}_2^r, \ldots, \boldsymbol{d}_{p_r}^r\}$. For any vector $\boldsymbol{d} \in \mathcal{D}_p^n$, the vector $\boldsymbol{d} - (1/p)\mathbf{1}$ is denoted by $\overline{\boldsymbol{d}}$. For any column $M \in \mathcal{M}^n$, we define $\overline{M} \equiv \{\overline{\boldsymbol{d}} \mid \boldsymbol{d} \in M\}$.

Given a pair of columns, $(M, M') \in \mathcal{M}_p^n \times \mathcal{M}_{p'}^n$, the dependency measure between them is defined by

$$\chi^2(M, M') \equiv \sum_{\boldsymbol{d} \in M} \sum_{\boldsymbol{d}' \in M'} \left( (\boldsymbol{d} \cdot \boldsymbol{d}') - \frac{n}{pp'} \right)^2 \Big/ \left( \frac{n}{pp'} \right),$$

where $(\boldsymbol{d} \cdot \boldsymbol{d}')$ is the inner product of vectors $\boldsymbol{d}$ and $\boldsymbol{d}'$. This definition directly implies that

$$\chi^2(M, M') = \sum_{\boldsymbol{d} \in M} \sum_{\boldsymbol{d}' \in M'} \left( \overline{\boldsymbol{d}} \cdot \overline{\boldsymbol{d}'} \right)^2 \Big/ \left( \frac{n}{pp'} \right).$$

The dependency measure for a design $\mathcal{F} = \{M^1, M^2, \ldots, M^q\}$ is defined by

$$\chi^2(\mathcal{F}) \equiv \sum_{1 \le r < s \le q} \chi^2(M^r, M^s).$$

From these notations and definitions, the problem of constructing a mixed-level design can be described as generating of columns $\{M^1, M^2, \ldots, M^q\}$ from set $\mathcal{M}^n$ while maintaining a low level of $\chi^2(M^r, M^s)$ $(1 \le r < s \le q)$.

**Remark** From the definition of family $\mathcal{D}_p^n$, the condition $d_1+\cdots+d_n = n/p$ implies the equal occurrence property of columns, *i.e.*, every column contains each level of $\{1, 2, \ldots, p\}$ exactly $n/p$ times. For example, when $p = 3$ and $n = 12$, each level of $\{1, 2, 3\}$ appears four times respectively. The equal occurrence property was assumed in previous studies (see Booth and Cox (1962) or Lin (1993) or Wu (1993), for example) and it is also assumed here.

# 3 Generalized theorems

## 3.1 Number of dimensions in orthogonal design

A design $\{M^1, M^2, \ldots, M^q\}$ consisting of mutually independent columns, *i.e.*, $\chi^2(M^r, M^s) = 0$ $(1 \le \forall r < \forall s \le q)$, is called an orthogonal design. Here we consider the maximum size of mixed-level orthogonal design. For any $S \subseteq \mathbf{R}^n$, $\mathrm{spn}(S)$ denotes the linear subspace spanned by $S$. The linear subspace $\{\boldsymbol{x} \in \mathbf{R}^n \mid \mathbf{1}^t\boldsymbol{x} = 0\}$ is denoted by $H$, where $\mathbf{1}$ is the all-one vector.

**Theorem 1** *Any orthogonal design* $\mathcal{F} = \{M^1, M^2, \ldots, M^q\}$ *satisfies the inequality*

$$\sum_{r=1}^{q}(p(M^r) - 1) \le n - 1. \tag{1}$$

Proof. From the definition, $\dim(\overline{M^r}) = p(M^r) - 1$ and $\mathrm{spn}(\overline{M^r}) \subseteq H$ for all $r \in \{1, 2, \ldots, q\}$. When $r \ne s$, the assumption $\chi^2(M^r, M^s) = 0$ implies $\mathrm{spn}(\overline{M^r}) \perp \mathrm{spn}(\overline{M^s})$. Thus, the equality $\dim(\overline{M^1} \cup \overline{M^2} \cup \cdots \cup \overline{M^q}) = \sum_{r=1}^{q}(p(M^r) - 1)$ holds, so $\sum_{r=1}^{q}(p(M^r) - 1) \le \dim(H) = n - 1$. //

From Equation (1), we define a measure for the degree of saturation by

$$v = \sum_{r=1}^{q}(p(M^r) - 1)/(n - 1),$$

where $v = 1$ implies saturated design and $v > 1$ implies supersaturated design. As defined by Booth and Cox (1962), in supersaturated design, the number of columns is greater than or equal to the number of rows. The term "supersaturated" means that it is impossible to estimate the effects of the assigned factors. In two-level design, the estimation is impossible if the number of columns is greater than or equal to the number of rows. In mixed-level design, however, this relationship between impossibility and the number is not maintained. We call a supersaturated design if $\sum_{r=1}^{q}(p(M^r) - 1) > n - 1$ based on impossibility on the estimation.

## 3.2 Lower bound of dependency

In this subsection, we propose a lower bound of the dependency.

**Theorem 2** *Any design $\mathcal{F} = \{M^1, M^2, \ldots, M^q\}$ satisfies the inequality*

$$\chi^2(\mathcal{F}) \geq (1/2)v(v-1)n(n-1), \tag{2}$$

*where $v$ is the degree of saturation defined by $\sum_{r=1}^{q}(p(M^r)-1)/(n-1)$.*

Proof. For any column $M^r \in \mathcal{F}$, we denote $p(M^r)$ by $p_r$ and $M^r$ by $\{\boldsymbol{d}_1^r, \boldsymbol{d}_2^r, \ldots, \boldsymbol{d}_{p_r}^r\}$. We set $p^* \equiv p_1 + \cdots + p_q$. Let $X$ be an $n \times p^*$ matrix defined by

$$X = \left[\sqrt{p_1}\overline{\boldsymbol{d}_1^1}, \sqrt{p_1}\overline{\boldsymbol{d}_2^1}, \ldots, \sqrt{p_1}\overline{\boldsymbol{d}_{p_1}^1}, \sqrt{p_2}\overline{\boldsymbol{d}_1^2}, \sqrt{p_2}\overline{\boldsymbol{d}_2^2}, \ldots, \sqrt{p_2}\overline{\boldsymbol{d}_{p_2}^2}, \ldots, \sqrt{p_q}\overline{\boldsymbol{d}_1^q}, \sqrt{p_q}\overline{\boldsymbol{d}_2^q}, \ldots, \sqrt{p_q}\overline{\boldsymbol{d}_{p_q}^q}\right].$$

We denote the positive semi-definite matrix $X^tX$ by Y and the ordered eigenvalues of $Y$ by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{p^*} \geq 0$. Because each column vector of $X$ is contained in the $(n-1)$-dimensional subspace $H$, the rank of $Y$ is less than or equal to $n-1$. Thus, we have $\lambda_n = \lambda_{n+1} = \cdots = \lambda_{p^*} = 0$.

Because $Y$ is symmetric, the multi-set of the eigenvalues of $Y^tY$ becomes $\{\lambda_1^2, \lambda_2^2, \ldots, \lambda_{p^*}^2\}$, so we have

$$\begin{aligned}
\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{n-1}^2 &= \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{p^*}^2 = \mathrm{tr}(Y^tY) = \sum_{i=1}^{p^*}\sum_{j=1}^{p^*}(y_{ij})^2 \\
&= \sum_{r=1}^{q}\sum_{i=1}^{p_r}\sum_{s=1}^{q}\sum_{j=1}^{p_s}\left(\sqrt{p_r}\overline{\boldsymbol{d}_i^r}\cdot\sqrt{p_s}\overline{\boldsymbol{d}_j^s}\right)^2 = \sum_{r=1}^{q}\sum_{i=1}^{p_r}\sum_{s=1}^{q}\sum_{j=1}^{p_s} p_r p_s\left(\overline{\boldsymbol{d}_i^r}\cdot\overline{\boldsymbol{d}_j^s}\right)^2 \\
&= \sum_{r=1}^{q}\sum_{s=1}^{q} n\chi^2(M^r, M^s) = 2n\sum_{1\leq r<s\leq q}\chi^2(M^r, M^s) + n\sum_{r=1}^{q}\chi^2(M^r, M^r) \\
&= 2n\sum_{1\leq r<s\leq q}\chi^2(M^r, M^s) + n\sum_{r=1}^{q} n(p_r-1) = 2n\sum_{1\leq r<s\leq q}\chi^2(M^r, M^s) + n^2v(n-1).
\end{aligned}$$

From this we obtain

$$\chi^2(\mathcal{F}) = \sum_{1\leq r<s\leq q}\chi^2(M^r, M^s) = (1/(2n))\left((\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{n-1}^2) - n^2v(n-1)\right).$$

A lower bound of the value $\lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{n-1}^2$ is obtained as the optimal value of the convex quadratic programming problem:

$$\begin{aligned}
\text{minimize} \quad & \lambda_1^2 + \lambda_2^2 + \cdots + \lambda_{n-1}^2 \\
\text{subject to} \quad & \lambda_1 + \lambda_2 + \cdots + \lambda_{n-1} = \mathrm{tr}(Y).
\end{aligned}$$

The definition of $Y$ implies that

$$\mathrm{tr}(Y) = \sum_{r=1}^{q}\sum_{i=1}^{p_r}\left(\sqrt{p_r}\overline{\boldsymbol{d}_i^r}\cdot\sqrt{p_r}\overline{\boldsymbol{d}_i^r}\right) = \sum_{r=1}^{q}\sum_{i=1}^{p_r} p_r(n/p_r^2)(p_r-1) = \sum_{r=1}^{q} n(p_r-1) = nv(n-1).$$

The above convex quadratic programming problem has a unique optimal solution:

$$\lambda_1^* = \lambda_2^* = \cdots = \lambda_{n-1}^* = \mathrm{tr}(Y)/(n-1) = nv.$$

The optimal value is equal to $(nv)^2(n-1)$, which implies

$$\chi^2(\mathcal{F}) \geq (1/(2n))\left((nv)^2(n-1) - n^2v(n-1)\right) = (1/2)n(n-1)v(v-1). //$$

### 3.3 Condition to attain lower bound

In this subsection, we consider a design that attains the lower bound proposed above. To obtain this design, we need the following two lemmas.

**Lemma 1** *For any column $M \in \mathcal{M}^n$, every vector $\boldsymbol{m} \in \mathrm{spn}(\overline{M})$ satisfies*

$$\sum_{\boldsymbol{d} \in M} (\overline{\boldsymbol{d}} \cdot \boldsymbol{m})\, \overline{\boldsymbol{d}} = \frac{n}{p(M)}\, \boldsymbol{m}.$$

Proof. For simplicity, we denote $p(M)$ by $p$. For any $\boldsymbol{d}' \in M$, we have the following equality:

$$\begin{aligned}
\left( \sum_{\boldsymbol{d} \in M} (\overline{\boldsymbol{d}} \cdot \boldsymbol{m})\, \overline{\boldsymbol{d}} \cdot \overline{\boldsymbol{d}'} \right) &= \sum_{\boldsymbol{d} \in M \setminus \{\boldsymbol{d}'\}} (\overline{\boldsymbol{d}} \cdot \boldsymbol{m}) \left( \overline{\boldsymbol{d}} \cdot \overline{\boldsymbol{d}'} \right) + \left( \overline{\boldsymbol{d}'} \cdot \boldsymbol{m} \right) \left( \overline{\boldsymbol{d}'} \cdot \overline{\boldsymbol{d}'} \right) \\
&= \left( - \sum_{\boldsymbol{d} \in M \setminus \{\boldsymbol{d}'\}} \overline{\boldsymbol{d}} \cdot \boldsymbol{m} \right) (n/p^2) + \left( \overline{\boldsymbol{d}'} \cdot \boldsymbol{m} \right) (n/p^2)(p-1) \\
&= \left( \overline{\boldsymbol{d}'} \cdot \boldsymbol{m} \right) (n/p^2) + \left( \overline{\boldsymbol{d}'} \cdot \boldsymbol{m} \right) (n/p^2)(p-1) = \left( (n/p)\boldsymbol{m} \cdot \overline{\boldsymbol{d}'} \right).
\end{aligned}$$

Because $\boldsymbol{m} \in \mathrm{spn}(\overline{M})$, the above equality implies the desired result. //

**Lemma 2** *Let $\mathcal{F} = \{M^1, M^2, \ldots M^q\}$ be a saturated and orthogonal design. Every column $M \in \mathcal{M}^n$ satisfies the equality*

$$\sum_{r=1}^{q} \chi^2(M, M^r) = n(p(M) - 1).$$

Proof. We denote $p(M)$ by $p$ and $p(M^r)$ by $p_r$. For each $\boldsymbol{d} \in M$ and $M^r \in \mathcal{F}$, we denote the projection of $\overline{\boldsymbol{d}}$ to $\mathrm{spn}(\overline{M^r})$ by $\overline{\boldsymbol{d}}^{(r)}$. Because $\mathcal{F}$ is saturated and orthogonal, vectors $\overline{\boldsymbol{d}}^{(1)}, \overline{\boldsymbol{d}}^{(2)}, \ldots, \overline{\boldsymbol{d}}^{(q)}$ are mutually orthogonal and $\overline{\boldsymbol{d}}^{(1)} + \overline{\boldsymbol{d}}^{(2)} + \ldots + \overline{\boldsymbol{d}}^{(q)} = \overline{\boldsymbol{d}}$. Thus, we have

$$\begin{aligned}
\sum_{r=1}^{q} \chi^2(M, M^r) &= \sum_{r=1}^{q} (p p_r / n) \sum_{\boldsymbol{d} \in M} \sum_{\boldsymbol{d}' \in M^r} \left( \overline{\boldsymbol{d}} \cdot \overline{\boldsymbol{d}'} \right)^2 \\
&= \sum_{r=1}^{q} (p p_r / n) \sum_{\boldsymbol{d} \in M} \sum_{\boldsymbol{d}' \in M^r} \left( \overline{\boldsymbol{d}}^{(r)} \cdot \overline{\boldsymbol{d}'} \right) \left( \overline{\boldsymbol{d}}^{(r)} \cdot \overline{\boldsymbol{d}'} \right) \\
&= \sum_{r=1}^{q} (p p_r / n) \sum_{\boldsymbol{d} \in M} \left( \left( \sum_{\boldsymbol{d}' \in M^r} \overline{\boldsymbol{d}'} \cdot \overline{\boldsymbol{d}}^{(r)} \right) \overline{\boldsymbol{d}'} \cdot \overline{\boldsymbol{d}}^{(r)} \right) \\
&= \sum_{r=1}^{q} (p p_r / n) \sum_{\boldsymbol{d} \in M} \left( (n/p_r) \overline{\boldsymbol{d}}^{(r)} \cdot \overline{\boldsymbol{d}}^{(r)} \right) = \sum_{r=1}^{q} p \sum_{\boldsymbol{d} \in M} \left( \overline{\boldsymbol{d}}^{(r)} \cdot \overline{\boldsymbol{d}}^{(r)} \right) \\
&= p \sum_{\boldsymbol{d} \in M} \sum_{r=1}^{q} \left( \overline{\boldsymbol{d}}^{(r)} \cdot \overline{\boldsymbol{d}}^{(r)} \right) = p \sum_{\boldsymbol{d} \in M} \left( \overline{\boldsymbol{d}} \cdot \overline{\boldsymbol{d}} \right) = p \sum_{\boldsymbol{d} \in M} (n/p^2)(p-1) = n(p-1). //
\end{aligned}$$

The above lemmas imply the following theorem.

**Theorem 3** *Let $\mathcal{F}$ be a design and $\{\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_v\}$ be a partition of $\mathcal{F}$ such that each member of the partition is a saturated and orthogonal design. Then we have the equality $\chi^2(\mathcal{F}) = (1/2)n(n-1)v(v-1)$.*

Proof. Because each member of the partition is saturated and orthogonal, $\chi^2(\mathcal{F}_1) = \chi^2(\mathcal{F}_2) = \cdots = \chi^2(\mathcal{F}_v) = 0$. Therefore, we have the following:

$$\begin{aligned}\chi^2(\mathcal{F}) &= \sum_{1\le r<s\le v}\sum_{M\in\mathcal{F}_r}\sum_{M'\in\mathcal{F}_s}\chi^2(M,M') = \sum_{1\le r<s\le v}\sum_{M\in\mathcal{F}_r} n(p(M)-1) \\ &= \sum_{1\le r<s\le v} n(n-1) = (1/2)v(v-1)n(n-1).//\end{aligned}$$

A saturated and orthogonal design satisfies Equation (2) by equality. A design matrix $\boldsymbol{F}^* = [\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_v]$ attains the lower bound, where $\boldsymbol{F}_1, \boldsymbol{F}_2, \ldots, \boldsymbol{F}_v$ are saturated and orthogonal designs.

### 3.4 Relationships to previous studies

The above theorems can be regarded as a generalization of previous studies to mixed-level supersaturated design, where the dependency is defined by the $\chi^2$ statistic. Specifically, Tang and Wu (1997) have shown a lower bound of the average squared inner product over all pairs in a design, where this average is sometimes denoted by $E(s^2)$. In addition, they have shown that a supersaturated design partitioned into saturated and orthogonal designs is optimum in terms of the $E(s^2)$ criterion. Based on this property, they developed a method for constructing a design by sequentially adding matrices generated by permutation of rows in an initial matrix that is a saturated and orthogonal design. Their algorithm is justified by the property on the $E(s^2)$ criterion. Theorems 2 and 3 can be regarded as a generalization of the results of Tang and Wu (1997).

Lemma 2 is an extension to mixed-level design of the property of multi-level supersaturated design proven by Niki, Hashiguchi, Yamada and Ikebe (1997). Yamada, Ikebe, Hashiguchi and Niki (1997) have shown a method for constructing three-level supersaturated designs by sequentially adding matrices generated by permutation of rows in an initial design matrix that is a three-level saturated and orthogonal design. Theorems 2 and 3 ensure the optimality of their three-level designs in terms of $\chi^2$-dependency, although they did not give a theoretical justification for their algorithm.

## 4 Construction and evaluation

### 4.1 Construction

In this section we describe the construction of a mixed-level supersaturated design consisting of two-level and three-level columns with $n$=12 runs based on the theorems shown in Section 3. A design consisting of $p$-level columns is called a $p$-level design. Theorems 2 and 3 imply that a supersaturated design partitioned into saturated and orthogonal designs is optimal in terms of $\chi^2$-dependency. However, there is no three-level orthogonal design with $n = 12$ runs, although there are two-level saturated and orthogonal designs with $n = 12$ runs. Therefore, we construct mixed-level design $\mathcal{F} = \mathcal{F}_2 \cup \mathcal{F}_3$, where $\mathcal{F}_2$ is a two-level saturated and orthogonal design and $\mathcal{F}_3$ is a three-level design.

Let $\mathcal{F}_2 = \{M_2^1, M_2^2, \ldots, M_2^{n-1}\}$ and $\mathcal{F}_3 = \{M_3^1, M_3^2, \ldots, M_3^q\}$. When $\mathcal{F}_2$ is a saturated and orthogonal design, Lemma 2 says that $\sum_{r=1}^{n-1} \chi^2(M_3, M_2^r)$ is a constant for any $M_3 \in \mathcal{M}_3^{12}$. We hereby introduce criterion

$$\max\{\chi^2(M_2^r, M_3^s) \mid 1 \le r \le n-1, 1 \le s \le q\}.$$

For two three-level columns, we introduce criterion

$$\max\{\chi^2(M_3^r, M_3^s) \mid 1 \le r < s \le q\}.$$

The following lemma is convenient for constructing two-level and three-level supersaturated designs.

**Lemma 3** *Let $\mathcal{F}_2 = \{M_2^1, M_2^2, \ldots, M_2^{n-1}\}$ be a two-level saturated and orthogonal design. There is no three-level design $\mathcal{F}_3 = \{M_3^1, M_3^2, \ldots, M_3^q\}$ such that* $\max\{\chi^2(M_2^r, M_3^s) \mid 1 \le r \le n-1, 1 \le s \le q\} < 2n/(n-1)$.

Proof. The condition $\max\{\chi^2(M_2^r, M_3^s) \mid 1 \le r \le n-1, 1 \le s \le q\} < 2n/(n-1)$ implies $\sum_{r=1}^{n-1} \chi^2(M_2^r, M_3^s) < 2n$, while Lemma 2 implies $\sum_{r=1}^{n-1} \chi^2(M_2^r, M_3^s) = 2n$. Contradiction.//

Lemma 3 shows that there is no column $M_3$ in $\mathcal{M}_3^{12}$ that satisfies $\max\{\chi^2(M_2^r, M_3) \mid 1 \le r \le n-1\} < 2n/(n-1) = 24/11$. Because the variations in $\chi^2(M_2, M_3)$ are $\{0, 2.0, 6.0, 8.0\}$ and 6.0 comes after 2.0, we hereby explore three-level columns $\{M_3^1, M_3^2, \ldots, M_3^q\}$ under the condition $\max\{\chi^2(M_2^r, M_3^s) \mid 1 \le r \le n-1, 1 \le s \le q\} = 6.0$.

The $(12 \times 11)$ matrix

$$\boldsymbol{F}_2 = \begin{bmatrix} 1 & 1 & 1 & 1 & 2 & 1 & 1 & 2 & 2 & 2 & 2 \\ 1 & 1 & 1 & 2 & 1 & 2 & 2 & 1 & 1 & 2 & 2 \\ 1 & 1 & 2 & 1 & 1 & 2 & 2 & 2 & 2 & 1 & 1 \\ 1 & 2 & 2 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & 2 \\ 1 & 2 & 1 & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 1 \\ 1 & 2 & 2 & 1 & 2 & 2 & 1 & 1 & 1 & 2 & 1 \\ 2 & 1 & 2 & 1 & 2 & 1 & 2 & 1 & 1 & 1 & 2 \\ 2 & 1 & 1 & 2 & 2 & 2 & 1 & 1 & 2 & 1 & 1 \\ 2 & 1 & 2 & 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 \\ 2 & 2 & 1 & 1 & 1 & 2 & 1 & 2 & 1 & 1 & 2 \\ 2 & 2 & 1 & 1 & 1 & 1 & 2 & 1 & 2 & 2 & 1 \\ 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 & 2 \end{bmatrix}$$

is a saturated and orthogonal design matrix. The algorithm for constructing design matrix $\boldsymbol{F} = [\boldsymbol{F}_2, \boldsymbol{F}_3]$ is as follows. At the start of the algorithm, we set design $\mathcal{F}_3 = \phi$. In each iteration, when a three-level column $M_3$ is found that satisfies

$$\max\{\chi^2(M_2^r, M_3) \mid 1 \le r \le n-1\} \le 6.0 \quad \text{and} \quad \max\{\chi^2(M_3^r, M_3) \mid \mathcal{M}_3^r \in \mathcal{F}_3\} \le \chi_c^2,$$

$M_3$ is added to the set $\mathcal{F}_3$. We examine all columns in the set $\mathcal{M}_3^{12}$ in an arbitrary order. The threshold value $\chi_c^2$ is selected from $\{1.5, 3.0, 6.0, 11.5, 12.0, 15.0, 24.0\}$ which is the set of variations of $\chi^2(M_3^r, M_3^s)$. Table 1 shows the matrices constructed using the algorithm. The left portion, consisting of five columns, is generated to satisfy the two conditions: $\max\{\chi^2(M_2^r, M_3^s) \mid 1 \le r \le n-1, 1 \le s \le q\} \le 6.0$ and $\max\{\chi^2(M_3^r, M_3^s) \mid 1 \le r < s \le q\} \le \chi_c^2 = 1.5$.

### 4.2 Evaluation

Theorem 2 is useful for evaluating the dependency because the lower bound can be applied to any type of supersaturated design. We define an index for measuring design efficiency for given design $\mathcal{F}$ by

$$\frac{(1/2)n(n-1)v(v-1)}{\chi^2(\mathcal{F})},$$

where we call $\chi^2$-efficiency. This index measures the degree of attainment compared with an optimal design in terms of $\chi^2$-dependency. It derives from an analogy to well-known design indices, such as $D$-efficiency and $G$-efficiency. For example, when $\chi^2$-efficiency is equal to 1, the design is optimal in terms of $\chi^2$-dependency. Table 2 shows the number of columns, the degree of saturation $v$, the lower bounds, $\chi^2(\mathcal{F})$ and $\chi^2$-efficiency for design matrices shown in Table 1. The frequencies of the $\chi^2$ values are shown in Table 3. For example, there are 55 pairs of columns where one column is selected from $\boldsymbol{F}_2$ and the other one from $\boldsymbol{F}_{3a}$. Also, 9 pairs of $\chi^2$ values are 0's, 39 pairs of $\chi^2$ values are 2's and 7 pairs of $\chi^2$ values are 6's. According to $\chi^2$-efficiency, $[\boldsymbol{F}_2, \boldsymbol{F}_{3b}]$ and $[\boldsymbol{F}_2, \boldsymbol{F}_{3c}]$ are relatively better than $[\boldsymbol{F}_{3b}]$ and $[\boldsymbol{F}_{3c}]$, respectively. This is because adding a saturated and orthogonal design is advantageous in terms of $\chi^2$-efficiency.

## 5 Concluding Remarks

We have presented generalized theorems that provide a theoretical background for supersaturated design. Specifically, a lower bound of $\chi^2$-dependency for any type of supersaturated design has been proposed, compared to previous studies which obtained a similar lower bound for two-level supersaturated designs. We showed that a mixed-level supersaturated design attains the lower bound if the design is partitioned into saturated and orthogonal designs. This suggests the possibility of construction of an optimal mixed-level supersaturated design in terms of $\chi^2$-dependency through the sequential addition of design matrices generated by permutation of rows in an initial mixed-level saturated and orthogonal design matrix.

We have constructed mixed-level supersaturated designs with $n = 12$ runs, and we have derived $\chi^2$-efficiency as an index for measuring the degree of attainment compared with an optimal design in terms of $\chi^2$-dependency. We evaluated the constructed designs based on $\chi^2$-efficiency. The construction of better designs in terms of $\chi^2$-efficiency may be possible and it is a topic for future research.

## References


[1] Booth, K.H.V. and Cox, D.R. (1962). Some systematic supersaturated designs. *Technometrics* **4** 489-495.

[2] Cheng, C.S. (1997). $E(s^2)$- optimal supersaturated designs. *Statist. Sinica* **7** 929-939.

[3] Deng, L.Y., Lin, D.K.J. and Wang, J.N. (1994). Supersaturated design using Hadamard matrix. *IBM Res. Rep.* RC19470 IBM Watson Research Center.

[4] Iida, T. (1994). A construction method of two-level supersaturated design derived from $L12$. *Japanese J. Appl. Statist.* **23** 147-153 (in Japanese).

[5] Li, W.W. and Wu, C.F.J. (1997). Columnwise-pairwise algorithms with applications to the construction of supersaturated designs. *Technometrics* **39** 171-179.

[6] Lin, D.K.J. (1993). A new class of supersaturated designs. *Technometrics* **35** 28-31.

[7] Lin, D.K.J. (1995). Generating systematic supersaturated designs. *Technometrics* **37** 213-225.

[8] Nguyen, N.K. (1996). An algorithmic approach to constructing supersaturated designs. *Technometrics* **38** 69-73.

[9] Niki, N., Hashiguchi, H., Yamada, S. and Ikebe, Y. (1997). A fundamental theorem on supersaturated design. (in preparation).

[10] Plackett, R.L. and Burman, J.P. (1946). The design of optimum multifactorial experiments. *Biometrika* **33** 303-325.

[11] Satterthwaite, F.E. (1959). Random balance experimentation (with discussion). *Technometrics* **1** 111-137.

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

[13] Wu, C.F.J. (1993). Construction of supersaturated designs through partially aliased interactions. *Biometrika* **80** 661-669.

[14] Yamada, S., Ikebe, Y., Hashiguchi, H. and Niki, N. (1997). Construction of three-level supersaturated design. (submitted).

[15] Yamada, S. and Lin, D.K.J. (1997). Supersaturated design including an orthogonal base. *Canadian J. Statist.* **25** 203-213.

[16] Yamada, S., Lin, D.K.J. and Yasunari, Y. (1997). Three-level supersaturated design. *Memoirs of Tokyo Metropolitan Institute of Technology* **11** (in printing).

Table 1: Explored three-level design matrices.

| $\boldsymbol{F}_{3a}(\chi_c^2 = 1.5)$ | | | | | $\boldsymbol{F}_{3b}(\chi_c^2 = 3.0)$ | | | | | | | $\boldsymbol{F}_{3c}(\chi_c^2 = 6.0)$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 2 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| 1 | 3 | 3 | 3 | 3 | 1 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 2 | 2 |
| 2 | 1 | 2 | 2 | 3 | 2 | 1 | 2 | 3 | 1 | 3 | 3 | 2 | 1 | 2 | 2 | 3 | 2 | 3 | 3 | 1 | 1 |
| 3 | 1 | 3 | 3 | 2 | 3 | 2 | 1 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 2 | 2 | 1 | 3 | 2 | 3 |
| 2 | 2 | 1 | 3 | 1 | 2 | 2 | 3 | 1 | 3 | 3 | 1 | 2 | 2 | 2 | 3 | 1 | 3 | 2 | 2 | 1 | 3 |
| 2 | 2 | 2 | 3 | 3 | 2 | 3 | 1 | 2 | 3 | 2 | 3 | 2 | 3 | 3 | 2 | 2 | 1 | 2 | 3 | 3 | 2 |
| 2 | 3 | 3 | 1 | 2 | 2 | 3 | 3 | 3 | 2 | 1 | 2 | 2 | 3 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 1 |
| 3 | 2 | 3 | 2 | 1 | 3 | 3 | 2 | 1 | 3 | 1 | 3 | 3 | 2 | 3 | 1 | 2 | 2 | 3 | 2 | 3 | 2 |
| 3 | 3 | 1 | 1 | 3 | 3 | 3 | 3 | 2 | 1 | 3 | 1 | 3 | 2 | 3 | 3 | 3 | 3 | 2 | 3 | 3 | 3 |
| 3 | 3 | 2 | 2 | 1 | 3 | 1 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 1 | 2 | 3 | 3 | 3 | 2 | 3 | 3 |

| $\boldsymbol{F}_{3c}(\chi_c^2 = 6.0)$ (continued) | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 3 | 3 |
| 2 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 2 | 2 | 2 | 3 | 3 | 3 | 1 | 1 | 3 | 3 | 1 | 1 |
| 2 | 2 | 2 | 3 | 1 | 2 | 2 | 2 | 3 | 2 | 3 | 3 | 1 | 2 | 3 | 1 | 3 | 1 | 1 | 1 | 2 |
| 1 | 3 | 3 | 3 | 2 | 1 | 2 | 2 | 3 | 3 | 1 | 1 | 3 | 2 | 2 | 3 | 2 | 1 | 3 | 1 | 1 |
| 3 | 1 | 3 | 3 | 2 | 2 | 3 | 1 | 2 | 1 | 2 | 3 | 2 | 3 | 3 | 3 | 3 | 3 | 1 | 2 | 3 |
| 3 | 2 | 1 | 1 | 2 | 1 | 3 | 3 | 2 | 3 | 1 | 3 | 1 | 2 | 1 | 2 | 1 | 2 | 3 | 3 | 2 |
| 3 | 3 | 2 | 3 | 3 | 3 | 1 | 3 | 3 | 1 | 2 | 1 | 3 | 3 | 1 | 2 | 1 | 3 | 2 | 2 | 3 |
| 1 | 3 | 3 | 2 | 1 | 3 | 1 | 3 | 2 | 2 | 3 | 3 | 2 | 1 | 2 | 3 | 3 | 3 | 1 | 3 | 3 |
| 2 | 1 | 3 | 1 | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 2 | 3 | 1 | 2 | 1 | 3 | 2 | 2 | 2 | 1 |
| 3 | 3 | 1 | 2 | 3 | 2 | 2 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 3 | 2 | 1 | 3 | 3 | 2 |

Table 2: Evaluation of constructed supersaturated design.

| | Three-level | | | Mixed-level | | |
|---|---|---|---|---|---|---|
| | $\boldsymbol{F}_{3a}$ | $\boldsymbol{F}_{3b}$ | $\boldsymbol{F}_{3c}$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3a}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3b}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3c}]$ |
| Number of columns | 5 | 7 | 31 | 15 | 18 | 42 |
| Degree of saturation | 0.91 | 1.27 | 5.64 | 1.91 | 2.27 | 6.64 |
| Lower bound | - | 22.9 | 1724.7 | 114.5 | 190.9 | 2468.7 |
| $\chi^2(\mathcal{F})$ | 15.0 | 48.0 | 1905.0 | 135.0 | 216.0 | 2649.0 |
| $\chi^2$-efficiency | - | 0.47 | 0.91 | 0.85 | 0.88 | 0.93 |

Table 3: Frequency of $\chi^2$ values.

(i) $\chi^2(M_2, M_3)$

| $\chi^2$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3a}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3b}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3c}]$ |
|---|---|---|---|
| 0 | 9 | 17 | 51 |
| 2 | 39 | 48 | 249 |
| 6 | 7 | 12 | 41 |

(ii) $\chi^2(M_3, M_3)$

| $\chi^2$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3a}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3b}]$ | $[\boldsymbol{F}_2, \boldsymbol{F}_{3c}]$ |
|---|---|---|---|
| 1.5 | 10 | 10 | 136 |
| 3.0 | 0 | 11 | 91 |
| 6.0 | 0 | 0 | 238 |