Statistica Sinica **22** (2012), 1253-1272
doi:http://dx.doi.org/10.5705/ss.2010.246

ISOMORPHISM EXAMINATION BASED ON THE COUNT VECTOR

Chang-Yun Lin and Shao-Wei Cheng

Yeshiva University and National Tsing-Hua University

Abstract: Isomorphism examination determines whether two design matrices are equivalent subject to some row, column, and level permutations. The purpose of this paper is to study the isomorphism problem from the viewpoint of the count vector. We find that two designs are isomorphic if and only if there exists a special type of linear transformation between their count vectors. The transformation can be characterized in terms of set operations for the subscripts of elements in the count vector. Besides, we propose an initial screening method based on the count vector, called the split-count matrix. We prove that the split-count matrix is more efficient than most existing initial screening methods. Some modified versions of the split-count matrix, including a projection version and some simplified versions, are discussed. Some examples and comparisons are given to demonstrate the power of the split-count matrix.

Key words and phrases: Generalized word length pattern, Hamming distance, indicator function, J-characteristics, power moment, projection, set operations, split-count matrix, squared centered L_2.

1. Introduction

Isomorphism examination determines whether two design matrices are equivalent subject to row and column permutations and level exchanges. It has played an important role in the enumeration of designs and the search of optimal designs, such as minimum aberration designs. However, isomorphism examination is a computationally intensive and time-consuming task. For instance, it requires $n!k!(s!)^k$ comparisons to examine two non-isomorphic designs that have n runs and k factors, each with s levels. Some methods have been suggested in the literature to accelerate the computation. They usually adopt some measure of designs that is easy to calculate to initially classify designs into groups according to their value(s) of the measure. The classification ensures that designs classified into different groups are non-isomorphic so that it is no longer necessary to apply a thorough examination for designs in different groups. These measures, which might have been originally proposed for other purposes than isomorphism examinations, include *generalized word length pattern* (abbreviated as $GWLP$) proposed in Tang and Deng (1999), *confounding frequency vector* (abbreviated as CFV)

proposed in Deng and Tang (1999), *Hamming distance* matrix (abbreviated as **HD**) proposed in Clark and Dean (2001), *squared centered L_2 discrepancy* (abbreviated as CD_2^2) proposed in Ma, Fang, and Lin (2001), and *uth power moment* (abbreviated as K_u) proposed in Xu (2003). For details about these measures, the reader is referred to the survey paper by Katsaounis and Dean (2008). Because these methods cannot guarantee that designs classified into the same group are isomorphic, we call them *initial screening methods*. For designs classified into the same group by these measures, it is needed to apply some *unique determination methods*, which are usually much more computation-intensive, to examine whether the designs are really equivalent. A unique determination method provides necessary and sufficient conditions for equivalence of designs. There exist two approaches in the statistical literature that provide unique determination methods: one is the approach based on *J-characteristics* and *indicator function* (Ye (2003); Cheng and Ye (2004); Stufken and Tang (2007)), and the other is the Hamming distance approach of Clark and Dean (2001).

The purpose of this paper is to study the isomorphism problem of 2-level designs from the viewpoint of the count vector, introduced later in this section. In Section 2, we prove that two designs are isomorphic if and only if there exists a special type of linear transformation between their count vectors. This property can be utilized to uniquely determine the isomorphism class of designs. In Section 3, we propose an initial screening method called the split-count matrix. We prove that the split-count matrix has better classification power than most initial screening measures mentioned above. In Section 4, we discuss some modified versions of the split-count matrix, including one incorporating the use of projected designs and several simplified versions that are especially suitable for designs with a large number of factors. In Section 5, some examples and comparisons are given to demonstrate the power of our method. A summary is given in Section 6.

In the remainder of this section, we introduce some notation and terminology. For a set A, let $\|A\|$ be the cardinality of the set. Let $\mathcal{T} = \{1, \ldots, k\}$, where k is the number of factors of a design. For any $\mathbf{m} \subseteq \mathcal{T}$, define a $1 \times k$ vector

$$\mathbf{x_m} = (x_{\mathbf{m}1}, \ldots, x_{\mathbf{m}k}), \text{ where } \begin{cases} x_{\mathbf{m}j} = -1, \text{ if } j \in \mathbf{m}, \\ x_{\mathbf{m}j} = +1, \text{ otherwise.} \end{cases} \tag{1.1}$$

By arranging $\mathbf{x_m}$'s in the Yates order of $\mathbf{m}$, we define a $2^k \times k$ matrix

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

where the superscript T denotes vector transpose and the subscripts denote the subsets $\mathbf{m}$'s. For instance, $\mathbf{x}_1$ represents $\mathbf{x}_{\{1\}}$, $\mathbf{x}_{12}$ represents $\mathbf{x}_{\{1,2\}}$, and so forth.

Notice that $\mathbf{X}$ can be regarded as a k-factor full factorial design with $\mathbf{m}$ as its row index. If $\mathbf{h}_j$ denotes the jth column of $\mathbf{X}$, $j = 1, \ldots, k$, then $\mathbf{X}$ can also be represented as

$$\mathbf{X} = (\mathbf{h}_1, \ldots, \mathbf{h}_k). \tag{1.2}$$

For any $\mathbf{t} \subseteq \mathcal{T}$, let $\mathbf{h_t}$ be the component-wise product of the columns $\mathbf{h}_j$, $j \in \mathbf{t}$. That is, $\mathbf{h_t}$ is a $2^k \times 1$ vector whose $\mathbf{m}$-th component is

$$h_{\mathbf{mt}} = \prod_{j \in \mathbf{t}} x_{\mathbf{m}j}. \tag{1.3}$$

Because $\|\mathbf{m} \cap \mathbf{t}\|$ is the number of negative $x_{\mathbf{m}j}$'s in (1.3), it is obvious that

$$h_{\mathbf{mt}} = \begin{cases} -1, & \text{if } \|\mathbf{m} \cap \mathbf{t}\| \text{ is odd,} \\ \;\;1, & \text{if } \|\mathbf{m} \cap \mathbf{t}\| \text{ is even.} \end{cases} \tag{1.4}$$

By arranging the $\mathbf{h_t}$ in the Yates order of $\mathbf{t}$, we obtain a $2^k \times 2^k$ matrix

$$\mathbf{H} = (\mathbf{h}_\phi, \mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_{12}, \mathbf{h}_3, \mathbf{h}_{13}, \mathbf{h}_{23}, \mathbf{h}_{123}, \mathbf{h}_4, \mathbf{h}_{14}, \ldots), \tag{1.5}$$

referred to as the *model matrix* of $\mathbf{X}$, with $\mathbf{m}$ as its row index and $\mathbf{t}$ as its column index. The matrix $\mathbf{H}$ is a Hadamard matrix.

Let $\mathcal{D} = (d_{ij})$, an $n \times k$ matrix, be a k-factor design matrix with n runs and levels coded as $+1$ and -1. Let $N_\mathbf{m}$ denote the number of appearances that a run $\mathbf{x_m}$ occurs in the design $\mathcal{D}$. Then, up to a row permutation, $\mathcal{D}$ is uniquely determined by the $2^k \times 1$ vector

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

where $N_\mathbf{m}$'s are arranged in the Yates order of $\mathbf{m}$. We call $\mathbf{N}$ the *count vector* of design $\mathcal{D}$ because it counts for $\mathcal{D}$ the numbers of appearance of all possible level combinations in $\mathbf{X}$. The concept of the count vector has appeared in Tang (2001) and Stufken and Tang (2007) concerning the study of J-characteristics, and in Fontana, Pistone, and Rogantin (2000), Ye (2003), and Cheng and Ye (2004) to define the indicator function.

Two count vectors are called isomorphic if their corresponding design matrices are isomorphic. 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. We combine $\mathbf{X}$ and $\mathbf{N}(\mathcal{D})$ to form a $2^k \times (k+1)$ matrix $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$. In each row of the matrix $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$, the first k components denote a run in the full factorial design and the last component denotes the number of appearances of the run in $\mathcal{D}$. Because the design matrix of $\mathcal{D}$ can be obtained from $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$,

$[\mathbf{X}|\mathbf{N}(\mathcal{D})]$ can fully characterize $\mathcal{D}$, and $\mathcal{D}$ is uniquely determined by $\mathbf{N}$ up to a row permutation.

2. Count Vector and Isomorphism

In this section, we first discuss how $[\mathbf{X}|\mathbf{N}(\mathcal{D})]$ is affected when the operations of sign switch (i.e., level exchange of 2-level designs), column, and row permutations are applied on $\mathcal{D}$. From the discussion, some necessary and sufficient conditions for two count vectors to be isomorphic are derived.

Let $\mathcal{D}^{\mathrm{r}}$ be the design obtained by permutating the rows of the design $\mathcal{D}$. Because the elements in $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}^{\mathrm{r}})$ are always sorted in Yate's order of $\mathbf{m}$, we have $[\mathbf{X}|\mathbf{N}(\mathcal{D})] = [\mathbf{X}|\mathbf{N}(\mathcal{D}^{\mathrm{r}})]$. In other words, for isomorphism examination based on the count vector, the row permutation operation can be ignored because it has no influence to the count vector.

Suppose that $\mathcal{D}^{\mathrm{s}}$ is obtained by switching the sign of factors $\kappa_1, \ldots, \kappa_g$ in $\mathcal{D}$. For $\kappa = \{\kappa_1, \ldots, \kappa_g\} \subseteq \mathcal{T}$, take $\nu^{\kappa}_{\mathbf{t}} = \|\kappa \cap \mathbf{t}\|$ for $\mathbf{t} \subseteq \mathcal{T}$. The design $\mathcal{D}^{\mathrm{s}}$ can be represented as $[\mathbf{X}^{\mathrm{s}}|\mathbf{N}(\mathcal{D})]$, where $\mathbf{X}^{\mathrm{s}} = ((-1)^{\nu^{\kappa}_1}\mathbf{h}_1, \ldots, (-1)^{\nu^{\kappa}_k}\mathbf{h}_k)$ is obtained by switching the signs of factors $\kappa_1, \ldots, \kappa_g$ in $\mathbf{X}$. Let $\mathbf{R}^{\mathrm{s}}$ be the $2^k \times 2^k$ matrix that represents the row permutation operation to transform $\mathbf{X}^{\mathrm{s}}$ into $\mathbf{X}$, i.e., $\mathbf{R}^{\mathrm{s}}_{ij} = 1$ if the jth row in $\mathbf{X}^{\mathrm{s}}$ is the ith row in $\mathbf{X}$ and $\mathbf{R}^{\mathrm{s}}_{ij} = 0$ otherwise. Then we have $\mathbf{R}^{\mathrm{s}}[\mathbf{X}^{\mathrm{s}}|\mathbf{N}(\mathcal{D})] = [\mathbf{X}|\mathbf{N}(\mathcal{D}^{\mathrm{s}})]$, where

$$\mathbf{R}^{\mathrm{s}}\mathbf{X}^{\mathrm{s}} = \mathbf{X}, \tag{2.1}$$

and $\mathbf{R}^{\mathrm{s}}\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}^{\mathrm{s}})$. Let $\lambda(\cdot)$ denote the procedure to transform a design matrix into its model matrix as shown in (1.2) to (1.5). Then, $\lambda(\mathbf{X}) = \mathbf{H}$ and $\lambda(\mathbf{X}^{\mathrm{s}}) = \mathbf{H}^{\mathrm{s}}$, where

$$\mathbf{H}^{\mathrm{s}} = ((-1)^{\nu^{\kappa}_{\phi}}\mathbf{h}_{\phi}, (-1)^{\nu^{\kappa}_1}\mathbf{h}_1, (-1)^{\nu^{\kappa}_2}\mathbf{h}_2, (-1)^{\nu^{\kappa}_{12}}\mathbf{h}_{12}, \ldots, (-1)^{\nu^{\kappa}_{1\cdots k}}\mathbf{h}_{1\cdots k}).$$

By applying λ on both sides of (2.1), we obtain

$$\mathbf{R}^{\mathrm{s}}\mathbf{H}^{\mathrm{s}} = \mathbf{H}, \tag{2.2}$$

because $\mathbf{R}^{\mathrm{s}}$ and λ are exchangeable operations. Let $\mathbf{S}^{\kappa}$ be the $2^k \times 2^k$ diagonal matrix with diagonal $\mathbf{h}_{\kappa}$. Then $\mathbf{H}^{\mathrm{s}} = \mathbf{H}\mathbf{S}^{\kappa}$, and (2.2) can be written as

$$\mathbf{R}^{\mathrm{s}}\mathbf{H}\mathbf{S}^{\kappa} = \mathbf{H}. \tag{2.3}$$

Because $\mathbf{H}^{-1} = 2^{-k}\mathbf{H}$ and $(\mathbf{S}^{\kappa})^{-1} = \mathbf{S}^{\kappa}$, we obtain from (2.3) that

$$\mathbf{R}^{\mathrm{s}} = \mathbf{H}(\mathbf{H}\mathbf{S}^{\kappa})^{-1} = \mathbf{H}(\mathbf{S}^{\kappa})^{-1}\mathbf{H}^{-1} = 2^{-k}\mathbf{H}\mathbf{S}^{\kappa}\mathbf{H}.$$

We call $\mathbf{S}^{\kappa}$ a *sign-switch matrix*. There are 2^k different $\mathbf{S}^{\kappa}$'s, each corresponding to a sign switch set $\kappa \subseteq \mathcal{T}$. Define $\mathcal{S}$ as the collection of these $\mathbf{S}^{\kappa}$'s. For any design $\mathcal{D}^{\mathrm{s}}$ that is obtained by sign-switching some factors in $\mathcal{D}$, there exists a matrix $\mathbf{S} \in \mathcal{S}$ such that, for the count vectors of $\mathcal{D}$ and $\mathcal{D}^{\mathrm{s}}$,

$$\mathbf{N}(\mathcal{D}^{\mathrm{s}}) = (2^{-k}\mathbf{HSH})\mathbf{N}(\mathcal{D}). \tag{2.4}$$

The reverse statement is also true because the count vector representation of a design is unique.

Let $(j_1, j_2, \cdots, j_k)$ be a permutation of $(1, 2, \cdots, k)$. Suppose that $\mathcal{D}^{\mathrm{c}}$ is obtained from $\mathcal{D}$ by the column permutation $(j_1, j_2, \cdots, j_k)$, i.e., the ith column in $\mathcal{D}^{\mathrm{c}}$ is the j_ith column in $\mathcal{D}$. By following an argument similar to that for sign switch operations, we can find a $2^k \times 2^k$ matrix $\mathbf{R}^{\mathrm{c}}$ that satisfies $\mathbf{R}^{\mathrm{c}}\mathbf{N}(\mathcal{D}) = \mathbf{N}(\mathcal{D}^{\mathrm{c}})$ and

$$\mathbf{R}^{\mathrm{c}}\mathbf{H}^{\mathrm{c}} = \mathbf{H}, \tag{2.5}$$

where

$$\mathbf{H}^{\mathrm{c}} = (\mathbf{h}_{\phi}, \mathbf{h}_{j_1}, \mathbf{h}_{j_2}, \mathbf{h}_{j_1 j_2}, \mathbf{h}_{j_3}, \cdots, \mathbf{h}_{j_1 \cdots j_k}). \tag{2.6}$$

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

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

Let

$$\mathbf{C}^{j_1 j_2 \cdots j_k} = (\mathbf{e}_{\phi}, \mathbf{e}_{j_1}, \mathbf{e}_{j_2}, \mathbf{e}_{j_1 j_2}, \mathbf{e}_{j_3}, \mathbf{e}_{j_1 j_3}, \mathbf{e}_{j_2 j_3}, \mathbf{e}_{j_1 j_2 j_3}, \mathbf{e}_{j_4}, \ldots, \mathbf{e}_{j_1 \cdots j_k}). \tag{2.7}$$

Then $\mathbf{H}^{\mathrm{c}}$ in (2.6) can be written as $\mathbf{H}^{\mathrm{c}} = \mathbf{H}\mathbf{C}^{j_1 \cdots j_k}$, and (2.5) becomes

$$\mathbf{R}^{\mathrm{c}}\mathbf{H}\mathbf{C}^{j_1 \cdots j_k} = \mathbf{H}. \tag{2.8}$$

Because $(\mathbf{C}^{j_1 \cdots j_k})^{-1} = (\mathbf{C}^{j_1 \cdots j_k})^T$, we obtain from (2.8) that

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

We call $\mathbf{C}^{j_1 \cdots j_k}$ a *column-permutation matrix.* There are $k!$ different $\mathbf{C}^{j_1 \cdots j_k}$'s, each corresponding to a permutation $(j_1, j_2, \cdots, j_k)$. Let $\mathcal{C}$ be the collection of them. For any design $\mathcal{D}^{\mathrm{c}}$ that is obtained by permuting columns of $\mathcal{D}$, there exists a matrix $\mathbf{C} \in \mathcal{C}$ such that for the count vectors of $\mathcal{D}$ and $\mathcal{D}^{\mathrm{c}}$,

$$\mathbf{N}(\mathcal{D}^{\mathrm{c}}) = (2^{-k}\mathbf{H}\mathbf{C}^T\mathbf{H})\mathbf{N}(\mathcal{D}). \tag{2.9}$$

The reverse statement is also true because the count vector representation of a design is unique.

Theorem 1. *Suppose $\mathcal{D}$ and $\mathcal{D}'$ are two k-factor designs with count vectors $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$, respectively. The designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic if and only if there exist a sign-switch matrix $\mathbf{S} \in \mathcal{S}$ and a column-permutation matrix $\mathbf{C} \in \mathcal{C}$ such that*

$$\mathbf{N}(\mathcal{D}') = (2^{-k}\mathbf{H}\mathbf{C}^T\mathbf{S}\mathbf{H})\mathbf{N}(\mathcal{D}). \tag{2.10}$$

Proof. The result follows directly from (2.4), (2.9), and

$$(2^{-k}\mathbf{H}\mathbf{C}^T\mathbf{H})(2^{-k}\mathbf{H}\mathbf{S}\mathbf{H}) = 2^{-2k}\mathbf{H}\mathbf{C}^T(\mathbf{H}\mathbf{H})\mathbf{S}\mathbf{H} = 2^{-k}\mathbf{H}\mathbf{C}^T\mathbf{S}\mathbf{H}.$$

Example 1. Let $\mathcal{A}_1$ be a 3-factor 2-level design with the count vector $\mathbf{N}(\mathcal{A}_1) = (r_1,\, r_2,\, r_3,\, r_4,\, r_5,\, r_6,\, r_7,\, r_8)^T$, where r_i's are non-negative integers. Let $\mathcal{A}_2$ be obtained from $\mathcal{A}_1$ by switching the sign of factors 1 and 2 and exchanging columns of factors 1 and 3. According to the definition of the count vector, it can be easily obtained that $\mathbf{N}(\mathcal{A}_2) = (r_4, r_8, r_2, r_6, r_3, r_7, r_1, r_5)^T$. By Theorem 1, $\kappa = \{1, 2\}$, so set $\mathbf{S} = \mathbf{S}^{12}$; the permutation is $(j_1, j_2, j_3) = (3, 2, 1)$, so set $\mathbf{C}^{j_1 j_2 j_3} = \mathbf{C}^{321}$, where $\mathbf{C}^{321}$ can be 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 matrices $\mathbf{S}^{12}$ and $\mathbf{C}^{321}$ are

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

from which we obtain

$$2^{-k}\mathbf{H}(\mathbf{C}^{321})^T\mathbf{S}^{12}\mathbf{H} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}. \tag{2.11}$$

It is then easy to check that $\mathbf{N}(\mathcal{A}_1)$ and $\mathbf{N}(\mathcal{A}_2)$ satisfy (2.10).

The matrix $2^{-k}\mathbf{H}\mathbf{C}^T\mathbf{S}\mathbf{H}$ in (2.10) is a permutation matrix and its function is to rearrange the order of elements in a count vector. The systematical reordering of elements in a count vector resulting from the sign switches and column permutations applied on the design matrix can be characterized by using some set operations of the subscripts, as shown in the next theorem. First we need some definitions. Let π denote the permutation $(j_1, j_2, \ldots, j_k)$ of $(1, 2, \ldots, k)$, and then for $\mathbf{m} \subseteq \mathcal{T}$, take

$$\mathbf{m}^\pi = \bigcup_{i \in \mathbf{m}} \{j_i\}. \tag{2.12}$$

For sets $\mathbf{a}$ and $\mathbf{b}$, their symmetric difference is $\mathbf{a} \ominus \mathbf{b} = (\mathbf{a} \cup \mathbf{b}) \setminus (\mathbf{a} \cap \mathbf{b})$.

Theorem 2. *Suppose that $\mathcal{D}$ and $\mathcal{D}'$ are two k-factor designs with count vectors $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$, respectively. The design $\mathcal{D}'$ can be obtained from $\mathcal{D}$ by switching sign of factors in $\kappa = \{\kappa_1, \ldots, \kappa_g\}$, permuting columns via the permutation $\pi = (j_1, j_2, \cdots, j_k)$, and permuting some rows, if and only if for any row indexes $\mathbf{m} \subseteq \mathcal{T}$, the components in $\mathbf{N}(\mathcal{D})$ and $\mathbf{N}(\mathcal{D}')$ satisfy $N_{\mathbf{m}}(\mathcal{D}') = N_{\mathbf{m}^\pi \ominus \kappa}(\mathcal{D})$.*

Proof. We first prove sufficiency. Suppose that the run $\mathbf{x}_{\mathbf{m}}$ in $\mathcal{D}'$ is obtained from the run $\mathbf{x}_{\mathbf{y}}$ in $\mathcal{D}$ through the sign switches, column permutations, and row permutations as stated. It is enough to show that $\mathbf{y} = \mathbf{m}^\pi \ominus \kappa$. Recall that the subscript set $\mathbf{m}$ of $\mathbf{x}_{\mathbf{m}}$ (or $N_{\mathbf{m}}$) indicates the factors whose levels are set to -1 in the run $\mathbf{x}_{\mathbf{m}}$. If $\mathcal{D}^{\mathrm{s}}$ is the design obtained from $\mathcal{D}$ by switching sign of factors in κ. Under the sign-switch operation, the run $\mathbf{x}_{\mathbf{y}}$ in $\mathcal{D}$ becomes the run $\mathbf{x}_{\mathbf{y}\ominus\kappa}$ in $\mathcal{D}^{\mathrm{s}}$. Next, if $\mathcal{D}'$ is the design obtained from $\mathcal{D}^{\mathrm{s}}$ through the column permutation $(j_1, \ldots, j_k)$, then the run $\mathbf{x}_{\mathbf{m}^\pi}$ in $\mathcal{D}^{\mathrm{s}}$ becomes the run $\mathbf{x}_{\mathbf{m}}$ in $\mathcal{D}'$. It is then obvious that $\mathbf{x}_{\mathbf{y}\ominus\kappa} = \mathbf{x}_{\mathbf{m}^\pi}$, and therefore $\mathbf{y} \ominus \kappa = \mathbf{m}^\pi$. Sufficiency holds because $\mathbf{y} = (\mathbf{y} \ominus \kappa) \ominus \kappa = \mathbf{m}^\pi \ominus \kappa$. Necessity is true because the count vector representation of a design is unique.

Example 2. (Example 1 cont.) Let $\kappa = \{1, 2\}$ and $(j_1, j_2, j_3) = (3, 2, 1)$. By Theorem 2, $\mathbf{N}(\mathcal{A}_2)$ can be obtained from $\mathbf{N}(\mathcal{A}_1)$ through the set operations of subscripts, as follows.

$$\begin{aligned}\mathbf{N}(\mathcal{A}_2) &= (N_{\phi\ominus\kappa}(\mathcal{A}_1), N_{j_1\ominus\kappa}(\mathcal{A}_1), N_{j_2\ominus\kappa}(\mathcal{A}_1), N_{j_1j_2\ominus\kappa}(\mathcal{A}_1),\\ &\qquad N_{j_3\ominus\kappa}(\mathcal{A}_1), N_{j_1j_3\ominus\kappa}(\mathcal{A}_1), N_{j_2j_3\ominus\kappa}(\mathcal{A}_1), N_{j_1j_2j_3\ominus\kappa}(\mathcal{A}_1))^T\\ &= (r_4, r_8, r_2, r_6, r_3, r_7, r_1, r_5)^T.\end{aligned}$$

Clark and Dean (2001) provided a necessary and sufficient condition for two designs to be isomorphic based on the Hamming distance. A distinction between their and our methods is that the number of iterations for distinguishing two non-isomorphic designs in their method is a function of n and k, while in our methods it is a function of k ($k!2^k$) as shown in Theorem 2.1 and 2.2.

An algorithm for checking equivalence of two designs is this. Step 1: select a sign-switch matrix with diagonal h_κ, i.e., $\mathbf{S}^\kappa$, and select a column-permutation matrix, $\mathbf{C}^\pi$. Step 2: calculate $\mathbf{R} = 2^{-k}\mathbf{H}(\mathbf{C}^\pi)^T\mathbf{S}^\kappa\mathbf{H}$ and check whether $\mathbf{RN}(\mathcal{D}) = \mathbf{N}(\mathcal{D}')$; if so, then $\mathcal{D}'$ is obtained from $\mathcal{D}$ by switching signs of factors in κ and then permutating columns by π; if not, repeat Steps 1 and 2 by replacing κ and π until all of the combinations of κ and π are considered. Designs $\mathcal{D}$ and $\mathcal{D}'$ are non-isomorphic if there exists no combination of κ and π such that $\mathbf{RN}(\mathcal{D}) = \mathbf{N}(\mathcal{D}')$.

Table 1. Split-count vectors of the designs with three factors.

$\mathbf{t}=\{1\}$	$\mathbf{N}_1^+ = \xi(\{N_\phi, N_2, N_3, N_{23}\})$	$\mathbf{N}_1^- = \xi(\{N_1, N_{12}, N_{13}, N_{123}\})$
$\mathbf{t}=\{2\}$	$\mathbf{N}_2^+ = \xi(\{N_\phi, N_1, N_3, N_{13}\})$	$\mathbf{N}_2^- = \xi(\{N_2, N_{12}, N_{23}, N_{123}\})$
$\mathbf{t}=\{1,2\}$	$\mathbf{N}_{12}^+ = \xi(\{N_\phi, N_{12}, N_3, N_{123}\})$	$\mathbf{N}_{12}^- = \xi(\{N_1, N_2, N_{13}, N_{23}\})$
$\mathbf{t}=\{3\}$	$\mathbf{N}_3^+ = \xi(\{N_\phi, N_1, N_2, N_{12}\})$	$\mathbf{N}_3^- = \xi(\{N_3, N_{13}, N_{23}, N_{123}\})$
$\mathbf{t}=\{1,3\}$	$\mathbf{N}_{13}^+ = \xi(\{N_\phi, N_2, N_{13}, N_{123}\})$	$\mathbf{N}_{13}^- = \xi(\{N_1, N_{12}, N_3, N_{23}\})$
$\mathbf{t}=\{2,3\}$	$\mathbf{N}_{23}^+ = \xi(\{N_\phi, N_1, N_{23}, N_{123}\})$	$\mathbf{N}_{23}^- = \xi(\{N_2, N_{12}, N_3, N_{13}\})$
$\mathbf{t}=\{1,2,3\}$	$\mathbf{N}_{123}^+ = \xi(\{N_\phi, N_{12}, N_{13}, N_{23}\})$	$\mathbf{N}_{123}^- = \xi(\{N_1, N_2, N_3, N_{123}\})$

3. Split-Count Matrix

Here we propose an initial screening measure, called the *split-count matrix*. It is based on the count vector. We prove that isomorphic designs have identical split-count matrices, and that the isomorphism screening using the split-count matrix is more efficient than most initial screening methods mentioned in Section 1.

Let $\xi(\cdot)$ be a sorting function that arranges, in descending order, the elements of a set of non-negative integers. For any non-empty column index $\mathbf{t} \subseteq \mathcal{T}$ and a count vector $\mathbf{N}$, let $\mathbf{n}_\mathbf{t}^+$ be the collection of $N_\mathbf{m}$'s, $\mathbf{m} \subseteq \mathcal{T}$ and $\|\mathbf{m} \cap \mathbf{t}\|$ is even, and $\mathbf{n}_\mathbf{t}^-$ be the collection of $N_\mathbf{m}$'s, $\mathbf{m} \subseteq \mathcal{T}$ and $\|\mathbf{m} \cap \mathbf{t}\|$ is odd. Notice that $\|\mathbf{n}_\mathbf{t}^+\| = \|\mathbf{n}_\mathbf{t}^-\|$ for any $\mathbf{t}$. We call $\mathbf{N}_\mathbf{t}^+ = \xi(\mathbf{n}_\mathbf{t}^+)$ the *positive split-count vector* of $\mathbf{t}$, and $\mathbf{N}_\mathbf{t}^- = \xi(\mathbf{n}_\mathbf{t}^-)$ the *negative split-count vector* of $\mathbf{t}$. Hereafter, they are referred to as the *split-count vectors* of $\mathbf{t}$ when there is no ambiguity. We list in Table 1 all the split-count vectors for the case $k = 3$.

Lemma 1. *Let $\mathcal{D}'$ be obtained from $\mathcal{D}$ by switching sign of factors in $\kappa = \{\kappa_1, \ldots, \kappa_g\}$, permuting columns via the permutation $\pi = (j_1, \ldots, j_k)$, and permuting some rows. For any column index $\mathbf{t} \subseteq \mathcal{T}$, let $\nu_\mathbf{t}^{\kappa,\pi} = \|\kappa \cap \mathbf{t}^\pi\|$, where $\mathbf{t}^\pi$ is as at* (2.12). *Then*

(i) $\mathbf{N}_\mathbf{t}^+(\mathcal{D}') = \mathbf{N}_{\mathbf{t}^\pi}^+(\mathcal{D})$, *and* $\mathbf{N}_\mathbf{t}^-(\mathcal{D}') = \mathbf{N}_{\mathbf{t}^\pi}^-(\mathcal{D})$, *if* $\nu_\mathbf{t}^{\kappa,\pi}$ *is even;*

(ii) $\mathbf{N}_\mathbf{t}^+(\mathcal{D}') = \mathbf{N}_{\mathbf{t}^\pi}^-(\mathcal{D})$, *and* $\mathbf{N}_\mathbf{t}^-(\mathcal{D}') = \mathbf{N}_{\mathbf{t}^\pi}^+(\mathcal{D})$, *if* $\nu_\mathbf{t}^{\kappa,\pi}$ *is odd.*

Proof. Because $N_\mathbf{m}(\mathcal{D}') = N_{\mathbf{m}^\pi \ominus \kappa}(\mathcal{D})$, Theorem 2, to prove Lemma 1 it is enough to show that when $\nu_\mathbf{t}^{\kappa,\pi}$ is even, $\|\mathbf{m} \cap \mathbf{t}\|$ is odd if and only if $\|(\mathbf{m}^\pi \ominus \kappa) \cap \mathbf{t}^\pi\|$ is odd, and when $\nu_\mathbf{t}^{\kappa,\pi}$ is odd, $\|\mathbf{m} \cap \mathbf{t}\|$ is even if and only if $\|(\mathbf{m}^\pi \ominus \kappa) \cap \mathbf{t}^\pi\|$ is odd. These follow immediately from

$$\begin{aligned}\|(\mathbf{m}^\pi \ominus \kappa) \cap \mathbf{t}^\pi\| &= \|\mathbf{m}^\pi \cap \mathbf{t}^\pi\| + \|\kappa \cap \mathbf{t}^\pi\| - 2\|\mathbf{m}^\pi \cap \kappa \cap \mathbf{t}^\pi\| \\ &= \|\mathbf{m} \cap \mathbf{t}\| + \nu_\mathbf{t}^{\kappa,\pi} - 2\|\mathbf{m}^\pi \cap \kappa \cap \mathbf{t}^\pi\|.\end{aligned}$$

Let $\mathbf{A}$ and $\mathbf{B}$ be vectors (of the same dimension) of non-negative integers. We say $\mathbf{A}$ is greater than $\mathbf{B}$ in *lexicographic order* (lex order, for short), denoted

$\mathbf{A} >_{lex} \mathbf{B}$, if in the vector difference $\mathbf{A} - \mathbf{B}$ the leftmost (or topmost) nonzero is positive. For example, $(2,1,0) >_{lex} (1,1,1)$ and $(2,1,1) >_{lex} (2,1,0)$. We say $\mathbf{A} \geq_{lex} \mathbf{B}$ if $\mathbf{A} >_{lex} \mathbf{B}$ or $\mathbf{A} = \mathbf{B}$.

For any non-empty column index $\mathbf{t} \subseteq \mathcal{T}$ and a count vector $\mathbf{N}$, let $\mathbf{N_t}$ be the $2^k \times 1$ vector such that

$$\mathbf{N_t} = \begin{cases} \begin{pmatrix} \mathbf{N_t^+} \\ \mathbf{N_t^-} \end{pmatrix}, \text{if } \mathbf{N_t^+} \geq_{lex} \mathbf{N_t^-}, \\ \begin{pmatrix} \mathbf{N_t^-} \\ \mathbf{N_t^+} \end{pmatrix}, \text{if } \mathbf{N_t^-} >_{lex} \mathbf{N_t^+}. \end{cases} \tag{3.1}$$

For $j = 1, \ldots, k$, sort the $\mathbf{N_t}$ with $\|\mathbf{t}\| = j$ in lex order and rename the sorted $\mathbf{N_t}$ as

$$\mathbf{N}^j_{(1)} \geq_{lex} \mathbf{N}^j_{(2)} \geq_{lex} \cdots \geq_{lex} \mathbf{N}^j_{(\binom{k}{j})}. \tag{3.2}$$

The split-count matrix, denoted as $\mathbf{N}^{sp}$, is then defined as

$$\mathbf{N}^{sp} = (\mathbf{N}^1_{(1)}, \ldots, \mathbf{N}^1_{(\binom{k}{1})}, \mathbf{N}^2_{(1)}, \ldots, \mathbf{N}^2_{(\binom{k}{2})}, \ldots, \mathbf{N}^k_{(\binom{k}{k})});$$

this combines the sorted $\mathbf{N_t}$'s by columns with j varying from 1 to k.

Example 3. (Example 1 cont.) For illustration, we set $r_i = i$ for $i = 1, \ldots, 8$. The split-count matrix of $\mathcal{A}_1$ is

$$\mathbf{N}^{sp}(\mathcal{A}_1) = \left(\begin{array}{ccc|ccc|c} \mathbf{N}_3 & \mathbf{N}_2 & \mathbf{N}_1 & \mathbf{N}_{23} & \mathbf{N}_{13} & \mathbf{N}_{12} & \mathbf{N}_{123} \\ 8 & 8 & 8 & 8 & 8 & 8 & 8 \\ 7 & 7 & 6 & 7 & 6 & 5 & 5 \\ 6 & 4 & 4 & 2 & 3 & 4 & 3 \\ 5 & 3 & 2 & 1 & 1 & 1 & 2 \\ \hline 4 & 6 & 7 & 6 & 7 & 7 & 7 \\ 3 & 5 & 5 & 5 & 5 & 6 & 6 \\ 2 & 2 & 3 & 4 & 4 & 3 & 4 \\ 1 & 1 & 1 & 3 & 2 & 2 & 1 \end{array}\right).$$

The first three columns are sorted $\mathbf{N_t}$ with $\|\mathbf{t}\| = 1$, columns 4 to 6 are sorted $\mathbf{N_t}$ with $\|\mathbf{t}\| = 2$, and the last column is $\mathbf{N}_{123}$. Each $\mathbf{N_t}$ is a combination of $\mathbf{N_t^+}$ and $\mathbf{N_t^-}$, with the one on top larger than the one on bottom in lex order.

For an initial screening measure for isomorphism examination, we have the following.

Theorem 3. *If two designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, then $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$.*

Proof. Suppose that $\mathcal{D}'$ can be obtained from $\mathcal{D}$ as stated in Lemma 1. From Lemma 1 and (3.1), we know that $\mathbf{N}_{\mathbf{t}}(\mathcal{D}') = \mathbf{N}_{\mathbf{t}^\pi}(\mathcal{D})$ for any non-empty $\mathbf{t} \subseteq \mathcal{T}$. Because $\|\mathbf{t}\| = \|\mathbf{t}^\pi\|$, the sorted $\mathbf{N}_{\mathbf{t}}(\mathcal{D}')$ with $\|\mathbf{t}\| = j$ portion of $\mathbf{N}^{sp}(\mathcal{D}')$ is identical to the sorted $\mathbf{N}_{\mathbf{t}^\pi}(\mathcal{D})$'s with $\|\mathbf{t}^\pi\| = j$ portion of $\mathbf{N}^{sp}(\mathcal{D})$ for $j = 1, \ldots, k$.

Theorem 3 indicates that no matter what operations of column permutation, row permutation, and sign switch are applied to a design matrix, its split-count matrix is invariant under the operations.

The split-count matrix is more powerful in isomorphism screening than most initial screening measures (except **HD**) mentioned in Section 1. This is shown in the next theorem. To prove the theorem, note that, given in Stufken and Tang (2007), for any column index $\mathbf{t} \subseteq \mathcal{T}$, the $\mathbf{t}$-th component of J-characteristics, denoted by $J_{\mathbf{t}}$, can be written as $J_{\mathbf{t}} = \sum_{\mathbf{m} \subseteq \mathcal{T}} h_{\mathbf{mt}} N_{\mathbf{m}}$. By (1.4) and the definition of split-count vectors, we can write

$$J_{\mathbf{t}} = \mathbf{1}_{2^{k-1}}^T \left(\mathbf{N}_{\mathbf{t}}^+ - \mathbf{N}_{\mathbf{t}}^-\right), \tag{3.3}$$

where $\mathbf{1}_{2^{k-1}}$ is the $2^{k-1} \times 1$ vector with all components one.

Theorem 4. *For designs $\mathcal{D}$ and $\mathcal{D}'$, the following hold.*

(i) *If $\mathbf{N}^{sp}(\mathcal{D}) = \mathbf{N}^{sp}(\mathcal{D}')$, then $CFV(\mathcal{D}) = CFV(\mathcal{D}')$.*

(ii) *If $CFV(\mathcal{D}) = CFV(\mathcal{D}')$, then $GWLP(\mathcal{D}) = GWLP(\mathcal{D}')$.*

(iii) *$GWLP(\mathcal{D}) = GWLP(\mathcal{D}')$ if and only if $K_u(\mathcal{D}) = K_u(\mathcal{D}')$ for $u = 1, \ldots, k$.*

(iv) *If $GWLP(\mathcal{D}) = GWLP(\mathcal{D}')$ (or $K_u(\mathcal{D}) = K_u(\mathcal{D}')$ for $u = 1, \ldots, k$), then $CD_2^2(\mathcal{D}) = CD_2^2(\mathcal{D}')$.*

Proof. For (i), it is enough to show that CFV is completely determined by $\mathbf{N}^{sp}$. For a k-factor design $\mathcal{D}$ with n runs, Deng and Tang (1999) defined CFV as

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

where $l_{i,j}$ is the number of $\mathbf{t}$'s such that $|J_{\mathbf{t}}| = j$ and $\|\mathbf{t}\| = i$. In $\mathbf{N}^{sp}(\mathcal{D})$, columns 1 to k contain $\mathbf{N}_{\mathbf{t}}(\mathcal{D})$ with $\|\mathbf{t}\| = 1$, where $\mathbf{N}_{\mathbf{t}}(\mathcal{D})$ is a combination of $\mathbf{N}_{\mathbf{t}}^+(\mathcal{D})$ and $\mathbf{N}_{\mathbf{t}}^-(\mathcal{D})$ as at (3.1). By (3.3), the vector $(l_{1,n}, \ldots, l_{1,1})$ in $CFV(\mathcal{D})$ is completely determined by the columns 1 to k in $\mathbf{N}^{sp}(\mathcal{D})$ because $\mathbf{N}_{\mathbf{t}}(\mathcal{D})$ uniquely determines $|J_{\mathbf{t}}|$. The same argument applies to the cases of $i = 2, \ldots, k$.

For (ii), it is enough to show that $GWLP$ is completely determined by CFV. Tang and Deng (1999) took $GWLP(\mathcal{D}) = (\alpha_1(\mathcal{D}), \ldots, \alpha_k(\mathcal{D}))$, where $\alpha_i(\mathcal{D}) = \sum_{\|\mathbf{t}\|=i} \left(\frac{J_{\mathbf{t}}}{n}\right)^2$, for $i = 1, \ldots, k$. The result (ii) holds because α_i, $i = 1, \ldots, k$, can be expressed as a function of $l_{i,j}$'s as $\alpha_i(\mathcal{D}) = \sum_{j=1}^n l_{i,j} \left(\frac{j}{n}\right)^2$.

For (iii), an equation given by Xu (2003, p.696) shows the relationship between $GWLP$ and K_u, $u = 1, \ldots, k$:

$$K_u(\mathcal{D}) = c_u\alpha_u(\mathcal{D}) + c_{u-1}\alpha_{u-1}(\mathcal{D}) + \cdots + c_1\alpha_1(\mathcal{D}) + c_0 - C,$$

where $c_i = c_i(u; n, k, s) = [n/(n-1)] \sum_{m=0}^{u} (-1)^{m+i} \binom{u}{m} k^{u-m} [\sum_{j=0}^{m} j! S(m,j) s^{-j} (s-1)^{j-i} \binom{k-i}{j-i}]$, $C = k^u/(n-1)$, and the $S(m,j)$ are Stirling numbers of the second kind. It is obvious that the K_u can be uniquely determined by $GWLP$, and vice versa.

The result (iv) follows directly from, Ye (2003, p.992),

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

which shows that $GWLP$ uniquely determines CD_2^2.

The theorem immediately provides, as a corollary, the rank order of these measures on classification power. The proof of the corollary is straightforward and is thus omitted here.

Corollary 1. *For an initial screening measure M and fixed n and k, let $\#(M)$ be the number of non-isomorphic groups distinguished by M. Then,*

$$\#(\mathbf{N}^{sp}) \geq \#(CFV) \geq \#(GWLP) = \#(K_u) \geq \#(CD_2^2).$$

Suppose that we define the screening efficiency by

$$eff(M) = \frac{\#(M)}{\text{total number of non-isomorphic classes}}.$$

According to Corollary 1, the screening efficiencies of the measures mentioned in the corollary can be ranked as

$$1 \geq eff(\mathbf{N}^{sp}) \geq eff(CFV) \geq eff(GWLP) = eff(K_u) \geq eff(CD_2^2) > 0.$$

For the efficiency comparison between **HD** and the split-count matrix, neither is superior to the other. As will be shown in Example 6 in Section 5, the split-count matrix has higher efficiencies than **HD** in most instances, but there are cases where **HD** has better classification power.

4. Projection and Simplified Methods

In isomorphism classification using initial screening measures, projection is a widely employed technique to improve efficiency. For a k-factor design and an integer $p \leq k$, there are $\binom{k}{p}$ p-factor projected designs. We can apply an initial screening measure M on each of the $\binom{k}{p}$ projected designs and obtain $\binom{k}{p}$ M-measures. An M-measure could be a value (such as CD_2^2), a vector (such as CFV, $GWLP$, or K_u), or a matrix (such as $\mathbf{HD}$ and $\mathbf{N}^{sp}$). The frequency of these M-measures is referred to as *p-dimensional projection frequency* of M. We call the collection of all p-dimensional projection frequencies, $p = 1, \ldots, k$, the *complete projection frequency* of M, and denote it by P_M. If two designs are isomorphic, their p-dimensional projection frequencies are identical for any p. Therefore, P_M can be used as an initial screening measure. The classification power of P_M is at least as high as M, i.e.,

$$\#(P_M) \geq \#(M) \text{ and } eff(P_M) \geq eff(M).$$

For the projection versions of the initial screening measures that appeared in Corollary 1, the ranking remains unchanged, i.e.,

$$\#(P_{\mathbf{N}^{sp}}) \geq \#(P_{CFV}) \geq \#(P_{GWLP}) = \#(P_{K_u}) \geq \#(P_{CD_2^2}),$$

$$1 \geq eff(P_{\mathbf{N}^{sp}}) \geq eff(P_{CFV}) \geq eff(P_{GWLP}) = eff(P_{K_u}) \geq eff(P_{CD_2^2}) > 0.$$

For a k-factor design, the dimension of its split-count matrix is $2^k \times (2^k - 1)$. Dimension increases dramatically when k becomes large. In the remainder of this section, some lower dimensional measures based on the split-count matrix are proposed for the isomorphism classification of designs with a large k.

Let us sum up the split-count vectors over $\mathbf{t}$'s with $\|\mathbf{t}\| = j$, i.e.,

$$\mathbf{SN}^j = \sum_{\|\mathbf{t}\|=j} (\mathbf{N}_{\mathbf{t}}^+ + \mathbf{N}_{\mathbf{t}}^-),$$

for $j = 1, \ldots, k$, and define the *sum of the split-count matrix* as

$$\mathbf{SN}^{sp} = (\mathbf{SN}^1, \mathbf{SN}^2, \ldots, \mathbf{SN}^k).$$

Take the design $\mathcal{A}_1$ in Example 3 as an example, its sum of the split-count matrix is

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

According to Theorem 3 and the property that $\mathbf{SN}^{sp}$ is determined by $\mathbf{N}^{sp}$, when designs $\mathcal{D}$ and $\mathcal{D}'$ are isomorphic, $\mathbf{SN}^{sp}(\mathcal{D}) = \mathbf{SN}^{sp}(\mathcal{D}')$. This implies that $\mathbf{SN}^{sp}$ can be used as an initial screening measure. The measure $\mathbf{SN}^{sp}$ reduces the dimension of $\mathbf{N}^{sp}$ from $2^k \times (2^k - 1)$ to $2^{k-1} \times k$. It is clear that the screening efficiency of $\mathbf{SN}^{sp}$ is not higher than that of $\mathbf{N}^{sp}$ but, as will be shown in Example 6 in Section 5, in many cases their efficiencies are either equal or very close.

For a highly fractional factorial design, i.e., a k-factor design with n distinct runs, where n is much smaller than 2^k, a large proportion of the components in its count vector are zero. These 0's appear in the bottom of the $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ because the components in $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ are sorted from large to small, causing the split-count matrix to contain many rows of 0's. These rows can be removed to reduce dimensionality. Because $\mathbf{N}_\mathbf{t}^+$'s and $\mathbf{N}_\mathbf{t}^-$'s contain at least $2^{k-1}-n$ zeros, in the split-count matrix the $(n+1)$-th to the (2^{k-1})-th rows and the $(2^{k-1}+n+1)$-th to the (2^k)-th rows are zero. We can at least remove these 2^k-2n rows to reduce the dimension of $\mathbf{N}^{sp}$ from $2^k \times (2^k-1)$ to $2n \times (2^k-1)$, and dimension can be further reduced if there are more rows of 0's in $\mathbf{N}^{sp}$. This technique can also be applied to $\mathbf{SN}^{sp}$ to at least reduce its dimension from $2^{k-1} \times k$ to $n \times k$.

For a design with single replicate (i.e., the components in its count vector are either 1 or 0), its $\mathbf{N}^{sp}$ can be uniquely determined by its CFV. Then, the isomorphism screening based on $\mathbf{N}^{sp}$ is equivalent to that based on CFV. Take, as an example, a k-factor design whose count vector contains n ones and 2^k-n zeros. Note that $l_{i,j}$ in its CFV is the number of $\mathbf{t}$'s with $|J_\mathbf{t}|=j$ and $\|\mathbf{t}\|=i$. Among $\mathbf{t}$'s with $\|\mathbf{t}\|=i$, there are $l_{i,j}$ of them with $|J_\mathbf{t}|=j$, $j=1,\ldots,n$, and $\binom{n}{i}-\sum_{j=1}^{n} l_{i,j}$, denoted by $l_{i,0}$, with $|J_\mathbf{t}|=0$. From (3.3), we know that for any of $l_{i,j}$ different $\mathbf{t}$'s with $|J_\mathbf{t}|=j$, $j=0,1,\ldots,n$, the absolute difference between the numbers of 1's in its corresponding $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ is j. Together with the condition that the sum of the numbers of 1's in $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ is n, we know that one of $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ must have $(n+j)/2$ ones on top followed by $2^{k-1}-(n+j)/2$ zeros, and the other with $(n-j)/2$ ones on top followed by $2^{k-1}-(n-j)/2$ zeros. Among all $\mathbf{N}_\mathbf{t}$ with $\|\mathbf{t}\|=i$, which are obtained by sorting $\mathbf{N}_\mathbf{t}^+$ and $\mathbf{N}_\mathbf{t}^-$ in lex order as presented in (3.1), $l_{i,j}$ of them, $j=0,1,\ldots,n$, are the vector whose components from top to bottom are $(n+j)/2$ ones, $2^{k-1}-(n+j)/2$ zeros, $(n-j)/2$ ones, and $2^{k-1}-(n-j)/2$ zeros. Therefore, CFV can fully determine $\mathbf{N}^{sp}$. Because CFV has a much lower dimension than $\mathbf{N}^{sp}$ when k is large, the former should replace the latter in the initial screening of designs with single replicate.

The above discussion also points out a relationship between the strength of an orthogonal array and its $\mathbf{N}^{sp}$. For an n-run orthogonal array of strength s, all its $J_\mathbf{t}$ with $\|\mathbf{t}\| \leq s$ are zero. Therefore, the first $\binom{n}{1}+\binom{n}{2}+\cdots+\binom{n}{s}$ columns in $\mathbf{N}^{sp}$ are the vector whose components from top to bottom are $n/2$ ones, $2^{k-1}-n/2$ zeros, $n/2$ ones, and $2^{k-1}-n/2$ zeros. There is no need to compute and compare these columns for orthogonal arrays of strength s.

5. Some Comparisons

We use examples to study and compare the classification power of the initial screening measures mentioned in this paper. The measures are $\mathbf{N}^{sp}$, $\mathbf{SN}^{sp}$, $\mathbf{HD}$, CFV, $GWLP$, K_u, and CD_2^2, and their projection versions $P_{\mathbf{N}^{sp}}$, $P_{\mathbf{SN}^{sp}}$, $P_{\mathbf{HD}}$, P_{CFV}, P_{GWLP}, P_{K_u}, and $P_{CD_2^2}$, respectively. The computational time for each method is shown in Tables 2 and 3 (processor: 2.4 GHz Intel Core 2 Duo; memory: 4GB 1067 MHz DDR3; code: R). In this section, we use $OA(n,k,s)$ to denote n-run orthogonal arrays of strength s for k factors each at 2 levels.

Example 4. Katsaounis and Dean (2008) gave a classification example of non-isomorphic 4-factor designs. The designs, they denote by $df1$ and $df5$, are not mean-orthogonal, i.e., their strengths are less than one. Among initial screening measures discussed in their paper only $Deseq1$, proposed by Clark and Dean (2001) and which can be regarded a simpler projection version of $\mathbf{HD}$, can classify $df1$ and $df5$ as non-isomorphic. The two designs are represented in terms of the count vector as

$$\mathbf{N}(df1) = (1,0,1,1,1,3,2,1,1,1,0,2,0,0,0,2)^T,$$
$$\mathbf{N}(df5) = (2,0,1,2,1,0,1,0,1,0,3,1,1,2,0,1)^T.$$

The split-count matrices are

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

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

Because $\mathbf{N}^{sp}(df1)$ and $\mathbf{N}^{sp}(df5)$ are different (e.g., check their second columns), the split-count matrix can quickly identify the two designs as non-isomorphic. Their sum of split-count 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 $\mathbf{SN}^{sp}(df1)$ and $\mathbf{SN}^{sp}(df5)$ are different (e.g., check their first columns), $\mathbf{SN}^{sp}$ can also classify them as non-isomorphic. It takes less than one second for $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ to distinguish the designs. This example demonstrates how to use $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ to perform isomorphism examination and shows that, although most initial screening measures fail to distinguish them, the methods based on the count vector work.

Table 2. Results of isomorphism classification for the eight non-isomorphic designs given in Example 5.

Methods	Classification results	Computation time (sec.)
$\mathbf{N}^{sp}$	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7\},\{\mathcal{B}_8\}$	1.6
$\mathbf{SN}^{sp}$	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7\},\{\mathcal{B}_8\}$	1.1
$\mathbf{HD}$	$\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$	0.6
CFV	$\{\mathcal{B}_1,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6\},\{\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_7,\mathcal{B}_8\}$	1.4
$GWLP$	$\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$	1.3
K_u	$\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$	1.8
CD_2^2	$\{\mathcal{B}_1,\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6,\mathcal{B}_7,\mathcal{B}_8\}$	1.5
$P_{\mathbf{HD}}$	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$	18.7
P_{CFV}	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$	42.0
P_{GWLP}	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$	40.4
P_{K_u}	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$	57.0
$P_{CD_2^2}$	$\{\mathcal{B}_1\},\{\mathcal{B}_2\},\{\mathcal{B}_3\},\{\mathcal{B}_4\},\{\mathcal{B}_5\},\{\mathcal{B}_6\},\{\mathcal{B}_7,\mathcal{B}_8\}$	47.2

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

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

The classification results of the eight designs, together with the computation time (in seconds) under different initial screening measures, are presented in Table 2. The first part of the table contains the classification results of $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$. They successfully classify the eight designs as non-isomorphic, even though the projection technique is not employed. The second part contains the classification results of $\mathbf{HD}$, CFV, $GWLP$, K_u, and CD_2^2. The measure CFV can only classify these designs as two groups $\{\mathcal{B}_1,\mathcal{B}_4,\mathcal{B}_5,\mathcal{B}_6\}$ and $\{\mathcal{B}_2,\mathcal{B}_3,\mathcal{B}_7,\mathcal{B}_8\}$, and the other measures fail to distinguish between any of the eight non-isomorphic designs. The classification results of the projection versions of $\mathbf{HD}$, CFV, $GWLP$, K_u, and CD_2^2 are given in the third part of Table 2. The technique of projection significantly improves the classification power of these measures, but they still cannot distinguish the designs $\mathcal{B}_7$ and $\mathcal{B}_8$.

Example 6. Although Corollary 1 shows that $\mathbf{N}^{sp}$ can achieve higher screening efficiency than CFV, $GWLP$, CD_2^2, and K_u, it is still worthwhile illustrating how well the measures based on the count vector perform compared to other initial screening measures. In this example, we provide a more thorough comparison. We investigate the screening efficiencies of the measures for various OA's. Sun, Li, and Ye (2002) provides a complete catalog of $OA(n, k, 2)$'s for $n = 12, 16$, and 20. We study the screening efficiencies for all the OA's given in their paper with $k \leq 6$. For the cases of OA's with $k = 4$ or $n = 12$, all measures can fully classify all non-isomorphic classes. For the other cases, screening efficiencies are reported in Table 3. We also include in Table 3 some OA's with $n > 20$ or $s > 2$ as given in Lin and Cheng (2011). In the table, the notation $OA(n, k, s) : w$ indicates that there are w non-isomorphic classes for $OA(n, k, s)$. For each case of the OA's in Table 3, the screening efficiencies and the computation time (in seconds) are given on the left-hand side and the right-hand side of the colon individually. The results for the measures not adopting and adopting projection are arranged on the top and bottom of the table individually.

For the OA's listed in Table 3, only the projection version of $\mathbf{N}^{sp}$ can successfully distinguish all non-isomorphic classes, i.e., reach 100% efficiency. For the rest of the measures, those based on the count vector generally perform better than the others. For example, $P_{\mathbf{N}^{sp}}$ and $P_{\mathbf{SN}^{sp}}$ have equal or higher efficiencies than other projection versions in every cases, and strictly higher in all 6 cases with larger w's ($w > 30$). Actually in many cases $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$, which do not utilize projection, already have better performance (with efficiencies at least 96%) than the projection versions of those measures not based on the count vector. For $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$, although $eff(\mathbf{SN}^{sp})$ is always no more than $eff(\mathbf{N}^{sp})$, we can see that their efficiencies are very close. They have identical efficiencies in most cases, and the largest efficiency difference between them is 2.3% (in the case of $OA(32, 5, 2)$). This closeness is more apparent in their projection versions, $P_{\mathbf{N}^{sp}}$ and $P_{\mathbf{SN}^{sp}}$. For the measures not based on the count vector, the use of projection can significantly improve efficiency. For example, in the case of $OA(36, 5, 2)$, the efficiencies of $\mathbf{HD}$, CFV, $GWLP$, K_u, and CD_2^2 are improved by from 28% to 44.9% after projection is applied. The improvement achieved by the use of projection is not so significant for $\mathbf{N}^{sp}$ and $\mathbf{SN}^{sp}$ because their efficiencies (no projection versions) are already high enough in most cases. Among the measures not based on the count vector, $P_{\mathbf{HD}}$ has the best performance. In the cases of $OA(20, 6, 2)$, $\mathbf{HD}$ has higher efficiency than $\mathbf{N}^{sp}$ although the situation is reversed in their projection versions. Neither $\mathbf{HD}$ nor $\mathbf{N}^{sp}$ is entirely superior to the other, as mentioned in Section 3.

6. Summary

In this paper, the count vector approach is employed for isomorphism examination. A count vector can be regarded as an alternative representation of

Table 3. Efficiency comparison of initial screening measures.

Method	$OA(16,5,2)^\flat : 11$	$OA(20,5,2)^\flat : 11$	$OA(24,5,2)^\natural : 63$	$OA(28,5,2)^\natural : 127$
$\mathbf{N}^{sp}$	100.0 % : 1.5 (sec.)	100.0 % : 1.5	100.0 % : 9.6	98.4 % : 21.0
$\mathbf{SN}^{sp}$	100.0 % : 1.5	100.0 % : 1.7	100.0 % : 8.2	98.4 % : 21.4
HD	90.9 % : 1.5	100.0 % : 0.7	74.6 % : 25.4	78.7 % : 46.2
CFV	100.0 % : 1.5	90.9 % : 2.4	77.8 % : 28.7	46.5 % : 145.9
$GWLP$	90.9 % : 2.3	90.9 % : 2.4	58.7 % : 51.7	40.2 % : 169.3
K_u	90.9 % : 3.3	90.9 % : 3.3	58.7 % : 148.5	40.2 % : 180.0
CD_2^2	90.9 % : 2.6	90.9 % : 2.6	58.7 % : 144.4	40.2 % : 171.1
$P_{\mathbf{N}^{sp}}$	100.0 % : 47.1	100.0 % : 45.3	100.0 % : 297.9	100.0 % : 595.5
$P_{\mathbf{SN}^{sp}}$	100.0 % : 45.4	100.0 % : 52.2	100.0 % : 254.9	100.0 % : 607.4
$P_{\mathbf{HD}}$	100.0 % : 19.5	100.0 % : 20.8	96.8 % : 106.4	93.7 % : 272.1
P_{CFV}	100.0 % : 46.4	90.9 % : 46.1	92.1 % : 270.1	69.3 % : 555.0
P_{GWLP}	100.0 % : 44.0	90.9 % : 47.7	92.1 % : 301.3	69.3 % : 570.6
P_{K_u}	100.0 % : 74.7	90.9 % : 76.3	92.1 % : 428.3	69.3 % : 901.3
$P_{CD_2^2}$	100.0 % : 52.5	90.9 % : 53.1	92.1 % : 299.4	69.3 % : 625.7
	$OA(32,5,2)^\natural : 491$	$OA(36,5,2)^\natural : 1242$	$OA(16,6,2)^\flat : 27$	$OA(20,6,2)^\flat : 75$
$\mathbf{N}^{sp}$	99.2 % : 71.5	99.3 % : 236.8	96.3 % : 38.9	68.0 % : 1562.6
$\mathbf{SN}^{sp}$	96.9 % : 81.9	98.4 % : 199.4	96.3 % : 38.9	68.0 % : 1560.8
HD	67.0 % : 328.8	67.0 % : 882.6	63.0 % : 600.7	69.3 % : 1484.3
CFV	45.2 % : 877.2	23.1 % : 5561.8	96.3 % : 38.7	56.0 % : 2158.4
$GWLP$	26.1 % : 1825.6	13.8 % : 9461.2	63.0 % : 602.4	56.0 % : 2158.5
K_u	26.1 % : 919.9	13.8 % : 9563.0	63.0 % : 605.0	56.0 % : 2164.7
CD_2^2	26.1 % : 886.5	13.8 % : 9479.8	63.0 % : 602.7	56.0 % : 2159.4
$P_{\mathbf{N}^{sp}}$	100.0 % : 2101.5	100.0 % : 7085.4	100.0 % : 229.6	100.0 % : 786.2
$P_{\mathbf{SN}^{sp}}$	99.2 % : 2029.1	99.9 % : 5584.8	100.0 % : 228.3	100.0 % : 673.0
$P_{\mathbf{HD}}$	97.1 % : 1155.5	95.0 % : 2717.3	100.0 % : 111.0	96.0 % : 435.9
P_{CFV}	75.6 % : 2189.0	58.7 % : 5885.6	100.0 % : 218.1	92.0 % : 839.3
P_{GWLP}	75.6 % : 2249.9	58.7 % : 5847.1	100.0 % : 218.2	92.0 % : 843.9
P_{K_u}	75.6 % : 3513.3	58.7 % : 9004.3	100.0 % : 386.6	92.0 % : 1236.3
$P_{CD_2^2}$	75.6 % : 2478.4	58.7 % : 6424.7	100.0 % : 242.0	92.0 % : 870.2
	$OA(32,6,3)^\natural : 10$	$OA(40,6,3)^\natural : 9$	$OA(48,6,3)^\natural : 45$	$OA(32,7,3)^\natural : 17$
$\mathbf{N}^{sp}$	100.0 % : 2.2	100.0 % : 1.5	97.8 % : 42.4	100.0 % : 2.9
$\mathbf{SN}^{sp}$	100.0 % : 1.5	100.0 % : 1.4	97.8 % : 42.1	100.0 % : 2.4
HD	90.0 % : 36.0	100.0 % : 0.7	71.1 % : 637.7	64.7 % : $> 10^5$
CFV	100.0 % : 1.3	88.9 % : 36.4	82.2 % : 499.2	100.0 % : 2.1
$GWLP$	90.0 % : 36.5	88.9 % : 36.4	60.0 % : 1168.3	64.7 % : $> 10^5$
K_u	90.0 % : 37.3	88.9 % : 37.1	60.0 % : 1172.0	64.7 % : $> 10^5$
CD_2^2	90.0 % : 36.6	88.9 % : 36.5	60.0 % : 1169.2	64.7 % : $> 10^5$
$P_{\mathbf{N}^{sp}}$	100.0 % : 137.2	100.0 % : 97.0	100.0 % : 453.7	100.0 % : 365.8
$P_{\mathbf{SN}^{sp}}$	100.0 % : 95.3	100.0 % : 87.8	100.0 % : 430.2	100.0 % : 310.8
$P_{\mathbf{HD}}$	100.0 % : 48.8	100.0 % : 44.6	93.3 % : 328.0	100.0 % : 120.9
P_{CFV}	100.0 % : 81.5	88.9 % : 107.0	88.9 % : 660.7	100.0 % : 269.7
P_{GWLP}	100.0 % : 79.6	88.9 % : 106.5	88.9 % : 643.2	100.0 % : 265.4
P_{K_u}	100.0 % : 130.0	88.9 % : 153.4	88.9 % : 875.4	100.0 % : 440.4
$P_{CD_2^2}$	100.0 % : 87.1	88.9 % : 113.9	88.9 % : 701.2	100.0 % : 289.3

$\flat$: OA's given in Sun, Li, and Ye (2002).

$\natural$: OA's given in Lin and Cheng (2011).

design matrix and, although the idea of the count vector has been adopted in several papers, the focus of the previous work was primarily on topics concerning the J-characteristics expression of the count vector (or, in the terminology of indicator function, the coefficients of monomials). We show that the count vector itself can be a useful tool in the study of design properties. For isomorphism examination, we provide necessary and sufficient conditions for two count vectors to be isomorphic. The operations of column, row, and level permutations on design matrix cause a systematic rearrangement of elements in the count vector. We identify the pattern of the rearrangement and characterize it in two different ways, linear transformation of vectors and set operations of subscripts. For the faster initial screening based on the count vector, we propose several measures and prove that they are invariant to column, row, and level permutations. We also prove or illustrate by examples that the initial screening measures based on the count vector generally have satisfactory classification power and better screening efficiency.

Acknowledgements

The research of C.-Y. Lin was supported in part by National Institutes of Health (NIH) P41 HG004222-01. The research of S.-W. Cheng was supported by the National Science Council of Taiwan, and the National Center for Theoretical Science of Taiwan. The authors are grateful to the Editor, an associate editor, and two referees for helpful comments.

References

Cheng, S.-W. and Ye, K. Q. (2004). Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. *Ann. Statist.* **32**, 2168-2185.

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

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

Fontana, R., Pistone, G. and Rogantin, M. P. (2000). Classification of two-level factorial fractions. *J. Statist. Plann. Inference* **87**, 149-172.

Katsaounis, T. I. and Dean, A. M. (2008). A survey and evaluation of methods for determination of combinatorial equivalence of factorial designs. *J. Statist. Plann. Inference* **138**, 245-258.

Lin, C.-Y. and Cheng, S.-W. (2011). Design enumeration through assembling count functions. Preprint.

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

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

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.

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

Tang, B. and Deng, L.-Y. (1999). Minimum G_2-aberration for nonregular fractional factorial designs. *Ann. Statist.* **27**, 1914-1926.

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

Ye, K. Q. (2003). Indicator function and its application in two-level factorial designs. *Ann. Statist.* **31**, 984-994.

McDermott Center of Human Growth and Development and Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX, U.S.A.

E-mail: Chang-Yun.Lin@utsouthwestern.edu

Institute of Statistics, National Tsing-Hua University , Hsinchu, Taiwan.

E-mail: swcheng@stat.nthu.edu.tw

(Received October 2010; accepted July 2011)