Southeast Asian Bulletin of Mathematics (2020) 44: 819–844

Southeast Asian
Bulletin of
Mathematics

Quality Engineering with Balanced Fractional Factorial Experimental Designs

Man V.M. Nguyen
Department of Mathematics, Faculty of Science, Mahidol University, 272, Rama VI, 10400, Bangkok, Thailand
Center of Excellence in Mathematics (CEM), Ministry of Education, 272, Rama VI, 10400, Bangkok, Thailand
Email: man.ngu@mahidol.edu

Received 13 July 2019
Accepted 15 May 2020

Communicated by Edy Tri Baskoro

AMS Mathematics Subject Classification(2000): 05B15, 05C15, 20B35, 62K15

Abstract. Balanced fractional factorial designs or Orthogonal arrays (OAs) are popular structures in Quality Engineering and Statistical Quality Control. Orthogonal arrays of strength at least 2 have useful properties that can be employed for experimental designs, manufacturing, quality technology and for scientific discoveries in general.

In this paper, we introduce a graph method and a group-theoretic approach for constructing mixed OAs of any strength, with a given parameter set of run-size and factor levels.

Keywords: Coloring of graph; Factorial designs; Orthogonal arrays; Subgroups of symmetric groups.

1. Introduction

Factorial experimental designs play essential roles in Quality Engineering, specifically in mass manufacturing and sustainable economic development, with a bunch of approaches, methods and techniques since the 1950s. Quality Engineering concerns about achieving *quality* and *productivity* at the same time, based on the fundamental idea of *continuous improvement* (called *kaizen* by the Japanese). In industrial manufacturing, the manufacturer would theoretically try to achieve 6-sigma quality level, corresponding to the ideal error ratio 3.4

PPM (parts per million, see [16]).

Table 1: Design of Experiments (DOE)- the quality ladder (Kenett [10])

Statistical methods	**Management approach**
DOE, TQM, Six-sigma	**Quality by Design** (1980s- now)
Statistical Quality Control	Process Improvement (1960-90's)
Sampling	Inspection (1950-60's)
Data Accumulation	Fire Fighting (before WW 2)

Statistical methods for quality engineering have been developed through few milestones, shown in the above table, by contributions of American scholars and engineers as Walter A. Shewhart, W. Edwards Deming, Joseph M. Juran, a few renown Indian statisticians as C.R. Rao, R.C. Bose, and many Japanese pioneers as Kaoru Ishikawa and Dr. Genichi Taguchi.

Six Sigma, particularly designed to work across all processes and industries, is a strategic engineering management paradigm originally developed at *Motorola*. However, Six Sigma draws heavily on the previous quality paradigms and methodologies, such as statistical quality control and total quality management-TQM, see [6].

We are interested in mathematical constructions of a combinatorial structure, called *balanced fractional factorial designs* or *orthogonal array*, a special kind of factorial experimental designs (a subclass of designs of experiments or DOE).

Orthogonal arrays (OAs) with strength $t > 1$ (or t-balanced fractional factorial designs) have statistically good features which can be employed not only in experimental designs, industries and services [16, 17], algebraic coding theory, software engineering [5], but also in emerging and fast-developed areas such as statistical disclosure control [4], computational biology, particularly DNA microarray experiments [7, 8], and in applications of statistical quality management and control, see Wu and Hamada [24] for more information.

Definition 1.1. *Formally, we fix d finite sets $Q_1, Q_2, \ldots, Q_d$ called factors, where $1 < d \in \mathbb{N}$. The elements of a factor are called its levels. The (full) factorial design (also factorial experiment design- FED) with respect to these factors is the Cartesian product $D = Q_1 \times Q_2 \times \ldots \times Q_d$.*

A fractional design or fraction F of D is a subset consisting of elements of D (possibly with multiplicities). Put $r_i := |Q_i|$ be the number of levels of the ith factor. We say that F is symmetric if $r_1 = r_2 = \cdots = r_d$, otherwise F is mixed.

Moreover, F is said to be strength t orthogonal array (OA) or t-balanced fractional designs if, for each choice of t coordinates (columns) from F, each combination of coordinate values from those columns occurs equally often; here t is a natural number.

The main aim of experimental design is to identify an unknown function $\phi : D \to \mathbb{R}$ on a full design D, which is a mathematical model of some quantity of interest (favor, usefulness, best-buy, quality, ...) that is be computed or optimized. OAs can provide smaller (and so more economic) fractional designs, which still allow us to identify the most important features of ϕ. Specifically, strength 3 OAs permit estimation of all the main effects of the experimental factors, without confounding them with the two-factor interactions. Strength 4 OAs, furthermore allow us to separately estimate all two-factor interactions.

A comprehensive reference on the use of orthogonal arrays (OAs) as factorial design in diverse problems of statistical parameter optimization is provided by Wu and Hamada [24]. Stufken and Tang [23] provided a complete solution to enumerating non-isomorphic two-level OAs of strength t with $t+2$ constraints for any t and any run size $N = \lambda\, 2^t$. Bulutoglu and Margot [3] recently formulated an integer linear programming (ILP) method for classifying OAs of strength 3 and 4 with run size at most 162.

A few specific construction methods of OAs have been proposed in Brouwer et al. [2], Nguyen [14] and [15]; and OAs with strength at least 2 are online reported by Sloane [22]. Moreover, a parallel computing approach can return lexicographically minimum column (LMC) matrices, more details can be found in Phan et al. [18, 19] and Schoen et al. [20].

The major motivation of this work is to combine a graph-coloring method and a group-theoretic approach for constructing mixed orthogonal arrays (OAs) with any strength. We will specifically discuss about describing OAs by colored graphs and then present a group theory-based solution for the factor extension problem of a given orthogonal array.

Section 2 recalls background and states the design extension problem. We define canonical orthogonal arrays using colored graphs in Section 3; and transformations (*isomorphism*) of an OA in Section 4. Next we present an integer linear formulation with the row permutation group of a design F to compute an extension $[F|X]$ in Section 5, and last but not least, employ localizing the formation of vector solutions X in Section 6. Section 7 concludes the paper with a few comments.

2. The Balanced Fractional Factorial Design Construction Problem

Some standard constructions of orthogonal arrays are reviewed in [9, 14].

Let $s_1 > s_2 > \cdots > s_m$ be the distinct factor sizes of an orthogonal array F, originally determined by the number of levels r_i. Assume that F has exactly a_i factors with s_i levels, where $s_i \neq s_j$ if $i \neq j = 1, \ldots, m$; now the total number of factors is $d = a_1 + a_2 + \ldots + a_m$. We rewrite $OA(N; r_1, r_2, \cdots, r_d; t)$ as $OA(N; s_1^{a_1} \cdot s_2^{a_2} \cdots s_m^{a_m}; t)$ for a mixed array of strength t, with N runs.

The *design type* T of F is described by either $r_1 \cdot r_2 \cdots r_d$ or $s_1^{a_1} \cdot s_2^{a_2} \cdots s_m^{a_m}$.

We take the r_i in nonincreasing order, so that they are related to the s_k by

$$s_1 = r_1 = \cdots = r_{a_1}, s_2 = r_{a_1+1} = \cdots = r_{a_1+a_2}, \ldots,$$
$$s_m = r_{a_1+a_2+\cdots+a_{m-1}+1} = \cdots = r_{a_1+a_2+\cdots+a_m} = r_d.$$

For example, the matrix F below is a $4 \cdot 2^3$ mixed OA of strength 3:

$$F = \begin{bmatrix} 0&0&0&0&1&1&1&1&2&2&2&2&3&3&3&3 \\ 0&1&0&1&0&1&0&1&1&0&1&0&1&0&1&0 \\ 0&0&1&1&0&0&1&1&1&1&0&0&1&1&0&0 \\ 0&1&1&0&0&1&1&0&1&0&0&1&1&0&0&1 \end{bmatrix}^T.$$

We can narrow down the set of candidate arrays by using the divisibility.

Lemma 2.1. (Divisibility) *In an $OA(N; r_1 \cdot r_2 \cdots r_d; t)$, the run size N must be divisible by the least common multiple* (lcm) *of all numbers $\prod_{i \in I} r_i$ where $|I| = t$.*

An efficient way to construct strength t arrays is by starting with a full array with t factors, then extending it column by column. So generally, we formulate the design extension problem as:

Given a strength t orthogonal array F_0 with N runs and d factors, extend it to a strength t orthogonal array $F = [F_0|X]$ with $d+1$ factors, where X is a new factor (or column).

A few specific construction methods for this problem are reported in [2, 15], among those are an arithmetic method giving the unique $OA(64; 4^4 \cdot 2^6; 3)$ (by the Rao bound [9]) and a Latin square construction to list $OA(96; 6 \cdot 4^2 \cdot 2^5; 3)$.

Lemma 2.2. (The Rao bound) *Assume that an $OA(N; r_1, r_2, \cdots r_d; t)$ exists.*

(i) *If strength t is even, then $N \geq \sum_{j=0}^{t/2} \sum_{|I|=j} \Pi_{i \in I}(r_i - 1)$.*

(ii) *If strength t is odd, then*

$$N \geq 1 + \sum_{j=1}^{(t-1)/2} \sum_{|I|=j} \Pi_{i \in I}(r_i - 1) + \max_j \left((r_j - 1) \sum_{|I|=\frac{t-1}{2}, j \notin I} \prod_{i \in I}(r_i - 1) \right).$$

Example 2.3.

(i) For an $OA(N; 3^5 \cdot 2; 3)$, N must be a multiple of $\text{lcm}(3 \cdot 3 \cdot 3, 2 \cdot 3 \cdot 3) = 54$, by the divisibility. This run size fulfills the Rao bound as well, since

$$\begin{aligned} N &\geq 1 + a_1.(r_1 - 1) + a_2.(r_2 - 1) + a_1.(r_1 - 1)(r_2 - 1) + (a_2 - 1)(r_2 - 1)^2 \\ &\geq 1 + 1.(2-1) + 5.(3-1) + 1.(2-1)(3-1) + (5-1)(3-1)^2 = 30. \end{aligned}$$

(ii) An $OA(96; 6 \cdot 4^b \cdot 2^a; 3)$ is valid only for cases of a, b satisfying few conditions that $a + b \geq 3, b \leq 2$, and $a + 3b \leq 15$. Does an $OA(96; 6 \cdot 4^2 \cdot 2^6; 3)$ exist?

The design resolution R (the length of the shortest word in the defining relation) of a design is a useful way to classify fractional factorial designs according to the alias patterns they produce (see [24]). For regular designs (one can be defined by generator words), their strength $t = R - 1$. However, orthogonal arrays include both regular designs and irregular designs!

Resolution III designs: designs in which *no main effects* are *aliased* with any other main effect, but main effects are aliased with two-factor interactions, and some two-factor interactions may be aliased with each other.

Resolution IV designs: designs in which *no main effect* is *aliased* with any other main effect or 2-factor interactions, but 2-factor interactions can be aliased with each other. Both resolution III, IV designs are useful in factor screening.

Resolution V designs: no main effect or two-factor interaction is aliased with any other main effect or two-factor interaction, but two-factor interactions are aliased with three-factor interactions.

3. Canonical Orthogonal Arrays with Colored Graphs

We introduce the concept of *canonical orthogonal arrays*, then use it to classify *non-isomorphic arrays* of given design type and run size. We first encode an array as a *colored graph*, then use the software package *nauty*, by B. Mckay [13] to find the *canonical labeling graph* of the colored graph and decode the result back to an array. Testing isomorphism between arrays is reduced to testing isomorphism between their colored graphs. Precisely, we describe a way to translate an array to a graph and show how to color that graph. Then we present a method to get back (demerge) an array from a colored graph. Thirdly, we find the canonical graph of a colored graph using *nauty*. We close this part by computing the *canonical orthogonal array* of a given orthogonal array.

3.1. The Graph of an Orthogonal Array

A design D with d factors is viewed as a set R of d-tuples $v = (p_1, \ldots, p_d)$, where $p_i \in Q_i$ for level sets $Q_1, \ldots, Q_d$. So each d-tuple from R represents a row of D. A (undirected) graph $G = (V, E)$ is constructed from this OA as

$$V = R \cup S \cup C; \tag{1}$$

where R is the set of row-vertices (one vertex per row), $C := \{x_1, \ldots, x_d\}$ is the set of columns (one vertex for each column-factor), and $S := \bigcup_{i=1}^{d} Q_i$ is the set of levels (symbols) per column (one vertex per level per column). Then

$$|V| = |R| + \big(\sum_i^d |Q_i|\big) + d = N + \sum_i^d r_i + d.$$

Let

$$E_1 := \bigcup_{1\le i\le d} \big\{\{v, p_i\} : v = (p_1, \dots, p_d) \in R \text{ and } p_i \in Q_i\big\},$$
$$E_2 := \bigcup_{1\le i\le d} \big\{\{s, x_i\} : s \in Q_i\big\}.$$

Then the edge set and its size respectively are

$$E = E_1 \cup E_2 \subseteq (R \times S) \cup (S \times C), \ |E| = d|R| + \sum_i^d |Q_i| = dN + \sum_i^d r_i. \quad (2)$$

Since R, S, C are cocliques (ie, vertices in each set are not adjacent with each other), G is a tripartite graph with the vertex partition $R \cup S \cup C$. Let $n_S := |S|$ be the number of symbols, and $N = |R|$ the run size of D. The adjacency matrix A of G has the following pattern:

$$A = \left[\begin{array}{c|c|c} 0 & RS & 0 \\ \hline SR & 0 & SC \\ \hline 0 & CS & 0 \end{array}\right]$$

where RS is the $N \times n_S$-adjacency matrix formed by the row-symbol adjacency, $SR = RS^T$, and $SC = CS^T$, where CS is the $d \times n_S$ adjacency matrix formed by the column-symbol adjacency. We call a vertex with valency i an *i-vertex*, and write $V(x)$ for the neighbors of a vertex $x \in V$.

To use the package *nauty*, we need to number the vertices of G. We number the row-vertices R first, then the symbol-vertices S and finally the column-vertices C. We color the resulting graph G by the following coloring rules:

(i) all vertices of R have color A; here A is called the row color;
(ii) all vertices of S have color B; here B is called the symbol color;
(iii) factors $x_1, \dots, x_d$ have the same color if and only if the corresponding level sets have the same cardinality: $\text{color}(x_i) = \text{color}(x_j) \iff |Q_i| = |Q_j|$. Figure 1 shows the colored graph of a 6 runs orthogonal array.

Denote by $\mathcal{F}_{T,N} = \text{OA}(N; T; t)$ the class of all orthogonal arrays with given type $T = s_1^{a_1} \cdot s_2^{a_2} \cdots s_m^{a_m}$, of strength $t \ge 1$, and run size N. If the array $D \in \mathcal{F}_{T,N}$, then the set of column-vertices C is a disjoint union of color classes $C_1, \dots, C_m$, called the *column-color classes*, and the total number of colors of G is $2+m$. Also note that each row-vertex is adjacent to precisely d symbol-vertices, and each symbol-vertex is adjacent to exactly one column-vertex. Remark that

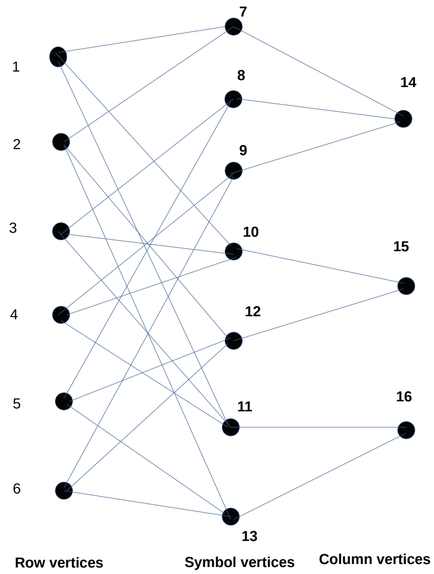

Figure 1: The colored graph of a 6 runs orthogonal array

the partition (R, S, C) is not a color partition, and $d = \sum_{i=1}^{m} |C_i|$. Recall that $n_S = |S|$. We write

$$f := \Big[[1, \ldots, N],\, [N+1, \ldots, N+n_S], \qquad (3)$$

$$[N+n_S+1, \ldots, N+n_S+a_1], \ldots, [N+n_S+1+\sum_{i=1}^{m-1} a_i, \ldots, |V|] \Big]$$

for the *color partition* (determining row, symbol and column-vertices, respectively); and denote the colored graph just obtained by G_D.

Example 3.1. Let D be the $OA(6; 3^1 \cdot 2^2; 1)$

$$D = \begin{bmatrix} 0 & 0 & 1 & 1 & 2 & 2 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 & 0 \end{bmatrix}^T .$$

Then $N = 6$, $n_S = 7$, $d = 3$, $m = 2$, and the vertices

$$V = R \cup S \cup C = \{1, 2, \ldots, 6, 7, \ldots 13, 14, 15, 16\}.$$

The color classes have sizes 6, 7, 1, 2, with corresponding vertices

$$f := \big\{\{1,2,3,4,5,6\},\{7,8,9,10,11,12,13\},\{14\},\{15,16\}\big\}.$$

The symbol permutation (0,1) on column 2 of array D is performed by its corresponding permutation $p_S = (10,11)$ on symbol-vertices 10, 11 of the colored graph G_D. Switching columns 2 and 3 of D has counterpart $p_C = (15,16)$ on column-vertices. And permuting rows 1 and 2 can be done by the permutations on row-vertices $p_R = (1,2)$.

Let $\mathcal{G}$ be the *set of all colored graphs*. Define the map

$$\Phi : \mathcal{F}_{T,N} \longrightarrow \mathcal{G}, \quad D \longmapsto \Phi(D) = G_D,$$

taking an array D to the corresponding colored graph G_D described above.

Lemma 3.2. Φ *is an injection.*

Now we characterize more clearly the image $\Phi(\mathcal{F}_{T,N}) \subseteq \mathcal{G}$. We write $v(u)$ for the valency of a vertex $u \in V$. Recall that $S = Q_1 \cup Q_2 \cup \ldots \cup Q_d$, where $|Q_i| = r_i$ for $i = 1,\ldots,d$; and $C = C_1 \cup \ldots \cup C_m$, where $|C_k| = a_k$, for $k = 1,\ldots,m$.

Lemma 3.3. *Let D be a design with factors Q_i and with run size N. Then*

(i) *G_D is tripartite with the vertex partition (R,S,C) given by (1) and with $|R| = N$, $|S| = \sum_{k=1}^m a_k s_k$, and $|C| = \sum_{k=1}^m a_k$.*

(ii) *Every vertex $r \in R$ has valency d.*

(iii) *The valency of a column-vertex c in C is s_k, where k is the unique element of $\{1,\ldots,m\}$ such that $c \in C_k$.*

(iv) *The valency of a symbol-vertex: if $s \in S$ then there is a unique $c \in C_k$ such that $\{s,c\} \in E$ for some $k \in \{1,\ldots,m\}$; then*

$$v(s) = \frac{N}{v(c)} + 1 = \frac{N}{s_k} + 1$$

[since $t \geq 1$, there are exactly $\frac{N}{s_k}$ rows in D having symbol s in column c].

(v) *Relationship between R and C: if $r \in R$, and $c \in C$, there exists a unique shortest path of length 2 from r to c through a vertex in S.*

Definition 3.4.

(i) *Given parameters U,N, the colored graphs which satisfy properties* (i)–(v) *of Lemma 3.3 are called the colored graphs of type U,N. They form a subset of $\mathcal{G}$, written $\mathcal{G}_{U,N}$.*

(ii) *By Lemma 3.3* (i)*, vertices of R, S, C in a graph in $\mathcal{G}_{U,N}$ are called the row-vertices, the symbol-vertices and the column-vertices respectively.*

3.2. Demerging a Colored Graph

How can we make an orthogonal array which is associated with a colored graph? This array-making process is called **demerging** a colored graph. What we want to do now is to demerge a colored graph $g \in \mathcal{G}_{U,N}$.

We firstly find the column-vertex set C of g. It may happen that some vertices have the same valency even if they belong to distinct colors (row and column colors, for instance). This can usually be solved by computing the intersection of their neighbor sets. More precisely, we have the following claim.

Lemma 3.5. *Suppose that $\frac{N}{s_k} \in \mathbb{N}$ for all $k \in \{1, \ldots, m\}$, in which $\frac{N}{s_k} > 1$ for at least a number k. Then, a subset C of the vertex set V of a graph g in $\mathcal{G}_{U,N}$ is the column-vertex set if and only if the valencies of vertices in C are $\{s_1, s_2, \ldots, s_m\}$ and their neighbor sets are mutually disjoint subsets of V.*

Proof. Use Lemma 3.3. ■

For instance, consider a strength 1 array $F := OA(4; 4^4; 1)$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 2 & 2 & 2 & 2 \\ 3 & 3 & 3 & 3 \end{bmatrix}$$

in which $\frac{N}{s_1} = 1$. The row and column vertices of the colored graph G_F are not distinguishable. We will see later that this kind of array requires a subtle treatment to demerge the colored graph.

Proposition 3.6. *(Constructing an array from a colored graph) Given parameters $T = s_1^{a_1} \cdot s_2^{a_2} \cdots s_m^{a_m}$ and run size N, such that $\frac{N}{s_k} \in \mathbb{N}$ for all $k \in \{1, \ldots, m\}$, and such that there is at least one k for which $\frac{N}{s_k} > 1$, we have $\Phi(\mathcal{F}_{T,N}) = \mathcal{G}_{U,N}$.*

Proof. See [14, Proposition 40]. ■

Corollary 3.7. *Provided that $\frac{N}{s_k} \in \mathbb{Z}^{\times}$ for all $k \in \{1, \ldots, m\}$, and that there is at least a number $\frac{N}{s_k} > 1$, the mapping Φ is already an injection (Lemma 3.2), now we have that Φ is a bijection between the set $\mathcal{F}_{T,N}$ of fractions of type U, N and the set $\mathcal{G}_{U,N}$ of colored graphs of type U, N.*

The inverse mapping Φ^{-1} from $\mathcal{G}_{U,N}$ to $\mathcal{F}_{T,N}$ is called the *demerging mapping* of $\mathcal{G}_{U,N}$. Any orthogonal array $D \in \mathcal{F}_{T,N}$ of strength $t \geq 2$ is determined uniquely by its companion graph $G_D \in \mathcal{G}_{U,N}$. Indeed, if strength $t \geq 2$ then $\frac{N}{s_i s_k} \geq 1$ for all $i, k = 1, \ldots, m$. So $\frac{N}{s_k} > 1$ for each $k = 1, \ldots, m$.

Lemma 3.8. *Let G_F, G_D be the two colored graphs which are formed by two* *fractions $F, D \in \mathbb{F} = \mathbb{F}_{T,N}$. Then F and D are isomorphic arrays if and only if G_F and G_D are isomorphic graphs.*

Example 3.9. We construct an $OA(6; 3 \cdot 2^2; 1)$ from the colored graph described by Figure 1 in Appendix C. Here $m = 2, d = 3, s_1 = 3, s_2 = 2$, the column vertex set $C = \{14, 15, 16\}$ since their neighbor sets $\{7, 8, 9\}, \{10, 12\}$, and $\{11, 13\}$ are mutually disjoint.

Vertices $1, 2, \ldots 6$, for instance, also have valency 3, but they cannot represent the first column-vertex (3-level column) since their neighbors are not disjoint. Now the first column-vertex is 14, its neighbor $V(14) = \{7, 8, 9\}$ (represent levels 0,1,2 in column 1) lead us to row-vertices 1,2; 3,5 and 4,6 respectively. The symbol vertices are $[[7, 8, 9], [10, 12], [11, 13]]$; those correspond to levels 0,1,2 in column 1, levels 0,1 in column 2 and levels 0,1 in column 3 of F. The array is obtained as

$$F = \begin{bmatrix} 0 & 0 & 1 & 2 & 1 & 2 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 \end{bmatrix}^T$$

3.3. Finding Canonical Graphs

For any colored graph G, denote by canon(G) the *canonical labeling graph* computed using the package *nauty* [13]. It consists of a vertex relabeling permutation, p, say and new adjacencies. Hence, canon(G) is determined fully by these adjacencies. The vertex-relabeling p is of the form $p = p_R\ p_S\ p_{C_1}\ p_{C_2} \cdots p_{C_m}$, where $p_R, p_S, p_{C_1},\ p_{C_2}, \ldots, p_{C_m}$ are permutations on sets $R, S, C_1, C_2, \ldots, C_m$ respectively. Indeed this fact follows from the requirement of preserving $m + 2$ color classes that we input to the *nauty* computation. We define $G_F := \Phi(F)$ and $G_D := \Phi(D)$ be the colored graphs of arrays F and D respectively.

As a result of Lemma 3.8, we have the following corollary.

Corollary 3.10. *F and D are isomorphic arrays* $\Longleftrightarrow$ canon$(G_F) =$ canon(G_D).

Notice that if $G \in \mathcal{G}_{U,N}$ then canon$(G) \in \mathcal{G}_{U,N}$. Let D^* be the *canonical labeling orthogonal array* of an orthogonal array D. Then $G_D \in \mathcal{G}_{U,N}$, and $G_{D^*} \in \mathcal{G}_{U,N}$. Now D^* can be constructed using the scheme below:

$$D \longrightarrow G_D \longrightarrow \text{canon}(G_D) \longrightarrow D^*,$$

in which the first arrow represents the mapping Φ. The third arrow computing D^*, is done by the demerging map Φ^{-1}. For orthogonal arrays of strength $t \geq 2$, the canonical array D^* is uniquely determined by canon(G_D).

3.4. Computing Canonical Orthogonal Arrays

We may build the canonical orthogonal array D^* from the adjacencies of the graph canon(G_D) that came from *nauty*. Since the relabeling permutation p preserves color classes, we do not need to rearrange vertices in the canonical graph canon(G_D). We can apply the demerging scheme (using the demerging mapping). But if we list adjacencies of vertices in G_D in the order: rows R, symbols S, columns C, then we can also do the following:

(i) Locate column-vertices: Column-vertices in canon(G), denoted by Cv, occupy rows from $N + n_S + 1$ to $n := |V|$ of B;
(ii) specify row-vertices: row-vertices occupy rows from 1 to N;
(iii) from row-vertices we are able to build up the array D^* row by row by tracking the symbol-vertices which are listed in the corresponding row. Notice that levels of each column must be numbered in the decreasing order, but not necessarily between columns.

Example 3.11. Let D be an the full $OA(16; 4^1 \cdot 2^2; 3)$. Then the run size $N = 16$, the number of factors $d = 3$, there are $n_S = 8$ symbol vertices, there are $m = 2$ distinct levels, so the vertices

$$V = R \cup S \cup C = \big\{\{1, 2, \ldots, 15, 16\}, \{17, \ldots 20, 21, 22, 23, 24\}, \{25, 26, 27\}\big\}.$$

The color classes have sizes 16, 8, 1, 2, with the corresponding vertices

$$\begin{aligned} f := \big\{&\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}, \\ &\{17, 18, 19, 20, 21, 22, 23, 24\}, \{25\}, \{26, 27\}\big\}. \end{aligned}$$

The relabeling permutation $p = (2, 3)(6, 9, 7, 13, 14, 8)(10, 11, 15, 12)(22, 23, 24)$, the column vertices $Cv = [25, 26, 27]$, and the symbol-vertices

$$Sv = \big[[17, 18, 19, 20], [21, 22], [23, 24]\big].$$

For the row $u = [17, 22, 24]$, we refer to symbol-vertices, ie, symbols 0 in column 1, symbol 1 in column 2, and symbol 1 in column 3. We obtain its companion run $[0, 1, 1] \in D^*$. The new adjacencies of the canonical graph are given in Table 2.

4. Transformations (Isomorphisms) of Orthogonal Arrays

It is not immediately obvious how to define isomorphisms of a factorial design, given in Definition 1.1. In fact, there is more than one sensible definition that could be made. We give the definition that is most useful for our purposes in this section, see Appendix B for generic concepts.

Recall that $T := r_1 \cdot r_2 \cdots r_d$ is a design type, equivalently we could group a_i factors with the same s_i levels in $T := s_1^{a_1} \cdot s_2^{a_2} \cdots s_m^{a_m}$, $s_i \neq s_j$ when $i \neq j$. Denote by $\mathrm{OA}(N; T)$ the set of all OAs with given type T and run size $N \in \mathbb{N}$. Set

Table 2: Adjacency relations of a colored graph

17	21	22						
17	22	24						
17	21	23						
17	23	24						
18	21	22						
19	21	22						
20	21	22						
18	21	23						
18	22	24						
19	22	24						
20	22	24						
19	21	23						
20	21	23						
18	23	24						
19	23	24						
20	23	24						
1	2	3	4	25				
5	8	9	14	25				
6	10	12	15	25				
7	11	13	16	25				
1	3	5	6	7	8	12	13	26
1	2	5	6	7	9	10	11	27
3	4	8	12	13	14	15	16	27
2	4	9	10	11	14	15	16	26
17	18	19	20					
21	24							
22	23							

$U := \{(i, j, x) \mid i = 1, \dots, N,\ j = 1, \dots, d,\ x \in Q_j\}$, and call it the *underlying set* of $\mathrm{OA}(N;T)$. In other words, U consists of all possible triples of a row i, a column j, and an entry F_{ij} for any matrix $F \in \mathrm{OA}(N;T)$. The k-th column index set $J_k \subseteq \mathbb{N}_d := \{1, 2, \cdots, d\}$ precisely consists of column indices of factors having s_k levels, for each $k = 1, ..., m$.

We can now encode any $F \in \mathrm{OA}(N;T)$ by its *lookup table*

$$Lt(F) := \{(i, j, F_{ij}) \mid i = 1, \dots, N,\ j = 1, \dots, d\} \subseteq U.$$

The *encoding map* Lt from $\mathrm{OA}(N;T)$ to the power set of U is clearly injective. The image of Lt consists of all sets $S \subseteq U$ with the following property:

$$\#\{x \mid (i, j, x) \in S\} = 1 \text{ for all } i = 1, \dots, N \text{ and } j = 1, \dots, d. \tag{4}$$

We next define group actions on the set U:

(i) The *row permutation group* is $R := \mathrm{Sym}_N$. It acts via $\phi_R : R \to \mathrm{Sym}(U)$ defined by

$$(i, j, x)^{\phi_R(r)} = (i^r, j, x).$$

(ii) The *column permutation group* is $C := \prod_{k=1}^m C_k$ where $C_k := \mathrm{Sym}(J_k)$. It acts via $\phi_C : C \to \mathrm{Sym}(U)$ defined by

$$(i, j, x)^{\phi_C(c)} = (i, j^c, x).$$

(iii) The *level permutation group* is $L := \prod_{j=1}^d L_j$, here $L_j = \mathrm{Sym}_{r_j}$, switching levels of all columns of F. Denote l_j by the projection of l onto L_j. Then L acts via $\phi_L : L \to \mathrm{Sym}(U)$ defined by

$$(i, j, x)^{\phi_L(l)} = (i, j, x^{l_j}).$$

Definition 4.1. *The full group G of fraction transformations of U is defined as*

$$G := \phi_R(R) \quad \phi_C(C) \quad \phi_L(L) \leq \mathrm{Sym}(U). \tag{5}$$

G is generated by the *isomorphism images* of the row, column and level permutation groups in $\mathrm{Sym}(U)$. By Property (4) we can prove that, for any array $F \in \mathrm{OA}(N;T)$ and $g \in G$, there exists a unique $F' \in \mathrm{OA}(N;T)$ with $Lt(F') = Lt(F)^g$. Hence, G acts faithfully on $\mathrm{OA}(N;T)$ via the mapping $\pi : G \to \mathrm{Sym}(\mathrm{OA}(N;T)) = \mathrm{Sym}(U)$ defined by

$$F^g = F^{\pi(g)} := Lt^{-1}\left(Lt(F)^g\right).$$

The newly defined G is a permutation group acting on the space $\mathrm{OA}(N;T)$.

To describe the structure of the group G, we need to know the relationship between the three types of permutations. It is clear that the column permutations $c \in C := \prod_{k=1}^m C_k$ and the level permutations $l \in L := \prod_{k=1}^m L_k$ act independently on distinct sections. As expected for isomorphisms, they preserve the strength of a fraction.

Proposition 4.2. [Properties of G] *Indeed we have the following properties:*

(i) *Column-Column relation. Column permutations in distinct sections commute, ie, $[C_k, C_h] = 1, \ \forall k \neq h$.*

(ii) *Level-Level relation. Level permutations of columns in distinct sections commute, ie, $[L_k, L_h] = 1, \ \forall k \neq h$.*

(iii) *Row-Column relation. Row permutations commute with column ones, ie,*

$$[R, C] = 1. \tag{6}$$

(iv) *Row-Level relation. Row permutations commute with level ones, ie,*

$$[R, L] = 1. \tag{7}$$

(v) *Column-Level relation. Let* $c \in C$ *be a column permutation and let* $l = l_1 \cdots l_d$ *be a level permutation. Then* c *commutes with* l *if, and only if,* $l_i = l_j$ *whenever* i *and* j *are in the same cycle of* c.

As a result, the subgroup generated by column permutations in section k *and level permutations in that section is a wreath product*

$$\mathrm{Sym}_{s_k} \wr C_k = L_k \rtimes \mathrm{Sym}_{a_k}.$$

Proof. See [14, Proposition 34]. ∎

Corollary 4.3. *We see that* G *is nearly a direct product of symmetric groups and a wreath product of symmetric groups. Precisely, the full group* G *or the permutation group acting on the space* $\mathrm{OA}(N;T)$ *can be identified with the wreath product*

$$G = R \times (C \ltimes L), \quad \textit{where } C \ltimes L = \prod_{k=1}^{m} \mathrm{Sym}_{s_k} \wr C_k. \tag{8}$$

As a result, the order of G *can be calculated from OA parameters, as*

$$|G| = N!\, a_1! \cdots a_m!\, (s_1!)^{a_1} \cdots (s_m!)^{a_m}.$$

The next concept plays a crucial role in the remaining parts.

Definition 4.4. *Let* F *and* F' *be in* $\mathrm{OA}(N;T)$.

(i) *An isomorphism from* F *to* F' *is* $g \in G$ *such that* $F^g = F'$.
(ii) *The automorphism group of an orthogonal array* $F \in \mathrm{OA}(N;T)$ *is the normalizer of* F *in the group* G, *i.e.,* $\mathrm{Aut}(F) := \{g \in G \mid F^g = F\}$.
(iii) *Any subgroup* $A \leq \mathrm{Aut}(F)$ *is called a group of automorphisms of* F.

5. An Integer Linear Formulation for the Design Extension

We now formulate necessary conditions for extending a known orthogonal array $F = OA(N; r_1 \cdots r_d; t)$ of strength t by a factor X to get a new design $[F|X]$ with the same strength. Assume $t = 3$, given an array $F = OA(N; r_1 \cdots r_d; 3)$ with columns $S_1, \ldots, S_d$, where S_i has r_i levels ($i = 1, \ldots, d$).

An s-level factor X is orthogonal to a pair of factors (S_i, S_j) of F, written $X \perp [S_i, S_j]$, if the frequency of all tuples $(a, b, x) \in [S_i, S_j, X]$ is $N/(r_i r_j s)$. Extending F by X means constructing an $OA(N; r_1 \cdots r_d \cdot s; 3)$, denoted by $[F|X]$. By the definition of OAs, $[F|X]$ exists if and only if X is orthogonal to any pair of columns of F. We can find a set P of necessary constraints for the existence of array $[F|X]$ in terms of polynomials in the coordinate indeterminates of X, by the following rules.

(i) Calculate frequencies of 3-tuples, and locate positions of pairs of (S_i, S_j).

(ii) Set the sums of coordinate indeterminates of X (corresponding to these positions) equal to the product of those frequencies with the constant $0+1+2+\ldots+s-1=\frac{s(s-1)}{2}$. The number of equations of P then is $\sum_{i\neq j}^{d} r_i r_j$, since each pair of (S_i, S_j) can be coded by a new factor with $r_i\ r_j$ levels. If $s=2$, the constraints P are in fact the sufficient conditions for the existence of X.

For instance, let $F = OA(16; 4\cdot 2^2; 3) = [S_1|S_2|S_3]$ be a full design. By transformation rule (ii), the sums of coordinates of X corresponding to the $Y=[S_1,S_2]$ symbols and the $Z=[S_2,S_3]$ symbols must equal a multiple of the appropriate frequencies. That means:

$$
\begin{aligned}
X \perp [S_1, S_2] &\Leftrightarrow X \perp Y \Leftrightarrow x_1+x_2 = \ldots = x_{15}+x_{16} = \lambda\cdot(0+1) = 1, \ldots,\\
X \perp [S_2, S_3] &\Leftrightarrow x_1+x_5+x_9+x_{13} = \ldots = x_4+x_8+x_{12}+x_{16}\\
&\qquad = \mu\cdot(0+1) = 2.
\end{aligned}
$$

One solution of P is given in the last row of the matrix below:

$$
\left[\begin{array}{cccccccccccccccc}
0&0&0&0&1&1&1&1&2&2&2&2&3&3&3&3\\
0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1\\
0&1&0&1&0&1&0&1&0&1&0&1&0&1&0&1\\
\hline
0&1&2&3&0&1&2&3&0&1&2&3&0&1&2&3
\end{array}\right]^{T}.
$$

Generally, the set P of linear constraints with integer coefficients is described by the matrix equation $AX=b$, in which $A\in \mathrm{Mat}_{\mathrm{m_1,N}}(\mathbb{N})$,

$$
X = (x_1,\ldots,x_N) \in \{0,1,\ldots,s-1\}^N \subseteq \mathbb{N}^N \tag{9}
$$

is a vector of unknowns, $b\in\mathbb{N}^{m_1}$, and $m_1 := \sum_{i\neq j}^{d} r_i r_j = |P|$. Since each orthogonal array is isomorphic to an array having the first row zero, we let $x_1=0$ throughout. By Gaussian elimination, we get the reduced system

$$
M\,X = c, \tag{10}
$$

where $M\in\mathrm{Mat}_{m,N}(\mathbb{Z})$, the set of all $m\times N$ ($m\le m_1$) matrices with integral entries, $c\in\mathbb{Z}^m$, and the vector of unknowns $X=(0,x_2,\ldots,x_N)\ \in\mathbb{Z}^N$.

The extension

$$
K := [F|X] = OA(N; r_1\cdots r_d\cdot s; t)
$$

clearly depends on solving the integer linear system (10) $M.X=c$ in terms of $X=(x_j)\in\{\,0,1,\ldots,s-1\,\}^N$ for $j=1,\ldots,N$. This approach is useful if a few constraints, structures or pruning techniques would be found and used to delete out some (not all) isomorphic vectors in each isomorphic class, and we then retain isomorph-free vectors. From that point, the search for all isomorph-free designs becomes feasible.

Fix an array $F\in\mathrm{OA}(N;T;t)$, recall from Definition 4.4 that the automorphism group of F is $\mathrm{Aut}(F) := \{g\in G \mid F^g = F\}$, where G is the full group of isomorphisms, see Equation (5). We first define the row permutation group of a fractional design F.

Let $g \in \mathrm{Aut}(F)$. Then g induces a permutation g_1 in the full group G_K of K, see Formula (8). Let g_R be the row permutation component of g, then g_R is also the row permutation component of g_1. Due to Definition 4.4, we have

Theorem 5.1. *For an automorphism $g \in \mathrm{Aut}(F)$, g induces a row permutation $g_1 \in G_K$ and generates the image K^{g_1} which is isomorphic to K.*

Proof. Formula (5) says any permutation g acting on F has the decomposition $g = g_R\, g_C\, g_S$ where g_C and g_S are the column and symbol permutations acting on F, respectively. Besides, the row permutation g_R induces a row permutation $g_1 \in G_K$, we furthermore have

$$K^{g_1} = [F|X]^{g_1} = [F^g|X^{g_R}] = [F|X^{g_R}] \tag{11}$$

since g already fixes F, and only g_R acts on the column X by moving its coordinates. As a result, $K^{g_1} = [F|X^{g_R}]$ is isomorphic to $K := [F|X]$. ■

Definition 5.2. *Let $H := \mathrm{Row}(\mathrm{Aut}(F))$ be the group of all row permutations g_R extracted from the group* $\mathrm{Aut}(F)$. *We call H the* row permutation group *of F.*

The *direct product* of H and τ is very useful for pruning later on, given by

$$\sigma := H \times \tau, \tag{12}$$

where $\tau := \mathrm{Sym}_s$, the symbol permutation group acting on the X's coordinates.

6. Localizing the Formation of Vector Solutions

It is now obvious that, by recursion, the process of building vector solution X can be brought back to strength 1 derived designs. We can effectively prune $Z(P)$ from those smallest sub-designs by searching for some subgroups of the row permutation group $H = \mathrm{Row}(\mathrm{Aut}(F))$ acting on strength 1 derived designs. Those subgroups, discussed in next parts, must have the property that they act separately on the row-index sets corresponding to the derived designs.

Fix $I_N := [1, 2, \ldots, N]$ the row-index list of F, and recall $r_1 \geq r_2 \geq \ldots \geq r_d$. We explicitly distinguish the list I_N with $\{1, 2, \ldots, N\}$ in this section. Then H acts naturally on X' indices. Furthermore, we employ the following.

Definition 6.1. *We say a row permutation $g_R \in H$* acts fixed-point free*, or globally on X if it moves every index from the whole set $\{1, 2, \ldots, N\}$.*

Otherwise, if the moved points of g_R form a proper subset J of $\{1, \ldots, N\}$, i.e., it fixes point-wise the complement 'list' of J in I_N, we say g_R acts locally *at that subset J.*

The first step is to localize the formation of a vector X of the form (9) by taking the derived designs of strength $t-1$. We get the r_1 derived designs $F_1, \ldots, F_{r_1}$, each of which is an $OA(r_1^{-1}N; r_2 \cdots r_d; t-1)$. Clearly, if a solution vector X exists, then it is formed by r_1 sub-vectors u_i of length $\frac{N}{r_1}$:

$$X = [u_1; u_2; \ldots; u_{r_1}], \text{ where } u_i = \left(x_{\frac{(i-1)N}{r_1}+1}, \ldots, x_{\frac{iN}{r_1}} \right). \tag{13}$$

Denote by V_i the set of all sub-vectors u_i which can be added to the ith derived design F_i to form an $OA(r_1^{-1}N; r_2 \cdots r_d \cdot s; t-1)$. Let $V = V_1 \times V_2 \times \ldots \times V_{r_1}$.

We propose an algorithm for finding all non isomorphic solution $X \in V$.

Algorithm 1 Find all non isomorphic vectors X in $[F|X]$

EXTEND-ONE-FACTOR(F)

Input: F is a strength t design;

Output: All non-isomorphic extensions of F to $[F|X]$

a/ Find all candidate sub-vectors $u_i \in V_i$, $i = 1, \ldots, r_1$, using associated permutation subgroups

b/ Discard (prune) them as many as possible by using subgroups of H

c/ Plug those u_is together, then compute the representatives of the $\sigma = H \times \tau$-orbits in V, the solution space $Z(P)$ of P.

6.1. Forming Permutation Subgroups of the Derived Designs

We viewed $F \in OA(N; r_1 \cdot r_2 \cdots r_d; 3)$ as an $N \times d$-matrix with the $[l,j]$-entry is written as $F[l,j]$. For each derived design F_i w. r. t. the first column of F, the row-index set of F_i, denoted by RowInd(F_i) for $1 \le i \le r_1$, is defined as

$$\text{RowInd}(F_i) := \big\{ l \in \{1, 2, \ldots, N\} : F[l,1] = i-1 \big\}.$$

Definition 6.2. *The stabilizer in H of F_i is defined by*

$$\begin{aligned} N_H(F_i) &:= \text{Normalizer}\big(H, \text{RowInd}(F_i)\big) \\ &= \big\{ h \in H : \text{RowInd}(F_i)^h = \text{RowInd}(F_i) \big\}. \end{aligned} \tag{14}$$

In this way, we find r_1 subgroups of H corresponding to the derived designs F_i. But it can happen that RowInd$(F_l)^h \ne$ RowInd(F_l) for some $h \in N_H(F_i)$ and $1 \le l \ne i \le r_1$. To make sure that the row permutations act independently on the F_i, we need the following structure.

Definition 6.3. *The group of row permutations acting locally on each F_i is defined as:*

$$L(F_i) := \text{Centralizer}\big(N_H(F_i), J(F_i)\big), \tag{15}$$

where $J(F_i) := I_N \setminus \text{RowInd}(F_i)$ is the sublist of I_N consisting of elements not in $\text{RowInd}(F_i)$. The $L_i := L(F_i)$ acts locally on $\text{RowInd}(F_i)$, i.e. it acts on the row-indices of F_i and fixes pointwise any row-index outside F_i. These subgroups L_i - of the group $H = \text{Row}(\text{Aut}(F))$ - are called the row permutation subgroups associated with strength 2 derived designs.

These subgroups can be determined further as follows.

For an integer $m = 1, 2, \ldots, t-1$ and for $j = 1, 2, \ldots m$, denote by

$$F_{i_1,\ldots,i_m} := OA\bigg(\frac{N}{r_1 r_2 \cdots r_m}; r_{m+1} \cdots r_d; t-m\bigg) \tag{16}$$

the derived designs of F taken with respect to symbols $i_1, \ldots, i_m$, where symbol i_j in column j and $i_j = 1, \ldots, r_j$. Define the row-index set of $F_{i_1,\ldots,i_m}$ by

$$\text{RowInd}(F_{i_1,\ldots,i_m}) := \bigcap_{j=1}^{m} \big\{l \in \{1, 2, \ldots, N\} : F[l, j] = i_j - 1\big\}. \tag{17}$$

Definition 6.4. *Let $J(F_{i_1,\ldots,i_m}) := I_N \setminus \text{RowInd}(F_{i_1,\ldots,i_m})$. Generalizing (14) and (15) gives:*

$$\begin{aligned} N_H(F_{i_1,\ldots,i_m}) &:= \text{Normalizer}\big(H, \text{RowInd}(F_{i_1,\ldots,i_m})\big), \\ L(F_{i_1,\ldots,i_m}) &:= \text{Centralizer}\big(N_H(F_{i_1,\ldots,i_m}), J(F_i)\big), \text{ for } 1 \le i_j \le r_j. \end{aligned} \tag{18}$$

$L(F_{i_1,\ldots,i_m})$ is called the subgroup associated with the derived design $F_{i_1,\ldots,i_m}$. We say $L(F_{i_1,\ldots,i_m})$ acts locally on the derived design $F_{i_1,\ldots,i_m}$, and write $L_{i_1,\ldots i_m} := L(F_{i_1,\ldots,i_m})$, for $1 \le i_j \le r_j$, $j = 1, 2, \ldots m$.

For $t = 3$, we compute these subgroups for $m = 1$ and $m = 2$. If $m = 1$, we have s_1 subgroups $L(F_i)$ acting locally on strength 2 derived designs; and if $m = 2$, then s_1 s_2 subgroups $L(F_{i,j})$ acting locally on strength 1 designs.

6.2. Using Permutation Subgroups of the Derived Designs

We now show how to use the subgroups $L_{i_1,\ldots,i_m}$. Recall that $Z(P)$ is the set of all natural solutions X. From Equation (11) in Theorem 5.1,K^g is an isomorphic array of $K = [F|X]$, hence the vector X^g can be pruned from $Z(P)$, for any solution X and any permutation $g \in \text{Aut}(F)$.

We use the following notations in the remaining parts. For a fixed m-tuple of symbols $i_1, \ldots, i_m$, let $V_{i_1,\ldots,i_m}$ be the set of solutions of fraction

$$F_{i_1,\ldots,i_m} = OA((r_1 r_2 \cdots r_m)^{-1} N; r_{m+1} \cdots r_d; t-m), \text{ for } 1 \le m \le t-1.$$

For any sub-vector $u \in V_{i_1,\dots,i_m}$, from (17) and (13), let

$$\begin{aligned} I(u) &:= \mathrm{RowInd}(F_{i_1,\dots,i_m}); \quad J(u) := I_N \setminus I(u), \\ Z(u) &:= \big\{(x_j) : j \in J(u) \text{ and } \exists X \in Z(P) \text{ s.t. } X[I(u)] = u\big\}, \end{aligned}$$

here $X[I(u)] := (x_i : i \in I(u))$. For instance, if $m = 1$ and $u \in V_1$ then

$$Z(u) \ = \big\{\, [u_2; \dots; u_{r_1}] : X = [u; u_2; \dots; u_{r_1}] \in Z(P) \,\big\}.$$

Theorem 6.5. (Main Theorem) *For any pair of sub-vectors $u, v \in V_{i_1,\dots,i_m}$, if $v = u^{g_R}$ for some row permutation $g_R \in L_{i_1,\dots,i_m}$, we have $Z(u) = Z(v)$.*

We prove this theorem in two claims. In Theorem 6.6, without loss of generality, it suffices to give the proof for the first strength 2 derived array. Then Theorem 6.7 shows the induction step.

Theorem 6.6. [Case $m = 1$] *Let u_1 and v_1 be two arbitrary sub-solutions in V_1, ie, they form strength 2 OAs $[F_1|u_1]$ and $[F_1|v_1]$ of the form $OA(r_1^{-1}N; r_2 \cdots r_d \cdot s; 2)$. Let*

$$\begin{aligned} Z_X(u_1) &= \big\{\, [u_2; \dots; u_{r_1}] : X = [u_1; u_2; \dots; u_{r_1}] \in Z(P) \,\big\}, \\ Z_Y(v_1) &= \big\{\, [v_2; \dots; v_{r_1}] : Y = [v_1; v_2; \dots; v_{r_1}] \in Z(P) \,\big\}. \end{aligned}$$

Suppose that there exists a nontrivial subgroup, say $L(F_1)$, and if $v_1 = u_1^h$ for some $h \in L_1$, we have $Z_X(u_1) = Z_Y(v_1)$.

Proof. Pick up a nontrivial permutation h in $L(F_1)$. Then it acts locally on $\mathrm{RowInd}(F_1)$. By symmetry, we just check that $Z_X(u_1) \subseteq Z_Y(v_1)$. We choose any sub-vector

$$\boldsymbol{u}^* := [u_2; \dots; u_{r_1}] \in Z_X(u_1).$$

Then $X = [u_1; u_2; \dots; u_{r_1}]$ is in $Z(P)$. We view $h \in \mathrm{Aut}(F)$, so

$$\begin{aligned} D^h &= [F|X]^h = \big[F^h|\, X^h\big] = \big[F|\, X^h\big] = \big[F|\, [u_1; u_2; \dots; u_{r_1}]^h\big] \\ &= \big[F|\, [u_1^h; u_2; \dots; u_{r_1}]\big] = \big[F|\, [v_1; u_2; \dots; u_{r_1}]\big]. \end{aligned}$$

This implies that $[v_1; u_2; \dots; u_{r_1}]$ is a solution. Hence $u^* \in Z_Y(v_1)$. ∎

As a result, we can wipe out all solutions $Y = [v_1; v_2; \dots; v_{r_1}] \in Z(P)$ if $v_1 \in \ u_1^{L_1}$, the L_1- orbit of u_1 in V_1. In other words, if we get $V_1 \neq \emptyset$, then it suffices to find the first sub-vector of vector X by selecting $|V_1|/|L_1|$ representatives u_1 from the L_1- orbits in V_1. Furthermore, the above proof is independent of the original choice of derived design. Hence it can be done simultaneously at all solution sets $V_1, V_2, \dots, V_{r_1}$, using the subgroups $L_1, \dots, L_{r_1}$.

We call this procedure, that results from the **Main Theorem**, the *local pruning process* using strength 2 derived designs. Next, if $t \geq 3$ we extend the proof of Theorem 6.6 to cases $2 \leq m \leq t-1$.

Theorem 6.7. [Case $m > 1$] *For any pair of sub-vectors $u, v \in V_{i_1,i_2}$, if $v = u^{g_R}$ for some $g_R \in L_{i_1,i_2}$, we have $Z(u) = Z(v)$.*

Proof. See Appendix A for a proof. ∎

6.3. Operations on Derived Designs

The above localizing idea can be enhanced further when we consider each derived design as an agent that receives data from its lower strength derived designs, make some appropriate operations, then pass the result to its parent design. We name this operation an agent-based localization. Specifically, notice that strength 1 and strength t designs require special operations. Precisely, at the global scale of strength t design, it suffices to find only the representatives of the $H \times \tau$-orbits [see Formula (12)] in the solution space $Z(P)$ of P.

We now formalize our new agent-based localization. Recall from Formula (16) that the symbols $i_1, \ldots, i_m$ $(1 \leq i_j \leq r_j)$ indicate the derived design having symbol i_j in column j, for $j = 1, \ldots, m$. From Equation (18), $L_{i_1,\ldots,i_m}$ are the subgroups associated with the derived designs $F_{i_1,\ldots,i_m}$ having strength $t-m$. When $m = t-1$, write $L_{i_1,\ldots,i_{t-1}}$ for the subgroup associated with the strength 1 derived design $F_{i_1,\ldots,i_{t-1}}$. The agents of derived designs can be described as follows.

At initial designs $F_{i_1,\ldots,i_{t-1}}$ (Initial step when $m = t-1$):
Input: $F_{i_1,\ldots,i_{t-1}}$;
Operation:
1. form $V_{i_1,\ldots,i_{t-1}}$, the set of all strength 1 vectors of length $(r_1 r_2 \cdots r_{t-1})^{-1}N)$ being appended to $F_{i_1,\ldots,i_{t-1}}$,
2. compute $L_{i_1,\ldots,i_{t-1}}$, and
3. find the representatives of $L_{i_1,\ldots,i_{t-1}}$- orbits in the set $V_{i_1,\ldots,i_{t-1}}$;

Output: these representatives, ie, solutions of $F_{i_1,\ldots,i_{t-1}}$.

At strength k derived designs $(1 < k \leq t-1)$: let $m := t-k$, we have
Input: the solutions having length $(r_1 r_2 \cdots r_m \cdot r_{m+1})^{-1}N$ of strength $k-1$ sub-designs; and the subgroup $L_{i_1,\ldots,i_m}$;
Operation:
1. form sub-solutions having length $(r_1 r_2 \cdots r_m)^{-1}N)$ of $F_{i_1,\ldots,i_m}$,
2. prune these solutions by $L_{i_1,\ldots,i_m}$;

Output: representatives of the $L_{i_1,\ldots,i_m}$- orbits in the set $V_{i_1,\ldots,i_m}$.

At the (global) design F:

Input: the sub-vectors from strength $t-1$ derived designs;

Operation: find the representatives of σ-orbits in the Cartesian product $V = V_1 \times V_2 \times \ldots \times V_{r_1} = \{\text{vectors } X \text{ of length } N\}$ where V_i had been already pruned by the subgroup L_i $(i = 1, 2, \ldots, m)$;

Output: Two steps

a/ (Isomorph-free test 1) returns solution vectors X which are non-isomorphic up to $\sigma = H \times \tau$, see Equation (12);

b/ (Isomorph-free test 2) forms orthogonal arrays $K = [F|X]$ of the same strength t, then select only non-isomorphic arrays, by computing their canonical arrays, (see Section 3).

We brief ideas in Algorithm 2, **Pruning-Uses-Symmetry**(F, d).

Algorithm 2 Pruning uses subgroups of derived designs

Pruning-Uses-Symmetry(F, d)

Input: F is a strength t design; d is the number of columns required

Output: All non-isomorphic extensions of F

⋄ STEP 1: *Local pruning at strength k derived designs.*

1a) Find sub-vectors of $F_{i_1,\ldots,i_m}$, for $m := t-k$, and $k = 1, \ldots, t-1$,

1b) prune these sub-vectors locally and simultaneously by using $L_{i_1,\ldots,i_m}$,

1c) concatenate these sub-vectors to get sub-vectors in $V_{i_1,\ldots,i_{m-1}}$.

Comment: For $t = 3$, in Step 1), form subvectors $u_{i,j} \in V_{i,j}$ simultaneously at the $r_1 r_2$ sets $V_{i,j}$, then concatenate $u_{i,j}$ $(1 \le i \le r_1, 1 \le j \le r_2)$ to get $u_i \in V_i$.

⋄ STEP 2: *Pruning at strength t design F.*

2a) Select the representative vectors X from the $\sigma = H \times \tau$-orbits of V
Comment: Each vector in V is formed by sub-vectors found from Step 1

2b) append non-isomorphic vectors X to F to get strength t OAs $[F|X]$,

2c) get back non-isomorphic arrays into a list Lf, return Lf (find non-isomorphic OAs by computing distinct canonical arrays, see Section).

⋄ STEP 3: *Repeating step.*

`If` # current columns $< d$ Call **Pruning-Uses-Symmetry**(f, d) for $f \in Lf$
`Else` Return Lf `EndIf`

Example 6.8. Let $U := \big[[3,1],[2,3]\big]$, $F = OA(24; 3.2^3; 3)$,

$$F = \begin{bmatrix} 0&0&0&0&0&0&0&0&1&1&1&1&1&1&1&1&2&2&2&2&2&2&2&2 \\ 0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1&0&0&0&0&1&1&1&1 \\ 0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1&0&0&1&1 \\ 0&0&1&1&1&1&0&0&0&1&0&1&0&1&0&1&1&1&0&0&0&0&1&1 \end{bmatrix}^T.$$

Aut(F) has order 12288. Compute the group H = Row(Aut(F)) (from Definition), and update H = Stabilizer(H, [1]), which is a permutation group of size 768. The three strength 2 derived designs give 8, 8, and 16 candidates respectively, so we must check $8.8.16 = |V| = 1024$ cases. The row permutation subgroups of these strength 2 derived designs with orders 8, 1, 16 are

$$
\begin{aligned}
L_0 &= [(), (7,8), (5,6), (5,6)(7,8), (3,4), (3,4)(7,8), (3,4)(5,6), (3,4)(5,6)(7,8)],\\
L_1 &= [()], L_2 = [(), (23,24), (21,22), (21,22)(23,24), (19,20), (19,20)(23,24),\\
&\quad (19,20)(21,22), (19,20)(21,22)(23,24), (17,18), (17,18)(23,24),\\
&\quad (17,18)(21,22), (17,18)(21,22)(23,24), (17,18)(19,20),\\
&\quad (17,18)(19,20)(23,24), (17,18)(19,20)(21,22),\\
&\quad (17,18)(19,20)(21,22)(23,24)].
\end{aligned}
$$

The subspaces are pruned to 1, 8, 1 vectors respectively; we then check 8 cases.

7. Summary And Closing Comments

A few unknown mixed OAs that previous well-known methods failed to construct (e.g. the strength 3 mixed balanced design $OA(96; 6 \cdot 4^2 \cdot 2^c; 3)$, with $c > 5$, see Nguyen [2, 15]), now can be found by our combined approach (of graph and group-theoretic methods). Some of their non-isomorphic arrays are listed in the following Table 3.

Table 3: **New strength 3 mixed OAs of sizes** $N \leq 100$.

N	Type; Strength t	#	Size of the group Aut(F)	Methods
⋮				
80	$5 \cdot 4 \cdot 2^6$; $t = 3$	≥ 5	$2^2, 4^3$	(IS)
96	$8 \cdot 2^{12}$	0		(Rao)
96	$6 \cdot 4^2 \cdot 2$	≥ 4	32, 64, 256, 9216	(IS)
96	$6 \cdot 4^2 \cdot 2^2$	≥ 249	2^{58}, 4^{65}, 8^{56}, 16^{42}, 32^{19}, 64^7, 128, 256	(IS)
96	$6 \cdot 4^2 \cdot 2^3$	≥ 29987		(IS)
96	$6 \cdot 4^2 \cdot 2^4$	≥ 7895	1^{1520}, 2^{3649}, 4^{2265}, 8^{403}, 16^{52}, 24, 32^5	(IS)
96	$6 \cdot 4^2 \cdot 2^5$; $t = 3$	≥ 1199	1^{411}, 2^{370}, 4^{250}, 8^{137}, 12, 16^{29}	(IS)
96	$6 \cdot 4^2 \cdot 2^6$; $t = 3$	≥ 8	2^2, 4^2, 8^4	(IS)

In [15], by the Latin squares method, only one $OA(80; 5 \cdot 4 \cdot 2^6; 3)$ and one $OA(96; 6 \cdot 4^2 \cdot 2^5; 3)$ were found. For the most interesting one with size at most 100, the series of $OA(96; 6 \cdot 4^2 \cdot 2^6; 3)$ can not be built up by the Latin squares and other methods, but thank to the group-theoretic approach we currently obtain at least 8 non-isomorphic OAs, and theirs distinct automorphism group sizes are 2, 4 and 8. We have used multiplicity notation for automorphism group orders. The (**IS**) construction means employing the **I**nteger linear formulation and **S**ymmetries of automorphism groups of OAs, fully developed in this paper.

We have discussed mathematical aspects of factor enlarging problem of OAs with strength at least 2, provided a fix number N of experiments. Our approach combining permutation groups and other formulations provides a generic framework for enumerating mixed OAs of any strength with all feasible factor levels and with run sizes N satisfying the Rao bound. However, we currently restrict checking the approach for sizes $N < 100$ experiments only.

The dual of the problem, namely fixing the factors and the strength, and try to find better lower bounds of the run sizes also is very interesting and challenging. Techniques from Bose-Mesner or Terwilliger algebras, in the excellent survey by Bannai et al. [1], and other approaches as semidefinite programming (see[11, 21]) could be promising leads to go.

Acknowledgement. The author appreciates partial supports by *Center of Excellence in Mathematics*, Thailand and *Faculty of Science*, Mahidol University, Thailand during completion of this article. He thanks the Organizers of the *Fifth Biannual International Conference on Group Theory* 2019 (5BIGTC- 2019, ITB, Indonesia) for motivating this write up.

Appendix A: A Proof of Theorem 6.7

For any pair of sub-vectors $u, v \in V_{i_1,i_2}$, if $v = u^{g_R}$ for some $g_R \in L_{i_1,i_2}$, we have $Z(u) = Z(v)$. We prove this result for $t = 3$ and $m = 2$ only. For arbitrary $t > 3$, and $m > 2$, the proof is a straightforward generalization.

(i) Similar to the proof of Theorem 6.6, without loss of generality, we consider the first derived design $F_1 = OA(n; r_2 \cdots r_d; 2)$ where $n = N/r_1$. Taking derived designs of F_1 with respect to the second column (having r_2 levels), we get r_2 strength 1 arrays, denoted by $f_1 := F_{1,1}$, $f_2 := F_{1,2}$, $\dots, f_{r_2} := F_{1,r_2}$, each is $OA(r_2^{-1}n; r_3 \cdots r_d; 1)$. Any u_1 in V_1 can be written as $u_1 = [u_{1,1}; u_{1,2}; \dots; u_{1,r_2}]$, a concatenation of r_2 sub-vectors $u_{1,j}$ of length $\frac{n}{r_2}$, where

$$u_{1,j} = \left(x_{\frac{(j-1)n}{r_2}+1}, \dots, x_{\frac{jn}{r_2}}\right) \quad \text{for } j = 1, \dots, r_2.$$

(ii) Known that the subgroup $L(f_j) := \text{Centralizer}\big(N_H(f_j), J(f_j)\big)$ [see from Equations (17) and (18)] consists of row permutations acting locally on

$$\text{RowInd}(f_j) = \left\{\frac{(j-1)n}{r_2} + 1, \dots, \frac{jn}{r_2}\right\}, \text{ for } j = 1, \dots, r_2.$$

Hence the subgroup $L(f_j)$ fixes $J(f_j) = [1, \dots, N] \setminus \text{RowInd}(f_j)$ pointwise.

(iii) Since V_1 is the Cartesian product of the subsets $V_{1,j} := \{ u_{1,j} \}$, we prove $V_{1,j}$ using $L(f_j)$, for all $j = 1, \ldots, r_2$. Start with $j = 1$. Let $u_{1,1}$, $v_{1,1}$ be two arbitrary sub-vectors in $V_{1,1}$. They can be used to make strength 1 arrays $[f_1|u_{1,1}]$ and $[f_1|v_{1,1}]$ being of the form $OA(r_2^{-1}n; r_3 \cdots r_d \cdot s; 1)$. Let

$$Z_X(u_{1,1}) := \big\{ \big[[u_{1,2}; \ldots; u_{1,r_2}]; u_2; \ldots; u_{r_1} \big] \big\},$$
$$Z_Y(v_{1,1}) := \big\{ \big[[v_{1,2}; \ldots; v_{1,r_2}]; v_2; \ldots; v_{r_1} \big] \big\},$$

for $X = [u_1; u_2; \ldots; u_{r_1}] \in Z(P)$, $Y = [v_1; v_2; \ldots; v_{r_1}] \in Z(P)$, where $v_1 = [v_{1,1}; v_{1,2}; \ldots; v_{1,r_2}] \in V_1$.

(iv) We prove if $v_{1,1} = u_{1,1}^h$ for some $h \in L(f_1)$, then $Z_X(u_{1,1}) = Z_Y(v_{1,1})$. In fact, we only need to have $Z_X(u_{1,1}) \subseteq Z_Y(v_{1,1})$. Let any sub-vector

$$\boldsymbol{u}^* := \big[\ [u_{1,2}; \ldots; u_{1,r_2}]; u_2; \ldots; u_{r_1} \big] \in Z_X(u_{1,1}),$$

and $h \in L(f_1)$. Then we have $X = [u_1; u_2; \ldots; u_{r_1}] \in Z(P)$, and

$$\begin{aligned} K^h &= [F|X]^h = F^h|\, X^h = F|\, X^h = F|\, [\, u_1^h; u_2; \ldots; u_{r_1} \,] \\ &= F|\, \big[[u_{1,1}^h; u_{1,2}; \ldots; u_{1,r_2}]; u_2; \ldots; u_{r_1} \big] \\ &= F|\, \big[[v_{1,1}; u_{1,2}; \ldots; u_{1,r_2}]; u_2; \ldots; u_{r_1} \big]. \end{aligned}$$

(v) $Y = \big[[v_{1,1}; u_{1,2}; \ldots; u_{1,r_2}]; u_2; \ldots; u_{r_1} \big]$ so is a solution and $\boldsymbol{u}^* \in Z_Y(v_{1,1})$. In F_1, the choice of f_j does not affect to the proof, so the pruning process can be applied at the same time for all f_j, $j = 1, \ldots, r_2$.

Appendix B: Group of Transformations of a Design

Given a set X, a *permutation* of X is a bijection from X to itself. We write $\mathrm{Sym}(X)$ for the *symmetric group* on X, ie, the group of all permutations of X. We denote Sym_N instead of $\mathrm{Sym}(\{1, 2, \ldots, N\})$, for a natural number N. We write elements of Sym_N in *cycle notation*, so the permutation $p = (1, 2, 3)(4, 5)$ is defined by $1^p = 2,\ 2^p = 3,\ 3^p = 1,\ 4^p = 5,\ 5^p = 4$. We say a group K *acts* on a set X if we have a group homomorphism $\phi : K \to \mathrm{Sym}(X)$. We abbreviate $x^{\phi(g)}$ by x^g. Let $p \in \mathrm{Sym}_N$. The *action of p on a subset* $B \subseteq \{1, 2, \ldots, N\}$ is given by $B^p := \{x^p : x \in B\}$. The *action of p on a list* of length N is given by

$$[y_1, y_2, \ldots, y_N]^p := [y_{1^{p^{-1}}}, y_{2^{p^{-1}}}, \ldots, y_{N^{p^{-1}}}].$$

In other words, we compute the ith position of Y^p by $Y^p[i] = y_{i^{p^{-1}}} = Y[i^{p^{-1}}]$.

Let X be the set of all structures of a particular combinatorial type built on an *underlying set* T. For example, X could be all the graphs with vertex set T.

The subgroup $G := G(T) \leq \mathrm{Sym}(T)$ which acts naturally on X is called the *(full) group of transformations* of X. Two elements A and B of X are *isomorphic* if they are in the same G*-orbit*, that is, there exists a permutation g in G such that $A = B^g$. The automorphism group of a design $S \in$ X is defined as

$$\mathrm{Aut}(S) := \{g \in G : S^g = S\}. \tag{19}$$

The number of distinct objects isomorphic to a structure S is the length of the G-orbit of S. By Lagrange's theorem [12], this number is $\frac{|G|}{|\operatorname{Aut}(S)|}$.

References

[1] E. Bannai, E. Bannai, H. Tanaka, Y. Zhu, Design theory from the viewpoint of algebraic combinatorics, *Graphs and Combinatorics* **33** (2017) 1–41.

[2] A. Brouwer, A. Cohen, M. Nguyen, Orthogonal arrays of strength 3 and small run sizes, *J. of Statistical Planning* & *Inference* **136** (2006) 3268–3280.

[3] D.A. Bulutoglu, F. Margot, Classification of orthogonal arrays by integer programming, *Journal of Statistical Planning and Inference* **138** (2008) 654–666.

[4] A. Chaudhuri, T.C. Christofides, C.R. Rao, *Handbook of Statistics*, Vol. **34**, North-Holland publications, Elsevier B.V., 2016.

[5] M.A. Fecko and M. Steinder, Combinatorial designs in multiple faults localization for battlefield networks, In: *IEEE MILCOM: Military Communications Conference*, IEEE Xplore, Vol. **2**, Vienna, 2001. doi: 10.1109/mil-com.2001.985975

[6] J.A. Feo, *Junran's Quality Manag. and Analysis*, McGraw-Hill, 2015.

[7] G.F.V. Glonek and P.J. Solomon, Factorial and time course designs for cDNA microarray experiments, *Biostatistics* **5** (2004) 89–111.

[8] S. Gupta, Balanced factorial designs for cDNA microarray experiments, *Communications in Statistics: Theory and Methods* **35** (8) (2006) 1469–1476.

[9] A.S. Hedayat, N.J.A. Sloane, J. Stufken, *Orthogonal Arrays*, Springer-Verlag, Germany, 1999.

[10] R.S. Kenett, S. Zacks, *Modern Industrial Statistics with Applications in R, Minitab*, **2nd** Ed., Wiley, 2014.

[11] M. Laurent, Strengthened semidefinite bounds for codes, *Journal Mathematical Programming* **109** (2-3) (2007) 239–261.

[12] M.B. Nathanson, *Elementary Methods in Number Theory*, Graduate Texts in Mathematics, Vol. **195**, Springer-Verlag, New York, 2000.

[13] B. McKay, Nauty, `cs.anu.edu.au/~bdm/nauty/`, *Australian Nat. Univ.* (2004)

[14] M.V.M. Nguyen, *Computer-Algebraic Methods for the Construction of Designs of Experiments*, Ph.D. Thesis, Technische Univ. Eindhoven, Netherlands, 2005.

[15] M.V.M. Nguyen, Some new constructions of strength 3 mixed orthogonal arrays, *Journal of Statistical Planning and Inference* **138** (1) (2008) 220–233.

[16] S.H. Park, *Six-Sigma for Quality and Productivity Promotion*, Asian Productivity Organization, Tokyo, 2003.

[17] M. Phadke, *Quality Engineering Using Robust Designs*, Prentice Hall, USA, 1989.

[18] H. Phan, B. Soh, M.V.M. Nguyen, A step-by-step extending parallelism approach for enumeration of combinatorial objects, In: *Algorithms and Architectures for Parallel Processing - 10th International Conference (ICA3PP)*, Springer- Verlag Berlin Heidelberg, 2010.

[19] H. Phan, B. Soh, M.V.M. Nguyen, A parallelism extended approach for the enumeration of orthogonal arrays, In: *Algorithms and Architectures for Parallel Processing - 11th International Conference (ICA3PP)*, Springer- Verlag Berlin Heidelberg, 2011.

[20] E.D. Schoen, P.T. Eendebak, M.V.M. Nguyen, Complete enumeration of pure-level and mixed-level orthogonal array, *Journal of Combinatorial Designs* **18** (2) (2010) 123–140.

[21] A. Schrijver, New code upper bounds from the Terwilliger algebra, *IEEE Transactions on Information Theory* **51** (8) (2005) 2859–2866.

[22] N.J.A. Sloane, *The On-Line Encyclopedia of Integer Sequences (OEIS)*, The OEIS Foundation Inc. `http://oeis.org/`

[23] J. Stufken and B. Tang, Complete enumeration of 2-level orthogonal arrays of strength D with $D+2$ constraints, *The Annals of Statistics* **35** (2) (2008) 793–814.

[24] C.F.J. Wu, M.S. Hamada, *Experiments: Planning, Analysis, and Parameter Design Optimization*, Wiley-Interscience, MR1780411, USA, 2000.