Statistica Sinica Preprint No: SS-2022-0200

Title	A More Efficient Isomorphism Check for Two-Level Nonregular Designs
Manuscript ID	SS-2022-0200
URL	http://www.stat.sinica.edu.tw/statistica/
DOI	10.5705/ss.202022.0200
Complete List of Authors	Chunyan Wang and Robert W. Mee
Corresponding Authors	Chunyan Wang
E-mails	2120150092@mail.nankai.edu.cn

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0200

A MORE EFFICIENT ISOMORPHISM CHECK FOR TWO-LEVEL NONREGULAR DESIGNS

Chunyan Wang[1] and Robert W. Mee[2]

[1] *Renmin University of China and* [2] *University of Tennessee*

Abstract: In this paper, we propose some new necessary and sufficient conditions for identifying isomorphism in two-level fractional factorial designs, using a parallel flats structure. A new algorithm for checking isomorphism is provided accordingly. The proposed algorithm is simple and general, and can be used for either regular or nonregular designs. By taking advantage of the parallel flats structure when it exists, the method is much faster than current methods for assessing the isomorphism of nonregular two-level designs. Examples are given to illustrate the results. An efficient implementation of the proposed algorithm in Matlab can be found in the online Supplementary Material.

Key words and phrases: Algorithm, equivalent group, parallel flats, two-level fractional factorial.

1. Introduction

In this study, we restrict our attention to two-level fractional factorial designs, which are extremely popular screening designs. Two fractional factorial designs are called isomorphic if and only if one design can be obtained from the other by row permutations, column permutations, and level permutations within columns. Two

Corresponding author: Chunyan Wang, Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China. E-mail: chunyanwang@ruc.edu.cn.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0200

isomorphic designs share the same statistical properties in some classical ANOVA models, and thus are considered essentially the same. Thus, determining isomorphism is important both in theory and in practice.

Given the simple algebraic structure of regular designs, the earliest studies on isomorphism checks focused on such designs. Draper and Mitchell (1967, 1968) proved that two isomorphic regular designs must have the same word length pattern. Draper and Mitchell (1970) further showed that two isomorphic regular designs must have the same letter pattern. The letter pattern counts the frequency of letters in words of different lengths. Agreement in letter pattern implies having the same word length pattern. Note that designs with the same letter pattern are not necessarily isomorphic; see Chen and Lin (1991), who disprove a conjecture of Draper and Mitchell (1970). Chen, Sun, and Wu (1993) first proposed necessary and sufficient conditions by "applying some algebraic and combinatorial methods" to identify isomorphic regular designs. By matching the factors using their delete-one-factor projections, Xu (2009) greatly improved the isomorphism checking procedure of Chen, Sun, and Wu (1993), and developed a sequential algorithm for constructing efficient two-level regular designs. Shrivastava and Ding (2010) provided a new approach for testing the isomorphism of two-level regular designs by modeling them as simple bipartite graphs. Liu, Yang, and Liu (2011) proposed the three-dimensional letter interaction pattern matrix (LIPM), showing that it can uniquely determine a design, and thus be an efficient tool for checking isomorphism.

For isomorphism in general two-level fractional factorial designs (which can be either regular or nonregular), Ma, Fang, and Lin (2001) introduced the NIU algorithm based on the centered L_2-discrepancy, and Fang and Zhang (2004) proposed the minimum aberration majorization criterion based on the generalized word length pattern. However, the centered L_2-discrepancy and generalized word length pattern do not uniquely determine an isomorphism class, and so can only be used for initial screening for non-isomorphism. Clark and Dean (2001) were the first to present necessary and sufficient conditions for any two designs to be isomorphic, using the Hamming distance matrices of their projection designs and providing a checking algorithm. Beyond the Hamming distance matrices, Ye (2003), Cheng and Ye (2004), and Pang and Liu (2011) developed other necessary and sufficient conditions, as well as algorithms for checking isomorphism using indicator functions. Lin and Cheng (2012) proposed several efficient initial screening methods for distinguishing designs based on the count vector. They proved that their split-count matrix N^{sp} is more efficient than initial screening methods based on $CFV, GWLP, K_u$, and CD_2^2. Further details about these measures can be found in Deng and Tang (1999), Tang and Deng (1999), Xu (2003), and Ma, Fang, and Lin (2001). For other developments related to design isomorphism and complete enumeration results, see Stufken and Tang (2007), Sun, Li, and Ye (2008), Shrivastava and Ding (2010), Schoen, Eendebak, and Nguyen (2010), Ke et al. (2023), and Weng, Fang, and Elsawah (2023).

Parallel flats designs (PFDs), introduced in Connor and Young (1961), are a class of nonregular designs that retain some of the simplicity of regular fractional factorial designs. PFDs have received widespread attention, because they enjoy many desirable properties; see, for example, Srivastava and Li (1996), Liao et al. (1996), Srivastava and Chopra (1973), and Jones et al. (2019). More recently, Wang and Mee (2021) give a comprehensive review of two-level PFDs and develop a general theory. Edwards and Mee (2023) systematically study the structure of nonregular two-level designs, and connect the block diagonal information matrix of nonregular designs to the parallel flats structure.

This study aims to provide some new necessary and sufficient conditions for identifying isomorphism in two-level fractional factorial designs, by incorporating the parallel flats structure of two-level nonregular designs. A new algorithm for checking isomorphism is provided that is simple and applicable to both regular and nonregular designs. For nonregular designs with a parallel flats structure, the method is much faster than existing methods. Two examples are given to illustrate the results.

The rest of the paper is organized as follows. Section 2 gives some preliminary notation. In Section 3, we propose some necessary and sufficient conditions for identifying isomorphism in two-level fractional factorial designs. In Section 4, we propose a new algorithm for checking isomorphism and present several examples. Section 5 concludes the paper.

2. Preliminary notation and results

A regular 2^{n-p} design is also known as a single flat design, in the sense that it consists of all treatment combinations $x = (x_1, x_2, \ldots, x_n)$ that satisfy the equation $A \odot x = c$, where $A = (a_{ij})$ is a $p \times n$ alias matrix over GF[2] of rank p (GF[2], the Galois Field of order 2, is a finite field consisting of two elements, where the operations of addition and multiplication are performed over the set {0,1} modulo 2 (Mukerjee and Wu, 2006, p.18), c is a $p \times 1$ vector with levels ± 1, and $A \odot x$ is defined as the $p \times 1$ vector with ith element $x_1^{a_{i1}} \cdots x_n^{a_{in}}$. For given A, one has 2^p different options for the vector c, corresponding to the 2^p disjoint single flats; the full 2^n design is the concatenation of these. Taking f single flats corresponding to $C = [c_1, \ldots, c_f]$, we obtain a PFD with f flats (f-PFD). Thus, an f-PFD is determined by the pair (A, C).

For even f, the matrix C can sometimes be reduced, such that the design is an $(f/2)$-PFD composed of flats of size $2^{n-(p-1)}$. If an f-PFD cannot be reduced in this way, it is said to be of minimal form (Edwards and Mee, 2023); without loss of generality, we assume the f-PFD is of this form. Following Edwards and Mee (2023), a two-level design defined by (A, C) with N runs and n factors is an f-PFD with $1 \leq f \leq N$. When $f = 1$, it is a regular design, and any 2-PFD reduces to $f = 1$. When $f = N$, it is an N-PFD composed of flats of size 1, with $p = n$; in this case, there is no special structure, because an N-PFD is determined by an (A, C) pair, where A is an identity matrix of order n and C is a transpose of the design. Cheng (2014, p.139) excludes the case of $f = N$, considering only $p < n$. The best gains in our isomorphism check occur for f-PFDs with $3 < f \leq N/2$.

Some new necessary and sufficient conditions for identifying isomorphism designs can be proposed based on their (A, C) pairs, and a new algorithm for checking isomorphism is provided accordingly. Using a parallel flats structure, the proposed algorithm can greatly reduce the computational effort compared with that of existing algorithms. This is the subject and motivation of our study.

3. Necessary and sufficient conditions for isomorphism

Benefiting from the parallel flats structure, in this section, we propose several necessary and sufficient conditions for identifying isomorphism in two-level fractional factorial designs.

Let D_1, D_2 be two 2-level designs with N runs and n factors. Without loss of generality, suppose that the first row of both designs consist entirely of $+1$. Then, for $i = 1, 2$, D_i must be determined by (A_i, C_i) with $N = f_i \times 2^{n-p_i}$, for some f_i and p_i. There is a column consisting entirely of $+1$ in C_i, and we denote the single flat corresponding to this column as D_{0i}. We focus only on the case $f_1 = f_2$; otherwise, the two designs are non-isomorphic.

Following Wang and Mee (2021), two f-PFDs are called equivalent if one f-PFD can be obtained from the other by row permutations and column sign switches. Thus, equivalent designs must be isomorphic. The following lemma is obvious.

Lemma 1. *If D_1 and D_2 are isomorphic, there exists a column permutation to make them equivalent.*

Then, we have the following result, which is taken from Wang and Mee (2021).

Proposition 1. *If $A_1 = A_2$, then D_1 and D_2 are equivalent if and only if C_1 and C_2 belong to the same group. Therein, the group of C_i is*

$$G_{C_i} = \{C_{ij} \circ C_i : j = 1, 2, \ldots, f\}, \text{ with } C_{ij} \circ C_i = \{C_{ij} * C_{i1}, \ldots, C_{ij} * C_{if}\}, \quad (3.1)$$

where C_{ij} represents the jth column of C_i, for $i = 1, 2$ and $j = 1, \ldots, f$, and $\alpha * \beta = (\alpha_1\beta_1, \ldots, \alpha_f\beta_f)^T$, for any two column vectors $\alpha = (\alpha_1, \ldots, \alpha_f)^T$ and $\beta = (\beta_1, \ldots, \beta_f)^T$.

In Proposition 1, the two f-PFDs are assumed to have the same A matrix. However, two equivalent designs can have different A matrices, because the (A, C) pair representing a PFD is not unique. Now, we propose a general theory for checking the equivalence of two f-PFDs.

Theorem 1. *Let D_1, D_2 be two f-PFDs with N runs and n factors, where D_i is determined by (A_i, C_i), both of minimal form, with $N = f \times 2^{n-p}$, for $i = 1, 2$. Then, D_1 and D_2 are equivalent if and only if (i) the row spaces of A_1 and A_2 are equal, and (ii) when (A_2, C_2) is re-expressed as (A_2', C_2'), so that $A_2' = A_1$, the corresponding C_2' belongs to the same group as C_1.*

Proof. We first consider the sufficiency of the conditions. If the row spaces of A_1 and A_2 are the same, then D_{01} and D_{02} are equal up to row permutations, and hence are equivalent. Thus, we can choose A_2' to equal A_1. By condition (ii), C_2' is in the same group as C_1, so by Wang and Mee (2021, Theorem 1), D_1 and D_2 are equivalent.

Next, we prove the necessity, in two parts. First, if (i) does not hold, then D_{01} and D_{02} are nonequivalent 2^{n-p} designs. They are based on at least one different generator, and thus must have at least 2^{p-1} different defining words. Let W_1 be the defining words for D_{01}, and W_2 be the set of defining words for D_{02}. Denote the set of words in W_1 but not in W_2 as $W_1 \setminus W_2$, and the set of words in W_2 but not in W_1 as $W_2 \setminus W_1$. The cardinality of each of these sets is at least 2^{p-1}. Suppose D_1 and D_2 are equivalent. Then, the words in $W_1 \setminus W_2$ can be removed in D_1, and the words in $W_2 \setminus W_1$ can be removed in D_2. Consider the words in $W_1 \setminus W_2$. There exist 2^{p-1} words in $W_1 \setminus W_2$, corresponding to p independent words of W_1 and all the odd-order interactions of these

p words. Let $\bar{W}_1$ be this set of 2^{p-1} words. We have $\bar{W}_1 \subset W_1 \setminus W_2$. Because all words in $W_1 \setminus W_2$ are removed from D_1, for each of the 2^{p-1} words in $\bar{W}_1$, the sum of the values of the word in all f flats of D_1 should be zero. This indicates that each row of $K(C_1)$ sums to zero, where $K(C_1)$ is the $2^{p-1} \times f$ matrix generated by the p rows of C_1 and all its odd-order interaction rows. Design C_1^T must be a foldover design, and thus D_1 can be reduced to a $(f/2)$-PFD. Similarly, D_2 can be reduced to a $(f/2)$-PFD by considering the words of $W_2 \setminus W_1$. Thus, if D_1 and D_2, based on nonequivalent single flats D_{01} and D_{02}, respectively, are equivalent, then both can be reduced. This contradicts our assumption that both (A_1, C_1) and (A_2, C_2) are of minimal form.

Now, we consider the second part of the necessity proof. If (i) holds, then D_{01} and D_{02} are equivalent, and so $D_{01} = D_{02}$ up to row permutation. Then, by Wang and Mee (2021, Theorem 1), if (ii) does not hold, then D_1 and D_2 are not equivalent.

Moreover, we obtain the detailed form of C_2' when (i) holds. Let Λ be the binary 0-1 matrix denoting a full 2^p factorial, sorted by columns from right to left, omitting the first row of all zeroes. Thus, the row number of $\lambda_i = [\lambda_{i1}, \ldots, \lambda_{in}]$, the ith row of Λ, is given by $[1, 2, 4, \ldots, 2^{p-1}]\lambda_i^T$, $(i = 1, \ldots, 2^p - 1)$. If the row spaces of A_1 and A_2 are equal, then the rows of A_1 are a subset of the rows of ΛA_2. Let $I = [i_1, \ldots, i_p]$ such that $A_1 = [\lambda_{i_1}; \ldots; \lambda_{i_p}]A_2 = A_2'$. Now, we determine the C matrix, say C_2', under A_2' for design D_2. According to the definition of (A, C) of an f-PFD, the p rows of A correspond to p independent words, and the p rows of C indicate the values of these words in f flats. Let $\Gamma(C_2)$ be a $(2^p - 1) \times f$ matrix with the ith row defined as $\prod_{j=1}^{p}(1 - 2\lambda_{ij})c_{2j}$, for $C_2 = (c_{21}^T, \ldots, a_{2p}^T)^T$, and let the product of two row vectors

$\alpha = (\alpha_1, \ldots, \alpha_f)$ and $\beta = (\beta_1, \ldots, \beta_f)$ be defined as $\alpha \times \beta = (\alpha_1\beta_1, \ldots, \alpha_f\beta_f)$. Then, the jth row of $\Gamma(C_2)$ indicates the values of the word $\lambda_j A_2$ in the f flats, for $j = 1, \ldots, 2^p - 1$. Thus, C_2' can be obtained by concatenating the p rows of $\Gamma(C_2)$ by index I, where $I = [i_1, \ldots, i_p]$, such that $A_1 = [\lambda_{i_1}; \ldots; \lambda_{i_p}]A_2 = A_2'$.

□

Based on Lemma 1 and Theorem 1, a new necessary and sufficient condition for identifying isomorphism can be obtained, as shown in the following theorem.

Theorem 2. *Let D_1, D_2 be two f-PFDs with N runs and n factors, where D_i is determined by (A_i, C_i), both of minimal form, with $N = f \times 2^{n-p}$, for $i = 1, 2$. Then, they are isomorphic if and only if there exists a permutation τ of integers $\{1, \ldots, n\}$ such that (i) the row spaces of A_1 and $A_2^{\{c.\tau\}}$ are equal, and (ii) when (A_2, C_2) is re-expressed as (A_2', C_2'), so that $A_2' = A_1$, the corresponding C_2' belongs to the same group as C_1. Therein, $A_2^{\{c.\tau\}}$ reorders the n columns of A_2 with index τ. We call τ the isomorphic map from D_2 to D_1.*

The proof of Theorem 2 is provided in Appendix B. By a similar proof to that for Theorem 1, we can obtain the form of C_2' when (i) holds. That is, C_2' can be obtained by concatenating the p rows of $\Gamma(C_2)$ by index I^*, where $I^* = [i_1^*, \ldots, i_p^*]$, such that $A_1 = [\lambda_{i_1^*}; \ldots; \lambda_{i_p^*}]A_2^{\{c.\tau\}} = A_2'$ and $A_2^{\{c.\tau\}}$ reorders the n columns of A_2 with index τ.

A full search of the $n!$ possible permutations of τ is very time consuming, and can be avoided by using the parallel flats structure. First, we have the following results.

Proposition 2. *Row spaces of A_1 and $A_2^{\{c\cdot\tau\}}$ are equal for some permutation τ if and only if D_{01} and $D_{02}^{\{c\cdot\tau\}}$ are equivalent for permutation τ. That is, τ is an isomorphic* *map from single flat* D_{02} *to* D_{01}.

Both D_{01} and D_{02} are 2^{n-p} designs. From Theorem 2 and Proposition 2, an isomorphic map from D_2 to D_1 must be an isomorphic map from D_{02} to D_{01}. This reduces the number of permutations we need to search from $n!$ to $n!/p!$, because all n columns in the regular designs can be generated by any $n-p$ independent columns. The number can be reduced further by using the row coincidence distributions of the delete-one-factor projections. See Appendix A for details about row coincidence distributions.

For any permutation τ of integers $\{1, \ldots, n\}$, if τ is an isomorphic map from D_{02} to D_{01}, $D_{01}(-i)$ and $D_{02}(-\tau(i))$ must be isomorphic, and thus must have the same row coincidence distribution, where $D(-i)$ is obtained from D by deleting the ith factor for any design D. Thus, τ cannot be an isomorphic map if $D_{01}(-i)$ and $D_{02}(-\tau(i))$ do not have the same row coincidence distribution, for some i. For convenience, we call a permutation τ feasible if $D_{01}(-i)$ and $D_{02}(-\tau(i))$ have the same row coincidence distribution for every i. The key idea of this insight is to entertain only feasible maps by matching the factors using the row coincidence distributions of the delete-one-factor projections (delete-one row coincidence distributions, for short). An analogous technique was previously used by Xu (2009), demonstrating significant computational advantages.

Thus, all we need to do is search from all feasible maps. Note that such an isomorphic map may not be unique; however, we care only about the existence of such a map, not its uniqueness. An algorithm for identifying isomorphism in two-level fractional factorial designs is proposed in the next section.

4. An algorithm for isomorphism check

In this section, we propose a new algorithm for testing for isomorphism in two-level fractional factorial designs, based on Theorem 2. Consider two 2-level designs, with N runs and n factors being compared. The isomorphism check method is given in the following algorithm.

Step 0. Multiply the N rows of each design by its first row, and denote the resulting designs as D_1 and D_2, respectively. Thus, the first rows of D_1 and D_2 have entries of $+1$.

Step 1. Compute the row coincidence distributions for D_1 and D_2. If the row coincidence distributions do not coincide, then the designs are not isomorphic. Otherwise, go to Step 2.

Step 2. For $i = 1$ and 2, obtain (A_i, C_i) from D_i using the algorithm of Edwards and Mee (2023), in which $N = f_i \times 2^{n-p_i}$, A_i and C_i are matrices of size $p_i \times n$ and $f_i \times p_i$, respectively, and C_i has d_i distinct columns, for $i = 1, 2$. If $f_1 \neq f_2$ or $d_1 \neq d_2$, D_1 and D_2 are non-isomorphic. If $f_1 = f_2 = N$, go to Step $4^{\star}$. Otherwise, let $f = f_1 = f_2$ and $p = p_1 = p_2$, obtain the single flats of D_1 and D_2, denoted as D_{01} and D_{02}, respectively, containing the row $(1, \ldots, 1)$, and go to Step 3.

Step 3. Compute the row coincidence distributions for D_{01} and D_{02}. If these differ, then D_1 and D_2 are not isomorphic. Otherwise, compute the n delete-one row coincidence distributions for D_{01} and D_{02}. If the sets of delete-one row coincidence distributions do not coincide, then D_1 and D_2 are not isomorphic. Otherwise, go to Step 4.

Step 4. For each column of D_{01}, count the frequency for each distinct delete-one row coincidence distribution that appears. Let k_i be the frequency for the ith column. Relabel the columns of D_{01} by selecting $q = n - p$ new independent columns so that their frequency numbers k_i are as small as possible sequentially, and denote the resulting design as D'_{01}. Select q independent columns from D_{02} that have the same delete-one row coincidence distributions as those of the q independent columns from D'_{01}, and relabel the columns. If D_{01} and D_{02} do not match after relabeling the independent columns, consider another choice of relabeling and/or another choice of independent columns from the feasible maps. If D_{01} and D_{02} match after relabeling the independent columns under the choice of independent columns, obtain the permutation τ and check whether C_1 and $\Gamma(C_2)^{\{r.I^*\}}$ belong to the same group, where $\Gamma(C_2)^{\{r.I^*\}}$ consists of the p rows of $\Gamma(C_2)$ with index I^*. $I^* = [i_1^*, \ldots, i_p^*]$, such that $A_1 = [\lambda_{i_1^*}; \ldots; \lambda_{i_p^*}]A_2^{\{c.\tau\}}$. If so, the algorithm stops, D_1 and D_2 are isomorphic, and it outputs the isomorphic map τ. If not, consider another choice of relabeling and/or another choice of independent columns. If no such isomorphic map τ can be found after an exhaustive search, the two designs are non-isomorphic.

Step 4*. With $f = N$, both A_1 and A_2 can be identity matrices of order n; then, $C_1 = D'_1$ and $C_2 = D'_2$. For each permutation τ of integers $\{1, \ldots, n\}$, check whether C_1 and $C_2^{\{r.\tau\}}$ belong to the same group, where $C_2^{\{r.\tau\}}$ reorders the n rows of C_2 with index τ. If so, the algorithm stops, D_1 and D_2 are isomorphic, and it outputs the isomorphic map τ. If not, consider another choice. If no such isomorphic map τ can be found after an exhaustive search, the two designs are non-isomorphic.

In theory, our Step 4 requires $O(nf^3\binom{n}{q}q!)$ operations for the worst case, because there are at most $\binom{n}{q}q!$ feasible maps for relabeling the n columns, and each permutation requires nf^3 operations. In most instances, far fewer feasible maps need to be considered, owing to mismatched delete-one row coincidence distributions.

Remark 1. Note that, theoretically, in Step 3, we can only detect non-isomorphism between D_{01} and D_{02}. In most instances, however, we can also verify whether two designs are isomorphic, because the row coincidence distribution (or, equivalently, the word length pattern) uniquely determines a regular design for the vast majority of cases; see the catalog of all regular designs for $n \leq 11$ of size 4 and 8 in Wang and Mee (2021, Supplement) and H. Xu's website http://www.stat.ucla.edu/∼hqxu/pub/ffd2r/ for all resolution III designs of size 16 and 32.

An efficient Matlab implementation of the proposed algorithm is given in the Supplementary Material.

We can easily see that the new algorithm presents a considerable time saving over the isomorphism checking procedures of Clark and Dean (2001), Ye (2003), and Pang and Liu (2011), as summarized in Table 1. In particular, rather than considering all 2^n sign switches, in Step $4^\star$, we consider only N possible sign switches, because we

Table 1: Computational efficiency of the proposed algorithm and related algorithms.

Source of the algorithm	The number of operations
Clark and Dean (2001)	$O(N!n(n!)^2)$
Ye (2003)	$O(n(n!)2^{2n})$
Pang and Liu (2011)	$O(N^2n!2^n)$
The new algorithm - Step 4	$O(nf^3\binom{n}{q}q!)$
The new algorithm - Step 4*	$O(N^3n!)$

$N = f \times 2^{n-p}$ holds for any (f, p) pair in the new algorithm.

always have the treatment combination $(1, \ldots, 1)$. Thus, even when there is an f-PFD structure with $f = N$, our algorithm is more efficient. However, the greatest gains in efficiency occur when there is an f-PFD structure with $f < N$, because then much of the computation depends on the isomorphism of a regular design of size $N/f \geq 2$.

Remark 2. Our proposed isomorphism checking method generalizes the method of Xu (2009), which corresponds to the special case of $f = 1$, and thus allows isomorphism checking for general two-level designs.

We now consider two examples: 1) confirming isomorphism for all strength-two designs with 10 factors and 16 runs; 2) determining the number of non-isomorphic designs among a set of 80-run, 10-factor 5-PFDs.

Example 1. Sun (1993) obtained all 78 non-isomorphic strength-two designs with 16 runs and 10 factors by checking the corresponding projections of all non-isomorphic Hadamard matrices of order 16. The 78 non-isomorphic designs are listed in Appendix B of Sun (1993), and we denote them in order as $D_1, \ldots, D_{78}$. This can also be achieved using our algorithm, with greatly reduced computational effort spent on testing isomorphism, because many design pairs correspond to different row coincidence

Table 2: Ten row coincidence distributions for the 78 designs of 16 runs and 10 factors.

Row coincidence moments $\mathcal{M} = (M_3, M_4)$	Frequency	f	Corresponding design
$\mathcal{M}_1$=(48, 712)	7	1	{4}
		4	{13, 16, 20}
		8	{48, 54, 57}
$\mathcal{M}_2$=(51, 688)	6	4	{15}
		8	{40, 45, 49, 77}
		16	{65}
$\mathcal{M}_3$ =(54, 664)	24	1	{3}
		4	{8, 9, 12, 18, 21}
		8	{24, 26, 29, 32, 39, 41, 42 46, 50, 51, 53, 68, 72, 76, 78}
		16	{60, 63, 64}
$\mathcal{M}_4$ =(54, 676)	3	8	{31, 71}
		16	{62}
$\mathcal{M}_5$ =(54, 688)	3	4	{7}
		8	{25, 27}
$\mathcal{M}_6$ =(55.5, 658)	8	4	{6}
		8	{23, 36, 56, 69, 75}
		16	{61, 66}
$\mathcal{M}_7$ =(57, 664)	9	4	{14}
		8	{30, 34, 38, 47, 70, 74}
		16	{58, 67}
$\mathcal{M}_8$ =(58.5, 658)	5	4	{19}
		8	{35, 52, 73}
		16	{59}
$\mathcal{M}_9$ =(60, 640)	4	1	{2}
		4	{11}
		8	{37, 44}
$\mathcal{M}_{10}$ =(60, 664)	9	1	{1}
		4	{5, 10, 17}
		8	{22, 28, 33, 43, 55}

The rth row coincidence moment M_r is defined as $M_r = \sum_{i=1}^{N}\sum_{j=1}^{N} t_{ij}^r/N^2$, where t_{ij} is the (i,j)th element of $T = DD'$. See Appendix A for more details.

Statistica Sinica: Preprint
doi:10.5705/ss.202022.0200

Table 3: The parallel flats structure of the 78 designs of 16 runs and 10 factors.

D is an f-PFD with row coincidence moment $\mathcal{M}$, where D_0 has row coincidence moment $\mathcal{M}_0$							
D	$\mathcal{M}$	f	$\mathcal{M}_0$	D	$\mathcal{M}$	f	$\mathcal{M}_0$
D4	$\mathcal{M}_1$	1	(0, 10, 48, 712)	D62	$\mathcal{M}_4$	16	(10, 100, 1000, 10000)
D20		4	(0, 36, 192, 2832)	D7	$\mathcal{M}_5$	4	(3, 26, 252, 2504)
D16			(1, 34, 196, 2824)	**D27**		8	(5, 50, 500, 5000)
D13			(2, 28, 248, 2512)	**D25**			(6, 52, 504, 5008)
D57		8	(2, 68, 392, 5648)	D6	$\mathcal{M}_6$	4	(3, 28, 252, 2512)
D54			(3, 58, 468, 5128)	D56		8	(3, 58, 468, 5128)
D48			(4, 52, 496, 5008)	**D36**			(4, 52, 496, 5008)
D15	$\mathcal{M}_2$	4	(2, 26, 248, 2504)	**D75**			
D40		8	(4, 52, 496, 5008)	**D23**			(6, 52, 504, 5008)
D49				**D69**			
D77				**D61**		16	(10, 100, 1000, 10000)
D45			(5, 50, 500, 5000)	**D66**			
D65		16	(10, 100, 1000, 10000)	D14	$\mathcal{M}_7$	4	(2, 26, 248, 2504)
D3	$\mathcal{M}_3$	1	(0, 10, 54, 664)	D34		8	(4, 52, 496, 5008)
D21		4	(0, 34, 216, 2632)	D38			
D18			(1, 30, 232, 2568)	D74			
D12			(2, 28, 248, 2512)	D30			(5, 50, 500, 5000)
D9			(2, 30, 236, 2568)	D47			
D8			(3, 26, 252, 2504)	D70			
D41		8	(3, 58, 468, 5128)	D58		16	(10, 100, 1000, 10000)
D78				D67			
D39			(4, 52, 496, 5008)	D19	$\mathcal{M}_8$	4	(1, 28, 244, 2512)
D50				**D35**		8	(4, 52, 496, 5008)
D51				**D52**			
D53				**D73**			
D76				D59		16	(10, 100, 1000, 10000)
D46			(5, 50, 500, 5000)	D2	$\mathcal{M}_9$	1	(0, 10, 60, 640)
D29				D11		4	(2, 28, 248, 2512)
D32				**D37**		8	(4, 52, 496, 5008)
D72				**D44**			(6, 52, 504, 5008)
D24			(6, 52, 504, 5008)	D1	$\mathcal{M}_{10}$	1	(0, 10, 60, 664)
D26				D17		4	(1, 30, 232, 2568)
D42				D10			(2, 28, 248, 2512)
D68				D5			(3, 30, 264, 2568)
D60		16	(10, 100, 1000, 10000)	D55		8	(3, 58, 468, 5128)
D63				D33			(4, 52, 496, 5008)
D64				D28			(5, 50, 500, 5000)
D31	$\mathcal{M}_4$	8	(5, 50, 500, 5000)	D43			(6, 52, 504, 5008)
D71				D22			(7, 58, 532, 5128)

$\mathcal{M}$ denotes the row coincidence moments (M_3, M_4) for D listed in Table 2.; $\mathcal{M}_0$ denotes the row coincidence moments (M_1, M_2, M_3, M_4) of the single flat D_0; the design in boldface means that the single flat of the corresponding design is not unique among those of all non-isomorphic 78 designs.

distributions, different numbers of flats, or non-isomorphic single flats. Accordingly, in general, our algorithm terminates before steps 4 or $4^{\star}$.

Details on all non-isomorphic 78 nonregular designs with 16 runs and 10 factors are summarized in Tables 2 and 3. In step 1, we found 10 different row coincidence distributions. In step 2, these designs were found to be regular designs or 4-PFDs, 8-PFDs, or 16-PFDs. In step 3, we discovered 22 non-isomorphic D_0, four of resolution II and 18 of resolution I. With 78 designs, there are 3003 pairs of designs. For 98.57% (2960) of these pairs, non-isomorphism is determined before steps 4 or $4^{\star}$. Of the remaining 43 pairs, 38 are distinguished in step 4, and five pairs are examined in Step $4^{\star}$, where $f = 16$.

Example 2. The variable neighborhood search algorithm in Edwards and Mee (2023) can be employed to generate D-efficient PFDs for estimating the two-factor interaction model. Nearly 1200 5-PFDs with 80 runs were constructed. Thirty-four of these 5-PFDs had D-efficiency of 88.8%, and the remainder all had lower D-efficiency. We are interested in how many non-isomorphic designs appear in this set of 34 designs. All have the same A-efficiency (74%) and maximum variance inflation factor (2.1875) for the two-factor interaction model. However, they are not all isomorphic. In step 1, we found three different row coincidence distributions. In step 2, all designs were found to be 5-PFDs, with no repeated flats. In step 3, we discovered two non-isomorphic D_0, one of resolution II and one of resolution I. In step 4, it was confirmed that there are exactly eight non-isomorphic designs, which occurred with frequencies between 1 and 13 times each. In Table 4, we list the characteristics of these eight designs in terms of their generalized resolution, GWLP, frequency among the 34 designs, and trace($\mathcal{A}'\mathcal{A}$), where $\mathcal{A}$ is the 56×120 alias matrix with columns corresponding to the 120 possible three-factor interactions. For more details, see Appendix A.

Table 4: Eight non-isomorphic 5-PFDs with 80 runs and 10 factors, with D-efficiency = 88.8%.

Design	Frequency	R [♯]	GWLP=$(B_1, B_2, B_3, B_4, B_5)$	tr$(\mathcal{A}'\mathcal{A})^{\dagger}$	$D_0 = (t@1_{n_0}, D_0')^{\ddagger}$	
					t	D_0'
D1	1	2.8	(0.00, 0.04, 0.32, 1.16, 5.84)	101.20	0	10-6.7
D2	5	2.8	(0.00, 0.04, 0.32, 1.16, 5.84)	105.74	0	10-6.7
D3	3	2.8	(0.00, 0.04, 0.32, 1.16, 5.84)	110.28	0	10-6.7
D4	13	2.8	(0.00, 0.04, 0.64, 0.84, 4.56)	107.52	0	10-6.7
D5	3	2.8	(0.00, 0.04, 0.64, 0.84, 4.56)	112.21	0	10-6.7
D6	2	2.8	(0.00, 0.04, 0.64, 0.84, 4.56)	116.71	0	10-6.7
D7	6	1.8	(0.04, 0.00, 0.56, 0.92, 4.56)	105.45	1	9-5.2
D8	1	1.8	(0.04, 0.00, 0.56, 0.92, 4.56)	109.64	1	9-5.2

[♯] R represents generalized resolution; [†] $\mathcal{A}$ represents the alias matrix $(X_1'X_1)^{-1}X_1'X_2$, where X_1 consists of the intercept, main effects, and two-factor interaction effects, and X_2 consists of three-factor interaction effects; [‡] Designs D_1–D_8 are 5-PFDs based on single flat $D_0 = (t@1_{n_0}, D_0')$; Design 10-6.7 is a resolution-II design that consists of columns of indices $\{1, 2, 4, 8, 1, 3, 5, 10, 12, 15\}$ of H_{16}, and design 9-5.2 is a resolution-III design that consists of columns of indices $\{1, 2, 4, 8, 3, 5, 10, 12, 15\}$ of H_{16}, where H_{16} is the Sylvester Hadamard matrix of order 16 (with columns labeled from 0 to 15).

5. Conclusion

Checking for isomorphism is vital for design construction, because, in general, we can ignore designs from the same isomorphism class. Clark and Dean (2001) provided the initial necessary conditions for checking isomorphism of nonregular two-level designs. Ye (2003) and Pang and Liu (2011) made subsequent improvements. In this paper, we propose new necessary and sufficient conditions, as well as a new algorithm for identifying isomorphism in two-level fractional factorial designs, using a parallel flats structure. The proposed algorithm is simple and general. In addition, by checking for and exploiting any parallel flats structure, the proposed algorithm is much faster than competing methods in the literature.

Acknowledgments

The authors gratefully acknowledge the help of Professor Hongquan Xu for his constructive suggestions. We also thank Professor David Edwards for the designs in Example 2 and his valuable suggestions. This work was supported by the MOE Project of Key Research Institute of Humanities and Social Sciences (22JJD110001).

Supplementary Material

In the online Supplementary Material, we provide an efficient Matlab implementation of the proposed algorithm for checking the isomorphism of two-level designs, called "isocheck". All 78 non-isomorphic strength-two designs with 16 runs and 10 factors in Example 1 are provided in the MATLAB .mat file "N16p10designs.mat". The eight non-isomorphic D-efficient 5-PFDs with 80 runs and 10 factors in Example 2 for a two-factor interaction model are provided in the MATLAB .mat file "N80p10f5PFDs.mat".

Appendix

Appendix A: Details for generalized word length pattern and row coincidence distributions

For a regular two-level design with levels ± 1, the word length pattern is the vector $\text{WLP} = (A_3, A_4, \ldots, A_n)$, where A_r is the number of r-factor interaction columns that sum to $\pm N$. For nonregular designs, Tang and Deng (1999) defined the generalized word length pattern GWLP $= (B_3, B_4, \ldots, B_n)$, where B_r is the sum of the squares of all r-factor interaction columns, divided by N. Note that for a regular design, $A_r = B_r$. The G_2-aberration criterion ranks designs based on GWLP. The G_2-aberration criterion is very cheap to compute, due to its connection to the moments of the row coincidence distribution, or equivalently, the moments of Hamming distances. For a two-level design D, $T = DD'$ gives the row coincidence distribution. The r^{th} moment of the row coincidence distribution, also called as the r^{th} row coincidence moment, is defined as $M_r = \sum_{i=1}^{N} \sum_{j=1}^{N} t_{ij}^r / N^2$; therein t_{ij} is the (i, j)-th element of T. Butler (2003) proved that ranking designs in terms of G_2-aberration is equivalent to sorting on the moments of their row coincidence distributions. Furthermore, Butler (2003) gave explicit formulae for the B_r's in terms of M_r's (see Mee, 2009, App. J).

Appendix B: Proof of Theorem 2

Proof. The sufficiency of the conditions is obvious. Next we prove the necessity by showing that two f-PFDs of minimal form, say D_1 and D_2, based on non-isomorphic single flats D_{01} and D_{02}, respectively, must be non-isomorphic.

As D_{01} and D_{02} are non-isomorphic 2^{n-p} designs, they must have at least 2^{p-1} different words for any permutation τ of D_{02}'s columns. Let W_1 be the defining words for D_{01} and W_2^τ be the set of defining words for D_{02} after the permutation τ. Given τ, the words in W_1 but not in W_2^τ form the set $W_1 \setminus W_2^\tau$, while the words in W_2^τ but not in W_1 form the set $W_2^\tau \setminus W_1$. The cardinality of each of these sets is at least 2^{p-1} for any permutation τ.

Suppose D_1 and D_2 are isomorphic. Then there must exist a permutation τ under which the words in $W_1 \setminus W_2^\tau$ can be removed in D_1 and the words in $W_2^\tau \setminus W_1$ can be removed in D_2. Consider the words in $W_1 \setminus W_2^\tau$. There exist 2^{p-1} words in $W_1 \setminus W_2^\tau$ corresponding to p independent words of W_1 and all the odd-order interactions of these p words. Let $\tilde{W}_1$ be this set of 2^{p-1} words. We have $\tilde{W}_1 \subset W_1 \setminus W_2^\tau$. As all words in $W_1 \setminus W_2^\tau$ should be removed in D_1, then for each of the 2^{p-1} words in $\tilde{W}_1$, the sum of the values of the word in all f flats of D_1 should be 0. This indicates that each row of $L(C_1)$ sums to zero, where $L(C_1)$ is the $2^{p-1} \times f$ matrix generated by the p rows of C_1 and all its odd-order interaction rows. Design C_1^T must be a foldover design, and thus D_1 can be reduced to a $(f/2)$-PFD. Similarly, we can obtain that D_2 can be reduced to a $(f/2)$-PFD by considering the words of $W_2^\tau \setminus W_1$. Thus, if two f-PFDs based on non-isomorphic single flat are isomorphic, then both can be reduced. This contradicts our assumption that both (A_1, C_1) and (A_2, C_2) are of minimal form.

In summary, two f-PFDs based on non-isomorphic single flat must be non-isomorphic. The proof is complete. □

References

Butler, N. A. (2003). Minimum aberration construction results for nonregular two-level fractional factorial designs. *Biometrika* **90**, 891–898.

Chen, J. and Lin, D. K. (1991). On the identity relationships of 2^{k-p} designs. *J. Statist. Plann. Inference* **28**, 95–98.

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/Revue Internationale de Statistique* 131–145.

Cheng, C. S. (2014). *Theory of Factorial Design: Single- and Multi-Stratum Experiments.* Chapman and Hall/CRC.

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.

Connor, W. S. and Young, S.(1961). *Fractional Factorial Designs for Experiments with Factors at Two and Three Levels. National Bureau of Standards Applied Mathematics Series (58)* National Bureau of Standards, Gaithersburg.

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

Draper, N. R. and Mitchell,T.J. (1967). The construction of saturated 2_R^{k-p} designs. *Ann. Math. Statist.* **38**, 1110–1126.

Draper, N. R. and Mitchell, T. J. (1968). Construction of the set of 256-run designs of resolution ≥ 5 and the set of even 512-run designs of resolution ≥ 6 with special refernce to the unique saturated designs. *Ann. Math. Statist.* **39**, 246–255.

Draper, N. R. and Mitchell,T. J. (1970). Construction of the set of 512-run designs of resolution ≥ 5 and the set of even 1024-run designs of resolution ≥ 6. *Ann. Math. Statist.* **41**, 876–887.

Edwards, D. J. and Mee, R. W. (2023). Structure of two-level nonregular designs. *J. Amer. Statist. Assoc.* **118**, 1222–1233.

Fang, K. T. and Zhang, A. (2004). Minimum aberration majorization in non-isomorphic saturated designs. *J. Statist. Plann. Inference* **126**, 337–346.

Jones, B. A., Lekivetz, R. Majumdar, D., Nachtsheim, C. J. and Stallrich, J. W. (2019). Construction, properties, and analysis of group-orthogonal supersaturated designs. *Technometrics* **61**, 1–12.

Ke, X., Fang, K. T., Elsawah, A. M. and Lin, Y. (2023). New non-isomorphic detection methods for orthogonal designs. *Comm. Statist. Simulation Comput.* **52**, 27–42.

Liao, C. T., Iyer, H. K. and Vecchia, D. F. (1996). Construction of orthogonal two-level designs of user specified resolution where $N \neq 2^k$. *Technometrics* **38**, 342–353.

Lin C.-Y. and Cheng S.-W. (2012). Isomorphism examination based on the count vector. *Statist. Sinica* **22**, 1253-1272.

Liu, Y., Yang, J. F. and Liu, M. Q. (2011). Isomorphism check in fractional factorial designs via letter interaction pattern matrix. *J. Statist. Plann. Inference* **141**, 3055–3062.

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

Mee, R. W. (2009). *A Comprehensive Guide to Factorial Two-Level Experimentation.* Springer, New York.

Mukerjee, R., and Wu, C. F. J. (2006). *A Modern Theory of Factorial Design.* Springer, New York.

Pang, F. and Liu, M. Q. (2011). Geometric isomorphism check for symmetric factorial designs. *J. Complexity* **27**, 441–448.

Schoen, E. D., Eendebak, P. T. and Nguyen, M. V. (2010). Complete enumeration of pure-level and mixed-level orthogonal arrays. *Journal of Combinatorial Designs* **18**, 123–140.

Shrivastava, A. K. and Ding, Y. (2010). Graph based isomorph-free generation of two-level regular fractional factorial designs. *J. Statist. Plann. Inference* 140, 169-179.

Srivastava, J. N. and Li, J. (1996). Orthogonal designs of parallel flats type. *J. Statist. Plann. Inference* **53**, 261–283.

Srivastava, J. N. and Chopra, D. A. (1973). Balanced arrays and orthogonal arrays, in *A Survey of Combinatorial Theory,* ed. J. N. Srivastava. Amsterdam, North-Holland

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. (1993). Estimation capacity and related topics in experimental designs. unpublished doctoral thesis, University of Waterloo, Dept. of Statistics and Actuarial Science.

Sun, D. X., Li, W. and Ye, K. Q. (2008). Algorithmic construction of catalogs of non-isomorphic two-level orthogonal designs for economic run sizes. *Statistics and Applications* 6, 141–155.

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

Wang, C. Y. and Mee, R. W. (2021). Two-level parallel flats designs. *Ann. Statist.* **49**, 3015–3042.

Weng, L. C., Fang, K. T. and Elsawah, A. M. (2023). Degree of isomorphism: a novel criterion for identifying and classifying orthogonal designs. *Statist. Papers (Berlin, Germany)* **64**, 93–116.

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

Xu, H. (2009). Algorithmic construction of efficient fractional factorial designs with large run sizes. *Technometrics* **51**, 262–277.

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

doi:10.5705/ss.202022.0200

Chunyan Wang

Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: chunyanwang@ruc.edu.cn

Robert W. Mee

Department of Business Analytics and Statistics, University of Tennessee, Knoxville, Tennessee 37996, USA.

E-mail: rmee@utk.edu