# Design Enumeration through Projection and Isomorphism Examination Based on Counting Vector

國立清華大學統計研究所

博士論文

指導教授：　鄭少為　博　士

博　士　生：　林　長　鋆

中　華　民　國　九十八　年　一　月

# 中文摘要

在實驗設計的研究中，設計完整列舉與同構檢驗是兩項很困難的工作。雖然這是兩項不同的工作，但彼此之間卻有很密切的關係。設計完整列舉是指生成所有不同構的設計，如所有的正規設計或所有的直交陣列等。而同構檢驗則是比較兩個設計是否可經由行列或是水準互換而得到相同結構。在列舉所有的設計時，經常在某階段需要透過同構檢驗來減少多餘的計算。本篇論文包含兩個主要的部份。在第一個部份，我們基於指標函數和投影理論，提出一個組裝法來列舉所有的兩水準設計。此組裝法利用一個層級的結構而逐步地生成所有的設計。我們亦對此組裝法發展了一個演算法，來完整地生成所有非同構的設計。在第二個部份，我們提出利用計數向量的同構檢驗法。我們亦證明了此方法比文獻中曾提出的許多同構檢驗法更有效率。我們亦使用設計投影的技巧來更進一步改善檢驗的效率。

**Abstract**

Design enumeration and isomorphism examination are two difficult tasks in the study of experimental design. The two tasks are different but often related to each other. Design enumeration is concerned with the complete generation of certain types of designs, such as regular designs or orthogonal arrays. Isomorphism examination, on the other hand, compares whether two designs have the same structure subject to some row and column operations. During the process of design enumeration, it is often required to perform isomorphism examination at some stage to reduce the possible redundant calculation. This dissertation includes two main parts. In the first part, we propose an assembly method based on the indicator function and the projection to enumerate all non-isomorphic two-level designs. The assembly method allows us to generate all designs sequentially in a hierarchical structure. We present an algorithm based on the assembly method and generate a complete catalogue of non-isomorphic designs for some cases. In the second part, we develop an initial screening method based on the counting vector for isomorphism examination. We prove that the method provides a more efficient examination than some methods proposed in other articles. The technique of projection is also applied to improve the examination efficiency.

# 誌謝辭

首先我要感謝鄭少為老師給我的指導及幫助，使得此篇論文可以順利地完成。並感謝所有口試委員的指正及建議，使得論文內容更加完整。另外，感謝所有教授我統計學理的老師，使我建立良好的統計基礎。也感謝研究所同學的協助，解決我許多生活及學習上的疑惑。當然，我要感謝我的父母及岳父母，培養我良好的品格及積極向上的進取心。還有，我要感謝我的太太筱晴，無條件的支持我，並陪我走過這些日子。最後我要感謝我兩個可愛的兒子和女兒，他們總是帶給我歡笑及對人生的希望。

# Contents

# Chapter 1

# Introduction

Design enumeration and isomorphism examination are two difficult and time-consuming tasks in the study of experimental design. The two tasks are different but often related to each other. Design enumeration is concerned with the *complete* generation of certain types of designs, such as regular designs or orthogonal arrays. Isomorphism examination, on the other hand, compares whether two designs have the same structure subject to some row and column permutations and level exchange. During the process of design enumeration, it is often required to perform isomorphism examination at some stage to reduce the possible redundant calculation. In practice, design enumeration and isomorphism examination are also important issues in the search of optimal designs, such as minimum aberration designs. Since isomorphic designs share the same statistical properties, it is enough to choose one design matrix to represent the whole group of isomorphic designs. It can significantly reduce the computation for finding optimal designs by only searching among these chosen designs. If a complete catalogue of all non-isomorphic designs is not available, we may get a design that is not globally optimal. Therefore, design enumeration and isomorphism examination play important roles in the exploration of optimal designs.

In Chapter 2, we study the methods about the design enumeration. For the regular design, which can be characterized in terms of a group structure, the complete enumeration can be achieved by considering all possible combinations of group generators. However, for the non-regular designs, the task of design enumeration becomes much more difficult because the group structure is no longer available. To tackle this problem, we employ the indicator function approach. The indicator function is a polynomial representative of a design matrix and can be applied on designs with or without a group structure. In addition, we employ the technique of projection to construct all possible designs. Based on the indicator function and projection technique, we propose an assembly method which allows us to generate the design by assembling the indicator functions of the projections. We find that there exists a hierarchical structure between the projections and the designs generated from them. This hierarchical structure makes it possible to sequentially construct all designs through the assembly method. By applying these methods, we can efficiently enumerate all non-isomorphic designs for many cases.

The projection index set is an accompaniment naturally appeared in the assembled method. We utilize it to perform an initial isomorphism examination for the designs constructed by the assembly method in Chapter 2 to save computation time. However, when the run size and the number of factors are lager, the isomorphism examination based on the projective index set becomes inefficient. It then calls for the development of better methods for isomorphism examination.

In Chapter 3, we propose an innovative method that can significantly improve the efficiency of the isomorphism examination. In this chapter, we study the relationship between two isomorphism designs in terms of their counting vectors, whose components represent the number of the replicates of the experimental runs. We find that the operations of sign switch, column and row permutations on the design matrix are related to the rearrangement of components of the counting vector. Some sufficient and necessary conditions for two counting vectors to be isomorphic are therefore developed. The conditions offer us a theoretical basis to present an isomorphism examination measure, called the *split-N matrix*. The split-N matrix is invariant to the sign switch, column and row permutations so that it can be a measure for the isomorphism examination. We also find that some existing measures for isomorphism examination can be expressed as a function of the split-N matrix. In other words, non-isomorphic designs that can be distinguished by those measures can be classified by the split-N matrix, but not the other way around. That is, split-N matrix has higher classification efficiency than those measures. The technique of projection is also applied in the examination method based on the split-N matrix, which greatly improves the examination efficiency. Some simplified methods are proposed for the designs with large number of factors. They may have lower efficiency than the split-N matrix but can greatly reduce the computation time and the storage memory when the number of factors is large.

# Chapter 2

# Design enumeration through projection

## 2.1 Introduction

Early works on design enumeration focused on regular designs. Because regular designs possess a group structure, its enumeration can be more easily implemented by considering all possible combinations of group generators. For a $2^{k-m}$ regular design, the number of its possible combinations of group generators is less than $\binom{2^k-1}{m}$. Although the number could be large for designs with larger run sizes and more factors, we at least know how to construct all designs in the case of regular designs. For non-regular designs, the task of design enumeration becomes much more difficult because the group structure is no longer available. Recently, a new mathematical framework for design theory called *indicator function* was developed for non-regular designs. An indicator function is a polynomial representative of a design matrix and it can be applied on designs with or without a group structure. Through indicator function approach, the enumeration of non-regular designs can be translated into a problem of finding all coefficients of a polynomial that can constitute an indicator function.

We denote by $OA(n, k, s, d)$ an $n \times k$ matrix which can represent an orthogonal array with $n$ runs, $k$ factors each with $s$ levels, and strength $d$. The run size $n$ of an $OA$ of strength $d$ must be a multiple of $s^d$. Two $OA$s are called *isomorphic* if one can be obtained from the other by row (run) permutation, column (factor) permutation and level exchange (or sign switch when $s = 2$). To use the definition to check whether two $OA(n, k, s, d)$'s are isomorphic, we must compare at most $n!k!(s!)^k$ design matrices. For instance, it requires $16!15!2^{15} \approx 8.97 \times 10^{29}$ comparisons for two non-isomorphic $OA(16, 15, 2, 2)$'s. The isomorphism examination of two designs based on the original definition is very time consuming when the run size or the number of factors is large. The indicator function approach is also useful in dealing with this issue. For example, suppose that design $\mathcal{D}'$ is derived from design $\mathcal{D}$ only through row permutation. Their indicator functions are identical. This property makes it easier to discuss the isomorphism of designs in terms of their indicator functions.

In this chapter, we propose a method to enumerate non-isomorphic two-level designs. The method includes two main parts. The first part is a projection approach to construct *all* indicator functions. Based on the relationship existing among the projections of a design, we propose an assembly method to sequentially generate indicator functions in a hierarchical order. The second part is an isomorphism examination that utilizes some projection properties.

The remainder of this chapter is organized as follows. Section 2.2 reviews some works on the subjects of isomorphism examination and design enumeration. Indicator function is also introduced in this section. Section 2.3 introduces the assembly method and a hierarchical structure for sequential generation of designs. Section 2.4 provides two methods for isomorphism examination. Section 2.5 gives an algorithm which is developed based on materials in Section 2.3 and Section 2.4. Some computational results are also given in this section.

## 2.2 Review

We review some works on design enumeration and isomorphic examination in this section. A recently developed mathematic framework for design theory, indicator function, and its application on design enumeration and isomorphism examination are also introduced.

### 2.2.1 Design enumeration and isomorphic examination

Most works of design enumeration in literature were accomplished case by case. For example, Seiden and Zemach (1966) gave a complete enumeration of $OA(n, d+1, 2, d)$. Fujii, Namikawa and Yamamoto (1989) completely enumerated $OA(2^{d+1}, d+2, 2, d)$ and $OA(2^{d+1}, d+3, 2, d)$. Chen, Sun and Wu (1993) explored the algebraic structure of regular designs and gave a collection of regular designs with 16, 32 and 64 runs through an exhaustive computer search. Lam and Tonchev (1996) completely enumerated all $OA(27, 12, 3, 2)$. Hedayat, Seiden and Stufken (1997) enumerated all $OA(54, 5, 3, 3)$. Yumiba, Hyodo and Yamamoto (1997) classified all $OA(24, 6, 2, 2)$. Yamamoto, Fujii, Hyodo and Yumiba (1992a, 1992b) and Hedayat, Sloane and Stufken (1999) enumerated all $OA(n, n-1, 2, 2)$'s for $n = 4, 8, 12, 16, 20, 24$. Sun, Li and Ye (2002) also succeeded in enumerating non-isomorphic $OA(n, k, 2, 2)$'s for $n = 12, 16, 20$ and arbitrary $k$ through a thorough computer search. Xu (2005) gave a complete catalogue of three-level regular fractional factorial designs with 27, 81, 243 and 279 runs. The method of enumerating non-isomorphic $OA(n, d+2, 2, d)$ was demonstrated by Stufken and Tang (2007). They applied the properties of $J$-characteristic to solve an equation under some constraints, of which each solution represents a class of isomorphic designs.

Some isomorphism examination methods were proposed for both regular and non-regular designs recently. Clark and Dean (2001) introduced the Hamming distance, which records the number of differences between two rows in a design. Through Hamming distance, one can distinguish two designs if they are non-isomorphic or find their permutation relationship if they are isomorphic. Clark and Dean (2001) also provided the necessary and sufficient conditions for isomorphism of two-level fractional factorial designs. Ma, Fang and Lin (2001) combined Hamming distance and the measures of uniformity, called $CD_2^2$, to develop a *p-dimensional $CD_2^2$ projection frequency*, which records the $CD_2^2$-values of all $p$-factor projections of a design. According to the projection frequency, some non-isomorphic designs can be distinguished. Although they conjectured that it is a sufficient condition for isomorphism, it has been shown that this method fails for some cases. Xu (2005) introduced the coding theory approach for three-level regular fractional factorial designs and defined the power moments, $K_u$, by calculating the number of coincidences between two rows. Similar to Ma, Fang and Lin (2001), Xu (2005) obtained *p-dimensional $K_u$ projection frequency* by projecting a design to arbitrary $p$ factors. Whenever two designs have different $p$-dimensional $K_u$ projection frequency for some $u$ and $p$, they must be non-isomorphic.

### 2.2.2 Indicator function and $J$-characteristics

Indicator function was first defined in Fontana, Pistone and Rogantin (2000) for studying two-level fractional factorial designs without replicates. Ye (2003) extended it to two-level fractional factorial designs with replicates. Recently, Cheng and Ye (2004) generalized it to designs with more than two levels. Indicator function for two-level designs can be briefly described as follows. Let $\mathcal{G}$ be a $2^k$ full factorial design with levels labeled by $-1$ and 1. The design points of $\mathcal{G}$ can be regarded as the solutions of the polynomial system $x_i^2 - 1 = 0, i = 1, 2, \cdots, k$. Let $\mathcal{T} = \{1, 2, \cdots, k\}$ denote the collection of factors of $\mathcal{G}$. For any non-empty subset $\mathbf{t}$ of $\mathcal{T}$, define

$$C_{\mathbf{t}}(\mathbf{x}) = \prod_{j \in \mathbf{t}} x_j$$

and $C_{\emptyset}(\mathbf{x}) = 1$ for $\mathbf{x} \in \mathcal{G}$. Let $\mathcal{D}$ be a $k$-factor fractional factorial design (i.e., $\mathcal{D} \subseteq \mathcal{G}$), in which a design point in $\mathcal{G}$ is allowed to appear more than once. The indicator function of $\mathcal{D}$, denoted by $F_{\mathcal{D}}(\mathbf{x})$, is defined to be the number of appearance of $\mathbf{x}$ in $\mathcal{D}$ for every $\mathbf{x} \in \mathcal{G}$. It was shown in Ye (2003) that $F_{\mathcal{D}}(\mathbf{x})$ can be expressed as:

$$\sum_{\mathbf{t} \subseteq \mathcal{T}} b_{\mathbf{t}} C_{\mathbf{t}}, \tag{2.1}$$

where

$$b_{\mathbf{t}} = 2^{-k} \sum_{\mathbf{x} \in \mathcal{D}} C_{\mathbf{t}}. \tag{2.2}$$

We denote the term with the highest order in $F_{\mathcal{D}}(\mathbf{x})$ by $C_{\mathcal{T}}$ and the coefficient of $C_{\mathcal{T}}$ by $b_{\mathcal{T}}$. The $C_{\mathcal{T}}$ is the only term in $F_{\mathcal{D}}$ with order $k$. Every indicator function can be expressed as a polynomial in which every term is of the form $x_1^{d_1} \cdot \cdots \cdot x_k^{d_k}$, where $d_1, \cdots, d_k \in \{0, 1\}$. In the following context, when we refer to polynomials, we mean polynomials of such form.

Another theoretical structure parallel to indicator function is $J$-characteristics, which were first introduced in Deng and Tang (1999) and Tang and Deng (1999) but explicitly defined in Tang (2001). $J$-characteristics are similar to the coefficients in an indicator function, i.e., $b_{\mathbf{t}}$'s. Although $J$-characteristics do not utilize the polynomial framework as indicator function approach does, they play very similar roles in design theory. Stufken and Tang (2007) applied $J$-characteristics to enumerate all non-isomorphic $OA(n, k, 2, d)$ for the case $k = d + 2$. From the viewpoint of indicator function, their method can be described as follows. The indicator function of an $OA$ of strength $d$ has the property that $b_{\phi} = n/2^k$ and $b_{\mathbf{t}} = 0$ for all nonempty $\mathbf{t}$ with $\|\mathbf{t}\| \le d$, where $\|\mathbf{t}\|$ is the number of elements in $\mathbf{t}$. In the indicator function of an $OA(n, k = d + 2, 2, d)$, there are $k$ $\mathbf{t}$'s with $\|\mathbf{t}\| = d + 1$ and only one $\mathbf{t}$ with $\|\mathbf{t}\| = d + 2$. We denote the $b_{\mathbf{t}}$'s with $\|\mathbf{t}\| \ge d + 1$

by $b_{\mathbf{t}_1}, b_{\mathbf{t}_2}, \cdots, b_{\mathbf{t}_k}, b_{\mathbf{t}_{k+1}}$, where $\mathbf{t}_p = \mathcal{T} - \{k+1-p\}$ for $p = 1, \cdots, k$ and $\mathbf{t}_{k+1} = \mathcal{T}$. Except $b_\phi$, the $b_{\mathbf{t}_1}, \cdots, b_{\mathbf{t}_{k+1}}$ are the only coefficients that can be non-zero in the indicator function of an $OA(n, k = d+2, 2, d)$. Stufken and Tang (2007) proved that:

1. when $k$ is odd, every class of isomorphic $OA(n, k = d+2, 2, d)$'s contains a unique array whose $b_{\mathbf{t}_p}$'s satisfy

$$b_{\mathbf{t}_1} \leq \cdots \leq b_{\mathbf{t}_{k-1}} \leq -|b_{\mathbf{t}_k}|, \ b_{\mathbf{t}_{k+1}} \leq 0; \tag{2.3}$$

2. when $k$ is even, every class of isomorphic $OA(n, k = d+2, 2, d)$'s contains a unique array whose $b_{\mathbf{t}_p}$'s satisfy either

$$b_{\mathbf{t}_1} \leq \cdots \leq b_{\mathbf{t}_{k-1}} \leq -|b_{\mathbf{t}_{k+1}}| \tag{2.4}$$

   or

$$b_{\mathbf{t}_1} \leq \cdots \leq b_{\mathbf{t}_{k-1}} \leq -|b_{\mathbf{t}_k}|, \ b_{\mathbf{t}_{k+1}} < -|b_{\mathbf{t}_k}|. \tag{2.5}$$

By substituting the design point $\mathbf{1} = (1, \cdots, 1)$ into the indicator function of an $OA(n, k = d+2, 2, d)$, we can get

$$b_\phi + b_{\mathbf{t}_1} + b_{\mathbf{t}_2} + \cdots + b_{\mathbf{t}_{k+1}} = w, \tag{2.6}$$

where $w$ is the number of appearance of the design point $\mathbf{1}$, which must be a non-negative integer. The complete set of non-isomorphic $OA(n, k = d+2, 2, d)$'s can be obtained by solving equation (2.6) under constraint (2.3) for an odd $k$ or under constraint either (2.4) or (2.5) for an even $k$.

The method in Stufken and Tang (2007) simultaneously achieves two objectives: complete enumeration of designs and isomorphism examination. First, because the solutions to equation (2.6) contain all possible combinations of coefficients, the method can construct all designs. Second, the constraints (2.3), (2.4), or (2.5) make it possible that every class of isomorphic designs only appears once in the search of the solutions. In other words, each solution of (2.6) under the constraints must represent a design and the solutions of non-isomorphic designs must be different. Hence, isomorphism examination becomes unnecessary, which saves a lot of computation.

Unfortunately, the generalization of this method to the cases $k > d+2$ is very difficult because of three reasons. First, when $k > d+2$, the number of the coefficients that could be non-zero increases dramatically. For instance, when $d = 2$ and $k = d + 4$, we would need to handle $\binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 42$ coefficients. It means that there would be 42 indeterminate terms in an equation like (2.6). Second, some solutions of an equation like (2.6) do not represent a design. Third, when $k > d+2$, the relationship among the non-zero coefficients becomes much more complicated. The constraints (2.3), (2.4), or (2.5) are not enough to describe the relationship so that we may generate more solutions than required. In other words, because some solutions are corresponding to designs that are isomorphic, an isomorphism examination is inevitable. In view of these problems, we take a different approach based on projection to tackle the problem of design enumeration and isomorphism examination for $k > d+2$, which will be presented in later sections.

## 2.3 Design enumeration

We will present in this section an innovative method based on projection for complete design enumeration. Let $\mathcal{D}$ be an $n \times k$ design matrix and $\mathbf{d}_j$ be the $jth$ column of $\mathcal{D}$. Let $\mathcal{D}_{(-j)}$ be the $n \times (k-1)$ design matrix obtained by excluding $\mathbf{d}_j$ from $\mathcal{D}$. We refer to $\mathcal{D}_{(-j)}$ as a leave-one-out (LOO) projection of $\mathcal{D}$. The Figure 2.1 presents an example of a 3-factor design $\mathcal{A}$ and its three LOO projections. Suppose that $F_{\mathcal{D}}(\mathbf{x}) = \sum_{t \subseteq \mathcal{T}} b_{\mathbf{t}} C_{\mathbf{t}}$ is the indicator function of $\mathcal{D}$. By

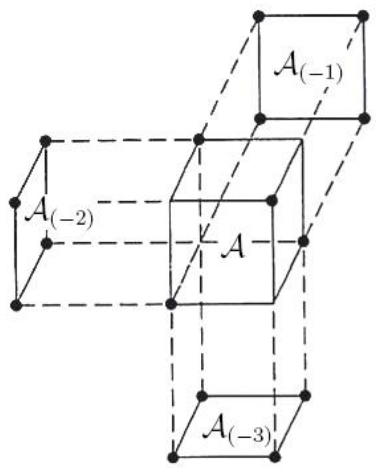


Figure 2.1: The 3-factor design $\mathcal{A}$ and its three LOO projections $\mathcal{A}_{(-1)}$, $\mathcal{A}_{(-2)}$ and $\mathcal{A}_{(-3)}$.

Theorem 1 in Ye (2003), the indicator function of $\mathcal{D}_{(-j)}$ is

$$F_{\mathcal{D}_{(-j)}}(x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_k) = \sum_{\mathbf{t} \subseteq \mathcal{T}_{(-j)}} 2b_{\mathbf{t}} C_{\mathbf{t}}, \tag{2.7}$$

where $\mathcal{T}_{(-j)} = \mathcal{T} - \{j\}$. That is, the indicator function of $\mathcal{D}_{(-j)}$ is obtained from $F_{\mathcal{D}}$ by doubling the coefficients and eliminating the terms whose suffixes include factor $j$. Obviously, given $F_{\mathcal{D}}$, we can obtain $k$ LOO indicator functions $F_{\mathcal{D}(-j)}$, $j = 1, \cdots, k$, and each of them represents a $(k-1)$-factor fractional factorial design. Because $\mathcal{D}_{(-j)}$ is a projection of $\mathcal{D}$, the strength of $\mathcal{D}_{(-j)}$ is at least as high as the strength of $\mathcal{D}$. Let $\mathcal{F}^*(n, k, 2, d)$ be the collection of all different indicator functions of $OA(n, k, 2, d)$'s. Suppose that $F_{\mathcal{D}}$ is an indicator function in $\mathcal{F}^*(n, k, 2, d)$. Its $F_{\mathcal{D}_{(-j)}}$'s must belong to $\mathcal{F}^*(n, k-1, 2, d)$. If we have $\mathcal{F}^*(n, k-1, 2, d)$, then by applying an assembly method given in Section 2.3.2 for all possible combinations of the indicator functions in $\mathcal{F}^*(n, k-1, 2, d)$, we can construct $\mathcal{F}^*(n, k, 2, d)$. This hierarchical structure makes it possible to sequentially generate *all* indicator functions.

### 2.3.1 Incomplete indicator function

A polynomial is called an *incomplete indicator function* if it is obtained from an indicator function by deleting the highest-order term, $b_{\mathcal{T}}C_{\mathcal{T}}$. Indicator functions of different designs may be corresponding to an identical incomplete indicator function because there exist indicator functions whose coefficients are identical except $b_{\mathcal{T}}$. Note that not every polynomial with $b_{\mathcal{T}} = 0$ is an incomplete indicator function. A necessary and sufficient condition for a polynomial with $b_{\mathcal{T}} = 0$ to be an incomplete indicator function is that there must exist at least one $b'_{\mathcal{T}}$ such that the sum of the polynomial and $b'_{\mathcal{T}}C_{\mathcal{T}}$ is a non-negative integer function.

Let $F'(\mathbf{x})$ be a polynomial with $b_{\mathcal{T}} = 0$. Let $\lfloor w \rfloor$ be the largest integer that is not more than $w$ and $\lceil w \rceil$ be the smallest integer that is not less than $w$. Let $y = \min_{\mathbf{x}\in\mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x})=1} F'(\mathbf{x})$ and $z = \min_{\mathbf{x}\in\mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x})=-1} F'(\mathbf{x})$. Define

$$d(F'(\mathbf{x})) = \begin{cases} \lceil F'(\mathbf{x}) \rceil - F'(\mathbf{x}), & for\ \mathbf{x} \in \mathcal{G} \text{ such that } C_{\mathcal{T}}(\mathbf{x}) = 1 \\ F'(\mathbf{x}) - \lfloor F'(\mathbf{x}) \rfloor, & for\ \mathbf{x} \in \mathcal{G} \text{ such that } C_{\mathcal{T}}(\mathbf{x}) = -1. \end{cases}$$

The following theorem offers the necessary and sufficient condition for $F'(\mathbf{x})$ to be an incomplete indicator function.

**Theorem 1.** *Let $F'(\mathbf{x})$ be a polynomial with $b_{\mathcal{T}} = 0$. The $F'(\mathbf{x})$ is an incomplete indicator function if and only if*
*(a) $d(F'(\mathbf{x}))$ is constant for all $\mathbf{x} \in \mathcal{G}$, and*
*(b) $y \geq -z$.*

*Proof.* Suppose that $F'(\mathbf{x})$ is an incomplete indicator function. There exists an indicator function $F(\mathbf{x})$ and a $b_{\mathcal{T}}$ such that

$$F(\mathbf{x}) = \begin{cases} F'(\mathbf{x}) + b_{\mathcal{T}} = F'(\mathbf{x}) + (b_{\mathcal{T}} - \lfloor b_{\mathcal{T}} \rfloor) + \lfloor b_{\mathcal{T}} \rfloor, & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = 1 \\ F'(\mathbf{x}) - b_{\mathcal{T}} = F'(\mathbf{x}) - (b_{\mathcal{T}} - \lfloor b_{\mathcal{T}} \rfloor) - \lfloor b_{\mathcal{T}} \rfloor, & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = -1. \end{cases}$$

It can be written as

$$b_{\mathcal{T}} - \lfloor b_{\mathcal{T}} \rfloor = \begin{cases} (F(\mathbf{x}) - \lfloor b_{\mathcal{T}} \rfloor) - F'(\mathbf{x}) = \lceil F'(\mathbf{x}) \rceil - F'(\mathbf{x}), & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = 1 \\ F'(\mathbf{x}) - (F(\mathbf{x}) + \lfloor b_{\mathcal{T}} \rfloor) = F'(\mathbf{x}) - \lfloor F'(\mathbf{x}) \rfloor, & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = -1. \end{cases}$$

Therefore, $d(F'(\mathbf{x})) = b_{\mathcal{T}} - \lfloor b_{\mathcal{T}} \rfloor$, which is constant for all $\mathbf{x} \in \mathcal{G}$. The condition (a) is proved. Because $F(\mathbf{x})$ is an indicator function, we have

$$F(\mathbf{x}) = \begin{cases} F'(\mathbf{x}) + b_{\mathcal{T}} \geq 0, & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = 1 \\ F'(\mathbf{x}) - b_{\mathcal{T}} \geq 0, & for\ \mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x}) = -1. \end{cases}$$

Therefore, we have $y + b_{\mathcal{T}} \geq 0$ and $z - b_{\mathcal{T}} \geq 0$, i.e., $y \geq -b_{\mathcal{T}} \geq -z$. The condition (b) follows. Conversely, if $y \leq 0$, then $z - (-y) \geq 0$ by (b). Let $b_{\mathcal{T}} = -y$. Then, $y + b_{\mathcal{T}} = 0$ and $z - b_{\mathcal{T}} \geq 0$. If $z \leq 0$, then $y + z \geq 0$ by (b). Let $b_{\mathcal{T}} = z$. Then, $z - b_{\mathcal{T}} = 0$ and $y + b_{\mathcal{T}} \geq 0$. If $y \geq 0$ and $z \geq 0$, let $b_{\mathcal{T}} = z$ . Then $z - b_{\mathcal{T}} = 0$ and $y + b_{\mathcal{T}} \geq 0$. Since $y = \min_{\mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x})=1} F'(\mathbf{x})$, $y + b_{\mathcal{T}} \geq 0$ ensures that $F'(\mathbf{x}) + b_{\mathcal{T}} \geq 0$ for all $\mathbf{x} \in \mathcal{G}$ such that $C_{\mathcal{T}}(\mathbf{x}) = 1$. Similarly, since $z = \min_{\mathbf{x} \in \mathcal{G} \text{ s.t. } C_{\mathcal{T}}(\mathbf{x})=-1} F'(\mathbf{x})$, $z - b_{\mathcal{T}} \geq 0$ ensures that $F'(\mathbf{x}) - b_{\mathcal{T}} \geq 0$ for all $\mathbf{x} \in \mathcal{G}$ such that $C_{\mathcal{T}}(\mathbf{x}) = -1$. Hence, under the condition (b), there must exist a $b_{\mathcal{T}}$ such that $F'(\mathbf{x}) + b_{\mathcal{T}} \geq 0$ for those $\mathbf{x}$ such that $C_{\mathcal{T}}(\mathbf{x}) = 1$s and $F'(\mathbf{x}) - b_{\mathcal{T}} \geq 0$ for those $\mathbf{x}$ such that $C_{\mathcal{T}}(\mathbf{x}) = -1$s. By (a), if $y + b_{\mathcal{T}}$ or $z - b_{\mathcal{T}}$ is an integer, it is clear that the $F'(\mathbf{x}) + b_{\mathcal{T}}$ are integers for those $\mathbf{x}$ such that $C_{\mathcal{T}}(\mathbf{x}) = 1$s and $F'(\mathbf{x}) - b_{\mathcal{T}}$ are integers for those $\mathbf{x}$ such that $C_{\mathcal{T}}(\mathbf{x}) = -1$s. Hence, if (a) and (b) hold, there must exist a $b_{\mathcal{T}}$ such that $F'(\mathbf{x}) + b_{\mathcal{T}} C_{\mathcal{T}}$ is an indicator function, i.e., $F'(\mathbf{x})$ is an incomplete indicator function. □

For instance, suppose that $\mathcal{T} = \{1, 2\}$. The polynomial $F'(\mathbf{x}) = 3/4 - 2/4C_1(\mathbf{x}) - 5/4C_2(\mathbf{x})$ is a polynomial with $b_{\mathcal{T}} = 0$ but not an incomplete indicator function because for $\mathbf{x} = (1, 1)$, $d(F'(\mathbf{x})) = \lceil F'(\mathbf{x}) \rceil - F'(\mathbf{x}) = -1 - (-1) = 0$ but for $\mathbf{x} = (1, -1)$, $d(F'(\mathbf{x})) = F'(\mathbf{x}) - \lfloor F'(\mathbf{x}) \rfloor = 6/4 - 1 = 2/4$, which violates the condition (a) in Theorem 1. For an incomplete indicator function, the following theorem gives the possible values of $b_{\mathcal{T}}$'s with which the incomplete indicator function can become an indicator function.

**Theorem 2.** *Let $y$ and $z$ be as defined in Theorem 1. Suppose that $F'(\mathbf{x})$ is an incomplete indicator function. A $b_{\mathcal{T}}$ can be used to form an indicator function $F(\mathbf{x}) = F'(\mathbf{x}) + b_{\mathcal{T}} C_{\mathcal{T}}$ if and only if $b_{\mathcal{T}}$ is a value in $\{-y, -y+1, \cdots, z-1, z\}$.*

*Proof.* First, let us prove that $z - (-y)$ is a non-negative integer. Suppose that there exist $\mathbf{x}_1$ and $\mathbf{x}_2$ such that $y = F'(\mathbf{x}_1)$ and $z = F'(\mathbf{x}_2)$. Because $F'(\mathbf{x})$ is an incomplete indicator function, $d(F'(\mathbf{x}_1)) = d(F'(\mathbf{x}_2))$ by Theorem 1 (a). Therefore,

$$\begin{aligned} z-(-y) &= \lfloor z \rfloor + (z - \lfloor z \rfloor) + \lceil y \rceil - (\lceil y \rceil - y) \\ &= \lfloor z \rfloor + d(F'(\mathbf{x}_2)) + \lceil y \rceil - d(F'(\mathbf{x}_1)) \\ &= \lfloor z \rfloor + \lceil y \rceil, \end{aligned}$$

which is an integer. By Theorem 1 (b), it is clear that $\lfloor z \rfloor + \lceil y \rceil \geq 0$. Any $b_{\mathcal{T}}$ such that both $y + b_{\mathcal{T}}$ and $z - b_{\mathcal{T}}$ are non-negative integers can be used to form an indicator function. Hence, $b_{\mathcal{T}}$ can be and must be any value in the intersection of $\{-y, -y+1, -y+2, \cdots\}$ and $\{z, z-1, z-2, \cdots\}$. Because $z - (-y)$ is a non-negative integer, we have $\{-y, -y+1, -y+2, \cdots\} \bigcap \{z, z-1, z-2, \cdots\} = \{-y, -y+1, \cdots, z-1, z\}$. $\square$

For instance, let $\mathcal{T} = \{1, 2\}$ and $F'(\mathbf{x}) = 7/4 - 1/4C_1(\mathbf{x}) + 1/4C_2(\mathbf{x})$. It is clear that $d(F'(\mathbf{x})) = 1/4$ for all $\mathbf{x} \in \mathcal{G}$ and $y > -z$ where $y = 7/4$ and $z = 5/4$. According to Theorem 1, $F'(\mathbf{x})$ is an incomplete indicator function. To obtain an indicator function from $F'(\mathbf{x})$, the $b_{12}$ could be chosen from the following values $-7/4(= -y)$, $-3/4$, $1/4$, $5/4(= z)$ by Theorem 2. The two theorems play important roles in the assembly method that will be introduced in next section. Theorem 1 will be applied to check whether an assembled polynomial is an incomplete indicator function. If the assembled polynomial passes the Theorem 1, the values of $b_{\mathcal{T}}$ given by Theorem 2 complement it to form an indicator function. For the application of the two theorems, reader is referred to Example 1 in

Section 2.3.3.

### 2.3.2 Assembly method

The indicator function of a design can be built from the indicator functions of its LOO projections as explained and illustrated below. Suppose that $\mathcal{A}$ is a three-factor fractional factorial design and its indicator function is

$$F_{\mathcal{A}} = b_{\phi} + b_1C_1 + b_2C_2 + b_3C_3 + b_{12}C_{12} + b_{13}C_{13} + b_{23}C_{23} + b_{123}C_{123}.$$

Its three LOO indicator functions are

$$F_{\mathcal{A}_{(-1)}} = 2b_{\phi} + 2b_2C_2 + 2b_3C_3 + 2b_{23}C_{23},$$

$$F_{\mathcal{A}_{(-2)}} = 2b_{\phi} + 2b_1C_1 + 2b_3C_3 + 2b_{13}C_{13},$$

and

$$F_{\mathcal{A}_{(-3)}} = 2b_{\phi} + 2b_1C_1 + 2b_2C_2 + 2b_{12}C_{12}.$$

In any two LOO indicator functions, the coefficients with the same suffix must be identical. Suppose that we only know the indicator functions of the three LOO projections. To rebuild $F_{\mathcal{A}}$ by using $F_{\mathcal{A}_{(-1)}}$, $F_{\mathcal{A}_{(-2)}}$ and $F_{\mathcal{A}_{(-3)}}$, we can first divide coefficients of each LOO indicator function by two and then assign them to the corresponding coefficients in $F_{\mathcal{A}}$. The reason of dividing by two is that the coefficients in $F_{\mathcal{A}_{(-j)}}$ are twice of their corresponding coefficients in $F_{\mathcal{A}}$. The relationship between $F_{\mathcal{A}}$ and its LOO indicator functions is given in Table 2.1. The coefficients in each indicator function are listed in each row and the column shows the relationship of the coefficients between indicator functions. Since none of the LOO indicator functions contains the term $b_{123}C_{123}$, this assembly method only allows us to construct an incomplete indicator function. By applying Theorem 2 on the incomplete indicator function, we soon obtain suitable values for $b_{123}$ and successfully rebuild an $F_{\mathcal{A}}$. This method can be used for the designs with

Table 2.1: The relationship between $F_{\mathcal{A}}$ and its LOO indicator functions.

| design | $C_\phi$ | $C_1$ | $C_2$ | $C_3$ | $C_{12}$ | $C_{13}$ | $C_{23}$ | $C_{123}$ |
|---|---|---|---|---|---|---|---|---|
| $F_{\mathcal{A}}$ | $b_\phi$ | $b_1$ | $b_2$ | $b_3$ | $b_{12}$ | $b_{13}$ | $b_{23}$ | $b_{123}$ |
| $\frac{1}{2}F_{\mathcal{A}_{(-3)}}$ | $b_\phi$ | $b_1$ | $b_2$ | | $b_{12}$ | | | |
| $\frac{1}{2}F_{\mathcal{A}_{(-2)}}$ | $b_\phi$ | $b_1$ | | $b_3$ | | $b_{13}$ | | |
| $\frac{1}{2}F_{\mathcal{A}_{(-1)}}$ | $b_\phi$ | | $b_2$ | $b_3$ | | | $b_{23}$ | |

more factors. For those indicator functions with identical incomplete indicator function, they would have the same LOO indicator functions. Therefore, it is possible that we may obtain more than one indicator function from a group of LOO indicator functions.

In general, we only have $k$ $(k-1)$-factor indicator functions and do not know whether they are LOO projections of a $k$-factor indicator function. The assembly method can be applied to check whether their combination can construct a $k$-factor indicator function. The procedure is given as follows. Suppose that $F_{\mathcal{D}}$ is a $k$-factor indicator function in $\mathcal{F}^*(n,k,2,d)$ and its coefficients are unknown. We can first select one of the $(k-1)$-factor indicator functions in $\mathcal{F}^*(n,k-1,2,d)$ and assign it to $F_{\mathcal{D}_{(-1)}}$, i.e., replace the factor labels $1,2,\cdots,k-1$ in the chosen indicator function by $2,\cdots,k$ respectively, divide its coefficients by two, and set them to the corresponding coefficients in $F_{\mathcal{D}}$. Next, select one of the $(k-1)$-factor indicator functions in $\mathcal{F}^*(n,k-1,2,d)$ and assign it to $F_{\mathcal{D}_{(-2)}}$, i.e., replace its factor labels $1,2,\cdots,k-1$ by $1,3,\cdots,k$ respectively. The indicator function for $F_{\mathcal{D}_{(-2)}}$ could be the one we have selected for $F_{\mathcal{D}_{(-1)}}$, i.e., repeated selection is allowed. Because the coefficients in $F_{\mathcal{D}}$ with suffixes not including 1 have already been determined by $F_{\mathcal{D}_{(-1)}}$, the assignment of $F_{\mathcal{D}_{(-2)}}$ must obey the predetermined condition of the coefficients. The procedure can be repeated sequentially to assign the other LOO indicator functions until $F_{\mathcal{D}_{(-k)}}$. Finally, a polynomial with unknown $b_{\mathcal{T}}$ is assembled from $k$ $(k-1)$-factor indicator functions. We can check whether the polynomial is an incomplete indicator function by using Theorem 1, and determine the value of $b_{\mathcal{T}}$ from Theorem 2 if the polynomial passes the examination of Theorem 1.

### 2.3.3 An example

Here we give a simple but clear example to demonstrate the assembly method.

**Example 1.** Suppose that we have chosen three two-factor indicator functions from $\mathcal{F}^*(6,3,2,0)$,

$$F_{\mathcal{A}_1} = \frac{3}{2} + \frac{1}{2}C_1 - \frac{1}{2}C_2 - \frac{1}{2}C_{12},$$

$$F_{\mathcal{A}_2} = \frac{3}{2} - \frac{1}{2}C_1 - \frac{1}{2}C_2 - \frac{1}{2}C_{12},$$

and

$$F_{\mathcal{A}_3} = \frac{3}{2} - \frac{1}{2}C_1 + \frac{1}{2}C_2 + \frac{1}{2}C_{12}.$$

Our purpose is to use them to generate a three-factor indicator function, say

$$F_{\mathcal{A}} = b_\phi + b_1C_1 + b_2C_2 + b_3C_3 + b_{12}C_{12} + b_{13}C_{13} + b_{23}C_{23} + b_{123}C_{123}.$$

We can first assign $F_{\mathcal{A}_1}$ to $F_{\mathcal{A}_{(-1)}}$ by replacing factors label 1 and 2 in $F_{\mathcal{A}_1}$ to 2 and 3, respectively. This assignment can be denoted by

$$F_{\mathcal{A}_1} \Rightarrow F_{\mathcal{A}_{(-1)}} = \frac{3}{2} + \frac{1}{2}C_2 - \frac{1}{2}C_3 - \frac{1}{2}C_{23}.$$

After dividing $F_{\mathcal{A}_{(-1)}}$ by two, setting its coefficients to the corresponding terms in $F_{\mathcal{A}}$, we obtain

$$F_{\mathcal{A}} = \frac{3}{4} + b_1C_1 + \frac{1}{4}C_2 - \frac{1}{4}C_3 + b_{12}C_{12} + b_{13}C_{13} - \frac{1}{4}C_{23} + b_{123}C_{123}.$$

For $F_{\mathcal{A}_{(-2)}}$, the predetermined coefficients by the assignment of $F_{\mathcal{A}_{(-1)}}$ are $b_\phi(=\frac{3}{4})$ and $b_3(=-\frac{1}{4})$ (see Table 2.1). Therefore, the two indicator functions, $F_{\mathcal{A}_1}$ and

$F_{\mathcal{A}_2}$, are eligible for the assignment of $F_{\mathcal{A}_{(-2)}}$. We denote $F_{\mathcal{A}_1} \Rightarrow F_{\mathcal{A}_{(-2)}}$ as case 1 and $F_{\mathcal{A}_2} \Rightarrow F_{\mathcal{A}_{(-2)}}$ as case 2; their corresponding $F_{\mathcal{A}}$'s are:

case 1: $\frac{1}{2}F_{\mathcal{A}_1} \Rightarrow \frac{1}{2}F_{\mathcal{A}_{(-2)}} = \frac{3}{4} + \frac{1}{4}C_1 - \frac{1}{4}C_3 - \frac{1}{4}C_{13}$

$$F_{\mathcal{A}} = \frac{3}{2} + \frac{1}{4}C_1 + \frac{1}{4}C_2 - \frac{1}{4}C_3 + b_{12}C_{12} - \frac{1}{4}C_{13} - \frac{1}{4}C_{23} + b_{123}C_{123}.$$

case 2: $\frac{1}{2}F_{\mathcal{A}_2} \Rightarrow \frac{1}{2}F_{\mathcal{A}_{(-2)}} = \frac{3}{4} - \frac{1}{4}C_1 - \frac{1}{4}C_3 - \frac{1}{4}C_{13}$

$$F_{\mathcal{A}} = \frac{3}{2} - \frac{1}{4}C_1 + \frac{1}{4}C_2 - \frac{1}{4}C_3 + b_{12}C_{12} - \frac{1}{4}C_{13} - \frac{1}{4}C_{23} + b_{123}C_{123}.$$

Now, let us consider the last assignment for both cases. In case 1, the predetermined coefficients for $F_{\mathcal{A}_{(-3)}}$ are

$$b_\phi(=\frac{3}{4}), b_1(=\frac{1}{4}), \text{ and } b_2(=\frac{1}{4}). \tag{2.8}$$

Because none of the coefficients in the $F_{\mathcal{A}_1}$, $F_{\mathcal{A}_2}$, and $F_{\mathcal{A}_3}$ can satisfy the condition in (2.8), this assembly fails. In case 2, the predetermined coefficients are $b_\phi(=\frac{3}{4})$ ,$b_1(=-\frac{1}{4})$, and $b_2(=\frac{1}{4})$. The coefficients in $F_{\mathcal{A}_3}$ are eligible for the assignment of $F_{\mathcal{A}_{(-3)}}$. Finally, we obtain

$$F_{\mathcal{A}} = \frac{3}{4} - \frac{1}{4}C_1 + \frac{1}{4}C_2 - \frac{1}{4}C_3 + \frac{1}{4}C_{12} - \frac{1}{4}C_{13} - \frac{1}{4}C_{23} + b_{123}C_{123}.$$

By letting $b_{123} = 0$, we obtain a polynomial $F'_{\mathcal{A}}$. The Table 2.2, which checks the conditions in Theorem 1, confirms that $F'_{\mathcal{A}}$ is an incomplete indicator function. By Theorem 2, we soon obtain $b_{123} = -1/4$. Through the procedure, one three-factor indicator function is successfully generated from three two-factor indicator functions. Actually, this three-factor indicator function is corresponding to the design matrix

$$\begin{pmatrix} 1 & 1 & -1 \\ 1 & 1 & -1 \\ -1 & 1 & -1 \\ -1 & 1 & 1 \\ -1 & -1 & 1 \\ -1 & -1 & -1 \end{pmatrix}.$$

Table 2.2: Incomplete indicator function examination

| $x_1$ | $x_2$ | $x_3$ | $C_{123}$ | $F'_{\mathcal{A}}(\mathbf{x})$ | Condition examination |
|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 1/4 | (a). $d(F'_{\mathcal{A}}(\mathbf{x}))$=3/4 |
| -1 | 1 | 1 | -1 | 3/4 | (b). $y = \min_{\mathbf{x}\in\mathcal{G}} \text{ s.t. } _{C_{123}(\mathbf{x})=1} F'_{\mathcal{A}}(\mathbf{x}) = 1/4$ |
| 1 | -1 | 1 | -1 | -1/4 | $z = \min_{\mathbf{x}\in\mathcal{G}} \text{ s.t. } _{C_{123}(\mathbf{x})=-1} F'_{\mathcal{A}}(\mathbf{x}) = -1/4$ |
| -1 | -1 | 1 | 1 | 5/4 | $y \geq (-z)$ |
| 1 | 1 | -1 | -1 | 7/4 | |
| -1 | 1 | -1 | 1 | 5/4 | |
| 1 | -1 | -1 | 1 | 1/4 | |
| -1 | -1 | -1 | -1 | 3/4 | |

### 2.3.4 Hierarchical structure

There exists a hierarchical structure between $\mathcal{F}^*(n, k-1, 2, d)$ and $\mathcal{F}^*(n, k, 2, d)$, for $k = d+1, d+2, \cdots$. This hierarchical structure makes it possible to sequentially generate all indicator functions. Recall that indicator functions of $OA(n, k, 2, d)$'s have the property that $b_\phi = n/2^k$ and $b_{\mathbf{t}} = 0$ for all nonempty $\mathbf{t}$ with $\|\mathbf{t}\| \leq d$. Hence, $\mathcal{F}^*(n, d, 2, d)$ contains only one indicator function $F(\mathbf{x}) = n/2^d$. By applying the assembly method, we can generate $\mathcal{F}^*(n, d+1, 2, d)$ as shown below. Let $F(\mathbf{x})$ be an indicator function in $\mathcal{F}^*(n, d+1, 2, d)$. Then, $F(\mathbf{x})$ must be of the form:

$$F(\mathbf{x}) = \frac{n}{2^{d+1}} + b_{\mathcal{T}} C_{\mathcal{T}}, \tag{2.9}$$

where $\mathcal{T} = \{1, 2, \cdots, d+1\}$. By Theorem 2, $b_{\mathcal{T}}$ can be any of the values $-\frac{n}{2^{d+1}}, -\frac{n}{2^{d+1}}+1, \cdots, \frac{n}{2^{d+1}}$. Because there are $(n/2^d+1)$ $b_{\mathcal{T}}$'s and $F(\mathbf{x}) = n/2^{d+1} + b_{\mathcal{T}} C_{\mathcal{T}}$ and $F(\mathbf{x}) = n/2^{d+1} - b_{\mathcal{T}} C_{\mathcal{T}}$ are isomorphic, we can conclude that there exist $\lceil n/2^{d+1} + 1/2 \rceil$ non-isomorphic designs in $\mathcal{F}^*(n, d+1, 2, d)$. The procedure can be repeated to sequentially generate $\mathcal{F}^*(n, d+2, 2, d)$ from $\mathcal{F}^*(n, d+1, 2, d)$, $\mathcal{F}^*(n, d+3, 2, d)$ from $\mathcal{F}^*(n, d+2, 2, d)$, and so on. The relationship between an $F_{\mathcal{D}}$ in $\mathcal{F}^*(n, d+2, 2, d)$ and its LOO indicator functions is shown in Table 2.3. Notice

Table 2.3: The relationship between $F_{\mathcal{D}}$ and its LOO indicator functions where $F_{\mathcal{D}}$ is in $\mathcal{F}^*(n, k = d+2, 2, d)$.

| design | $C_\phi$ | $C_{12\cdots(k-1)}$ | $C_{12\cdots(k-2)k}$ | $\cdots$ | $C_{2\cdots k}$ | $C_{12\cdots k}$ |
|---|---|---|---|---|---|---|
| $F_{\mathcal{D}}$ | $b_\phi$ | $b_{12\cdots(k-1)}$ | $b_{12\cdots(k-2)k}$ | $\cdots$ | $b_{2\cdots k}$ | $b_{12\cdots k}$ |
| $\frac{1}{2}F_{\mathcal{D}_{(-k)}}$ | $b_\phi$ | $b_{12\cdots(k-1)}$ | | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-(k-1))}}$ | $b_\phi$ | | $b_{12\cdots(k-2)k}$ | | | |
| $\vdots$ | $\vdots$ | | $\vdots$ | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-1)}}$ | $b_\phi$ | | | | $b_{2\cdots k}$ | |

*$b_{\mathbf{t}}$'s for $1 \leq \|\mathbf{t}\| \leq d$ are omitted because they are equal to zero.

that none of the non-zero coefficients in these LOO indicator functions has the same suffix except $b_\phi$. This makes the assembly method easier to be implemented because the later assigned LOO indicator functions do not have any restriction on their coefficients imposed by the LOO indicator functions. For $d = 2$, the relationship between $F_{\mathcal{D}}(\mathbf{x})$ in $\mathcal{F}^*(n, d+3, 2, d)$ and its LOO indicator functions is as shown in Table 2.4. When $k \geq d+3$, these LOO indicator functions have overlapped coefficients with each other.

Some issues should be highlighted here. Although all indicator functions in $\mathcal{F}^*(n, k, 2, d)$ can be generated from $\mathcal{F}^*(n, k-1, 2, d)$ by the assembly method, not every indicator functions in $\mathcal{F}^*(n, k-1, 2, d)$ can be obtained by projecting indicator functions in $\mathcal{F}^*(n, k, 2, d)$. For instance, $F_1(\mathbf{x}) = 3/4 - 1/2C_{123} - 1/2C_{124} - 1/2C_{134} - 1/2C_{234} + 1/2C_{1234}$ is the unique isomorphic design in $\mathcal{F}^*(12, 4, 2, 2)$, but $F_2(\mathbf{x}) = 3/2 + 3/2C_{123}$, which is an indicator function in $\mathcal{F}^*(12, 3, 2, 2)$, cannot be obtained from $F_1(\mathbf{x})$ through projection.

The method in Stufken and Tang (2007) can be regarded as a special case of our assembly method. Their method first obtains all possible $b_{\mathbf{t}}$'s for $\|\mathbf{t}\| = d+1$

Table 2.4: The relationship between $F_{\mathcal{D}}$ and its LOO indicator functions where $F_{\mathcal{D}}$ is in $\mathcal{F}^*(n, k = d+3, 2, d)$.

| design | $C_\phi$ | $C_{123}$ | $C_{124}$ | $C_{134}$ | $C_{234}$ | $C_{125}$ | $C_{135}$ | $C_{235}$ | $C_{145}$ | $C_{245}$ | $C_{345}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $F_{\mathcal{D}}$ | $b_\phi$ | $b_{123}$ | $b_{124}$ | $b_{134}$ | $b_{234}$ | $b_{125}$ | $b_{135}$ | $b_{235}$ | $b_{145}$ | $b_{245}$ | $b_{345}$ |
| $\frac{1}{2}F_{\mathcal{D}_{(-5)}}$ | $b_\phi$ | $b_{123}$ | $b_{124}$ | $b_{134}$ | $b_{234}$ | | | | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-4)}}$ | $b_\phi$ | $b_{123}$ | | | | $b_{125}$ | $b_{135}$ | $b_{235}$ | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-3)}}$ | $b_\phi$ | | $b_{124}$ | | | $b_{125}$ | | | $b_{145}$ | $b_{245}$ | |
| $\frac{1}{2}F_{\mathcal{D}_{(-2)}}$ | $b_\phi$ | | | $b_{134}$ | | | $b_{135}$ | | $b_{145}$ | | $b_{345}$ |
| $\frac{1}{2}F_{\mathcal{D}_{(-1)}}$ | $b_\phi$ | | | | $b_{234}$ | | | $b_{235}$ | | $b_{245}$ | $b_{345}$ |

| design | $C_{1234}$ | $C_{1235}$ | $C_{1245}$ | $C_{1345}$ | $C_{2345}$ | $C_{12335}$ |
|---|---|---|---|---|---|---|
| $F_{\mathcal{D}}$ | $b_{1234}$ | $b_{1235}$ | $b_{1245}$ | $b_{1345}$ | $b_{2345}$ | $b_{12345}$ |
| $\frac{1}{2}F_{\mathcal{D}_{(-5)}}$ | $b_{1234}$ | | | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-4)}}$ | | $b_{1235}$ | | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-3)}}$ | | | $b_{1245}$ | | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-2)}}$ | | | | $b_{1345}$ | | |
| $\frac{1}{2}F_{\mathcal{D}_{(-1)}}$ | | | | | $b_{2345}$ | |

and $\|\mathbf{t}\| = d+2$. This step is equivalent to using Theorem 2 to obtain the $b_{\mathcal{T}}$ for the indicator functions in $\mathcal{F}^*(n, d+1, 2, d)$ and $\mathcal{F}^*(n, d+2, 2, d)$. Their method then checks whether the combination of $b_{\mathbf{t}}$'s satisfies some constraints. This step is equivalent to constructing $\mathcal{F}^*(n, d+2, 2, d)$ from $\mathcal{F}^*(n, d+1, 2, d)$ by the assembly method and checking the conditions in Theorem 1.

## 2.4 Isomorphism examination

Since isomorphic designs share the same statistical properties, we only need one of them to represent a class of isomorphic designs. By reducing a class of isomorphic designs to one design, we can save a lot of calculation in design enumeration. In this section, we present two methods for examining isomorphism and reducing the number of designs in $\mathcal{F}^*(n, k, 2, d)$.



### 2.4.1 Method based on group structure

Let $\mathcal{F}(n, k, 2, d)$ be a subsect of $\mathcal{F}^*(n, k, 2, d)$ such that every class of isomorphic designs in $\mathcal{F}^*(n, k, 2, d)$ only has one representative indicator function in $\mathcal{F}(n, k, 2, d)$. To reduce $\mathcal{F}^*(n, k, 2, d)$ to $\mathcal{F}(n, k, 2, d)$ for the case $k = d+2$, Stufken and Tang (2007) developed some constraints on the coefficients of indicator functions as introduced in Section 2.2.2. We will show that these constraints still hold for $k > d+2$. Although for $k > d+2$, these constraints no longer have the ability to fully reduce $\mathcal{F}^*(n, k, 2, d)$ to $\mathcal{F}(n, k, 2, d)$, it is still an efficient method to use the constraints to discard some isomorphic indicator functions in $\mathcal{F}^*(n, k, 2, d)$.

Let $\delta = (\delta_1, \cdots, \delta_k)$ where $\delta_j = \pm 1$ be a vector acting on $\mathcal{D} = (\mathbf{d}_1, \cdots, \mathbf{d}_k)$. The $\delta$ is used to indicate whether the signs of the columns of $\mathcal{D}$ are switched.

When $\delta_j = -1$, the sign of column $j$ is switched, and not switched if $\delta_j = 1$. Let $\mathcal{D}'$ be the design matrix obtained by applying $\delta$ on $\mathcal{D}$, i.e., $\mathcal{D}' = (\delta_1 \mathbf{d}_1, \cdots, \delta_k \mathbf{d}_k)$. The coefficients of the indicator function of $\mathcal{D}'$ are $b'_{\mathbf{t}} = \delta_{\mathbf{t}} b_{\mathbf{t}}$ for $\mathbf{t} \subseteq \mathcal{T}$, where $\delta_{\mathbf{t}} = \prod_{j \in \mathbf{t}} \delta_j$. Let us now pay attention on the $\delta_{\mathbf{t}}$'s with $\|\mathbf{t}\| \geq k-1$. Recall that there exist $k$ $\mathbf{t}$'s with $\|\mathbf{t}\| = k-1$ and only one $\mathbf{t}$ with $\|\mathbf{t}\| = k$. We again denote the $\mathbf{t}$'s with $\|\mathbf{t}\| \geq k-1$ by $\mathbf{t}_1, \mathbf{t}_2 \cdots, \mathbf{t}_k, \mathbf{t}_{k+1}$, where $\mathbf{t}_p = \mathcal{T} \setminus \{k+1-p\}$ for $p = 1, \cdots, k$ and $\mathbf{t}_{k+1} = \mathcal{T}$. For a sign switch indicator $\delta$, the $\delta_{\mathbf{t}}$'s with $\|\mathbf{t}\| \geq k-1$ is denoted by $\delta^* = (\delta_{\mathbf{t}_1}, \cdots, \delta_{\mathbf{t}_k}, \delta_{\mathbf{t}_{k+1}})$. Since there are $2^k$ different $\delta$'s, we have $2^k$ different $\delta^*$'s. If we regard each $\delta^*$ as a run, then the $2^k$ different $\delta^*$'s form a half fractional factorial design, denoted by $\Delta^*$. We regard the $jth$ column of $\Delta^*$ as a factor and label the factor by $j$. Notice that when $k$ is odd, $\Delta^*$ is a half fractional factorial design with the defining relation $I = 12 \cdots k$ and when $k$ is even, $\Delta^*$ is a half fractional factorial design with the defining relation $I = 12 \cdots k(k+1)$. When $k$ is odd, let $\Delta'$ be a design obtained from $\Delta^*$ by deleting any of the first $k$ columns. Then $\Delta'$ forms a full factorial design, i.e., its runs contain all possible level combinations. Let $F_{\mathcal{D}}(\mathbf{x})$ be the indicator function of $\mathcal{D}$. Denote the coefficients with $\|\mathbf{t}\| \geq k-1$ in $F_{\mathcal{D}}(\mathbf{x})$ by $b_{\mathbf{t}_1}, b_{\mathbf{t}_2}, \cdots, b_{\mathbf{t}_k}, b_{\mathbf{t}_{k+1}}$. Suppose that $|b_{\mathbf{t}_{p'}}|$ is the smallest absolute coefficient among all $|b_{\mathbf{t}_p}|$'s, where $1 \leq p \leq k$. Because $\Delta'$ contains all level combinations, we can always find a run $\delta'_0$ in $\Delta'$ such that $b'_{\mathbf{t}_p} = \delta_{\mathbf{t}_p} b_{\mathbf{t}_p} \leq 0$ for $p \neq p'$. Because there exists a one-to-one correspondence between the column permutation of $\mathcal{D}$ and the order of $\{\mathbf{t}_1, \mathbf{t}_2, \cdots, \mathbf{t}_k\}$, we can always find a column permutation to $b'_{\mathbf{t}_p}$'s in an ascending order. This result is summarized in the following proposition.

**Proposition 1.** *When $k$ is odd, every class of isomorphic designs contains a design whose indicator function satisfies*

$$b_{\mathbf{t}_1} \leq \cdots \leq b_{\mathbf{t}_{k-1}} \leq -|b_{\mathbf{t}_k}|, \;\; b_{\mathbf{t}_{k+1}} \leq 0. \tag{2.10}$$

According to Proposition 1, we only need to preserve the designs whose indicator functions satisfy (2.10) and discard the others. Proposition 1 ensures that we will preserve at least one design for each class of isomorphic designs. The similar result for even $k$ is given below.

When $k$ is even, $\Delta^*$ satisfies $I = 12\cdots k(k+1)$. Deleting any column of $\Delta^*$ gives us a full factorial design. By a similar argument as for odd $k$, we obtain the following proposition.

**Proposition 2.** *When $k$ is even, every class of isomorphic designs contains a design whose indicator function satisfies either*

$$b_{\mathbf{t}_1} \le \cdots \le b_{\mathbf{t}_{k-1}} \le -|b_{\mathbf{t}_{k+1}}| \tag{2.11}$$

*or*

$$b_{\mathbf{t}_1} \le \cdots \le b_{\mathbf{t}_{k-1}} \le -|b_{\mathbf{t}_k}|,\ b_{\mathbf{t}_{k+1}} < -|b_{\mathbf{t}_k}|. \tag{2.12}$$

### 2.4.2 Projective index set

We can assign an index to each non-isomorphic designs in $\mathcal{F}(n, k-1, 2, d)$. We refer to the index as *isomorphism index.* An $OA(n, k, 2, d)$ has $k$ LOO projections and each LOO projection has its own isomorphism index. The collection of the isomorphism indices of the $k$ LOO projections can be used to examine isomorphism. The collection is referred to as *projective index set.*

**Theorem 3.** *If $\mathcal{D}$ and $\mathcal{D}'$ are two isomorphic designs, then their projective index sets must be identical.*

*Proof.* Let $\mathcal{D} = (\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_k)$ be an $OA(n, k, 2, d)$. Let $\mathbf{R}$ be an $n \times n$ matrix with elements 0 or 1, where 1 appears exactly one time in each row and each column of $\mathbf{R}$. Any row permutation of $\mathcal{D}$ is corresponding to an $\mathbf{R}$ and can be denoted by $\mathbf{R}\mathcal{D}$. Let $\Psi : \{1, 2, \cdots, k\} \rightarrow \{1, 2, \cdots, k\}$ be a function representing column permutation. Let $\mathbf{S} = (\delta_{ij})$ be a $k \times k$ diagonal matrix where $\delta_{jj} = -1$ when the sign of column $j$ is switched, and $\delta_{jj} = 1$ otherwise. If $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, there must exist an $\mathbf{R}$, an $\mathbf{S}$ and a $\Psi$ such that

$$\mathcal{D}' = \mathbf{R}[\mathbf{d}_{\Psi(1)}, \mathbf{d}_{\Psi(2)}, \cdots, \mathbf{d}_{\Psi(k)}]\mathbf{S}.$$

Let $\mathbf{S}_{(-j)}$ be a $(k-1) \times (k-1)$ matrix obtained by deleting the $jth$ column and the $jth$ row of $\mathbf{S}$. Then the LOO projections of $\mathcal{D}$ and $\mathcal{D}'$ have the relation

$$\mathcal{D}'_{(-j)} = \mathbf{R}\mathcal{D}_{(-\Psi(j))}\mathbf{S}_{(-\Psi(j))}, \ j = 1, 2, \cdots, k. \tag{2.13}$$

When $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, for each LOO projection of $\mathcal{D}$, there must exist a corresponding isomorphic LOO projection of $\mathcal{D}'$. $\square$

Note that the converse of Theorem 3 is not true in general. For instance, $F_{\mathcal{D}}(\mathbf{x}) = 5/2 - 1/2C_{123} - 1/2C_{124} - 1/2C_{134} - 1/2C_{234} + 1/2C_{1234}$ and $F_{\mathcal{D}}(\mathbf{x}) = 5/2 - 1/2C_{123} - 1/2C_{124} - 1/2C_{134} - 1/2C_{234} - 1/2C_{1234}$ have the identical projective index set, but $\mathcal{D}$ and $\mathcal{D}'$ are non-isomorphic. By Theorem 3, we know that any two designs with different projective index sets must be non-isomorphic. This property is useful to divide $\mathcal{F}^*(n, k, 2, d)$ into several groups of non-isomorphic designs. Moreover, if two $k$-factor designs have the same projective index set but different $|b_{\mathcal{T}}|$, they must be non-isomorphic. We will use the projective index set and $|b_{\mathcal{T}}|$ for isomorphism examination as presented in the next section.

## 2.5 Algorithm and results

Based on the materials in Sections 2.3 and 2.4, we develop an algorithm to enumerate non-isomorphic designs. The hierarchical structure in our method allows us to first generate $\mathcal{F}^*(n,k,2,d)$ for $k = d+1$ and then sequentially generate $\mathcal{F}^*(n,k,2,d)$ for higher $k$ by the assembly method. During the generation process, we use the method given in Section 2.4.1 to discard some isomorphic designs. The projective index set and $|b_{\mathcal{T}}|$ are then applied to divide the generated indicator functions into several non-isomorphic groups. Within each group, an exhaustive examination based on the definition of isomorphism, i.e., all column and row permutations and sign switches, is required. Recall that the examination conditions given in Section 2.4.1 are different for odd and even $k$'s. Here we just give the case for odd $k$. The case for even $k$ is similar.

### 2.5.1 Algorithm

Suppose that $\mathcal{F}^*(n,k-1,2,d)$ is known. Let $F_{\mathcal{D}}$ be any indicator function in $\mathcal{F}^*(n,k,2,d)$ and its LOO indicator functions be $F_{\mathcal{D}_{(-1)}},\cdots,F_{\mathcal{D}_{(-k)}}$. Let $b_{\mathcal{T}_{(-j)}}$ denote the highest order coefficient in $F_{\mathcal{D}_{(-j)}}$.

**Step 1.** Assign an indicator function in $\mathcal{F}^*(n,k-1,2,d)$ to $F_{\mathcal{D}_{(-k)}}$. The highest order coefficient of this indicator function should be non-positive by Proposition 1 and it results in $b_{\mathcal{T}_{(-k)}} \leq 0$. After dividing the coefficients of $F_{\mathcal{D}_{(-k)}}$ by two, assign them to the corresponding coefficients in $F_{\mathcal{D}}$, and record the isomorphism index of $F_{\mathcal{D}_{(-k)}}$.

**Step 2.** Sequentially assign the indicator function in $\mathcal{F}^*(n,k-1,2,d)$ to $F_{\mathcal{D}_{(-j)}}$ for $j = k-1, k-2, \cdots 2$. Each assignment must satisfy:

1. the coefficients with the same suffixes among $F_{\mathcal{D}_{(-k)}}, F_{\mathcal{D}_{(-(k-1))}}, \cdots, F_{\mathcal{D}_{(-j)}}$ are consistent;

2. $b_{\mathcal{T}_{(-k)}} \leq b_{\mathcal{T}_{(-(k-1))}} \leq \cdots \leq b_{\mathcal{T}_{(-j)}} \leq 0$.

After dividing the coefficients by two, assign them to the corresponding coefficients in $F_{\mathcal{D}}$, and record the isomorphism index of each assignment simultaneously.

**Step 3.** Assign an indicator function in $\mathcal{F}^*(n, k-1, 2, d)$ to $F_{\mathcal{D}_{(-1)}}$. This assignment must satisfy:

1. the coefficients with the same suffixes among $F_{\mathcal{D}_{(-k)}}, F_{\mathcal{D}_{(-(k-1))}}, \cdots, F_{\mathcal{D}_{(-1)}}$ are consistent;

2. $-|b_{\mathcal{T}_{(-1)}}| \geq b_{\mathcal{T}_{(-2)}}$.

After dividing coefficients by two, assign them to the corresponding coefficients in $F_{\mathcal{D}}$, and record the isomorphism index of $F_{\mathcal{D}_{(-1)}}$.

**Step 4.** An assembly polynomial is obtained if Steps 1 to 3 are successfully performed. Examine whether the polynomial is an incomplete indicator function by Theorem 1. If it passes Theorem 1, obtain the values of $b_{\mathcal{T}}$ by Theorem 2. By Proposition 1, we can only keep $F_{\mathcal{D}}$ with $b_{\mathcal{T}} \leq 0$. If the assembly polynomial does not pass Theorem 1, discard it and go back to Step 1.

**Step 5.** Repeat Step 1 to Step 4 for all possible assignments to obtain a set of indicator functions. Each indicator function in the set has a projective index set. Group together the indicator functions with the same projective index set and $|b_{\mathcal{T}}|$. For the indicator functions in the same group, check isomorphism by the definition.

### 2.5.2 Some results

Using the algorithm given in Section 2.5.1, we completely enumerate non-isomorphic designs for some cases. The numbers of non-isomorphic $OA(n,k,s,d)$ for different $n$, $k$ and $d$ are given in Table 2.5. In these tables, on the right of the colon is the number of the non-isomorphic $OA(n,k,s,d)$. The percentage in the bracket denotes the isomorphism examination efficiency according to projective index set and $|b_{\mathcal{T}}|$. The efficiency is defined by

$$\text{efficiency} = \frac{\text{the number of distinguishable non-isomorphism } OA(n,k,s,d)}{\text{the number of non-isomorphism } OA(n,k,s,d)}.$$

For the cases of $OA(n,4,2,2)$, $OA(n,5,2,3)$, and $OA(n,6,2,4)$, i.e., $k=d+2$, our results are consistent with those in Stufken and Tang (2007). For the cases of $k>d+2$, which are not included in Stufken and Tang (2007), our results show that there exist more and more non-isomorphic $OA(n,k,2,d)$'s when $k-d$ is larger.



Table 2.5: The numbers of non-isomorphic $OA(n,k,s,d)$

| $OA(n,k,s,d)$ | : | # | ( efficiency ) | $OA(n,k,s,d)$ | : | # | ( efficiency ) |
|---|---|---|---|---|---|---|---|
| $OA(4,3,2,2)$ | : | 1 | ( 100.0 % ) | $OA(12,5,2,2)$ | : | 2 | ( 100.0 % ) |
| $OA(8,3,2,2)$ | : | 2 | ( 100.0 % ) | $OA(16,5,2,2)$ | : | 11 | ( 100.0 % ) |
| $OA(12,3,2,2)$ | : | 2 | ( 100.0 % ) | $OA(20,5,2,2)$ | : | 11 | ( 90.9 % ) |
| $OA(16,3,2,2)$ | : | 3 | ( 100.0 % ) | $OA(24,5,2,2)$ | : | 63 | ( 92.1 % ) |
| $OA(20,3,2,2)$ | : | 3 | ( 100.0 % ) | $OA(28,5,2,2)$ | : | 127 | ( 69.3 % ) |
| $OA(24,3,2,2)$ | : | 4 | ( 100.0 % ) | $OA(32,5,2,2)$ | : | 491 | ( 75.6 % ) |
| $OA(28,3,2,2)$ | : | 4 | ( 100.0 % ) | $OA(36,5,2,2)$ | : | 1242 | ( 58.7 % ) |
| $OA(32,3,2,2)$ | : | 5 | ( 100.0 % ) | $OA(40,5,2,2)$ | : | 3919 | ( 58.4 % ) |
| $OA(36,3,2,2)$ | : | 5 | ( 100.0 % ) | $OA(16,5,2,2)$ | : | 27 | ( 100.0 % ) |

Continued...

| $OA(n,k,s,d)$ | : | # | ( efficiency ) | $OA(n,k,s,d)$ | : | # | ( efficiency ) |
|---|---|---|---|---|---|---|---|
| $OA(40,3,2,2)$ | : | 6 | ( 100.0 % ) | $OA(20,5,2,2)$ | : | 75 | ( 93.3 % ) |
| $OA(8,4,2,2)$ | : | 2 | ( 100.0 % ) | $OA(24,5,2,2)$ | : | 1350 | ( 97.7 % ) |
| $OA(12,4,2,2)$ | : | 1 | ( 100.0 % ) | $OA(8,4,2,3)$ | : | 1 | ( 100.0 % ) |
| $OA(16,4,2,2)$ | : | 5 | ( 100.0 % ) | $OA(16,4,2,3)$ | : | 2 | ( 100.0 % ) |
| $OA(20,4,2,2)$ | : | 3 | ( 100.0 % ) | $OA(24,4,2,3)$ | : | 2 | ( 100.0 % ) |
| $OA(24,4,2,2)$ | : | 10 | ( 100.0 % ) | $OA(32,4,2,3)$ | : | 3 | ( 100.0 % ) |
| $OA(28,4,2,2)$ | : | 7 | ( 100.0 % ) | $OA(40,4,2,3)$ | : | 3 | ( 100.0 % ) |
| $OA(32,4,2,2)$ | : | 19 | ( 100.0 % ) | $OA(48,4,2,3)$ | : | 4 | ( 100.0 % ) |
| $OA(36,4,2,2)$ | : | 15 | ( 100.0 % ) | $OA(56,4,2,3)$ | : | 4 | ( 100.0 % ) |
| $OA(40,4,2,2)$ | : | 32 | ( 96.9 % ) | $OA(64,4,2,3)$ | : | 5 | ( 100.0 % ) |
| $OA(8,5,2,2)$ | : | 1 | ( 100.0 % ) | $OA(72,4,2,3)$ | : | 5 | ( 100.0 % ) |
| $OA(80,4,2,3)$ | : | 6 | ( 100.0 % ) | $OA(32,5,2,4)$ | : | 2 | ( 100.0 % ) |
| $OA(16,5,2,3)$ | : | 2 | ( 100.0 % ) | $OA(48,5,2,4)$ | : | 2 | ( 100.0 % ) |
| $OA(24,5,2,3)$ | : | 1 | ( 100.0 % ) | $OA(64,5,2,4)$ | : | 3 | ( 100.0 % ) |
| $OA(32,5,2,3)$ | : | 5 | ( 100.0 % ) | $OA(80,5,2,4)$ | : | 3 | ( 100.0 % ) |
| $OA(40,5,2,3)$ | : | 3 | ( 100.0 % ) | $OA(96,5,2,4)$ | : | 4 | ( 100.0 % ) |
| $OA(48,5,2,3)$ | : | 10 | ( 100.0 % ) | $OA(112,5,2,4)$ | : | 4 | ( 100.0 % ) |
| $OA(56,5,2,3)$ | : | 7 | ( 100.0 % ) | $OA(128,5,2,4)$ | : | 5 | ( 100.0 % ) |
| $OA(64,5,2,3)$ | : | 19 | ( 100.0 % ) | $OA(144,5,2,4)$ | : | 5 | ( 100.0 % ) |
| $OA(72,5,2,3)$ | : | 15 | ( 100.0 % ) | $OA(160,5,2,4)$ | : | 6 | ( 100.0 % ) |
| $OA(80,5,2,3)$ | : | 33 | ( 97.0 % ) | $OA(32,6,2,4)$ | : | 2 | ( 100.0 % ) |
| $OA(16,6,2,3)$ | : | 1 | ( 100.0 % ) | $OA(64,6,2,4)$ | : | 5 | ( 100.0 % ) |
| $OA(24,6,2,3)$ | : | 2 | ( 100.0 % ) | $OA(80,6,2,4)$ | : | 1 | ( 100.0 % ) |
| $OA(32,6,2,3)$ | : | 10 | ( 100.0 % ) | $OA(96,6,2,4)$ | : | 9 | ( 100.0 % ) |
| $OA(40,6,2,3)$ | : | 9 | ( 88.9 % ) | $OA(112,6,2,4)$ | : | 3 | ( 100.0 % ) |
| $OA(48,6,2,3)$ | : | 45 | ( 88.9 % ) | $OA(128,6,2,4)$ | : | 17 | ( 100.0 % ) |

Continued...

| $OA(n,k,s,d)$ | : | # | ( efficiency ) | $OA(n,k,s,d)$ | : | # | ( efficiency ) |
|---|---|---|---|---|---|---|---|
| $OA(64,6,2,3)$ | : | 358 | ( 74.3 % ) | $OA(64,7,2,4)$ | : | 7 | ( 100.0 % ) |
| $OA(24,6,2,3)$ | : | 1 | ( 100.0 % ) | $OA(96,7,2,4)$ | : | 4 | ( 100.0 % ) |
| $OA(32,6,2,3)$ | : | 17 | ( 100.0 % ) | $OA(128,7,2,4)$ | : | 123 | ( 94.3 % ) |
| $OA(16,5,2,4)$ | : | 1 | ( 100.0 % ) | $OA(144,7,2,4)$ | : | 35 | ( 28.6 % ) |

We can find in Table 2.5 that for the cases of $k = d+1$, the numbers of non-isomorphic designs are consistent with our discussion about $OA(n, d+1, 2, d)$ given in Section 2.3.4. From equation (2.9), the number of non-isomorphic $OA(n, d+1, 2, d)$ is $\lceil n/2^{d+1} + 1/2 \rceil$. A more detailed list of these non-isomorphic designs is given in Table 2.6, together with the optimal designs based on minimum aberration criterion. The minimum aberration criterion and the resolution shown in the table were defined in Deng and Tang (1999). For the cases of $k \geq d+2$, the indicator functions of non-isomorphic designs do not follow a systematic structure as in the case of $k = d+1$. We therefore only present the numbers of non-isomorphic designs and the minimum aberration designs for the case $OA(n, 5, 2, 2)$ in Table 2.7 and the case $OA(n, 6, 2, 3)$ in Table 2.8, respectively. In Table 2.5, the 100% efficiency shows that the projective index set and $|b_T|$ can completely distinguish $OA(n, k, 2, d)$ when $k = d+1$. We also find that when $n$ and $k$ are lager, the isomorphism examination by the projective index set and $|b_T|$ becomes less efficient. The efficiency for $OA(36, 5, 2, 2)$ is 58.7%, and for $OA(114, 7, 2, 4)$ is only 28%. In next chapter, we will propose a more efficient method which has better power to classify designs into non-isomorphic groups. More accurate the non-isomorphism classification is, less time the procedure would spend on the definition-based isomorphism examination within each group, which is extremely time consuming.

Table 2.6: The classes of non-isomorphic $OA(n, k = d+1, 2, d)$ and the minimum aberration design in each class

| | | minimum aberration designs | |
|---|---|---|---|
| $n$ | non-isomorphic designs | *Resolution* | $F(\mathbf{x})$ |
| $1 \times 2^d$ | $0.5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} = 0.5$ | 3 | $0.5 + 0.5C_{\mathcal{T}}$ |
| $2 \times 2^d$ | $1 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0, 1\}$ | 4 | 1 |
| $3 \times 2^d$ | $1.5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0.5, 1.5\}$ | 3.67 | $1.5 + 0.5C_{\mathcal{T}}$ |
| $4 \times 2^d$ | $2 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0, 1, 2\}$ | 4 | 2 |
| $5 \times 2^d$ | $2.5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0.5, 1.5, 2.5\}$ | 3.8 | $2.5 + 0.5C_{\mathcal{T}}$ |
| $6 \times 2^d$ | $3 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0, 1, 2, 3\}$ | 4 | 3 |
| $7 \times 2^d$ | $3.5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0.5, 1.5, 2.5, 3.5\}$ | 3.86 | $3.5 + 0.5C_{\mathcal{T}}$ |
| $8 \times 2^d$ | $4 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0, 1, 2, 3, 4\}$ | 4 | 4 |
| $9 \times 2^d$ | $4.5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0.5, 1.5, 2.5, 3.5, 4.5\}$ | 3.88 | $4.5 + 0.5C_{\mathcal{T}}$ |
| $10 \times 2^d$ | $5 + b_{\mathcal{T}}C_{\mathcal{T}}$, $b_{\mathcal{T}} \in \{0, 1, 2, 3, 4, 5\}$ | 4 | 5 |

Table 2.7: The classes of non-isomorphic $OA(n, k = d+3, 2, d = 2)$ and the minimum aberration design in each class

| $n$ | # of non-isomorphic designs | minimum aberration designs | |
|---|---|---|---|
| | | *Resolution* | $F(\mathbf{x})$ |
| 12 | 2 | 3.67 | 0.375, −0.125, −0.125, 0.125, −0.125, 0.125, 0.125, 0.125, −0.125, −0.125, 0.125, −0.125, −0.125, −0.125, −0.125, 0.125, 0. |
| 16 | 11 | 5 | 0.5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, −0.5. |
| 20 | 11 | 3.8 | 0.625, −0.125, 0.125, 0.125, −0.125, 0.125, −0.125, 0.125, −0.125, −0.125, 0.125, −0.125, −0.125, −0.125, −0.125, −0.125, 0. |
| 24 | 63 | 4.67 | 0.75, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, −0.25, −0.25, −0.25, −0.25, 0.25, 0. |
| 28 | 127 | 3.86 | 0.875, −0.125, 0.125, −0.125, −0.125, 0.125, −0.125, −0.125, 0.125, −0.125, −0.125, −0.125, −0.125, −0.125, −0.125, 0.125, 0. |
| 32 | 491 | 5 | 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0. |
| 36 | 1242 | 3.89 | 1.125, −0.125, − 0.125, 0.125, 0.125, −0.125, −0.125, −0.125, −0.125, −0.125, 0.125, −0.125, −0.125, −0.125, −0.125, −0.125, 0. |
| 40 | 3919 | 4.8 | 1.25, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, −0.25, −0.25, −0.25, −0.25, −0.25, 0. |

*The numbers in the column of $F(\mathbf{x})$ are $b_\phi, b_{123}, b_{124}, b_{134}, b_{234}, b_{125}, b_{135}, b_{235}, b_{145}, b_{245}, b_{345}, b_{1234}, b_{1235}, b_{1245}, b_{1345}, b_{2345}, b_{12345}$.

Table 2.8: The classes of non-isomorphic $OA(n, k = d+3, 2, d = 3)$ and the minimum aberration design in each class

| $n$ | # of non-isomorphic designs | minimum aberration designs | |
|---|---|---|---|
| | | resolution | $F(\mathbf{x})$ |
| 16 | 1 | 4 | 0.25, 0, 0, −0.25, 0, 0, 0, 0, −0.25, 0, 0, 0, 0.25, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0. |
| 24 | 2 | 4.67 | 0.375, −0.125, −0.125, −0.125, 0.125, −0.125, −0.125, 0.125, −0.125, −0.125, 0.125, 0.125, 0.125, 0.125, −0.125, −0.125, 0, 0, 0, 0, 0, 0, 0. |
| 32 | 10 | 6 | 0.5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, −0.5. |
| 40 | 9 | 4.8 | 0.625, 0.125, 0.125, 0.125, −0.125, 0.125, −0.125, 0.125, −0.125, −0.125, −0.125, −0.125, 0.125, 0.125, 0.125, 0.125, 0, 0, 0, 0, 0, 0, 0. |
| 48 | 45 | 4.67 | 0.75, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0.25, −0.25, −0.25, −0.25, −0.25, 0, 0, 0. |

*The numbers in the column of $F(\mathbf{x})$ are $b_\phi, b_{1234}, b_{1235}, b_{1245}, b_{1345}, b_{2345}, b_{1236}, b_{1246}, b_{1346}, b_{2346}, b_{1256}, b_{1356}, b_{2356}, b_{1456}, b_{2456}, b_{3456}, b_{12345}, b_{12346}, b_{12356}, b_{12456}, b_{13456}, b_{23456}, b_{123456}$.

# Chapter 3

# Isomorphism examination based on counting vector

## 3.1 Introduction

There exist some methods in the literature to reduce the computation in isomorphism examination. They transfer designs into some examination measures that are easy to calculate and initially separate the designs into several groups according to the values of the examination measure. The initial separation ensures that designs in different groups are non-isomorphic. It then takes much less time to examine the isomorphism within each group. Because these methods cannot guarantee that designs within the same group are isomorphic, we call them *initial screening methods* for isomorphism examination. Draper and Mitchell (1968) proposed a method for isomorphism examination of regular designs by comparing their *word length patterns* ($WLP$s), which record the numbers of letters of the words in their defining contrast subgroups. However, it can be shown by counterexamples that word length pattern is not sufficient to fully distinguish non-isomorphic designs. For example, there exit two $2^{12-3}$ fractional factorial designs which are non-isomorphic but have identical word length pattern. Draper and Mitchell (1970) proposed *letter pattern* comparison for isomorphism examination. Letter pattern records the times of each letter appearing in different lengths of words. They conjectured that non-isomorphic designs should have different letter patterns. Nevertheless, Chen and Lin (1991) showed that the conjecture is not true by giving two non-isomorphic $2^{31-16}$ regular designs with identical letter pattern. These works only focus on regular designs. Recently, *$J$-characteristics* was introduced in Deng and Tang (1999) and Tang and Deng (1999) but explicitly defied in Tang (2001). Through $J$-characteristics, they defined *general word length pattern*, denoted by $GWLP$, and *confounding frequency vector*, denoted by $CFV$. Both of them extend the concept of $WLP$ from regular to non-regular designs. The $GWLP$ and $CFV$ were treated as initial screening methods for isomorphism examination in Katsaounis and Dean (2008). Based on Hamming distance, which records the number of differences between two rows in a design, Clark and Dean (2001) developed the distance matrix, denoted by **HD**, and recommended a two-step algorithm for the isomorphism examination. Its first step proposed an initial screening method, called $Deseq1$. Ma, Fang and Lin (2001) combined Hamming distance and the measures of uniformity to develop *squared centered $L_2$ discrepancy*, denoted by $CD_2^2$. Based on the coding theory approach, Xu (2003) defined the *power moments*, denoted by $K_u$. Because isomorphic designs have the same $CD_2^2$ and $K_u$, both are also useful examination measures for initial screening of non-isomorphic designs.

The technique of projection is widely employed in the initial screening methods of the isomorphism examination. For a $k$-factor design, there are $\binom{k}{p}$ ways to project the design onto $p$ factors. These projected designs are called the $p$-dimensional projections. We can apply the examination measures mentioned above on each of the $\binom{k}{p}$ projections and calculate their frequency. The frequency is referred to as the *$p$-dimensional projection frequency* corresponding to the specific measure and the measure calculated from the whole design matrix is referred to as the overall measure. Notice that the measure could be a single quantity, a vector, or a matrix. For instance, the measure for $CD_2^2$ is a single quantity and for $CFV$ is a vector, and the frequency of $p$-dimensional projections obtained from them are called the $p$-dimensional $CD_2^2$ projection frequency and the $p$-dimensional $CFV$ projection frequency, respectively. If designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, then for any $p$-dimensional projection in $\mathcal{D}$, there must exist a corresponding isomorphic $p$-dimensional projection in $\mathcal{D}'$. The one-to-one correspondence between the projections of $\mathcal{D}$ and $\mathcal{D}'$ ensures that their $p$-dimensional projection frequency must be identical. In other words, if there exists some $p$ where $1 \leq p \leq k$ such that the $p$-dimensional projection frequencies of $\mathcal{D}$ and $\mathcal{D}'$ are different, then $\mathcal{D}$ and $\mathcal{D}'$ must be non-isomorphic. Because two designs with the same overall measure may have different $p$-dimensional projection frequency but the designs with the same projection frequency always have the same overall measure, the $p$-dimensional projection frequency presents more detailed information than the overall measure. Therefore, the initial screening method with projection is more efficient than the method without projection. Ma, Fang and Lin (2001) and Xu (2005) applied the technique of projection to $CD_2^2$ and $K_u$, respectively. We denote their projection frequencies (the collections of $p$-dimensional projection frequencies for $p = 1, \cdots, k$) by $P_{CD_2^2}$ and $P_{K_u}$. In this chapter, we also apply the technique of projection to $CFV$ and $GWLP$, and call their projection frequencies $P_{CFV}$ and $P_{GWLP}$, respectively. The application of projection for **HD** was adopted in the algorithm of $Deseq1$ in Clark and Dean (2001).

In addition to these initial screening methods, there are the *complete classification methods* in the literature. A complete classification method for non-isomorphic designs was developed by Stufken and Tang (2007). Through searching all the solutions over equations of $J$-characteristics, the complete set of non-isomorphic $OA(n, d + 2, 2, d)$'s can be obtained. Nevertheless, this method is restricted to the case of $k = d+2$ only. Based on Hamming distance and projection, Clark and Dean (2001) derived a theorem which offered a complete classification method for isomorphism examination. Based on the *indicator function*, Cheng and Ye (2004) also provided a complete classification method. Katsaounis and Dean (2008) applied the method based on the indicator function to perform the isomorphism examination and found that this method was much slower than that based on the Hamming distance. Katsaounis and Dean (2008) provides a survey and evaluation for these methods of isomorphism examination.

The purpose of this chapter is to propose an efficient initial screening method for isomorphism examination based on *counting vector*, which will be introduced latter. We find that the operations of sign switch, column and row permutations on the design matrix are related to the rearrangement of components of the counting vector. Some sufficient and necessary conditions for two counting vectors to be isomorphic are therefore developed. The conditions offer us a theoretical basis to present an isomorphism examination measure, called the *split-N matrix*. The split-N matrix is invariant to the sign switch, column and row permutations so that it can be a measure for the isomorphism examination. We also find that some existing measures for isomorphism examination can be expressed as a function of the split-N matrix. In other words, non-isomorphic designs that can be distinguished by these measures can be classified by the split-N matrix, but not the other way around. That is, split-N matrix has higher classification efficiency than these measures. The technique of projection is also applied in the examination method based on the split-N matrix, which greatly improves the examination efficiency. Some simplified methods are proposed for the cases of design with large $k$. They may have lower efficiency than the split-N matrix but can greatly reduce the computation time and the storage memory when $k$ is large.

In Section 3.2, we discuss the transformation of the counting vectors between the isomorphic designs and provide necessary and sufficient conditions for them.

In Section 3.3, we define the split-N vectors and use them to construct the split-N matrix. In Section 3.3.2, we prove that the split-N matrix is more efficient for isomorphic examination than some other existing measures. In Section 3.3.3, we apply split-N matrix to the projections of the design and also prove that the method with projection based on split-N matrix is more efficient than that based on some other measures. Several simplified methods are proposed in Section 3.3.4. Some examples and computational comparisons are given in Section 3.4.

In the remainder of this section, we will first introduce some notation and terminology and then review some examination measures mentioned above. Let $\mathcal{T} = \{1, \cdots, k\}$, where $k$ is the number of factors of a design. For any $\mathbf{m} \subseteq \mathcal{T}$, define a $1 \times 2^k$ vector

$$\mathbf{x_m} = (x_{\mathbf{m}1}, \cdots, x_{\mathbf{m}k}), \text{ where } \begin{cases} x_{\mathbf{m}j} = -1, \text{ if } j \in \mathbf{m}, \\ x_{\mathbf{m}j} = +1, \text{ otherwise.} \end{cases} \tag{3.1}$$

Each $\mathbf{x_m}$ can be regarded as a run in the full factorial design. We rank $\mathbf{x_m}$'s in Yates order for all subsets $\mathbf{m}$ of $\mathcal{T}$. The $k$-factor full factorial design can then be represented by the $2^k \times k$ matrix

$$\mathbf{X} = (\mathbf{x}_\phi^T, \mathbf{x}_1^T, \mathbf{x}_2^T, \mathbf{x}_{12}^T, \mathbf{x}_3^T, \mathbf{x}_{13}^T, \mathbf{x}_{23}^T, \mathbf{x}_{123}^T, \mathbf{x}_4^T, \mathbf{x}_{14}^T, \cdots)^T, \tag{3.2}$$

where the superscript $T$ denotes vector transpose and the suffices denote subsets. For instance, $\mathbf{x}_1$ represents $\mathbf{x}_{\{1\}}$, $\mathbf{x}_{12}$ represents $\mathbf{x}_{\{1,2\}}$ and so forth.

Let $\mathbf{h}_j$ denote the $j$th column of $\mathbf{X}$, where $j = 1, \cdots, k$. Thus, $\mathbf{X}$ can also be represented by

$$\mathbf{X} = (\mathbf{h}_1, \cdots, \mathbf{h}_k). \tag{3.3}$$

For any $\mathbf{t} \subseteq \mathcal{T}$, we use $\mathbf{h_t}$ to represent the component-wise product of the columns $\mathbf{h}_j$'s, where $j \in \mathbf{t}$. That is, $\mathbf{h_t}$ is a $2^k \times 1$ vector whose $\mathbf{m}$-th component is

$$h_{\mathbf{mt}} = \prod_{j \in \mathbf{t}} x_{\mathbf{m}j}. \tag{3.4}$$

Ranking $\mathbf{h_t}$'s in Yates order for all subsets $\mathbf{t}$ of $\mathcal{T}$, we obtain the $2^k \times 2^k$ matrix

$$\mathbf{H} = (\mathbf{h}_\phi, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_{12}, \mathbf{h}_3, \mathbf{h}_{13}, \mathbf{h}_{23}, \mathbf{h}_{123}, \mathbf{h}_4, \mathbf{h}_{14}, \cdots), \tag{3.5}$$

which is referred to as the *model matrix* of the $k$-factor full factorial design $\mathbf{X}$. Notice that $\mathbf{m}$ is the row index and $\mathbf{t}$ is the column index of $\mathbf{H}$.

Now let $\mathcal{D} = (d_{ij})$, an $n \times k$ matrix, be a $k$-factor 2-level design with $n$ runs and levels coded as $+1$ and $-1$. Each row of $\mathcal{D}$ represents a run and each column represents a factor. Let $N_\mathbf{m}$ denote the number of replicate that a run $\mathbf{x_m}$ occurs in design $\mathcal{D}$. Design $\mathcal{D}$ can then be equivalently described by the $2^k \times 1$ vector

$$\mathbf{N} = (N_\phi, N_1, N_2, N_{12}, N_3, N_{13}, N_{23}, N_{123}, N_4, N_{14}, \cdots)^T, \tag{3.6}$$

where $N_\mathbf{m}$ is ranked in Yates order. The vector $\mathbf{N}$ is the counting vector of design $\mathcal{D}$ mentioned in the previous paragraph. Because it counts the number of the appearance of the run $\mathbf{x_m}$ in $\mathcal{D}$, it can be used to represent a design.

For any $\mathbf{t} \subseteq \mathcal{T}$, let

$$J_\mathbf{t} = \sum_{i=1}^{n} \prod_{j \in \mathbf{t}} d_{ij} = \sum_{\mathbf{m} \subseteq \mathcal{T}} h_{\mathbf{mt}} N_\mathbf{m}, \tag{3.7}$$

where $h_{\mathbf{mt}}$ denotes the $\mathbf{m}$-th row and the $\mathbf{t}$-th column of $\mathbf{H}$. The $J_\mathbf{t}$ values for all subsets $\mathbf{t}$ of $\mathcal{T}$ is referred to as $J$-characteristics of design $\mathcal{D}$. The $2^k \times 1$ vector that ranks $J$-characteristics in Yates order, i.e.,

$$\mathbf{J} = (J_\phi, J_1, J_2, J_{12}, J_3, J_{13}, J_{23}, J_{123}, J_4, J_{14}, \cdots)^T, \tag{3.8}$$

is called $J$-vector. For the details of $J$-characteristics, readers is referred to Stufken and Tang (2007). Tang (2001) showed that there exist the following relationships between counting vector and $J$-vector:

$$\mathbf{J} = \mathbf{HN}, \tag{3.9}$$

$$\mathbf{N} = 2^{-k}\mathbf{HJ}. \tag{3.10}$$

Another tool that is related to the counting vector and $J$-characteristics is the indicator function. Indicator function is a polynomial representative of the equation (3.10). The value of the polynomial is the number of the replicates that a run occurs in a design and the coefficients of its polynomial terms are equivalent to the normalized $J$-characteristics defined in Tang and Deng (1999).

In the following paragraph, we briefly express some examination measures in terms of the notation mentioned above. Let $\mathcal{D}$ be a $k$-factor 2-level design with $n$ runs. For $\mathbf{t} \subseteq \mathcal{T}$, let $\|\mathbf{t}\|$ denote the number of components in $\mathbf{t}$. Deng and Tang(1999) defined the $CFV$ by

$$CFV(\mathcal{D}) = ((l_{1,1}, \cdots, l_{1,n}), (l_{2,1}, \cdots, l_{2,n}), \cdots, (l_{k,1}, \cdots, l_{k,n})), \tag{3.11}$$

where $l_{ij}$ is the number of $\mathbf{t}$'s such that $|J_{\mathbf{t}}| = (n+1-i)$ and $\|\mathbf{t}\| = j$. Note that $|J_{\mathbf{t}}|$ is always an integer satisfying $0 \le |J_{\mathbf{t}}| \le n$. Tang and Deng (1999) defined $GWLP$ of $\mathcal{D}$ by

$$GWLP(\mathcal{D}) = (\alpha_1(\mathcal{D}), \cdots, \alpha_k(\mathcal{D})), \tag{3.12}$$

where

$$\alpha_j(\mathcal{D}) = \sum_{\|\mathbf{t}\|=j} \left(\frac{J_{\mathbf{t}}}{n}\right)^2, j = 1, \cdots, k. \tag{3.13}$$

A connection between $GWLP$ and $CD_2^2$ was given in Ye (2003) as follows:

$$CD_2^2(\mathcal{D}) = \left(\frac{13}{12}\right)^k - 2\left(\frac{35}{32}\right)^k + \left(\frac{9}{8}\right)\left\{1 + \sum_{j=1}^{k} \frac{\alpha_j(\mathcal{D})}{9^j}\right\}. \tag{3.14}$$

Xu (2003) showed that $K_u$ is a linear combination of $\alpha_1(\mathcal{D}), \cdots, \alpha_u(\mathcal{D})$. For positive integers $u$,

$$K_u(\mathcal{D}) = c_u\alpha_u(\mathcal{D}) + c_{u-1}\alpha_{u-1}(\mathcal{D}) + \cdots + c_1\alpha_1(\mathcal{D}) + \alpha_0(\mathcal{D}) - C_0, \tag{3.15}$$

where $c_i = c_i(u; n, k, s) = [n/(n-1)] \sum_{m=0}^{u} (-1)^{m+i} \binom{u}{m} k^{u-m} [\sum_{j=0}^{m} j! S(m, j) s^{-j} (s-1)^{j-i} \binom{k-i}{j-i}]$, $C_0 = k^u/(n-1)$ and $S(m, j)$ are Stirling numbers of the second kind. Based on Hamming distance, Clark and Dean (2001) defined distance matrix **HD** as follows. Let $\zeta[\mathcal{D}]^j_{i_1,i_2}$ be 1 if, in the $j$th column of $\mathcal{D}$, the sign in the $i_1$th and $i_2$th rows are different, and 0 if they are the same. Then, $\sum_{j=1}^{k} \zeta[\mathcal{D}]^j_{i_1,i_2}$ counts the number of columns in which the signs of the $i_1$th and $i_2$th rows fail to coincide. The distance matrix **HD** is defined as a matrix $(\pi_{i_1,i_2})$ where the $(i_1, i_2)$th component equals

$$\pi_{i_1,i_2} = \begin{cases} \sum_{j=1}^{k} \zeta[\mathcal{D}]^j_{i_1,i_2}, & \text{for } i_1 \neq i_2, \\ 0, & \text{for } i_1 = i_2. \end{cases} \tag{3.16}$$

Clark and Dean (2001) pointed out that **HD** is invariant to the sign switch and column permutation of $\mathcal{D}$, but not to row permutation.

## 3.2 Isomorphism of counting vectors

In this section, we discuss how the counting vector is affected by the operation of the sign switch, column permutation and row permutation. We obtain some necessary and sufficient conditions for two counting vectors to be isomorphic. Let $\mathcal{D} = (d_{ij})$ be a $k$-factor $n$-run design matrix with counting vector $\mathbf{N}(\mathcal{D})$ where $i = 1, \cdots, n$ and $j = 1, \cdots, k$. Let $\mathbf{H}$ be the model matrix of the $k$-factor full factorial design $\mathbf{X}$ where $\mathbf{X}$ and $\mathbf{H}$ are given in the equations (3.3) and (3.5), respectively. For two matrices $\mathbf{U}_1$ and $\mathbf{U}_2$ with the same number of rows, let $[\mathbf{U}_1|\mathbf{U}_2]$ denote the $l \times (m_1 + m_2)$ matrix formed by arranging the $l \times m_1$ matrix $\mathbf{U}_1$ in the first $m_1$ columns and the $l \times m_2$ matrix $\mathbf{U}_2$ in the last $m_2$ columns. Now let us combine $\mathbf{X}$ and $\mathbf{N}(\mathcal{D})$ to form the $2^k \times (k+1)$ matrix

$$[\mathbf{X}|\mathbf{N}(\mathcal{D})]. \tag{3.17}$$

In each row of the matrix $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$, the first $k$ components denote a run and the last component denotes the number of the replicates of the run in $\mathcal{D}$. Because $\mathbf{X}$

is the $k$-factor full factorial design, which contains all possible level combinations of a $k$-factor experiment, the $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$ can be used to fully characterize the design $\mathcal{D}$. In the following paragraphs, we will discuss how the sign switch, column and row permutation that are performed on $\mathcal{D}$ affect $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$ and the counting vector.

Suppose that design $\mathcal{D}'$ is obtained by permutating the rows of the design $\mathcal{D}$. Because row permutation only changes the order of the runs in $\mathcal{D}$ and $\mathcal{D}'$, no matter how the rows are permutated, the numbers of the replicates of each runs in $\mathcal{D}$ and in $\mathcal{D}'$ are identical, i.e. $\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}')$. In other words, the row permutation has no influence to the counting vector. Therefore, row permutation can be ignored when we discuss the isomorphism examination from the perspective of counting vectors.

Let us now focus on the sign switch operation on $\mathcal{D}$. Let $\mathcal{T} = \{1, \cdots, k\}$. Recall that $\mathbf{X} = (\mathbf{h}_1, \cdots, \mathbf{h}_k)$ and $\mathbf{H} = (\mathbf{h}_\phi, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_{12}, \mathbf{h}_3, \mathbf{h}_{13} \cdots, \mathbf{h}_{1\cdots k})$. Suppose that $\mathcal{D}^*$ is obtained by switching the sign of factors $\kappa_1, \cdots, \kappa_g$ of $\mathcal{D}$. Denote $\kappa = \{\kappa_1, \cdots, \kappa_g\}$ and $\delta_{\mathbf{t}} = \|\kappa \cap \mathbf{t}\|$ for $\mathbf{t} \subseteq \mathcal{T}$. Let us characterize design $\mathcal{D}$ as $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$. After performing this sign switch operation, the design becomes

$$[\mathbf{X}^*|\mathbf{N}(\mathcal{D})], \tag{3.18}$$

where

$$\mathbf{X}^* = ((-1)^{\delta_1}\mathbf{h}_1, \cdots, (-1)^{\delta_k}\mathbf{h}_k). \tag{3.19}$$

Because in each row of the matrix in (3.18), the last component records the number of the replicates of the run after sign switch, (3.18) indicates the design $\mathcal{D}^*$. If we rearrange the rows of $[\mathbf{X}^*|\mathbf{N}(\mathcal{D})]$ so that the $\mathbf{X}^*$ part is transformed into $\mathbf{X}$, we will obtain $[\mathbf{X}|\mathbf{N}(\mathcal{D}^*)]$. Suppose that we use matrix operation to represent the row rearrangement transforming $\mathbf{X}^*$ into $\mathbf{X}$. Then, there exists a $2^k \times 2^k$

matrix $\mathbf{R}^*$ such that

$$\mathbf{R}^*[\mathbf{X}^*|\mathbf{N}(\mathcal{D})] = [\mathbf{X}|\mathbf{N}(\mathcal{D}^*)], \tag{3.20}$$

where

$$\mathbf{R}^*\mathbf{X}^* = \mathbf{X}, \tag{3.21}$$

and

$$\mathbf{R}^*\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}^*). \tag{3.22}$$

If we can solve the equation (3.21) to obtain $\mathbf{R}^*$ , then $\mathbf{N}(\mathcal{D}^*)$ can be obtained from $\mathbf{N}(\mathcal{D})$ by equation (3.22). However, to obtain $\mathbf{R}^*$ directly from equation (3.21) is difficult because $\mathbf{X}^*$ is not a square matrix so that its inverse matrix does not exist. Therefore, we turn our attention from the design matrices $\mathbf{X}$ and $\mathbf{X}^*$ to their model matrices $\mathbf{H}$ and $\mathbf{H}^*$, respectively. Let $\lambda$ be a matrix operator which expands a design matrix to its model matrix. Take $\lambda$ on both sides of equation (3.21) to obtain

$$\lambda(\mathbf{R}^*\mathbf{X}^*) = \lambda(\mathbf{X}). \tag{3.23}$$

Because $\lambda(\mathbf{R}^*\mathbf{X}^*) = \mathbf{R}^*\lambda(\mathbf{X}^*)$, equation (3.23) can be written as

$$\mathbf{R}^*\mathbf{H}^* = \mathbf{H}, \tag{3.24}$$

where

$$\mathbf{H}^* = ((-1)^{\delta_\phi}\mathbf{h}_\phi, (-1)^{\delta_1}\mathbf{h}_1, (-1)^{\delta_2}\mathbf{h}_2, (-1)^{\delta_{12}}\mathbf{h}_{12}, \cdots, (-1)^{\delta_{1\cdots k}}\mathbf{h}_{1\cdots k}). \tag{3.25}$$

The multiplicator $(-1)^{\delta_\mathbf{t}}$ in equation (3.25) controls the sign switch of the column $\mathbf{h}_\mathbf{t}$. Collect all $(-1)^{\delta_\mathbf{t}}$'s and rank them in Yates order of $\mathbf{t}$ to form the vector

$$((-1)^{\delta_\phi}, (-1)^{\delta_1}, (-1)^{\delta_2}, (-1)^{\delta_{12}}, \cdots, (-1)^{\delta_{1\cdots k}})^T. \tag{3.26}$$

Notice that the vector in (3.26) is the column $\mathbf{h}_\mathbf{t}$ in $\mathbf{H}$ with $\mathbf{t} = \kappa$. Let $\mathbf{S}^\kappa$ be the $2^k \times 2^k$ diagonal matrix with diagonal being $\mathbf{h}_\kappa$. Because $\mathbf{H}^* = \mathbf{H}\mathbf{S}^\kappa$ from equation (3.25), equation (3.24) can be written as

$$\mathbf{R}^*\mathbf{H}\mathbf{S}^\kappa = \mathbf{H}. \tag{3.27}$$

Because $\mathbf{H}^{-1} = 2^{-k}\mathbf{H}$ and $(\mathbf{S}^\kappa)^{-1} = \mathbf{S}^\kappa$ ($\mathbf{S}^\kappa$ is an orthonormal and symmetric matrix), we obtain from equation (3.27) that

$$\mathbf{R}^* = \mathbf{H}(\mathbf{H}\mathbf{S}^\kappa)^{-1} = \mathbf{H}(\mathbf{S}^\kappa)^{-1}\mathbf{H}^{-1} = 2^{-k}\mathbf{H}\mathbf{S}^\kappa\mathbf{H}. \tag{3.28}$$

We call $\mathbf{S}^\kappa$ the *sign switch matrix*. For a $k$-factor design, there are $2^k$ different ways to perform sign switch. Define $\mathcal{S}$ as the collection of $\mathbf{S}^\kappa$'s for all subsets $\kappa$ of $\mathcal{T}$.

Let us now consider the column permutation operation on $\mathcal{D}$. Let $(j_1, j_2, \cdots, j_k)$ be a permutation of $(1, 2, \cdots, k)$. Suppose that $\mathcal{D}^{**}$ is obtained from $\mathcal{D}$ by the column permutation $(j_1, j_2, \cdots, j_k)$. After performing this column permutation operation, the design becomes

$$[\mathbf{X}^{**}|\mathbf{N}(\mathcal{D})], \tag{3.29}$$

where

$$\mathbf{X}^{**} = (\mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \cdots, \mathbf{h}_{j_k}). \tag{3.30}$$

Because in each row of the matrix in (3.29), the last component records the number of the replicates of the run after column permutation, (3.29) indicates the design $\mathcal{D}^{**}$. If we rearrange the rows of $[\mathbf{X}^{**}|\mathbf{N}(\mathcal{D})]$ so that the $\mathbf{X}^{**}$ part is transformed into $\mathbf{X}$, we will obtain $[\mathbf{X}|\mathbf{N}(\mathcal{D}^{**})]$. Suppose that we use matrix operation to represent the row rearrangement transforming $\mathbf{X}^{**}$ into $\mathbf{X}$. Then, there exists a $2^k \times 2^k$ matrix $\mathbf{R}^{**}$ such that

$$\mathbf{R}^{**}[\mathbf{X}^{**}|\mathbf{N}(\mathcal{D})] = [\mathbf{X}|\mathbf{N}(\mathcal{D}^{**})], \tag{3.31}$$

where

$$\mathbf{R}^{**}\mathbf{X}^{**} = \mathbf{X}, \tag{3.32}$$

and

$$\mathbf{R}^{**}\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}^{**}). \tag{3.33}$$

Similarly, to obtain $\mathbf{R}^{**}$, let us take $\lambda$ on both sides of equation (3.32), i.e.,

$$\lambda(\mathbf{R}^{**}\mathbf{X}^{**}) = \lambda(\mathbf{X}). \tag{3.34}$$

Because $\lambda(\mathbf{R}^{**}\mathbf{X}^{**}) = \mathbf{R}^{**}\lambda(\mathbf{X}^{**})$, equation (3.34) can be written as

$$\mathbf{R}^{**}\mathbf{H}^{**} = \mathbf{H}, \tag{3.35}$$

where

$$\mathbf{H}^{**} = (\mathbf{h}_\phi, \mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \mathbf{h}_{j_{12}}, \mathbf{h}_{j_3} \cdots, \mathbf{h}_{j_1 \cdots j_k}). \tag{3.36}$$

Let $\mathbf{I}_{2^k}$ be the $2^k \times 2^k$ identity matrix. Let us denote the columns of $\mathbf{I}_{2^k}$ by

$$\mathbf{I}_{2^k} = (\mathbf{e}_\phi, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_{12}, \mathbf{e}_3, \mathbf{e}_{13}, \mathbf{e}_{23}, \mathbf{e}_{123}, \mathbf{e}_4, \ldots, \mathbf{e}_{1\cdots k}),$$

that is, $\mathbf{e}_\phi = (1,0,0,0,0,\cdots,0)^T$, $\mathbf{e}_1 = (0,1,0,0,0,\cdots,0)^T$, $\mathbf{e}_2 = (0,0,1,0,0,\cdots,0)^T$, $\mathbf{e}_{12} = (0,0,0,1,0\cdots,0)^T$, and so forth. Let

$$\mathbf{C}^{j_1 j_2 \cdots j_k} = (\mathbf{e}_\phi, \mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \mathbf{e}_{j_1 j_2}, \mathbf{e}_{j_3}, \mathbf{e}_{j_1 j_3}, \mathbf{e}_{j_2 j_3}, \mathbf{e}_{j_1 j_2 j_3}, \mathbf{e}_{j_4}, \cdots, \mathbf{e}_{j_1 \cdots j_k}). \tag{3.37}$$

Because $\mathbf{H}^{**} = \mathbf{H}\mathbf{C}^{j_1 \cdots j_k}$ from equation (3.36), equation (3.35) can be written as

$$\mathbf{R}^{**}\mathbf{H}\mathbf{C}^{j_1 \cdots j_k} = \mathbf{H}. \tag{3.38}$$

Because $\mathbf{H}^{-1} = 2^{-k}\mathbf{H}$ and $(\mathbf{C}^{j_1 \cdots j_k})^{-1} = (\mathbf{C}^{j_1 \cdots j_k})^T$ ($\mathbf{C}^{j_1 \cdots j_k}$ is an orthonormal matrix), we obtain from equation (3.38) that

$$\mathbf{R}^{**} = \mathbf{H}(\mathbf{H}\mathbf{C}^{j_1 \cdots j_k})^{-1} = \mathbf{H}(\mathbf{C}^{j_1 \cdots j_k})^{-1}\mathbf{H}^{-1} = 2^{-k}\mathbf{H}(\mathbf{C}^{j_1 \cdots j_k})^T\mathbf{H}. \tag{3.39}$$

We call $\mathbf{C}^{j_1 \cdots j_k}$ the *column permutation matrix*. For a $k$-factor design, there are $k!$ different ways to perform column permutation. Define $\mathcal{C}$ as the collection of $\mathbf{C}^{j_1 j_2 \cdots j_k}$'s for all possible permutations $(j_1, j_2, \cdots, j_k)$ of $(1, 2, \cdots, k)$.

Let us combine the sign switch and the column permutation operations together. Suppose that $\mathcal{D}'$ is obtained from $\mathcal{D}$ by the sign switch of factors

$\kappa_1, \cdots, \kappa_g$ and the column permutation $(j_1, \cdots, j_k)$ of $(1, \cdots, k)$. After performing these sign switch and column permutation operations, the design becomes

$$[\mathbf{X}'|\mathbf{N}(\mathcal{D})], \tag{3.40}$$

where

$$\mathbf{X}' = ((-1)^{\delta_{j_1}} \mathbf{h}_{j_1}, \cdots, (-1)^{\delta_{j_k}} \mathbf{h}_{j_k}). \tag{3.41}$$

Similar to the previous discussion, the matrix in equation (3.40) indicates the design $\mathcal{D}'$ and there exists a $2^k \times 2^k$ matrix $\mathbf{R}$ such that

$$\mathbf{R}[\mathbf{X}'|\mathbf{N}(\mathcal{D})] = [\mathbf{X}|\mathbf{N}(\mathcal{D}')], \tag{3.42}$$

where

$$\mathbf{R}\mathbf{X}' = \mathbf{X}, \tag{3.43}$$

and

$$\mathbf{R}\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}'). \tag{3.44}$$

To obtain $\mathbf{R}$, let us take $\lambda$ on both sides of equation (3.43), i.e.,

$$\lambda(\mathbf{R}\mathbf{X}') = \lambda(\mathbf{X}). \tag{3.45}$$

Because $\lambda(\mathbf{R}\mathbf{X}') = \mathbf{R}\lambda(\mathbf{X}')$, equation (3.45) can be written as

$$\mathbf{R}\mathbf{H}' = \mathbf{H}, \tag{3.46}$$

where

$$\mathbf{H}' = ((-1)^{\delta_\phi} \mathbf{h}_\phi, (-1)^{\delta_{j_1}} \mathbf{h}_{j_1}, (-1)^{\delta_{j_2}} \mathbf{h}_{j_2}, (-1)^{\delta_{j_1 j_2}} \mathbf{h}_{j_1 j_2}, \cdots, (-1)^{\delta_{j_1 \cdots j_k}} \mathbf{h}_{j_1 \cdots j_k}). \tag{3.47}$$

Let $\mathbf{S} = \mathbf{S}^\kappa$ and $\mathbf{C} = \mathbf{C}^{j_1 \cdots j_k}$ where $\mathbf{S} \in \mathcal{S}$ and $\mathbf{C} \in \mathcal{C}$. Because $\mathbf{H}' = \mathbf{HSC}$ from equation (3.47), equation (3.46) can be written as

$$\mathbf{RHSC} = \mathbf{H}. \tag{3.48}$$

Because $\mathbf{H}^{-1} = 2^{-k}\mathbf{H}$, $\mathbf{S}^{-1} = \mathbf{S}$ and $\mathbf{C}^{-1} = \mathbf{C}^T$, we can obtain from equation (3.48) that

$$\mathbf{R} = \mathbf{H}(\mathbf{HSC})^{-1} = \mathbf{HC}^{-1}\mathbf{S}^{-1}\mathbf{H}^{-1} = 2^{-k}\mathbf{HC}^T\mathbf{SH}. \tag{3.49}$$

Notice that design $\mathcal{D}^*$ is a special case of $\mathcal{D}'$ with $(j_1, j_2, \cdots, j_k) = (1, 2, \cdots, k)$ and $\mathcal{D}^{**}$ is a special case of $\mathcal{D}'$ with $\kappa = \phi$. The above discussion is summarized in the following theorem.

**Theorem 4.** *Suppose that $\mathcal{D}$ and $\mathcal{D}'$ are two $k$-factor designs with counting vectors $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$, respectively. Two designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic if and only if there exists a matrix $\mathbf{R} = 2^{-k}\mathbf{HC}^T\mathbf{SH}$, where $\mathbf{S} \in \mathcal{S}$ and $\mathbf{C} \in \mathcal{C}$, such that*

$$\mathbf{RN}(\mathcal{D}) = \mathbf{N}(\mathcal{D}').$$

**Example 2.** For the demonstration purpose, let $\mathcal{A}_1$ be a 3-factor design with counting vector $\mathbf{N}(\mathcal{A}_1) = (1, 2, 3, 4, 5, 6, 7, 8)^T$. Assigning different numbers of replicates to the runs is intended to make it clear the rearrangement of components in the counting vector in later calculations. The model matrix of the 3-factor full factorial design is

$$\mathbf{H} = (\mathbf{h}_\phi, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_{12}, \mathbf{h}_3, \mathbf{h}_{13}, \mathbf{h}_{23}, \mathbf{h}_{123}) = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 & 1 & -1 \end{pmatrix}.$$

Suppose that $\mathcal{A}_2$ is obtained from $\mathcal{A}_1$ by exchanging factors 1 and 3 and switching the sign of factors 1 and 2. Because we switch the sign of factors 1 and 2, set

$\mathbf{S} = \mathbf{S}^{12}$. Exchanging factors 1 and 3 is to permute the factors with $(j_1, j_2, j_3) = (3, 2, 1)$, so set $\mathbf{C}^{j_1 j_2 j_3} = \mathbf{C}^{321}$ obtained from $\mathbf{I}_8 = (\mathbf{e}_\phi, \mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_{12}, \mathbf{e}_3, \mathbf{e}_{13}, \mathbf{e}_{23}, \mathbf{e}_{123})$ by exchanging $\mathbf{e}_1$ and $\mathbf{e}_3$ and exchanging $\mathbf{e}_{12}$ and $\mathbf{e}_{23}$. The $\mathbf{S}^{12}$ and $\mathbf{C}^{321}$ are shown below.

$$\mathbf{S}^{12} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \text{ and } \mathbf{C}^{321} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

From Theorem 4, we obtain

$$\mathbf{R} = 2^{-k}\mathbf{H}(\mathbf{C}^{321})^T\mathbf{S}^{12}\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

Therefore, $\mathbf{N}(\mathcal{D}') = \mathbf{R}\mathbf{N}(\mathcal{D}) = (4, 8, 2, 6, 3, 7, 1, 5)^T$.

Theorem 4 also infers that there exists a relationship between the $J$-vectors of two isomorphic designs. Corollary 1 below provides the transformation matrix for them.

**Corollary 1.** *Suppose that $\mathcal{D}$ and $\mathcal{D}'$ are two $k$-factor designs with $J$-vectors $\mathbf{J}(\mathcal{D})$ and $\mathbf{J}(\mathcal{D}')$, respectively. The two designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic if and only if there exists a matrix $\mathbf{Q} = \mathbf{C}^T\mathbf{S}$ such that $\mathbf{Q}\mathbf{J}(\mathcal{D}) = \mathbf{J}(\mathcal{D}')$, where $\mathbf{C} \in \mathcal{C}$ and $\mathbf{S} \in \mathcal{S}$.*

*Proof.* Let $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$ be the counting vectors of $\mathcal{D}$ and $\mathcal{D}'$, respectively. By Theorem 4, there exists an $\mathbf{R}$ such that

$$\mathbf{R}\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}'). \tag{3.50}$$

Multiplying $\mathbf{H}$ on both sides of equation (3.50), we obtain

$$\mathbf{HRN}(\mathcal{D}) = \mathbf{HN}(\mathcal{D}'). \tag{3.51}$$

Because $\mathbf{HH} = 2^k \mathbf{I}_{2^k}$, equation (3.51) can be written as

$$\mathbf{HR}(2^{-k}\mathbf{HH})\mathbf{N}(\mathcal{D}) = \mathbf{HN}(\mathcal{D}'). \tag{3.52}$$

By substituting $2^{-k}\mathbf{HC}^T\mathbf{SH}$ for $\mathbf{R}$ and $\mathbf{J}$ for $\mathbf{HN}$ according to equation (3.9), equation (3.52) becomes

$$2^{-2k}\mathbf{HHC}^T\mathbf{SHHJ}(\mathcal{D}) = \mathbf{J}(\mathcal{D}'). \tag{3.53}$$

Equation (3.53) can be reduced to

$$\mathbf{C}^T\mathbf{SJ}(\mathcal{D}) = \mathbf{J}(\mathcal{D}'). \tag{3.54}$$

The result follows. □

In Theorem 4, we provide a method that utilizes the the operations of matrices to find a transformation matrix $\mathbf{R}$ for the counting vectors of two isomorphic designs. However, readers may not want to work on such complex operations of matrices. The following theorem provides an alternative that uses the suffix's operation to find the transformation between counting vectors.

**Theorem 5.** *Let $\mathcal{D}$ and $\mathcal{D}'$ be two $k$-factor designs with counting vector $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$, respectively. Let $\mathcal{T} = \{1, \cdots, k\}$. Let $\kappa = \{\kappa_1, \cdots, \kappa_g\}$, which is a subset of $\mathcal{T}$. Let $(j_1, \cdots, j_k)$ be a permutation of $(1, \cdots, k)$. Define $\mathbf{m}' = \bigcup_{i\in\mathbf{m}}\{j_i\}$ for all subsets $\mathbf{m}$ of $\mathcal{T}$ and $\mathbf{a} \vee \mathbf{b} = \mathbf{a} \cup \mathbf{b} - \mathbf{a} \cap \mathbf{b}$ where $\mathbf{a}$ and $\mathbf{b}$ are the subsets of $\mathcal{T}$. Design $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic if and only if there exist a permutation $(j_1, \cdots, j_k)$ of $(1, \cdots, k)$ and a set $\kappa$ such that the components of $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$ have the following relationship:*

$$N_{\mathbf{m}}(\mathcal{D}') = N_{\mathbf{m}'\vee\kappa}(\mathcal{D}) \tag{3.55}$$

*for all subsets $\mathbf{m}$ of $\mathcal{T}$.*

*Proof.* Suppose that the run $\mathbf{x_m}$ in $\mathcal{D}'$ is obtained from the run $\mathbf{x_y}$ in $\mathcal{D}$ through the sign switch of factors in $\kappa$ and the column permutation $(j_1, \cdots, j_k)$. Then

$$N_{\mathbf{y}}(\mathcal{D}) = N_{\mathbf{m}}(\mathcal{D}'). \tag{3.56}$$

To find $\mathbf{y}$, let us separate the operations that transform $\mathcal{D}$ to $\mathcal{D}'$ into two parts. First, suppose that after switching sign of factors in $\kappa$ of $\mathcal{D}$, we obtain another design, say $\mathcal{D}^*$. The run $\mathbf{x_y}$ in $\mathcal{D}$ changes into the run $\mathbf{x}_{\mathbf{y}\vee\kappa}$ in $\mathcal{D}^*$, i.e.,

$$N_{\mathbf{y}}(\mathcal{D}) = N_{\mathbf{y}\vee\kappa}(\mathcal{D}^*). \tag{3.57}$$

Now, suppose that after permutating the columns of $\mathcal{D}^*$ by $(j_1, \cdots, j_k)$, we obtain the design $\mathcal{D}'$. The run $\mathbf{x}_{\mathbf{y}\vee\kappa}$ in $\mathcal{D}^*$ changes into the run $\mathbf{x_m}$ in$\mathcal{D}'$, i.e.,

$$N_{\mathbf{y}\vee\kappa}(\mathcal{D}^*) = N_{\mathbf{m}}(\mathcal{D}'). \tag{3.58}$$

Because $\mathcal{D}'$ is obtained from permutating the columns of $\mathcal{D}^*$ by $(j_1, \cdots, j_k)$, it is obvious that

$$\mathbf{x}_{\mathbf{y}\vee\kappa} = \mathbf{x}_{\mathbf{m}'}, \tag{3.59}$$

and therefore

$$\mathbf{y} \vee \kappa = \mathbf{m}'. \tag{3.60}$$

From equation (3.60), it is clear that

$$(\mathbf{y} \vee \kappa) \vee \kappa = \mathbf{m}' \vee \kappa. \tag{3.61}$$

By expanding the left-hand side of equation (3.61),

$$\begin{aligned}
&(\mathbf{y} \vee \kappa) \vee \kappa \\
&= (\mathbf{y} \vee \kappa) \cup \kappa - (\mathbf{y} \vee \kappa) \cap \kappa \\
&= (\mathbf{y} \cup \kappa - \mathbf{y} \cap \kappa) \cup \kappa - (\mathbf{y} \cup \kappa - \mathbf{y} \cap \kappa) \cap \kappa \\
&= \mathbf{y} \cup \kappa - (\kappa - \mathbf{y} \cap \kappa) \\
&= \mathbf{y},
\end{aligned} \tag{3.62}$$

the result follows. $\square$

**Example 3.** We use the designs $\mathcal{A}_1$ and $\mathcal{A}_2$ in Example 2 to illustrate the Theorem 5. The counting vector of $\mathcal{A}_1$ is $\mathbf{N}(\mathcal{A}_1) = (1, 2, 3, 4, 5, 6, 7, 8)^T = (N_\phi(\mathcal{A}_1), N_1(\mathcal{A}_1), N_2(\mathcal{A}_1), N_{12}(\mathcal{A}_1), N_3(\mathcal{A}_1), N_{13}(\mathcal{A}_1), N_{23}(\mathcal{A}_1), N_{123}(\mathcal{A}_1))^T$. Design $\mathcal{A}_2$ is obtained from $\mathcal{A}_1$ by switching the sign of factors 1 and 2 and exchanging factors 1 and 3, i.e., $\kappa = \{\kappa_1, \kappa_2\} = \{1, 2\}$ and $(j_1, j_2, j_3) = (3, 2, 1)$. Take the run $\mathbf{x}_{13} = (-1, 1, -1)$ in $\mathcal{A}_1$ as an example. The number of replicate of $\mathbf{x}_{13}$ in $\mathcal{A}_1$ is $N_{13}(\mathcal{A}_1) = 6$. First, the sign switch of factors 1 and 2 is to switch the sign of the first two components in $(-1, 1, -1)$, i.e., $(1, -1, -1)$, and then the permutation of factors 1 and 3 is to exchange the first and third elements in $(1, -1, -1)$. Finally we obtain $(-1, -1, 1)$, which is the run $\mathbf{x}_{12}$ in $\mathcal{A}_2$, i.e., $N_{12}(\mathcal{A}_2) = 6$. Working on the sign switch and the column permutation above is equivalent to implementing the following suffix's operations,

$$N_{12}(\mathcal{A}_2) = N_{\mathbf{m}}(\mathcal{A}_2) = N_{\mathbf{m}'\vee\kappa}(\mathcal{A}_1) = N_{j_1j_2\vee\kappa}(\mathcal{A}_1) = N_{32\vee12}(\mathcal{A}_1) = N_{13}(\mathcal{A}_1).$$

The relationship of other runs between two isomorphic designs can be easily found by the same rule. The $\mathbf{N}(\mathcal{A}_2)$ can be obtained by

$$\begin{aligned}
\mathbf{N}(\mathcal{A}_2) &= (N_\phi(\mathcal{A}_2), N_1(\mathcal{A}_2), N_2(\mathcal{A}_2), N_{12}(\mathcal{A}_2), N_3(\mathcal{A}_2), N_{13}(\mathcal{A}_2), N_{23}(\mathcal{A}_2), N_{123}(\mathcal{A}_2))^T \\
&= (N_{\phi\vee\kappa}(\mathcal{A}_1), N_{j_1\vee\kappa}(\mathcal{A}_1), N_{j_2\vee\kappa}(\mathcal{A}_1), N_{j_1j_2\vee\kappa}(\mathcal{A}_1), N_{j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_3\vee\kappa}(\mathcal{A}_1), \\
&\quad N_{j_2j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_2j_3\vee\kappa}(\mathcal{A}_1))^T \\
&= (N_{\phi\vee12}(\mathcal{A}_1), N_{3\vee12}(\mathcal{A}_1), N_{2\vee12}(\mathcal{A}_1), N_{32\vee12}(\mathcal{A}_1), N_{1\vee12}(\mathcal{A}_1), N_{31\vee12}(\mathcal{A}_1), \\
&\quad N_{21\vee12}(\mathcal{A}_1), N_{321\vee12}(\mathcal{A}_1))^T \\
&= (N_{12}(\mathcal{A}_1), N_{123}(\mathcal{A}_1), N_1(\mathcal{A}_1), N_{13}(\mathcal{A}_1), N_2(\mathcal{A}_1), N_{23}(\mathcal{A}_1), N_\phi(\mathcal{A}_1), N_3(\mathcal{A}_1))^T \\
&= (4, 8, 2, 6, 3, 7, 1, 5)^T.
\end{aligned}$$

## 3.3 Examination measure based on counting vectors

In section 3.2, we present some relationship between the counting vectors of isomorphic designs. However, it still a time-consuming task to examine isomorphism through identifying the sign switch and column permutation operations in the relationship. Based on the counting vector, we propose an examination measure, called the *split-N matrix*, which provides a fast and powerful initial screening for non-isomorphism designs. In the following sections, we prove that isomorphic designs have the same split-N matrix and isomorphism examination by split-N matrix is more efficient than most initial screening methods mentioned in Section 3.1.

### 3.3.1 Split-N vectors

To introduce split-N matrix, we define the positive and negative split-N vector first.

**Definition 1.** *For a design with counting vector* $\mathbf{N} = (N_\phi, N_1, N_2, N_{12}, N_3, \cdots, N_{1\cdots k})^T$*, let* $\xi$ *be a set operator which ranks the components in a set of non-negative integers from large to small in a vector. Let* $\mathcal{T} = \{1, \cdots, k\}$*. For any non-empty subset* $\mathbf{t}$ *of* $\mathcal{T}$*, let* ${\mathbf{n_t}}^+$ *denote the collection of* $N_\mathbf{m}$*'s where* $\mathbf{m} \subseteq \mathcal{T}$ *such that* $\|\mathbf{m} \cap \mathbf{t}\|$ *is even. On the other hand, for any non-empty subset* $\mathbf{t}$ *of* $\mathcal{T}$*, let* ${\mathbf{n_t}}^-$ *denote the collection of* $N_\mathbf{m}$*'s where* $\mathbf{m} \subseteq \mathcal{T}$ *such that* $\|\mathbf{m} \cap \mathbf{t}\|$ *is odd. Then for any non-empty subset* $\mathbf{t}$ *of* $\mathcal{T}$*, define the* positive split-N vector *of* $\mathbf{t}$ *by* $\mathbf{N}^+_\mathbf{t} = \xi {\mathbf{n_t}}^+$ *and the* negative split-N vector *of* $\mathbf{t}$ *by* $\mathbf{N}^-_\mathbf{t} = \xi {\mathbf{n_t}}^-$*. The* $\mathbf{N}^+_\mathbf{t}$ *and* $\mathbf{N}^-_\mathbf{t}$ *are referred to as* split-N vectors *of* $\mathbf{t}$.

Table 3.1: Split-N vectors for the design with $k = 3$

| | | | | | |
|---|---|---|---|---|---|
| $\mathbf{N}_1^+$ | = | $\xi(N_\phi, N_2, N_3, N_{23})^T$ | $\mathbf{N}_1^-$ | = | $\xi(N_1, N_{12}, N_{13}, N_{123})^T$ |
| $\mathbf{N}_2^+$ | = | $\xi(N_\phi, N_1, N_3, N_{13})^T$ | $\mathbf{N}_2^-$ | = | $\xi(N_2, N_{12}, N_{23}, N_{123})^T$ |
| $\mathbf{N}_{12}^+$ | = | $\xi(N_\phi, N_{12}, N_3, N_{123})^T$ | $\mathbf{N}_{12}^-$ | = | $\xi(N_1, N_2, N_{13}, N_{23})^T$ |
| $\mathbf{N}_3^+$ | = | $\xi(N_\phi, N_1, N_2, N_{12})^T$ | $\mathbf{N}_3^-$ | = | $\xi(N_3, N_{13}, N_{23}, N_{123})^T$ |
| $\mathbf{N}_{13}^+$ | = | $\xi(N_\phi, N_2, N_{13}, N_{123})^T$ | $\mathbf{N}_{13}^-$ | = | $\xi(N_1, N_{12}, N_3, N_{23})^T$ |
| $\mathbf{N}_{23}^+$ | = | $\xi(N_\phi, N_1, N_{23}, N_{123})^T$ | $\mathbf{N}_{23}^-$ | = | $\xi(N_2, N_{12}, N_3, N_{13})^T$ |
| $\mathbf{N}_{123}^+$ | = | $\xi(N_\phi, N_{12}, N_{13}, N_{23})^T$ | $\mathbf{N}_{123}^-$ | = | $\xi(N_1, N_2, N_3, N_{123})^T$ |

Table 3.2: Split-N vectors for $\mathcal{A}_1$

| | | | | | |
|---|---|---|---|---|---|
| $\mathbf{N}_1^+(\mathcal{A}_1)$ | = | $(7,5,3,1)^T$ | $\mathbf{N}_1^-(\mathcal{A}_1)$ | = | $(8,6,4,2)^T$ |
| $\mathbf{N}_2^+(\mathcal{A}_1)$ | = | $(6,5,2,1)^T$ | $\mathbf{N}_2^-(\mathcal{A}_1)$ | = | $(8,7,4,3)^T$ |
| $\mathbf{N}_{12}^+(\mathcal{A}_1)$ | = | $(8,5,4,1)^T$ | $\mathbf{N}_{12}^-(\mathcal{A}_1)$ | = | $(7,6,3,2)^T$ |
| $\mathbf{N}_3^+(\mathcal{A}_1)$ | = | $(4,3,2,1)^T$ | $\mathbf{N}_3^-(\mathcal{A}_1)$ | = | $(8,7,6,5)^T$ |
| $\mathbf{N}_{13}^+(\mathcal{A}_1)$ | = | $(8,6,3,1)^T$ | $\mathbf{N}_{13}^-(\mathcal{A}_1)$ | = | $(7,5,4,2)^T$ |
| $\mathbf{N}_{23}^+(\mathcal{A}_1)$ | = | $(8,7,2,1)^T$ | $\mathbf{N}_{23}^-(\mathcal{A}_1)$ | = | $(6,5,4,3)^T$ |
| $\mathbf{N}_{123}^+(\mathcal{A}_1)$ | = | $(7,6,4,1)^T$ | $\mathbf{N}_{123}^-(\mathcal{A}_1)$ | = | $(8,5,3,2)^T$ |

Take a design with $k = 3$ as an example. Its counting vector is $\mathbf{N} = (N_\phi, N_1, N_2, N_{12}, N_3, N_{13}, N_{23}, N_{123})^T$. Let $\mathcal{T} = \{1, 2, 3\}$. For $\mathbf{t} = 12$, the subsets $\mathbf{m}$ of $\mathcal{T}$ such that $\|\mathbf{m} \cap \mathbf{t}\|$ is even are $\phi$, $\{3\}$, $\{1, 2\}$, and $\{1, 2, 3\}$. Therefore, the positive split-N vector of $\mathbf{t} = 12$ is $\mathbf{N}_{12}^+ = \xi(N_\phi, N_{12}, N_3, N_{123})^T$. On the other hand, the subsets $\mathbf{m}$ of $\mathcal{T}$ such that $\|\mathbf{m} \cap \mathbf{t}\|$ is odd are $\{1\}$, $\{2\}$, $\{1, 3\}$, and $\{2, 3\}$. Therefore, the negative split-N vector of $\mathbf{t} = 12$ is $\mathbf{N}_{12}^- = \xi(N_1, N_2, N_{13}, N_{23})^T$. For the design with $k = 3$, we list in Table 3.1 the positive and negative split-N vectors for all non-empty subsets $\mathbf{t}$ of $\mathcal{T}$. In Example 2, $\mathbf{N}(\mathcal{A}_1) = (1, 2, 3, 4, 5, 6, 7, 8)^T$. The split-N vectors for $\mathcal{A}_1$ are given in Table 3.2.

**Lemma 1.** *Let $\mathbf{1}_n$ be the $n \times 1$ vector with all components being one, i.e. $\mathbf{1}_n = (1, 1, \cdots, 1)^T$. Let $J_\mathbf{t}$ be as defined in equation (3.7). For a k-factor design,*

$$J_\mathbf{t} = \mathbf{1}_{2^{k-1}}^T \mathbf{N}_\mathbf{t}^+ - \mathbf{1}_{2^{k-1}}^T \mathbf{N}_\mathbf{t}^- \tag{3.63}$$

*where $\mathbf{t} \subseteq \mathcal{T}$ but $\mathbf{t} \neq \phi$.*

*Proof.* In equation (3.1), $x_{\mathbf{m}j} = -1$ if $j \in \mathbf{m}$ and $x_{\mathbf{m}j} = +1$ otherwise. From equation (3.4), it is obvious that $h_{\mathbf{mt}} = 1$ for those subsets $\mathbf{m}$ of $\mathcal{T}$ such that $\|\mathbf{m} \cap \mathbf{t}\|$ is even and $h_{\mathbf{mt}} = -1$ for those subsets $\mathbf{m}$ of $\mathcal{T}$ such that $\|\mathbf{m} \cap \mathbf{t}\|$ is odd. By applying the result to equation (3.7), the result follows. □

**Theorem 6.** *Let $\mathcal{D}$ and $\mathcal{D}'$ be two k-factor designs with counting vectors $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$, respectively. Suppose that $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic and $\mathcal{D}'$ can be obtained from $\mathcal{D}$ by switching signs of factors $\kappa_1, \cdots, \kappa_g$ where $g \leq k$ and permuting columns with the permutation $(j_1, \cdots, j_k)$ of $(1, \cdots, k)$. Let $\mathcal{T} = \{1, \cdots, k\}$ and define $\mathbf{t}' = \bigcup_{i \in \mathbf{t}} \{j_i\}$ for $\mathbf{t} \subseteq \mathcal{T}$. Let $\kappa = \{\kappa_1, \cdots, \kappa_g\}$ and $\delta_{\mathbf{t}'} = \|\kappa \cap \mathbf{t}'\|$. The split-N vectors of $\mathcal{D}$ and $\mathcal{D}'$ have the relationship as given below.*
*(1). If $\delta_{\mathbf{t}'}$ is even, then $\mathbf{N_t}^+(\mathcal{D}') = \mathbf{N_{t'}}^+(\mathcal{D})$ and $\mathbf{N}_\mathbf{t}^-(\mathcal{D}') = \mathbf{N_{t'}}^-(\mathcal{D})$.*
*(2). If $\delta_{\mathbf{t}'}$ is odd, then $\mathbf{N_t}^+(\mathcal{D}') = \mathbf{N_{t'}}^-(\mathcal{D})$ and $\mathbf{N}_\mathbf{t}^-(\mathcal{D}') = \mathbf{N}_{\mathbf{t}'}^+(\mathcal{D})$.*

*Proof.* From equations (3.1) and (3.4), it is obvious that

$$h_{\mathbf{mt}} = \begin{cases} -1, & \text{if } \|\mathbf{m} \cap \mathbf{t}\| \text{ is odd,} \\ 1, & \text{if } \|\mathbf{m} \cap \mathbf{t}\| \text{ is even.} \end{cases} \tag{3.64}$$

Let $\mathbf{m}' = \bigcup_{i \in \mathbf{m}} \{j_i\}$ for all subsets $\mathbf{m}$ of $\mathcal{T}$ and $\mathbf{t}' = \bigcup_{i \in \mathbf{t}} \{j_i\}$ for all subsets $\mathbf{t}$ of $\mathcal{T}$. Notice that $\mathbf{m}$ is the row index and $\mathbf{t}$ is the column index. By substituting $\mathbf{m}' \vee \kappa$ for $\mathbf{m}$ and $\mathbf{t}'$ for $\mathbf{t}$ in equations (3.1) and (3.4), we obtain

$$\mathbf{x}_{\mathbf{m}' \vee \kappa} = (x_{(\mathbf{m}' \vee \kappa)1}, \cdots, x_{(\mathbf{m}' \vee \kappa)k}), \tag{3.65}$$

where

$$\begin{cases} x_{(\mathbf{m}' \vee \kappa)j} = -1, & \text{if } j \in \mathbf{m}' \vee \kappa, \\ x_{(\mathbf{m}' \vee \kappa)j} = +1, & \text{otherwise,} \end{cases} \tag{3.66}$$

and

$$h_{(\mathbf{m}' \vee \kappa)\mathbf{t}'} = \prod_{j \in \mathbf{t}'} x_{(\mathbf{m}' \vee \kappa)j}. \tag{3.67}$$

It is obvious that by equations (3.65) and (3.67),

$$h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'} = \begin{cases} -1, & \text{if } \|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\| \text{ is odd,} \\ 1, & \text{if } \|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\| \text{ is even.} \end{cases} \tag{3.68}$$

The subset $(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'$ can be expanded as follows,

$$\begin{aligned} (\mathbf{m}'\vee\kappa)\cap\mathbf{t}' &= (\mathbf{m}'\cup\kappa - \mathbf{m}'\cap\kappa)\cap\mathbf{t}' \\ &= (\mathbf{m}'\cap\mathbf{t}' - \mathbf{m}'\cap\kappa\cap\mathbf{t}') + (\kappa\cap\mathbf{t}' - \mathbf{m}'\cap\kappa\cap\mathbf{t}'). \end{aligned} \tag{3.69}$$

From equation (3.69), the number of the components in $(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'$ is

$$\|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\| = \|\mathbf{m}'\cap\mathbf{t}'\| + \|\kappa\cap\mathbf{t}'\| - 2\|\mathbf{m}'\cap\kappa\cap\mathbf{t}'\|. \tag{3.70}$$

Now, let us consider the case that $\delta_{\mathbf{t}'}$ is even, i.e., $\|\kappa\cap\mathbf{t}'\|$ is even. By equation (3.64), when $h_{\mathbf{mt}} = -1$, $\|\mathbf{m}\cap\mathbf{t}\|$ is odd and hence $\|\mathbf{m}'\cap\mathbf{t}'\|$ is odd. According to equation (3.70), $\|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\|$ is odd (odd + even − even) and hence $h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'} = -1$ by equation (3.68). When $h_{\mathbf{mt}} = 1$, $\|\mathbf{m}\cap\mathbf{t}\|$ is even and hence $\|\mathbf{m}'\cap\mathbf{t}'\|$ is even. According to equation (3.70), $\|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\|$ is even (even + even − even) and hence $h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'} = 1$ by equation (3.68). Therefore, when $\delta_{\mathbf{t}'}$ is even, $h_{\mathbf{mt}} = h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'}$ for all subsets $\mathbf{m}$ of $\mathcal{T}$. Because $N_{\mathbf{m}}(\mathcal{D}') = N_{(\mathbf{m}'\vee\kappa)}(\mathcal{D})$ by Theorem 5, the result (1) follows by separating $\mathbf{N}(\mathcal{D}')$ into $\mathbf{N}^+_{\mathbf{t}}(\mathcal{D}')$ and $\mathbf{N}^-_{\mathbf{t}}(\mathcal{D}')$ according to the signs of the components in the column $\mathbf{h_t}$ and separating $\mathbf{N}(\mathcal{D})$ into $\mathbf{N}^+_{\mathbf{t}'}(\mathcal{D})$ and $\mathbf{N}^-_{\mathbf{t}'}(\mathcal{D})$ according to the signs of the components in the column $\mathbf{h}_{\mathbf{t}'}$. Let us now consider the case that $\delta_{\mathbf{t}'}$ is odd, i.e., $\|\kappa\cap\mathbf{t}'\|$ is odd. When $h_{\mathbf{mt}} = -1$, $\|\mathbf{m}\cap\mathbf{t}\|$ is odd and hence $\|\mathbf{m}'\cap\mathbf{t}'\|$ is odd. Therefore $\|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\|$ is even (odd + odd − even) and hence $h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'} = 1$ by equation (3.68). When $h_{\mathbf{mt}} = 1$, $\|\mathbf{m}\cap\mathbf{t}\|$ is even and hence $\|\mathbf{m}'\cap\mathbf{t}'\|$ is even. We have that $\|(\mathbf{m}'\vee\kappa)\cap\mathbf{t}'\|$ is odd (even + odd − even) and hence $h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'} = -1$ by equation (3.68). Therefore, when $\delta_{\mathbf{t}'}$ is odd, $h_{\mathbf{mt}} = -h_{(\mathbf{m}'\vee\kappa)\mathbf{t}'}$ for all subsets $\mathbf{m}$ of $\mathcal{T}$. Because $N_{\mathbf{m}}(\mathcal{D}') = N_{(\mathbf{m}'\vee\kappa)}(\mathcal{D})$ by Theorem 5, the result (2) follows by separating $\mathbf{N}(\mathcal{D}')$ into $\mathbf{N}^+_{\mathbf{t}}(\mathcal{D}')$ and $\mathbf{N}^-_{\mathbf{t}}(\mathcal{D}')$ according to the signs of the components in the column $\mathbf{h_t}$ and separating $\mathbf{N}(\mathcal{D})$ into $\mathbf{N}^+_{\mathbf{t}'}(\mathcal{D})$ and $\mathbf{N}^-_{\mathbf{t}'}(\mathcal{D})$ according to the signs of the components in the column $\mathbf{h}_{\mathbf{t}'}$. $\square$

**Example 4.** We use the designs $\mathcal{A}_1$ and $\mathcal{A}_2$ in Example 2 to illustrate the Theorem 6. Notice that $\mathbf{N}(\mathcal{A}_1) = (N_\phi(\mathcal{A}_1), N_1(\mathcal{A}_1), N_2(\mathcal{A}_1), N_{12}(\mathcal{A}_1), N_3(\mathcal{A}_1), N_{13}(\mathcal{A}_1), N_{23}(\mathcal{A}_1), N_{123}(\mathcal{A}_1))^T = (1,2,3,4,5,6,7,8)^T$. Let $\kappa = \{\kappa_1, \kappa_2\} = \{1,2\}$ and $(j_1, j_2, j_3) = (3,2,1)$. First, take $\mathbf{t} = 23$ as an example. When $\mathbf{t} = 23$, $\mathbf{t}' = j_2 j_3 = 12$ and $\delta_{\mathbf{t}'} = 2$ (even). According to Theorem 5, when $\mathbf{t} = 23$, $\mathbf{N}^+_{23}(\mathcal{A}_2)$ can be obtained from Table 3.1 by

$$\begin{aligned}
\mathbf{N}^+_{23}(\mathcal{A}_2) &= \xi(N_{\phi\vee\kappa}(\mathcal{A}_2), N_{1\vee\kappa}(\mathcal{A}_2), N_{23\vee\kappa}(\mathcal{A}_2), N_{123\vee\kappa}(\mathcal{A}_2))^T \\
&= \xi(N_{\phi\vee\kappa}(\mathcal{A}_1), N_{j_1\vee\kappa}(\mathcal{A}_1), N_{j_2j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_2j_3\vee\kappa}(\mathcal{A}_1))^T \\
&= \xi(N_{\phi\vee 12}(\mathcal{A}_1), N_{3\vee 12}(\mathcal{A}_1), N_{21\vee 12}(\mathcal{A}_1), N_{321\vee 12}(\mathcal{A}_1))^T \\
&= \xi(N_{12}(\mathcal{A}_1), N_{123}(\mathcal{A}_1), N_\phi(\mathcal{A}_1), N_3(\mathcal{A}_1))^T \\
&= (8,5,4,1)^T,
\end{aligned}$$

By Table 3.2, it is obvious that $\mathbf{N}^+_{23}(\mathcal{A}_2)$ equals $\mathbf{N}^+_{12}(\mathcal{A}_1)$. Moreover,

$$\begin{aligned}
\mathbf{N}^-_{23}(\mathcal{A}_2) &= \xi(N_{2\vee\kappa}(\mathcal{A}_2), N_{12\vee\kappa}(\mathcal{A}_2), N_{3\vee\kappa}(\mathcal{A}_2), N_{13\vee\kappa}(\mathcal{A}_2))^T \\
&= \xi(N_{j_2\vee\kappa}(\mathcal{A}_1), N_{j_1j_2\vee\kappa}(\mathcal{A}_1), N_{j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_3\vee\kappa}(\mathcal{A}_1))^T \\
&= \xi(N_{2\vee 12(\mathcal{A}_1)}, N_{32\vee 12(\mathcal{A}_1)}, N_{1\vee 12(\mathcal{A}_1)}, N_{31\vee 12}(\mathcal{A}_1))^T \\
&= \xi(N_1(\mathcal{A}_1), N_{13}(\mathcal{A}_1), N_2(\mathcal{A}_1), N_{23}(\mathcal{A}_1))^T \\
&= (7,6,3,2)^T.
\end{aligned}$$

By Table 3.2, it is clear that $\mathbf{N}^-_{23}(\mathcal{A}_2)$ equals $\mathbf{N}^-_{12}(\mathcal{A}_1)$. The results verify the Theorem 6 (1). Let us take $\mathbf{t} = 3$ as another example. When $\mathbf{t} = 3$, $\mathbf{t}' = j_3 = 1$ and $\delta_{\mathbf{t}'} = 1$ (odd). Similarly, the $\mathbf{N}^+_3(\mathcal{A}_2)$ and $\mathbf{N}^-_3(\mathcal{A}_2)$ can be obtained by

$$\begin{aligned}
\mathbf{N}^+_3(\mathcal{A}_2) &= \xi(N_{\phi\vee\kappa}(\mathcal{A}_1), N_{j_1\vee\kappa}(\mathcal{A}_1), N_{j_2\vee\kappa}(\mathcal{A}_1), N_{j_1j_2\vee\kappa}(\mathcal{A}_1))^T \\
&= \xi(N_{\phi\vee 12}(\mathcal{A}_1), N_{3\vee 12}(\mathcal{A}_1), N_{2\vee 12}(\mathcal{A}_1), N_{32\vee 12}(\mathcal{A}_1))^T \\
&= \xi(N_{12}(\mathcal{A}_1), N_{123}(\mathcal{A}_1), N_1(\mathcal{A}_1), N_{13}(\mathcal{A}_1))^T \\
&= (8,6,4,2)^T \\
&= \mathbf{N}^-_1(\mathcal{A}_1),
\end{aligned}$$

Table 3.3: Relationship of split-N vectors between $\mathcal{A}_1$ and $\mathcal{A}_2$

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $\mathbf{N}_1^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_3^+(\mathcal{A}_1)$ | $\mathbf{N}_1^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_3^-(\mathcal{A}_1)$ |
| $\mathbf{N}_2^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_2}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_2^-(\mathcal{A}_1)$ | $\mathbf{N}_2^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_2}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_2^+(\mathcal{A}_1)$ |
| $\mathbf{N}_{12}^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_2}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{23}^-(\mathcal{A}_1)$ | $\mathbf{N}_{12}^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_2}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{23}^+(\mathcal{A}_1)$ |
| $\mathbf{N}_3^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_3}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_1^-(\mathcal{A}_1)$ | $\mathbf{N}_3^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_3}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_1^+(\mathcal{A}_1)$ |
| $\mathbf{N}_{13}^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_3}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{13}^-(\mathcal{A}_1)$ | $\mathbf{N}_{13}^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_3}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{13}^+(\mathcal{A}_1)$ |
| $\mathbf{N}_{23}^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_2j_3}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{12}^+(\mathcal{A}_1)$ | $\mathbf{N}_{23}^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_2j_3}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{12}^-(\mathcal{A}_1)$ |
| $\mathbf{N}_{123}^+(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_2j_3}^+(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{123}^+(\mathcal{A}_1)$ | $\mathbf{N}_{123}^-(\mathcal{A}_2)$ | $=$ | $\mathbf{N}_{j_1j_2j_3}^-(\mathcal{A}_1)$ | $=$ | $\mathbf{N}_{123}^-(\mathcal{A}_1)$ |

and

$$\begin{aligned}
\mathbf{N}_3^-(\mathcal{A}_2) &= \xi(N_{j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_3\vee\kappa}(\mathcal{A}_1), N_{j_2j_3\vee\kappa}(\mathcal{A}_1), N_{j_1j_2j_3\vee\kappa}(\mathcal{A}_1))^T \\
&= \xi(N_{1\vee12}(\mathcal{A}_1), N_{31\vee12}(\mathcal{A}_1), N_{21\vee12}(\mathcal{A}_1), N_{321\vee12}(\mathcal{A}_1))^T \\
&= \xi(N_2(\mathcal{A}_1), N_{23}(\mathcal{A}_1), N_\phi(\mathcal{A}_1), N_3(\mathcal{A}_1))^T \\
&= (7,5,3,1)^T \\
&= \mathbf{N}_1^+(\mathcal{A}_1).
\end{aligned}$$

The results verify the Theorem 6 (2). The relationship of split-N vectors for all non-empty subsets $\mathbf{t}$ of $\mathcal{T}$ between $\mathcal{A}_1$ and $\mathcal{A}_2$ is listed in Table 3.3.

### 3.3.2 Split-N matrix

**Definition 2.** *Define the* priority rule *for two vectors as follows. For two $n \times 1$ vectors, $\mathbf{A} = (a_1, \cdots, a_n)^T$ and $\mathbf{B} = (b_1, \cdots, b_n)^T$, $\mathbf{A}$ is said to be prior to $\mathbf{B}$, denoted by $\mathbf{A} \succ \mathbf{B}$, if there exists a $v$ $(\leq n)$ such that $a_i = b_i$ for $i < v$ and $a_i > b_i$ for $i = v$. When $\mathbf{A}$ is prior or equal to $\mathbf{B}$, it is denoted by $\mathbf{A} \succeq \mathbf{B}$ .*

**Definition 3.** *Let $\mathbf{N_t} = (\mathbf{N_t}^{+T}, \mathbf{N_t}^{-T})^T$ if $\mathbf{N_t^+} \succeq \mathbf{N_t^-}$ and $\mathbf{N_t} = (\mathbf{N_t}^{-T}, \mathbf{N_t}^{+T})^T$ otherwise. For all $\mathbf{N_t}$'s with $\|\mathbf{t}\| = j$, arrange them in the order defined in Definition 2 and write them as $\mathbf{N}^j_{(1)} \succeq \mathbf{N}^j_{(2)} \succeq \cdots \succeq \mathbf{N}^j_{(\binom{k}{j})}$. The* split-N matrix *is then defined by*

$$\mathbf{N}^{sp} = (\mathbf{N}^1_{(1)}, \cdots, \mathbf{N}^1_{(\binom{k}{1})}, \mathbf{N}^2_{(1)}, \cdots, \mathbf{N}^2_{(\binom{k}{2})}, \cdots, \mathbf{N}^k_{(\binom{k}{k})}). \qquad (3.71)$$

**Theorem 7.** *If $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, then* $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$.

*Proof.* Suppose that $\mathcal{D}'$ is obtained from $\mathcal{D}$ by switching the sign of factors $\kappa_1, \cdots, \kappa_g$, and permuting columns by $(j_1, \cdots, j_k)$. Let $\mathcal{T} = \{1, \cdots, k\}$ and $\mathbf{t}' = \bigcup_{i \in \mathbf{t}} \{j_i\}$ for $\mathbf{t} \subseteq \mathcal{T}$. Let $\mathcal{G}_j(\mathcal{D})$ be the collection of all split-N vectors of design $\mathcal{D}$ with $\|\mathbf{t}\| = j$ and $\mathcal{G}_j(\mathcal{D}')$ be the collection of all split-N vectors of design $\mathcal{D}'$ with $\|\mathbf{t}'\| = j$. By Theorem 6, $\mathcal{G}_j(\mathcal{D}) = \mathcal{G}_j(\mathcal{D}')$ for $j = 1, \cdots, k$. Because $\mathbf{N}^{sp}(\mathcal{D})$ ($\mathbf{N}^{sp}(\mathcal{D}')$) is obtained by arranging the split-N vectors in $\mathcal{G}_j(\mathcal{D})$ ($\mathcal{G}_j(\mathcal{D}')$) according to the order in Definition 2 and then merging them into a matrix, $\mathbf{N}^{sp}(\mathcal{D})$ must be identical to $\mathbf{N}^{sp}(\mathcal{D}')$. $\square$

**Example 5.** The split-N matrices of $\mathcal{A}_1$ and $\mathcal{A}_2$ in Example 2 are

$$\mathbf{N}^{sp}(\mathcal{A}_1) = \mathbf{N}^{sp}(\mathcal{A}_2) = \begin{pmatrix} 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 7 & 7 & 6 & 7 & 6 & 5 & 5 \\ 6 & 4 & 4 & 2 & 3 & 4 & 2 \\ 5 & 3 & 2 & 1 & 1 & 1 & 2 \\ 4 & 6 & 7 & 6 & 7 & 7 & 7 \\ 3 & 5 & 5 & 5 & 5 & 6 & 6 \\ 2 & 2 & 3 & 4 & 4 & 3 & 4 \\ 1 & 1 & 1 & 3 & 2 & 2 & 1 \end{pmatrix}.$$

The first three columns correspond to $\mathbf{N}_{\mathbf{t}}(\mathcal{A}_1)$'s (or $\mathbf{N}_{\mathbf{t}}(\mathcal{A}_2)$'s) with $\|\mathbf{t}\| = 1$, column 4 to column 6 correspond to $\mathbf{N}_{\mathbf{t}}(\mathcal{A}_1)$'s (or $\mathbf{N}_{\mathbf{t}}(\mathcal{A}_2)$'s) with $\|\mathbf{t}\| = 2$ and the last column is $\mathbf{N}_3(\mathcal{A}_1)$ (or $\mathbf{N}_3(\mathcal{A}_2)$).

Notice that in the split-N matrices of two isomorphic designs, the difference in their design matrices caused by row permutation, column permutation, and sign switch vanishes. The effect of the row permutation on design matrix is vanished in the split-N matrix because of the use of counting vector. When the counting vector is divided into the negative and positive split-N vectors according to Definition 1 and arranged into $\mathbf{N_t}$'s according to Definition 3, the effect of sign switch disappears. Finally, by ordering the $\mathbf{N_t}$'s to form the split-N matrices according to Definition 3, the effect of the column permutation is eliminated. In other words, split-N matrix is invariant to row permutation, column permutation and sign switch operations. Theorem 7 shows that when two designs are isomorphic, their split-N matrices are identical. It implies that for two designs with different split-N matrices, they must be non-isomorphic. Therefore, split-N matrix can be used as an initial screening measure for isomorphism examination. Compared to most initial screening methods mentioned in Section 3.1, the method based on split-N matrix is more powerful, as shown by the following theorem.

**Theorem 8.** *For two designs $\mathcal{D}$ and $\mathcal{D}'$,*
*(a). if $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$, then $CFV(\mathcal{D}) = CFV(\mathcal{D}')$;*
*(b). if $CFV(\mathcal{D}) = CFV(\mathcal{D}')$, then $GWLP(\mathcal{D}) = GWLP(\mathcal{D}')$;*
*(c). if $GWLP(\mathcal{D}) = GWLP(\mathcal{D}')$, then $CD_2^2(\mathcal{D}) = CD_2^2(\mathcal{D}')$ and $K_u(\mathcal{D}) = K_u(\mathcal{D}')$ for $u = 1, \cdots, k$.*

*Proof.* Suppose that $\mathcal{D}$ and $\mathcal{D}'$ are $k$-factor designs with $n$ runs. Recall that $l_{ji}$ in $CFV$ records the number of $\mathbf{t}$'s such that $|J_{\mathbf{t}}| = (n+1-i)$ for all subsets $\mathbf{t}$ of $\mathcal{T}$ with $\|\mathbf{t}\| = j$. Columns 1 to $k$ in $\mathbf{N}^{sp}(\mathcal{D})$ are the ordered $\mathbf{N_t}(\mathcal{D})$'s with $\|\mathbf{t}\| = 1$. Because $\mathbf{N_t}(\mathcal{D})$ is composed of $\mathbf{N_t}^+(\mathcal{D})$ and $\mathbf{N_t}^-(\mathcal{D})$, and $|J_{\mathbf{t}}| = |\mathbf{1}_{2^{k-1}}^T \mathbf{N}_{\mathbf{t}}^+(\mathcal{D}) - \mathbf{1}_{2^{k-1}}^T \mathbf{N}_{\mathbf{t}}^-(\mathcal{D})|$ by Lemma 1, the $l_{1,1}, \cdots, l_{1,n}$ of $CFV(\mathcal{D})$ can be obtained from columns 1 to $k$ in $\mathbf{N}^{sp}(\mathcal{D})$. Other $l_{ji}$'s for $j = 2, \cdots, k$ and $i = 1, \cdots, n$ can be similarly derived from $\mathbf{N}^{sp}(\mathcal{D})$. Therefore, $CFV(\mathcal{D})$ is completed determined by $\mathbf{N}^{sp}(\mathcal{D})$. Similarly, $CFV(\mathcal{D}')$ is determined by $\mathbf{N}^{sp}(\mathcal{D}')$. When $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$, the result (a) follows. If $CFV(\mathcal{D}) = CFV(\mathcal{D}')$, the frequencies of $J_{\mathbf{t}}$'s where $\|\mathbf{t}\| = j$ are identical for $\mathcal{D}$ and $\mathcal{D}'$. Because $\alpha_j$ is a function of square of $J_{\mathbf{t}}$'s where $\|\mathbf{t}\| = j$ by equation (3.13), $\alpha_j(\mathcal{D})$ equals to $\alpha_j(\mathcal{D}')$ for $j = 1, \cdots, k$. The result (b) follows. By equations (3.14) and (3.15), $CD_2^2$ and $K_u$ are functions of $GWLP$. Therefore, two designs with the same $GWLP$ have the same $CD_2^2$ and same $K_u$. Results (c) and (d) follow. $\square$

Theorem 8 implies that $\mathbf{N}^{sp}$ is more powerful in terms of the ability to classify non-isomorphic designs than other measures appeared in this theorem. For instance, Theorem 8 (a) shows that if $CFV(\mathcal{D}) \neq CFV(\mathcal{D}')$, $\mathbf{N}^{sp}(\mathcal{D})$ must be different from $\mathbf{N}^{sp}(\mathcal{D}')$, but if $\mathbf{N}^{sp}(\mathcal{D}) \neq \mathbf{N}^{sp}(\mathcal{D}')$, $CFV(\mathcal{D})$ may still equal $CFV(\mathcal{D}')$. In other words, non-isomorphic designs that can be distinguished by $CFV$ can be classified by $\mathbf{N}^{sp}$, but not the other way around. Define the examination efficiency for measure $M$ by

$$eff(M) = \frac{\textit{number of distinguishable non-isomorphism groups by } M}{\textit{number of total non-isomorphic groups}}. \tag{3.72}$$

According to Theorem 8, the examination efficiencies of the initial screening measures mentioned in Theorem 8 can be ranked as:

$$eff(\mathbf{N}^{sp}) \geq eff(CFV) \geq eff(GWLP) \geq eff(CD_2^2), eff(K_u). \tag{3.73}$$



### 3.3.3 Projection

The application of the projection to the isomorphism examination has been introduced in Section 3.1. Briefly speaking, for a $k$-factor design, there are $\binom{k}{p}$ $p$-dimensional projections. We can apply the measures mentioned above on each of the $\binom{k}{p}$ projections to obtain a $p$-dimensional projection frequency. If designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, their $p$-dimensional projection frequency are identical. In other words, if there exists some $p$ where $1 \leq p \leq k$ such that the $p$-dimensional projection frequencies of $\mathcal{D}$ and $\mathcal{D}'$ are different, then $\mathcal{D}$ and $\mathcal{D}'$ must be non-isomorphic. Notice that the classification measures obtained from the projected matrices reveal more information than that calculated from the original design matrix. Two designs may have the same value of a classification measure but different $p$-dimensional projection frequencies. Therefore, for any measure, the initial screening method with projection is more efficient than the method without projection. The projection frequencies for the measures $CFV$, $GWLP$, $CD_2^2$, $K_u$, and **HD** are denoted by $P_{CFV}$, $P_{GWLP}$, $P_{CD_2^2}$, $P_{K_u}$, and $Deseq1$, respectively. In this dissertation, we also apply the technique of projection on the split-N matrix to increase its efficiency for isomorphism examination. We define the $p$-dimensional $\mathbf{N}^{sp}$ projection frequency in Definition 4. The collection of these $p$-dimensional $\mathbf{N}^{sp}$ projection frequencies for $p = 1, \cdots, k$ is referred to as the projection frequency of $\mathbf{N}^{sp}$ and denoted by $P_{N^{sp}}$.

**Definition 4.** *Let $\mathcal{D}$ be a $k$-factor design. Let $\mathcal{T} = \{1, \cdots, k\}$ and $\mathbf{w} \subset \mathcal{T}$. Let $\mathcal{T}_{(-\mathbf{w})} = \mathcal{T} - \mathbf{w}$ and $p = k - \|\mathbf{w}\|$. For all $\mathbf{N_t}$'s where $\mathbf{t} \subseteq \mathcal{T}_{(-\mathbf{w})}$ and $\|\mathbf{t}\| = j$, arrange them in the order defined in Definition 2 and write them as $\mathbf{N}^j_{(1)} \succeq \mathbf{N}^j_{(2)} \succeq \cdots \succeq \mathbf{N}^j_{(\binom{p}{j})}$. Define the* leave-$\mathbf{w}$-out split-N matrix *of design $\mathcal{D}$ by*

$$\mathbf{N}^{sp}_{(-\mathbf{w})}(\mathcal{D}) = (\mathbf{N}^1_{(1)}, \cdots, \mathbf{N}^1_{(\binom{p}{1})}, \mathbf{N}^2_{(1)}, \cdots, \mathbf{N}^2_{(\binom{p}{2})}, \cdots, \mathbf{N}^p_{(\binom{p}{p})}), \tag{3.74}$$

*where $\mathbf{N}^{sp}_{(-\mathbf{w})}$ is referred to as a* $p$-dimensional split-N matrix. *For a given $p$, there are $\binom{k}{p}$ $p$-dimensional projections and hence there are $\binom{k}{p}$ $p$-dimensional split-N matrices. The frequency of these $p$-dimensional split-N matrices is called the $p$-dimensional $\mathbf{N}^{sp}$ projection frequency for $p = 1, \cdots, k$. When $\mathbf{w} = \phi$, i.e., $p = k$, equation (3.74) is the split-N matrix of design $\mathcal{D}$.*

**Theorem 9.** *If $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, then $P_{N^{sp}}(\mathcal{D}) = P_{N^{sp}}(\mathcal{D}')$.*

*Proof.* When $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, there exists a one-to-one correspondence between their projections. Therefore, the $p$-dimensional $\mathbf{N}^{sp}$ projection frequency of $\mathcal{D}$ and $\mathcal{D}'$ must be identical. $\square$

**Corollary 2.** *For two designs $\mathcal{D}$ and $\mathcal{D}'$,*
*(a). if $P_{N^{sp}}(\mathcal{D}) = P_{N^{sp}}(\mathcal{D}')$, then $P_{CFV}(\mathcal{D}) = P_{CFV}(\mathcal{D}')$;*
*(b). if $P_{CFV}(\mathcal{D}) = P_{CFV}(\mathcal{D}')$, then $P_{GWLP}(\mathcal{D}) = P_{GWLP}(\mathcal{D}')$;*
*(c). if $P_{GWLP}(\mathcal{D}) = P_{GWLP}(\mathcal{D}')$, then $P_{CD_2^2}(\mathcal{D}) = P_{CD_2^2}(\mathcal{D}')$ and $P_{K_u}(\mathcal{D}) = P_{K_u}(\mathcal{D}')$ for $u = 1, \cdots, k$.*

*Proof.* The results directly follow by Theorem 8. □

According to Theorem 9 and Corollary 2, it is clear that

$$eff(P_{N^{sp}}) \geq eff(P_{CFV}) \geq eff(P_{GWLP}) \geq eff(P_{CD_2^2}), eff(P_{K_u}). \tag{3.75}$$

### 3.3.4 Simplified methods

To use the split-N matrix to preform the isomorphism examination for two $k$-factor designs, it is required to compare two $2^k \times (2^k - 1)$ matrices. The computation time and the storage memory dramatically increase when $k$ becomes large. In this section, we propose some simplified methods based on the split-N matrix for the designs with large $k$.

For designs with single replicate, all components in the counting vector are either 1 or 0. In this case, the isomorphism examination based on $\mathbf{N}^{sp}$ is equivalent to the examination based on the $CFV$. When the components in the counting vector are either 1 or 0, $\mathbf{N}^{sp}$ is completely determined by $CFV$ and vice versa. Take the design with counting vector $\mathbf{N} = (1,1,0,0,1,0,1,1)^T$ as an example. When $|J_\mathbf{t}| = 1$, because $|J_\mathbf{t}| = |(\mathbf{1}_{2^{k-1}}^T \mathbf{N}_\mathbf{t}^+ - \mathbf{1}_{2^{k-1}}^T \mathbf{N}_\mathbf{t}^-)|$ by Lemma 1, the split-N vectors of $\mathbf{t}$, i.e., $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$, must meet one of the following conditions: (a). $\mathbf{N}_\mathbf{t}^+ = (1,1,1,0)^T$ and $\mathbf{N}_\mathbf{t}^- = (1,1,0,0)^T$, or (b). $\mathbf{N}_\mathbf{t}^+ = (1,1,0,0)^T$

and $\mathbf{N}_{\mathbf{t}}^{-} = (1,1,1,0)^T$. According to Definition 3, either conditions would give $\mathbf{N}_{\mathbf{t}} = (1,1,1,0,1,1,0,0)$. Similar argument can be applied for other values of $J_{\mathbf{t}}$ to obtain $\mathbf{N}_{\mathbf{t}}$. Because the $CFV$ records the frequency of values of $|J_{\mathbf{t}}|$'s and $\mathbf{N}^{sp}$ is obtained from $\mathbf{N}_{\mathbf{t}}$'s according to Definition 3, the split-N matrix is completed determined by $CFV$ for the design with single replicate. In other words, the examination efficiencies of the $\mathbf{N}^{sp}$ and $CFV$ are identical in this case. Because $CFV$ requires less storage memory and less computation time, it can be used to replace $\mathbf{N}^{sp}$ to perform the isomorphism examination when the components of counting vector are either 1 or 0.

The split-N vectors for isomorphic designs have the relationship as given in Theorem 6. Let us sum up the split-N vectors over $\mathbf{t}$ with $\|\mathbf{t}\| = j$, i.e., $\mathbf{SN}^j(\mathcal{D}) = \sum_{\mathbf{t}\subseteq\mathcal{T};\|\mathbf{t}\|=j}(\mathbf{N}_{\mathbf{t}}^{+}(\mathcal{D}) + \mathbf{N}_{\mathbf{t}}^{-}(\mathcal{D}))$ for $j = 1,\cdots,k$, and define the *sum of split-N matrix* as

$$\mathbf{SN}^{sp}(\mathcal{D}) = (\mathbf{SN}^1(\mathcal{D}),\cdots,\mathbf{SN}^k(\mathcal{D})). \tag{3.76}$$

According to Theorem 7, when designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$ and hance $\mathbf{SN}^{sp}(\mathcal{D}) = \mathbf{SN}^{sp}(\mathcal{D}')$ by Theorem 6. It implies that $\mathbf{SN}^{sp}$ can also be used as a measure for the isomorphism examination. In Example 2, the sum of split-N matrices of $\mathcal{A}_1$ and $\mathcal{A}_2$ are

$$\mathbf{SN}^{sp}(\mathcal{A}_1) = \mathbf{SN}^{sp}(\mathcal{A}_2) = \begin{pmatrix} 41 & 44 & 15 \\ 33 & 34 & 11 \\ 21 & 20 & 7 \\ 13 & 10 & 3 \end{pmatrix}.$$

Because $\mathbf{SN}^{sp}$ greatly reduces the dimension of $\mathbf{N}^{sp}$ from $2^k \times (2^k - 1)$ to $2^{k-1} \times k$, the isomorphism examination based on $\mathbf{SN}^{sp}$ can significantly save the comparison time and the storage memory. However, because different $\mathbf{N}^{sp}$'s may generate the same $\mathbf{SN}^{sp}$, the examination efficiency of $\mathbf{SN}^{sp}$ is lower than than that of $\mathbf{N}^{sp}$. In practice, we find that the efficiencies of $\mathbf{SN}^{sp}$ and $\mathbf{N}^{sp}$ are very closed (see the examples given in Section 3.4). When $\mathbf{SN}^{sp}$ is applied together with the technique of projections, we denote its projection frequency as $P_{SN^{sp}}$.

For the highly fractional factorial designs, a lot of components in their counting vector are zero. These zero components appear in the bottoms of $\mathbf{N}_{\mathbf{t}}^{+}$'s and $\mathbf{N}_{\mathbf{t}}^{-}$'s, which make the split-N matrix contain many rows with components all zero. These rows can be ignored to reduce the redundant comparison. For instance, let $\mathcal{D}$ be a highly fractional factorial designs. Let $\mathbf{N}_{\mathbf{t}^*}{}^{+}(\mathcal{D})$ (or $\mathbf{N}_{\mathbf{t}^*}{}^{+}(\mathcal{D})$) be the split-N vector that contains the least number, say $r$, of zero components among all $\mathbf{N}_{\mathbf{t}}{}^{+}(\mathcal{D})$'s and $\mathbf{N}_{\mathbf{t}}{}^{-}(\mathcal{D})$'s. Then from Definition 3, it is clear that in the split-N matrix, the $(2^{k-1}-r+1)$th to the $2^{k-1}$th rows and the $(2^k-r+1)$th to the $2^k$th rows are all zero. Let us denote by the split-N* matrix the $(2^k-2r)\times(2^k-1)$ matrix obtained from the split-N matrix by deleting these zero rows. The result of Theorem 6 still holds when the split-N matrix is replaced by the split-N* matrix. Therefore, the split-N* matrix can also be used as a measure for isomorphism examination. It reduces the dimension of the split-N matrix from $2^k \times (2^k-1)$ to $(2^k-2r)\times(2^k-1)$. This technique can also be applied to the $\mathbf{SN}^{sp}$ to reduce the dimension from $2^{k-1}\times k$ to $(2^{k-1}-r)\times k$.

## 3.4 Some comparisons

In this section, we use three examples to study the classification efficiency of the methods we propose in the previous sections and compare their performance with some initial screening methods existing in the literature.

**Example 6.** Katsaounis and Dean (2008) used two 4-factor designs, denoted in their paper by $df1$ and $df5$, to illustrate the power of their method for isomorphism examination. Among all the initial screening methods introduced in their paper, such as $Deseq1$, $P_{CD_2^2}$, $P_{K_u}$, $P_{CFV}$, $CFV_4$, 4R-prof, 8R-prof, gen-4R and ext-4R, only $Desq1$ can classify $df1$ and $df5$ as non-isomorphic. The two designs can be represent in terms of counting vectors as follows:

$$\mathbf{N}(df1) = (1,0,1,1,1,3,2,1,1,1,0,2,0,0,0,2)^T,$$

and

$$\mathbf{N}(df5) = (2,0,1,2,1,0,1,0,1,0,3,1,1,2,0,1)^T.$$

The split-N matrices of $df1$ and $df5$ are

$$\mathbf{N}^{sp}(df1) = \begin{pmatrix}
3&3&3&3&3&3&3&3&3&3&3&3&3&3&3\\
2&2&2&1&2&2&2&2&2&1&2&2&2&2&2\\
2&2&1&1&2&2&1&1&1&1&2&2&1&1&2\\
1&1&1&1&1&1&1&1&1&1&2&1&1&1&1\\
1&1&1&1&1&1&1&1&0&0&1&1&1&1&1\\
1&0&1&0&1&0&1&0&0&0&1&1&1&1&1\\
0&0&1&0&1&0&0&0&0&0&0&0&1&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
2&2&2&2&2&2&2&2&2&2&1&2&2&2&2\\
1&1&2&2&1&1&2&2&2&2&1&1&2&2&1\\
1&1&1&2&1&1&1&1&1&2&1&1&1&1&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&1&0&1&0&1&1&1&1&1&1&1&0&1&1\\
0&1&0&1&0&1&0&1&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&1&1&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{pmatrix},$$

and

$$\mathbf{N}^{sp}(df5) = \begin{pmatrix}
3&3&3&3&3&3&3&3&3&3&3&3&3&3&3\\
2&2&2&2&2&2&2&2&2&1&2&2&2&1&2\\
2&1&1&1&2&2&1&1&1&1&2&2&2&1&2\\
1&1&1&1&1&1&1&1&1&1&2&1&1&1&1\\
1&1&1&1&1&1&1&0&0&1&1&1&1&1&1\\
1&1&1&1&1&1&0&0&0&0&0&1&1&0&1\\
0&1&0&0&1&0&0&0&0&0&0&1&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0\\
2&2&2&2&2&2&2&2&2&2&1&2&2&2&2\\
1&2&2&2&1&1&2&2&2&2&1&1&1&2&1\\
1&1&1&1&1&1&1&1&1&2&1&1&1&2&1\\
1&1&1&1&1&1&1&1&1&1&1&1&1&1&1\\
1&0&1&1&0&1&1&1&1&1&1&0&1&1&1\\
0&0&0&0&0&0&1&1&1&1&1&0&0&1&0\\
0&0&0&0&0&0&0&1&1&0&0&0&0&0&0\\
0&0&0&0&0&0&0&0&0&0&0&0&0&0&0
\end{pmatrix}.$$

Because the second columns of $\mathbf{N}^{sp}(df1)$ and $\mathbf{N}^{sp}(df5)$ are different, the split-N method can quickly identify the two designs as non-isomorphic. The sum of split-N matrices are:

$$\mathbf{SN}^{sp}(df1) = \begin{pmatrix} 20 & 30 & 19 & 5 \\ 13 & 21 & 14 & 3 \\ 11 & 15 & 10 & 3 \\ 8 & 12 & 9 & 2 \\ 7 & 9 & 7 & 2 \\ 4 & 6 & 4 & 1 \\ 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \text{ and } \mathbf{SN}^{sp}(df5) = \begin{pmatrix} 20 & 30 & 19 & 5 \\ 15 & 21 & 12 & 3 \\ 19 & 15 & 12 & 3 \\ 8 & 12 & 9 & 2 \\ 7 & 9 & 7 & 2 \\ 4 & 6 & 4 & 1 \\ 1 & 3 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

Because the first and the third columns of $\mathbf{SN}^{sp}(df1)$ and $\mathbf{SN}^{sp}(df5)$ are apparently different, the $\mathbf{SN}^{sp}$ measure can also classify them as non-isomorphic designs. This example demonstrates how to use our methods to perform the isomorphism examination and shows that although most initial screening methods fail to distinguish the two designs, our methods still have good performance.

**Example 7.** Designs $\mathcal{B}_1$ to $\mathcal{B}_8$ are eight non-isomorphic $OA(32, 5, 2, 2)$'s with counting vectors

$$\begin{aligned}
\mathbf{N}(\mathcal{B}_1) &= (1,3,1,1,0,0,0,2,1,0,1,0,2,1,2,1,1,0,1,0,2,1,2,1,1,1,1,3,0,2,0,0)^T,\\
\mathbf{N}(\mathcal{B}_2) &= (1,1,3,0,0,1,0,2,1,2,0,0,1,1,2,1,1,0,1,1,2,2,0,1,1,1,0,3,1,0,2,0)^T,\\
\mathbf{N}(\mathcal{B}_3) &= (0,0,3,0,2,1,1,1,1,2,1,1,0,2,0,1,1,2,0,2,1,1,0,1,2,0,0,1,1,0,3,1)^T,\\
\mathbf{N}(\mathcal{B}_4) &= (1,1,2,0,0,0,1,3,2,1,1,0,1,2,0,1,1,2,0,1,2,1,1,0,0,0,1,3,1,1,2,0)^T,\\
\mathbf{N}(\mathcal{B}_5) &= (0,0,3,1,1,1,0,2,1,1,1,1,2,2,0,0,2,2,0,0,1,1,1,1,1,1,0,2,0,0,3,1)^T,\\
\mathbf{N}(\mathcal{B}_6) &= (1,0,2,1,2,1,1,0,1,1,0,2,0,2,1,1,1,1,0,2,0,2,1,1,3,0,0,1,0,1,3,0)^T,\\
\mathbf{N}(\mathcal{B}_7) &= (1,0,2,1,2,1,1,0,1,1,0,2,0,2,1,1,1,1,0,2,0,2,1,1,3,0,0,1,0,1,3,0)^T,\\
\mathbf{N}(\mathcal{B}_8) &= (0,1,0,1,3,0,1,2,2,2,1,1,0,0,1,1,1,1,2,2,1,1,0,0,1,0,1,0,0,3,2,1)^T.
\end{aligned}$$

Besides $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$, we also use the methods $CFV$, $GWLP$, $CD_2^2$, $K_u$, and **HD** and their projection versions $P_{CFV}$, $P_{GWLP}$, $P_{CD_2^2}$, $P_{K_u}$, and $Deseq1$ to perform the isomorphism examination. The classification results are presented in

Table 3.4. The first part of Table 3.4 contains the examination results of $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$. It shows that even though the projection technique is not applied, the $\mathbf{N}^{sp}$ can still 100% discriminate the eight non-isomorphic designs, and the simplified method $\mathbf{SN}^{sp}$ is as powerful as $\mathbf{N}^{sp}$ in this case. The second part contains the results of $CFV$, $GWLP$, $CD_2^2$, $K_u$ and **HD**. We find that when the projection technique is not applied, the overall $CFV$ can only separate these designs into two groups, $\{\mathcal{B}_1, \mathcal{B}_4, \mathcal{B}_5, \mathcal{B}_6\}$ and $\{\mathcal{B}_2, \mathcal{B}_3, \mathcal{B}_7, \mathcal{B}_8\}$. The other overall measures completely fail to distinguish any of these designs as non-isomorphic. Even thought the projection technique is adopted, the projection frequency, $P_{CFV}$, $P_{GWLP}$, $P_{CD_2^2}$, $P_{K_u}$, and $Deseq1$, still cannot fully distinguish the eight designs. The third part on the bottom of the table gives the results of these methods. It shows that projection frequencies cannot discriminate $\mathcal{B}_7$ and $\mathcal{B}_8$. In Section 3.3.2, equation (3.73) shows that $\mathbf{N}^{sp}$ has higher examination efficiency than $CFV$, $GWLP$, $CD_2^2$, and $K_u$, but does not indicate how much higher. In this example, we find that the efficiency of $\mathbf{N}^{sp}$ is eight times higher than the efficiencies of $GWLP$, $CD_2^2$, and $K_u$. We also find that the efficiency of **HD** is only 12.5% which is mush lower than the efficiency of $\mathbf{N}^{sp}$ for the eight designs. The power of the projection technique can also be found in this example. With projection, the efficiency greatly increases from 12.5% to 87.5%.

**Example 8.** In this example, we perform the isomorphism examination for various $OA(n, k, s, d)$'s using the initial screening methods introduced in this article. The number of the non-isomorphic $OA(n, k, s, d)$'s can be found in Stufken and Tang (2007), Sun, Li, and Ye (2002) and the results in Section 2.5.2. In Tables 3.5 and 3.6, the notation $OA(n, k, s, d) : w$ represents that there are $w$ non-isomorphic designs in total for the case of $OA(n, k, s, d)$. The following rows show the examination efficiency and the number of non-isomorphism groups identified by each method (on the left-hand side and right-hand side of the colon separately). From the two tables, we find that in most cases, the efficiencies of $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ are

Table 3.4: Result of isomorphism examination for the designs in example 7

| Methods | Examining result | Efficiency |
|---|---|---|
| $\mathbf{N}^{sp}$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7\},\{\mathcal{B}_8\}$ | 100% |
| $\mathbf{SN}^{sp}$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7\},\{\mathcal{B}_8\}$ | 100% |
| $CFV$ | $\{\mathcal{B}_1,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6\},\{\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_7,\mathcal{B}_8\}$ | 25% |
| $GWLP$ | $\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$ | 12.5% |
| $CD_2^2$ | $\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$ | 12.5% |
| $K_u$ | $\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$ | 12.5% |
| **HD** | $\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$ | 12.5% |
| $P_{CFV}$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$ | 87.5% |
| $P_{GWLP}$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$ | 87.5% |
| $P_{CD_2^2}$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$ | 87.5% |
| $P_{K_u}$ | $\{\mathcal{B}y_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$ | 87.5% |
| $Deseq1$ | $\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$ | 87.5% |

close to 100% and much higher than the other methods. For the cases that $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ cannot fully classify the non-isomorphic designs (i.e., efficiencies are lower than 100%), the projection frequencies of $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ can significantly improve the efficiencies to almost 100%, which is much higher than the efficiencies of the other methods. For instance, in $OA(36, 5, 2, 2)$, the efficiencies of $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ are 99.3% and 98.4%, respectively, but it is only 67.0% for **HD**, 23.1% for $CFV$, and 13.8% for the other methods. After the projection technique is applied, the efficiencies of $P_{N^{sp}}$ and $P_{SN^{sp}}$ can reach to 100% and 99.9%, respectively, but it is only 95.0% for $Deseq1$, and even less than 60.0% for the other methods. In Section 3.3, equations (3.73) and (3.75) indicate that the initial screening methods based on $\mathbf{N}^{sp}$ and $P_{N^{sp}}$ dominate the methods based on the other measures except **HD** and $Deseq1$. In the two tables, we find that only in the case of $OA(20, 6, 2, 2)$, the efficiency of **HD** is 1.3% higher than $\mathbf{N}^{sp}$. In the other cases, $\mathbf{N}^{sp}$ and $P_{N^{sp}}$ has better (or same) performance than **HD** and $Deseq1$, respectively. For instance, in the case of $OA(16, 6, 2, 2)$, the efficiency of $\mathbf{N}^{sp}$ is 33.3% higher than the efficiency of **HD** and in the case of $OA(28, 5, 2, 2)$, the efficiency of $P_{N^{sp}}$ is 6.3% higher than the efficiency of $Deseq1$. From the two tables, we also find that the projection technique can greatly enhances the examination efficiency. In the case of $OA(20, 6, 2, 2)$, all methods without projection have poor efficiencies: 69.3% for **HD**, 68.0% for $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$, and 56% for the other methods. However, the efficiencies of $P_{N^{sp}}$ and $P_{SN^{sp}}$ reach 100%, and it is 96% for $Deseq1$, and higher than 90% for the other methods. Although the simplified method based on $\mathbf{SN}^{sp}$ is less efficient than the method based on $\mathbf{N}^{sp}$ as mentioned above, we find in the two tables that the efficiencies of $\mathbf{SN}^{sp}$ and $\mathbf{N}^{sp}$ are very close. Even though in the case of $OA(32, 4, 2, 2)$, the efficiencies of $\mathbf{SN}^{sp}$ and $P_{SN^{sp}}$ are lower than the efficiencies of $CFV$ and $P_{CFV}$, the $\mathbf{SN}^{sp}$ and $P_{SN^{sp}}$ usually have better performance than the other methods when the number of non-isomorphic $OA(n, k, d, s)$'s increases.

Table 3.5: The isomorphism examination of the initial screening methods

| | $OA(24,4,2,2):10$ | $OA(32,4,2,2):19$ | $OA(36,4,2,2):15$ | $OA(40,4,2,2):32$ | $OA(44,4,2,2):28$ | $OA(48,4,2,2):54$ | $OA(52,4,2,2):48$ |
|---|---|---|---|---|---|---|---|
| $\mathbf{N}^{sp}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 32 | 100.0 % : 28 | 100.0 % : 54 | 100.0 % : 48 |
| $\mathbf{SN}^{sp}$ | 100.0 % : 10 | 94.7 % : 18 | 100.0 % : 15 | 100.0 % : 32 | 96.4 % : 27 | 94.4 % : 51 | 97.9 % : 47 |
| $CFV$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 96.9 % : 31 | 100.0 % : 28 | 96.3 % : 52 | 100.0 % : 48 |
| $GWLP$ | 90.0 % : 9 | 89.5 % : 17 | 93.3 % : 14 | 84.4 % : 27 | 82.1 % : 23 | 77.8 % : 42 | 79.2 % : 38 |
| $CD_2^2$ | 90.0 % : 9 | 89.5 % : 17 | 93.3 % : 14 | 84.4 % : 27 | 82.1 % : 23 | 77.8 % : 42 | 77.1 % : 37 |
| $K_u$ | 90.0 % : 9 | 89.5 % : 17 | 93.3 % : 14 | 84.4 % : 27 | 82.1 % : 23 | 77.8 % : 42 | 79.2 % : 38 |
| **HD** | 100.0 % : 10 | 94.7 % : 18 | 100.0 % : 15 | 100.0 % : 32 | 96.4 % : 27 | 100.0 % : 54 | 100.0 % : 48 |
| $P_{N^{sp}}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 32 | 100.0 % : 28 | 100.0 % : 54 | 100.0 % : 48 |
| $P_{SN^{sp}}$ | 100.0 % : 10 | 94.7 % : 18 | 100.0 % : 15 | 100.0 % : 32 | 96.4 % : 27 | 94.4 % : 51 | 97.9 % : 47 |
| $P_{CFV}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 96.9 % : 31 | 100.0 % : 28 | 96.3 % : 52 | 100.0 % : 48 |
| $P_{GWLP}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 96.9 % : 31 | 100.0 % : 28 | 96.3 % : 52 | 100.0 % : 48 |
| $P_{CD_2^2}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 96.9 % : 31 | 100.0 % : 28 | 96.3 % : 52 | 100.0 % : 48 |
| $P_{K_u}$ | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 96.9 % : 31 | 100.0 % : 28 | 96.3 % : 52 | 100.0 % : 48 |
| $Deseq1$ | 100.0 % : 10 | 94.7 % : 18 | 100.0 % : 15 | 100.0 % : 32 | 96.4 % : 27 | 100.0 % : 54 | 100.0 % : 48 |

| | $OA(16,5,2,2):11$ | $OA(20,5,2,2):11$ | $OA(24,5,2,2):63$ | $OA(28,5,2,2):127$ | $OA(32,5,2,2):491$ | $OA(36,5,2,2):1242$ | $OA(16,6,2,2):27$ |
|---|---|---|---|---|---|---|---|
| $\mathbf{N}^{sp}$ | 100.0 % : 11 | 100.0 % : 11 | 100.0 % : 63 | 98.4 % : 125 | 99.2 % : 487 | 99.3 % : 1233 | 96.3 % : 26 |
| $\mathbf{SN}^{sp}$ | 100.0 % : 11 | 100.0 % : 11 | 100.0 % : 63 | 98.4 % : 125 | 96.9 % : 476 | 98.4 % : 1222 | 96.3 % : 26 |
| $CFV$ | 100.0 % : 11 | 90.9 % : 10 | 77.8 % : 49 | 46.5 % : 59 | 45.2 % : 222 | 23.1 % : 287 | 96.3 % : 26 |
| $GWLP$ | 90.9 % : 10 | 90.9 % : 10 | 58.7 % : 37 | 40.2 % : 51 | 26.1 % : 128 | 13.8 % : 171 | 63.0 % : 17 |
| $CD_2^2$ | 90.9 % : 10 | 90.9 % : 10 | 58.7 % : 37 | 40.2 % : 51 | 26.1 % : 128 | 13.8 % : 171 | 63.0 % : 17 |
| $K_u$ | 90.9 % : 10 | 90.9 % : 10 | 58.7 % : 37 | 40.2 % : 51 | 26.1 % : 128 | 13.8 % : 171 | 63.0 % : 17 |
| **HD** | 90.9 % : 10 | 100.0 % : 11 | 74.6 % : 47 | 78.7 % : 100 | 67.0 % : 329 | 67.0 % : 832 | 63.0 % : 17 |
| $P_{N^{sp}}$ | 100.0 % : 11 | 100.0 % : 11 | 100.0 % : 63 | 100.0 % : 127 | 100.0 % : 491 | 100.0 % : 1242 | 100.0 % : 27 |
| $P_{SN^{sp}}$ | 100.0 % : 11 | 100.0 % : 11 | 100.0 % : 63 | 100.0 % : 127 | 99.2 % : 487 | 99.9 % : 1241 | 100.0 % : 27 |
| $P_{CFV}$ | 100.0 % : 11 | 90.9 % : 10 | 92.1 % : 58 | 70.1 % : 89 | 75.6 % : 371 | 58.7 % : 729 | 100.0 % : 27 |
| $P_{GWLP}$ | 100.0 % : 11 | 90.9 % : 10 | 92.1 % : 58 | 69.3 % : 88 | 75.6 % : 371 | 58.7 % : 729 | 100.0 % : 27 |
| $P_{CD_2^2}$ | 100.0 % : 11 | 90.9 % : 10 | 90.5 % : 57 | 65.4 % : 83 | 75.6 % : 371 | 58.7 % : 729 | 100.0 % : 27 |
| $P_{K_u}$ | 100.0 % : 11 | 90.9 % : 10 | 92.1 % : 58 | 69.3 % : 88 | 75.6 % : 371 | 58.7 % : 729 | 100.0 % : 27 |
| $Deseq1$ | 100.0 % : 11 | 100.0 % : 11 | 96.8 % : 61 | 93.7 % : 119 | 97.1 % : 477 | 95.0 % : 1180 | 100.0 % : 27 |

Table 3.6: The isomorphism examination of the initial screening methods

| | $OA(20,6,2,2):75$ | $OA(48,5,2,3):10$ | $OA(64,5,2,3):19$ | $OA(72,5,2,3):15$ | $OA(80,5,2,3):33$ | $OA(88,5,2,3):28$ | $OA(96,5,2,3):56$ |
|---|---|---|---|---|---|---|---|
| $\mathbf{N}^{sp}$ | 68.0 % : 51 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| $\mathbf{SN}^{sp}$ | 68.0 % : 51 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| $CFV$ | 56.0 % : 42 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 97.0 % : 32 | 100.0 % : 28 | 96.4 % : 54 |
| $GWLP$ | 56.0 % : 42 | 90.0 % : 9 | 84.2 % : 16 | 93.3 % : 14 | 75.8 % : 25 | 82.1 % : 23 | 67.9 % : 38 |
| $CD_2^2$ | 56.0 % : 42 | 90.0 % : 9 | 84.2 % : 16 | 93.3 % : 14 | 75.8 % : 25 | 82.1 % : 23 | 67.9 % : 38 |
| $K_u$ | 56.0 % : 42 | 90.0 % : 9 | 84.2 % : 16 | 93.3 % : 14 | 75.8 % : 25 | 82.1 % : 23 | 67.9 % : 38 |
| **HD** | 69.3 % : 52 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| $P_{N^{sp}}$ | 100.0 % : 75 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| $P_{SN^{sp}}$ | 100.0 % : 75 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| $P_{CFV}$ | 92.0 % : 69 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 97.0 % : 32 | 100.0 % : 28 | 96.4 % : 54 |
| $P_{GWLP}$ | 92.0 % : 69 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 97.0 % : 32 | 100.0 % : 28 | 96.4 % : 54 |
| $P_{CD_2^2}$ | 93.3 % : 69 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 97.0 % : 32 | 100.0 % : 28 | 96.4 % : 54 |
| $P_{K_u}$ | 92.0 % : 69 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 97.0 % : 32 | 100.0 % : 28 | 96.4 % : 54 |
| $Deseq1$ | 96.0 % : 72 | 100.0 % : 10 | 100.0 % : 19 | 100.0 % : 15 | 100.0 % : 33 | 100.0 % : 28 | 100.0 % : 56 |
| | $OA(32,6,2,3):10$ | $OA(40,6,2,3):9$ | $OA(48,6,2,3):45$ | $OA(32,7,2,3):17$ | $OA(64,7,2,4):7$ | $OA(96,7,2,4):4$ | $OA(256,7,2,5):17$ |
| $\mathbf{N}^{sp}$ | 100.0 % : 10 | 100.0 % : 9 | 97.8 % : 44 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $\mathbf{SN}^{sp}$ | 100.0 % : 10 | 100.0 % : 9 | 97.8 % : 44 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $CFV$ | 100.0 % : 10 | 88.9 % : 8 | 82.2 % : 37 | 100.0 % : 17 | 100.0 % : 7 | 75.0 % : 3 | 100.0 % : 17 |
| $GWLP$ | 90.0 % : 9 | 88.9 % : 8 | 60.0 % : 27 | 64.7 % : 11 | 85.7 % : 6 | 75.0 % : 3 | 94.1 % : 16 |
| $CD_2^2$ | 90.0 % : 9 | 88.9 % : 8 | 60.0 % : 27 | 64.7 % : 11 | 85.7 % : 6 | 75.0 % : 3 | 94.1 % : 16 |
| $K_u$ | 90.0 % : 9 | 88.9 % : 8 | 60.0 % : 27 | 64.7 % : 11 | 85.7 % : 6 | 75.0 % : 3 | 94.1 % : 16 |
| **HD** | 90.0 % : 9 | 100.0 % : 9 | 71.1 % : 32 | 64.7 % : 11 | 85.7 % : 6 | 75.0 % : 3 | 100.0 % : 17 |
| $P_{N^{sp}}$ | 100.0 % : 10 | 100.0 % : 9 | 100.0 % : 45 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $P_{SN^{sp}}$ | 100.0 % : 10 | 100.0 % : 9 | 100.0 % : 45 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $P_{CFV}$ | 100.0 % : 10 | 88.9 % : 8 | 88.9 % : 40 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $P_{GWLP}$ | 100.0 % : 10 | 88.9 % : 8 | 88.9 % : 40 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $P_{CD_2^2}$ | 100.0 % : 10 | 88.9 % : 8 | 88.9 % : 40 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $P_{K_u}$ | 100.0 % : 10 | 88.9 % : 8 | 88.9 % : 40 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |
| $Deseq1$ | 100.0 % : 10 | 100.0 % : 9 | 93.3 % : 42 | 100.0 % : 17 | 100.0 % : 7 | 100.0 % : 4 | 100.0 % : 17 |

# Chapter 4

# Summary

In Chapter 2, we adopt the indicator function approach to tackle the problems of design enumeration and isomorphism examination. For design enumeration, we propose an assembly method which generates a design from its LOO projections. Based on a hierarchical structure existing between $\mathcal{F}^*(n, k-1, 2, d)$ and $\mathcal{F}^*(n, k, 2, d)$ for $k = d+1, d+2, \cdots$, we can sequentially construct all indicator functions by the assembly method. To save the computation time, we generalize a method for isomorphism examination in Stufken and Tang (2007) to reduce the number of indicator functions in $\mathcal{F}^*(n, k, 2, d)$. Although for $k > d+2$, this generalization does not allow us to fully classify non-isomorphic $OA$s, it is still an efficient method to discard most isomorphic indicator functions in $\mathcal{F}^*(n, k, 2, d)$. We also propose a new method for isomorphism examination that utilizes some projection properties such as the projective index set. By applying these methods, we can efficiently enumerate all non-isomorphic designs for many cases.

In the assembly method, we currently consider all possible combinations of indicator functions in $\mathcal{F}^*(n, k-1, 2, d)$ to generate designs in $\mathcal{F}^*(n, k, 2, d)$. However, because there exist some constraints between the coefficients of LOO projections, many combinations cannot form an incomplete indicator function. One of our future work is to identify these constraints so that we can further reduce the computation by quickly eliminating combinations that do not satisfy the constraints. In Chapter 2, we use two methods, one based on group structure and the other based on projective index set and $|b_{\mathcal{T}}|$, to distinguish non-isomorphic designs. The two methods are only efficient for small designs. When the run size or number of factors is large, their performances get worse, i.e., many non-isomorphic designs cannot be distinguished by the two methods.

In Chapter 3, we adopt the counting vector approach, which is different from the conventional tools used to study designs, to characterize the fractional factorial designs. Because the components in the counting vector are ranked in Yates order, we obtain the relationship between the counting vectors of isomorphic designs. The effects of sign switch, column and row permutations on design matrix are corresponding to the matrix transformation on counting vectors. We obtain some sufficient and necessary conditions for counting vectors to be isomorphic, which offer a theoretical basis for the split-N method. We also develop an alternative method which uses suffix's operation to quickly obtain the counting vector from another isomorphic design.

In the split-N method, we split the counting vector according to the signs in each column of the model matrix. We find that any sign switch, column and row permutations working on the designs can be regarded as the permutation of the split-N vectors. We therefore propose the split-N matrix, which is invariant to the sign switch column and row permutations. The split-N matrix is a powerful measure for the isomorphism examination. Because the measures $CFV$, $GWLP$, $CD_2^2$, and $K_u$ are functions of split-N matrix, our method outperform the others as an initial screening method for isomorphism examination. From some examples in Section 3.4, we observe that the classification power of $\mathbf{N}^{sp}$ is much higher than the others in many cases. Although there does not exist a functional relationship between $\mathbf{N}^{sp}$ and $\mathbf{HD}$, most cases in Example 8 show that $\mathbf{N}^{sp}$ is more efficient than **HD**.

We also find that the projection technique developed for isomorphism examination can usually improve the classification efficiency. With projection, some measures can even enhance their examination efficiency for about 50% (see $OA(32, 5, 2, 2)$ in Example 8). From the examples in Section 3.4, we can find that among the methods with projection, our method $P_{N^{sp}}$ have much better performance for isomorphism examination. Unfortunately, because in the case of $OA(20, 7, 2, 2)$, the efficiency of $P_{N^{sp}}$ can only reach 98.7% (468/474), it is not a complete classification method.

There are some issues we would like to point out in the end of this dissertation. When the run size of the designs $n$ becomes large, the dimension of the hamming distance matrix, $\binom{n}{2} \times \binom{n}{2}$, increases quickly so that we need more computer memory to store the matrix of **HD** and have to spend more time on the comparison work. The dimension of split-N matrix, $2^k \times (2^k - 1)$, is not affected by $n$. Therefore, when $n$ is very large, split-N matrix seems to be a better method for isomorphism examination. However, the dimension of split-N matrix would dramatically increase when the number of factors $k$ becomes large. Under such a situation, **HD** becomes a better choice for isomorphism examination because the dimension of **HD** is not affected by $k$. Some simplified methods based on split-N matrix proposed in Section 3.3.4 can also be considered when $k$ is large.

# Bibliography


[1] Chen, J. and Lin, D. K. J. (1991). On the identity of $2^{k-p}$ designs. *Journal of Statistical Planning and Inference*, 28, 95-98.

[2] Chen, J., Sun, D. X. and Wu, C. F. J. (1993). A catalogue of two-level and three-level fractional factorial designs with small runs. *International Statistical Review*, 61, 131-145.

[3] Cheng, S. W. and Ye, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Annals of Statistics*, 32, 2168-2185.

[4] Clark, J. B. and Dean, A. M. (2001). Equivalence of fractional factorial designs. *Statistica Sinica*, 11, 537-547.

[5] Deng, L. Y. and Tang, B. (1999). Generalized resolution and minimum aberration criteria for Plackett-Burman and other nonregular factorial designs. *Statistica Sinica*, 9, 1071-1082.

[6] Draper, N. R. and Mitchell, T. J. (1968). Construction of the set of 256-run designs of resolution $\geq 5$ and the set of even 512-run designs of resolution $\geq 6$ with special reference to the unique saturated designs. *Annals of Mathematical Statistics*, 39, 246-255.

[7] Draper, N. R. and Mitchell, T. J. (1970). Construction of a set of 512-run designs of resolution $\geq 5$ and the set of even 1024-run designs of resolution $\geq 6$. *Annals of Mathematical Statistics*, 41, 876-887.

[8] Fontana, R., Pistone, G. and Rogantin, M. P. (2000). Classification of two-level factorial fractions. *Journal of Statistical Planning and Inference*, 87, 149-172.

[9] Fujii, Y., Namikawa, T. and Yamamoto, S. (1989). Classification of two-symbol orthogonal arrays of strength $t$, $t+3$ constraints and index 4, II. *SUT Journal of Mathematics* 25, 161-177.

[10] Hedayat, A. S., Seiden, E. and Stufken, J. (1997) On the maximum number of factors and the enumeration of 3-symbol orthogonal arrays of strength 3 and index 2. *Journal of Statistical Planning and Inference*, 58, 43-63.

[11] Hedayat, A. S., Sloane, N. J. A. and Stufken, J. (1999) *Orthogonal Arrays: Theory and Applications*. Springer, New York.

[12] Katsaounis, T. I. and Dean, A. M. (2008). A survey and evaluation of methods for determination of combinatorial equivalence of factorial designs. *Journal of Statistical Planning and Inference*, 138, 245-258.

[13] Lam, C. W. H. and Tonchev, V. D. (1996). Classification of affine resolvable 2-(27,9,4) designs. *Journal of Statistical Planning and Inference*, 56, 187-202.

[14] Ma, C. X., Fang, K. T. and Lin, D. K. J. (2001). On the isomorphism of fractional factorial designs. *Journal of Complexity*, 17, 86-97.

[15] Seiden, E. and Zemach, R. (1966). On orthogonal arrays. *Annals of Mathematical Statistics*, 37, 1355-1370.

[16] Stufken, J. and Tang, B. (2007). Complete enumeration of two-level orthogonal arrays of strength $d$ with $d+2$ constraints. *Annals of Statistics*, 35, 793-814.

[17] Sun, D. X., Li, W. and Ye, K. Q. (2002). An algorithm for sequentially constructing non-isomorphic orthogonal designs and its applications. *Technical*

*Report, Department of Applied Mathematics and Statistics, State University of New York at Stony Brook*, SUNYSB-AMS-01-13.

[18] Tang, B. (2001). Theory of $J$-characteristic for fractional factorial designs and projection justification of minimum $G_2$-aberration. *Biometrika*, 88, 401-407.

[19] Tang, B. and Deng, L. Y. (1999). Minimum $G_2$-aberration for nonregular fractional factorial designs. *Annals of Statistics*, 27, 1914-1926.

[20] Xu, H. (2003). Minimum moment aberration for nonregular designs and supersaturated designs. *Statistica Sinica*, 13, 691-708.

[21] Xu, H. (2005). A catalogue of three-level regular fractional factorial designs. *Metrika*, 62, 259-281.

[22] Yamamoto, S., Fujii, Y., Hyodo, Y. and Yumiba, H. (1992a). Classification of two-symbol orthogonal arrays of strength 2, size 16, 15 (maximal) constraints and index 4. *SUT Journal of Mathematics*, 28, 47-59.

[23] Yamamoto, S., Fujii, Y., Hyodo, Y. and Yumiba, H. (1992b). Classification of two-symbol orthogonal arrays of strength 2, size 20, 19 (maximal) constraints. *SUT Journal of Mathematics*, 28, 191-209.

[24] Ye, K. Q. (2003). Indicator function and its application in two-level factorial designs. *Annals of Statistics*, 31, 984-994.

[25] Yumiba, H., Hyodo, Y. and Yamamoto, S. (1997). Classification of two-symbol orthogonal arrays of size 24, strength 2, 6 constraints and index 6 derivable from saturated orthogonal arrays having 23 constraints. *SUT Journal of Mathematics*, 33, 47-63.