Minimum-Size Mixed-Level Orthogonal Fractional Factorial Designs Generation: A SAS-Based

Journal of Statistical Software

May 2013, Volume 53, Issue 10. http://www.jstatsoft.org/

Minimum-Size Mixed-Level Orthogonal Fractional Factorial Designs Generation: A SAS-Based Algorithm

Roberto Fontana
Politecnico di Torino

Sabrina Sampò
Politecnico di Torino

Abstract

Orthogonal fractional factorial designs (OFFDs) are frequently used in many fields of application, including medicine, engineering and agriculture. In this paper we present a software tool for the generation of minimum size OFFDs. The software, that has been implemented in SAS/IML, puts neither a restriction on the number of levels of each factor nor on the orthogonality constraints. The algorithm is based on the joint use of polynomial counting function and complex coding of levels and follows the approach presented in Fontana (2013).

Keywords: design of experiments, orthogonal fractional factorial design, orthogonal array, SAS.

1. Introduction

In this paper we present a software tool for generating minimum size orthogonal fractional factorial designs (OFFDs). The algorithm is based on the approach described in Fontana (2013). A preliminary version of it has been developed in the unpublished master's degree thesis Sampò (2011).

In this Section we give an overview of the use of OFFDs in practical applications and of the problems related to their generation. In Section 2 we briefly review the algebraic theory of OFFDs based on polynomial counting functions and *strata*. In Section 3 and Section 4 we describe the algorithm. Some applications of the algorithm are presented in Section 5. Finally, concluding remarks are given in Section 6. Section 1 and Section 2 are closely based on Section 1 and Section 2 of Fontana (2013). We include them here for completeness.

OFFDs are frequently used in many fields of application, including medicine, engineering and

agriculture. They offer a valuable tool for dealing with problems where there are many factors involved and each run is expensive. They also keep the statistical analysis of the data quite simple. The literature on the subject is extremely rich. A non-exhaustive list of references, mainly related to the theory of the design of experiments, includes the fundamental paper of Bose (1947) and the following books: Raktoe, Hedayat, and Federer (1981), Collombier (1996), Dey and Mukerjee (1999), Wu and Hamada (2000), Mukerjee and Wu (2006) and Bailey (2008).

Orthogonal arrays (OAs) represent an important class of OFFDs, see, for example, Hedayat, Sloane, and Stufken (1999) and Schoen, Eendebak, and Nguyen (2010). Indeed an OA of appropriate strength can be used to solve the wide range of problems related to the study of the size of the main effects and interactions up to a given order of interest.

It is evident that in many real-life experiments, finding an OFFD with the smallest possible number of runs is of great importance. This is particularly true in the case where the cost of each experiment is high in terms of resources and/or time, such as in the study of the relationship between fuel consumption and the design parameters of a new car engine.

A large number of techniques are known to generate OFFDs and, in particular, OAs. For example:

- The case where all factors have the same number p of levels and p is a prime number or a power of a prime number, is commonly studied using Galois Fields and finite geometries.
- Hadamard matrices are used for OAs where all factors have 2 levels and where strength is less than or equal to 3.
- Difference schemes are a tool for constructing mixed OAs of strength 2.

From the above it is clear that there are several different methods covering different situations. When different methods are applied to certain problems the solutions that are found can be significantly different. For example minimum size OAs with eleven 2-level factors and strength 2 obtained using the Galois field $GF(2) \equiv \mathbb{Z}_2$ have 16 runs while those obtained using Hadamard matrices have 12 runs. Thus the problem of finding minimum size OFFDs can be difficult for the non-expert user due to the difficulty of selecting the most appropriate method.

The joint use of polynomial indicator functions and complex coding of levels provides a general theory for mixed level OFFDs (see Pistone and Rogantin 2008). This theory puts neither a restriction on the number of levels of each factor nor on the orthogonality constraints. It also makes use of commutative algebra (see Pistone and Wynn 1996), and generalizes the approach to two-level designs as discussed in Fontana, Pistone, and Rogantin (2000). The definition of *strata* provided in Section 2 makes it possible to transform each OFFD into a solution of a homogeneous system of linear equations where the unknowns are positive integers.

The aim of this work is to develop a general software tool for minimum size OFFDs generation, where *general* means that it puts neither a restriction on the number of levels of each factor nor on the orthogonality constraints.

2. Algebraic generation of OAs

In this Section, for ease in reference, we report some relevant results of the algebraic theory of OFFDs from Fontana (2013). The interested reader can find further information, including the proofs of the propositions in Fontana (2013) itself and in Fontana *et al.* (2000), Pistone and Rogantin (2008), Fontana and Pistone (2010a) and Fontana and Pistone (2010b).

2.1. Fractions of a full factorial design

Let us consider an experiment which includes m factors $\mathcal{D}_j,\ j = 1, \ldots, m$. Let us code the n_j levels of the factor $\mathcal{D}_j$ by the n_j roots of the unity

$$\mathcal{D}_j = \left\{\omega_0^{(n_j)}, \ldots, \omega_{n_j-1}^{(n_j)}\right\},$$

where $\omega_k^{(n_j)} = \exp\left(\sqrt{-1}\,\frac{2\pi}{n_j}\,k\right)$, $k = 0, \ldots, n_j - 1,\ j = 1, \ldots, m$.

The *full factorial design* with *complex coding* is $\mathcal{D} = \mathcal{D}_1 \times \cdots \mathcal{D}_j \cdots \times \mathcal{D}_m$. We denote its cardinality by $\#\mathcal{D}$, with $\#\mathcal{D} = \prod_{j=1}^m n_j$.

Definition 2.1. A fraction $\mathcal{F}$ is a multiset $(\mathcal{F}_*, f_*)$ whose underlying set of elements $\mathcal{F}_*$ is contained in $\mathcal{D}$ and f_* is the multiplicity function $f_* : \mathcal{F}_* \to \mathbb{N}$ that for each element in $\mathcal{F}_*$ gives the number of times it belongs to the multiset $\mathcal{F}$.

We recall that the underlying set of elements $\mathcal{F}_*$ is the subset of $\mathcal{D}$ that contains all the elements of $\mathcal{D}$ that appear in $\mathcal{F}$ at least once. We denote the number of elements of a fraction $\mathcal{F}$ by $\#\mathcal{F}$, with $\#\mathcal{F} = \sum_{\zeta\in\mathcal{F}_*} f_*(\zeta)$.

Example 2.1. Let us consider $m = 1$, $n_1 = 3$. We get

$$\mathcal{D} = \left\{1, \exp\left(\sqrt{-1}\,\frac{2\pi}{3}\right), \exp\left(\sqrt{-1}\,\frac{4\pi}{3}\right)\right\}.$$

The fraction $\mathcal{F} = \left\{1, 1, \exp\left(\sqrt{-1}\,\frac{2\pi}{3}\right)\right\}$ is the multiset $(\mathcal{F}_*, f_*)$ where $\mathcal{F}_* = \left\{1, \exp\left(\sqrt{-1}\,\frac{2\pi}{3}\right)\right\}$, $f_*(1) = 2$, and $f_*(\exp\left(\sqrt{-1}\,\frac{2\pi}{3}\right)) = 1$. We get $\#\mathcal{F} = f_*(1) + f_*(\exp\left(\sqrt{-1}\,\frac{2\pi}{3}\right)) = 2 + 1 = 3$.

In order to use polynomials to represent all the functions defined over $\mathcal{D}$, including multiplicity functions, we define

- X_j, the j-th component function, which maps a point $\zeta = (\zeta_1, \ldots, \zeta_m)$ of $\mathcal{D}$ to its j-th component,
$$X_j \colon \mathcal{D} \ni (\zeta_1, \ldots, \zeta_m) \ \longmapsto\ \zeta_j \in \mathcal{D}_j\ .$$
The function X_j is called *simple term* or, by abuse of terminology, *factor*.

- $X^\alpha = X_1^{\alpha_1} \cdot \ldots \cdot X_m^{\alpha_m}$, $\alpha \in L = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_m}$, i.e., the monomial function
$$X^\alpha : \mathcal{D} \ni (\zeta_1, \ldots, \zeta_m) \ \mapsto\ \zeta_1^{\alpha_1} \cdot \ldots \cdot \zeta_m^{\alpha_m}\ .$$
The function X^α is called *interaction term*.

We observe that $\{X^\alpha : \alpha \in L = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_m}\}$ is a basis of all the complex functions defined over $\mathcal{D}$. We use this basis to represent the counting function of a fraction according to Definition 2.2.

Definition 2.2. The *counting function* R of a fraction $\mathcal{F}$ is a complex polynomial defined over $\mathcal{D}$ so that for each $\zeta \in \mathcal{D}$, $R(\zeta)$ equals the number of appearances of ζ in the fraction. A 0/1 valued counting function is called an *indicator function* of a single replicate fraction $\mathcal{F}$. We denote by c_α the coefficients of the representation of R on $\mathcal{D}$ using the monomial basis $\{X^\alpha,\ \alpha \in L\}$:

$$R(\zeta) = \sum_{\alpha \in L} c_\alpha X^\alpha(\zeta),\ \zeta \in \mathcal{D},\ c_\alpha \in \mathbb{C}\ .$$

With Proposition 2.1 from Pistone and Rogantin (2008), we link the orthogonality of two interaction terms with the coefficients of the polynomial representation of the counting function.

Proposition 2.1. If $\mathcal{F}$ is a fraction of a full factorial design $\mathcal{D}$, $R = \sum_{\alpha \in L} c_\alpha X^\alpha$ is its counting function and $[\alpha - \beta]$ is the m-tuple made by the componentwise difference in the rings $\mathbb{Z}_{n_j}$, $\Big([\alpha_1 - \beta_1]_{n_1}, \ldots, [\alpha_j - \beta_j]_{n_j}, \ldots, [\alpha_m - \beta_m]_{n_m}\Big)$, then

1. the coefficients c_α are given by $c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)}$;
2. the term X^α is centered on $\mathcal{F}$, i.e., $\frac{1}{\#\mathcal{F}} \sum_{\zeta \in \mathcal{F}} X^\alpha(\zeta) = 0$, if and only if, $c_\alpha = c_{[-\alpha]} = 0$;
3. the terms X^α and X^β are orthogonal on $\mathcal{F}$, if and only if, $c_{[\alpha-\beta]} = 0$.

We now define projectivity and, in particular, its relation with OAs. Given $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}, i_1 < \ldots < i_k$ we define the projection π_I as

$$\pi_I : \mathcal{D} \ni \zeta = (\zeta_1, \ldots, \zeta_m) \mapsto \zeta_I \equiv (\zeta_{i_1}, \ldots, \zeta_{i_k}) \in \mathcal{D}_{i_1} \times \ldots \times \mathcal{D}_{i_k}\ .$$

Definition 2.3. A fraction $\mathcal{F}$ *factorially projects* onto the I-factors, $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$, $i_1 < \ldots < i_k$, if the projection $\pi_I(\mathcal{F})$ is a multiple full factorial design, i.e., the multiset $(\mathcal{D}_{i_1} \times \ldots \times \mathcal{D}_{i_k}, f_*)$ where the multiplicity function f_* is constantly equal to a positive integer over $\mathcal{D}_{i_1} \times \ldots \times \mathcal{D}_{i_k}$.

Example 2.2. Let us consider $m = 2, n_1 = n_2 = 2$ and the fraction $\mathcal{F} = \{(-1,1), (-1,1), (1,-1), (1,1)\}$. We obtain $\pi_1(\mathcal{F}) = \{-1,-1,1,1\}$ and $\pi_2(\mathcal{F}) = \{-1,1,1,1\}$. It follows that $\mathcal{F}$ projects on the first factor and does not project on the second factor.

Definition 2.4. A fraction $\mathcal{F}$ is a *mixed orthogonal array* of strength t if it factorially projects onto any I-factors with $\#I = t$.

Proposition 2.2. A fraction *factorially projects onto the I-factors*, $I = \{i_1, \ldots, i_k\} \subset \{1, \ldots, m\}$, $i_1 < \ldots < i_k$, if and only if, all the coefficients of the counting function involving the I-factors only are 0.

Proposition 2.2 can be immediately stated for mixed OAs.

Proposition 2.3. A fraction is an *orthogonal array of strength t*, if and only if, all the coefficients c_α of the counting function up to the order t are 0.

2.2. Counting functions and strata

It follows from Propositions 2.1, 2.2 and 2.3 that the problem of finding OFFDs can be written as a polynomial system in which the indeterminates are the complex coefficients c_α of the counting polynomial fraction.

Let us now introduce a different way to describe the full factorial design $\mathcal{D}$ and all its subsets. We consider the indicator functions 1_ζ of all the single points of $\mathcal{D}$. The counting function R of a fraction $\mathcal{F}$ can be written as $\sum_{\zeta\in\mathcal{D}} y_\zeta 1_\zeta$ with $y_\zeta \equiv R(\zeta) \in \{0, 1, \ldots, n, \ldots\}$. The particular case in which R is an indicator function corresponds to $y_\zeta \in \{0, 1\}$. From Proposition 2.1 we obtain that the values of the counting function over $\mathcal{D}$, y_ζ, are related to the coefficients c_α by $c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta\in\mathcal{D}} y_\zeta \overline{X^\alpha(\zeta)}$. As described in Section 2.1, we consider m factors, $\mathcal{D}_1, \ldots, \mathcal{D}_m$ where $\mathcal{D}_j \equiv \Omega_{n_j} = \left\{\omega_0^{(n_j)}, \ldots, \omega_{n_j-1}^{(n_j)}\right\}$, for $j = 1, \ldots, m$. From Pistone and Rogantin (2008), we recall two basic properties which hold true for the full design $\mathcal{D}$. Note that the notation gcd stands for *greatest common divisor* and lcm stands for *least common multiple*.

Proposition 2.4. Let X_j be the simple term with level set $\mathcal{D}_j = \Omega_{n_j} = \left\{\omega_0^{(n_j)}, \ldots, \omega_{n_j-1}^{(n_j)}\right\}$, $j = 1, \ldots, m$. Let $X^\alpha = X_1^{\alpha_1} \cdots X_m^{\alpha_m}$ be an interaction.

1. Over $\mathcal{D}$, the term X_j^r takes all the values of Ω_{s_j} equally often, where $s_j = 1$ if $r = 0$ and $s_j = n_j/\gcd(r, n_j)$ if $r > 0$.
2. Over $\mathcal{D}$, the term X^α takes all the values of Ω_{s_α} equally often, where $s_\alpha = \mathrm{lcm}(s_1, \ldots, s_m)$ and s_i, that is determined according to the previous Item 1, corresponds to $X_i^{\alpha_i}, i = 1, \ldots, m$.

We refer to Ω_{s_α} as the level set of X^α. Sometimes we also write s in place of s_α to simplify the notation. Let us now define the *strata* that are associated with simple and interaction terms.

Definition 2.5. Given a term $X^\alpha, \alpha \in L = \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_m}$, the full design $\mathcal{D}$ is partitioned into the strata $D_h^\alpha = \left\{\zeta \in \mathcal{D} : \overline{X^\alpha(\zeta)} = \omega_h^{(s_\alpha)}\right\}$, where $\omega_h^{(s_\alpha)} \in \Omega_{s_\alpha}$ and s_α is determined according to the previous Proposition 2.4.

We use $n_{\alpha,h}$ to denote the number of points of the fraction $\mathcal{F}$ that are in the stratum D_h^α, $n_{\alpha,h} = \sum_{\zeta\in D_h^\alpha} y_\zeta$, $h = 0, \ldots, s-1$. Proposition 2.5 links the coefficients c_α with $n_{\alpha,h}$.

Proposition 2.5. Let $\mathcal{F}$ be a fraction of $\mathcal{D}$ with counting function $R = \sum_{\alpha\in L} c_\alpha X^\alpha$. Each $c_\alpha, \alpha \in L$, depends on $n_{\alpha,h}, h = 0, \ldots, s-1$, as $c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{h=0}^{s-1} n_{\alpha,h} \omega_h^{(s)}$, where s is determined by X^α according to Proposition 2.4.

We now use a part of Proposition 3 from Pistone and Rogantin (2008) to obtain conditions on $n_{\alpha,h}$ that make X^α centered on the fraction $\mathcal{F}$.

Proposition 2.6. Let X^α be a term with level set Ω_s on full design $\mathcal{D}$ and let $P_s(\zeta)$ be the complex polynomial associated with the sequence $(n_{\alpha,h})_{h=0,\ldots,s-1}$ so that $P_s(\zeta) = \sum_{h=0}^{s-1} n_{\alpha,h}\zeta^h$ and Φ_s the cyclotomic polynomial of the s-roots of the unity.

1. Let s be prime. The term X^α is centered on the fraction $\mathcal{F}$, if and only if, its s levels appear equally often $n_{\alpha,0} = n_{\alpha,1} = \ldots = n_{\alpha,s-1}$.

2. Let $s = p_1^{h_1} \dots p_d^{h_d}$, p_i prime, $i = 1, \dots, d$. The term X^α is centered on the fraction $\mathcal{F}$, if and only if, the remainder $H_s(\zeta) = P_s(\zeta) \bmod \Phi_s(\zeta)$, whose coefficients are integer linear combinations of $n_{\alpha,h}, h = 0, \dots, s-1$, is identically zero.

If we recall that $n_{\alpha,h}$ are related to the values of the counting function R of a fraction $\mathcal{F}$ by $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, Proposition 2.6 allows us to express the condition X^α *is centered on* $\mathcal{F}$ as integer linear combinations of the values $R(\zeta)$ of the counting function over the full design $\mathcal{D}$. In Section 2.3, we will show the use of this property to generate fractional factorial designs.

2.3. Generation of fractions

We use strata to generate fractions that satisfy a given set of constraints on the coefficients of their counting functions. Formally, we give the following definition.

Definition 2.6. Given $\mathcal{C} \subseteq \mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$, a counting function $R = \sum_\alpha c_\alpha X^\alpha$ associated with a fraction $\mathcal{F}$ is a $\mathcal{C}$-compatible counting function if $c_\alpha = 0$, $\forall \alpha \in \mathcal{C}$.

The set of all the fractions of $\mathcal{D}$ whose counting functions are $\mathcal{C}$-compatible is denoted by $OF(n_1 \dots n_m, \mathcal{C})$.

For example let us consider $OA(n, s^m, t)$, i.e., OAs with n rows and m columns, where each column has s symbols, s prime and with strength t. Using Proposition 2.3 the coefficients of the corresponding counting functions must satisfy the conditions $c_\alpha = 0$ for all $\alpha \in \mathcal{C}$ where $\mathcal{C} = \{\alpha \in L \equiv (\mathbb{Z}_s)^m : 0 < \|\alpha\| \leq t\}$ and $\|\alpha\|$ is the number of non-null elements of α. It follows that $OF(s^m, \mathcal{C}) = \bigcup_n OA(n, s^m, t)$. Now using Proposition 2.6, we can express these conditions using strata. If we consider $\alpha \in \mathcal{C}$, we can write the condition $c_\alpha = 0$ as:

$$\begin{cases} -n_{\alpha,0} + n_{\alpha,1} = 0 \\ -n_{\alpha,0} + n_{\alpha,2} = 0 \\ \quad\vdots \\ -n_{\alpha,0} + n_{\alpha,s-1} = 0 \end{cases} \iff B_{s_\alpha} N_\alpha = \begin{bmatrix} -1 & 1 & 0 & 0 & \dots & 0 \\ -1 & 0 & 1 & 0 & \dots & 0 \\ \vdots & & & \vdots & & \vdots \\ -1 & 0 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} n_{\alpha,0} \\ n_{\alpha,1} \\ \vdots \\ n_{\alpha,s-1} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

We observe that the $((s-1) \times s)$ matrix B_{s_α} is the same for all α because s_α is constant and equal to s.

To obtain all the conditions it is enough to vary $\alpha \in \mathcal{C}$. We therefore obtain the homogeneous system of linear equations $AY = \underline{0}$ where A is the $(\#\mathcal{C} \cdot (s-1) \times s^m)$ matrix whose rows contain the values, over $\mathcal{D}$, of the difference between two indicator functions of strata, $1_{D_h^\alpha} - 1_{D_0^\alpha}$ $h = 1, \dots, s-1$; Y is the s^m column vector whose entries are the values of the counting function over $\mathcal{D}$; $\underline{0}$ is the $(\#\mathcal{C} \cdot (s-1))$ column vector whose entries are all equal to 0.

Let us now consider the general case in which there are no restrictions on the number of levels and show our method for $OA(n, 4^2, 1)$. In this case the number of levels is a power of a prime, $4 = 2^2$. Using Proposition 2.3 the coefficients of the corresponding counting functions must satisfy the conditions $c_\alpha = 0$ for all $\alpha \in \mathcal{C}$ where $\mathcal{C} = \{\alpha \in L \equiv \mathbb{Z}_4 \times \mathbb{Z}_4 : \|\alpha\| = 1\}$. Let us consider $c_{1,0}$. From Item (1) of Proposition 2.4, X_1 takes the values in $\Omega_{s_{(1,0)}}$ where $s_{(1,0)} = 4$. From Proposition 2.6, X_1 will be centered on $\mathcal{F}$, if and only if, the remainder $H_4(\zeta) = P_4(\zeta) \bmod \Phi_4(\zeta)$ is identically zero. We have $\Phi_4(\zeta) = 1 + \zeta^2$ (see Lang 1965) and so we can compute the remainder $H_4(\zeta) = n_{(1,0),0} - n_{(1,0),2} + (n_{(1,0),1} - n_{(1,0),3})\zeta$. The condition that $H_4(\zeta)$ must be identically zero translates into

$$\begin{cases} n_{(1,0),0} - n_{(1,0),2} = 0 \\ n_{(1,0),1} - n_{(1,0),3} = 0 \end{cases} \iff B_4 N_{(1,0)} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} n_{(1,0),0} \\ n_{(1,0),1} \\ n_{(1,0),2} \\ n_{(1,0),3} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Let us now consider $c_{2,0}$. From Item (1) of Proposition 2.4, X_1^2 takes the values in $\Omega_{s_{(2,0)}}$ where $s_{(2,0)} = 2$. From Proposition 2.6, X_1^2 will be centered on $\mathcal{F}$, if and only if

$$-n_{(2,0),0} + n_{(2,0),1} = 0 \iff B_2 N_{(2,0)} = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} n_{(2,0),0} \\ n_{(2,0),1} \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}.$$

If we repeat the same procedure for all α, where $\|\alpha\| = 1$, and if we recall that $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, and so, for example, $n_{(1,0),0} = y_{1,1} + y_{1,\mathrm{i}} + y_{1,-1} + y_{1,-\mathrm{i}}$ with $\mathrm{i} = \sqrt{-1}$, the OAs $OA(n, 4^2, 1)$ become the positive integer solutions of the following integer linear homogeneous system:

$$AY = \underline{0}$$

where

$$A = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix},$$

$$Y = \begin{bmatrix} y_{1,1} \\ y_{\mathrm{i},1} \\ y_{-1,1} \\ y_{-\mathrm{i},1} \\ y_{1,\mathrm{i}} \\ y_{\mathrm{i},\mathrm{i}} \\ y_{-1,\mathrm{i}} \\ y_{-\mathrm{i},\mathrm{i}} \\ y_{1,-1} \\ y_{\mathrm{i},-1} \\ y_{-1,-1} \\ y_{-\mathrm{i},-1} \\ y_{1.-\mathrm{i}} \\ y_{\mathrm{i},-\mathrm{i}} \\ y_{-1,-\mathrm{i}} \\ y_{-\mathrm{i},-\mathrm{i}} \end{bmatrix}$$

and $\underline{0}$ is the 10-rows column vector whose entries are all equal to 0.

It should be noted that the matrix A of the coefficients does not have full rank, e.g., the first and the fourth rows are equal. This aspect is discussed in Fontana and Pistone (2010b). In any case the solution method used here does not require a reduction to a full rank matrix.

2.4. The optimization problem

When the tests are very expensive or require a substantial amount of time, experimenters are interested in finding minimum size orthogonal fractional designs, i.e., fractional factorial designs that satisfy some orthogonality requirements *and* consist of the minimum number of points.

According to our formalization, the problem is equivalent to extracting one fraction $\mathcal{F}_*$ from $OF(n_1 \dots n_m, \mathcal{C})$, such that the size of the fraction $\#\mathcal{F}_*$ is minimum.

The problem can be written as

$$\begin{aligned} &\text{minimize } \underline{1}^\top Y \\ &\text{subject to } AY = \underline{0}, \end{aligned} \tag{1}$$

where A is the matrix built as explained in Section 2.3, $\underline{0}$ is the column vector that has the same numer of rows of A and whose entries are all equal to 0, $\underline{1}$ is the $\#\mathcal{D}$ column vector whose entries are all equal to 1, $\underline{1}^\top$ is its transpose and Y is a vector of positive integer numbers that contains the unknown counting function values, $Y^\top \neq [0, \dots, 0]$.

3. The algorithm and its SAS implementation

We now focus on the details of the algorithm that makes use of the theoretical results described in Section 2. The implementation used SAS software: in particular SAS/IML (SAS Institute, Inc. 2008a), for the construction of the constraints $AY = \underline{0}$ and SAS/OR (SAS Institute, Inc. 2008b), to solve the optimization problem as stated in Equation 1 of Section 2.4.

As described in the previous section, we state the problem of finding OFFDs that satisfy a set of conditions in terms of orthogonality as a homogeneous system of linear equations, $AY = \underline{0}$, in which the indeterminates are positive integers that contain the unknown counting values. For the sake of simplicity, we work with mixed level OAs of strength t, that is $OA(n, n_1 \cdot \dots \cdot n_m, t)$; the extension to mixed level OFFDs, regardless of the orthogonality constraints they must satisfy, is straightforward, as we will explain in Section 6.1.

With reference to Figure 1, the algorithm

- takes the levels n_j for each factor, $j = 1, \dots, m$, and the required strength t as input;
- generates the constraint matrix A;
- performs the optimization phase;
- provides a minimum size OA as output.

Constraints generation is done by the following major steps:

1. We generate the full factorial $\mathbb{Z}_{n_1} \times \dots \times \mathbb{Z}_{n_m}$. It represents both the set L of multi-indexes and the full factorial design with integer coding, i.e., the full factorial design where the levels of the factor $\mathcal{D}_j$ are coded with the integer $0, \dots, n_j - 1$, $j = 1, \dots, m$. We denote by α a generic element of L.
2. Let us denote the subset of L that contains all the α such that the norm of $\|\alpha\|$ is less than or equal to t, $\alpha \neq [0, \dots, 0]$ by $\mathcal{C}$; for all $\alpha \in \mathcal{C} \subseteq L$ we build the column vector $\left[\overline{X^\alpha(\zeta)} : \zeta \in \mathcal{D}\right]$ and, coherently with Proposition 2.4, we also compute s_α.
3. According to Proposition 2.6, for all $s \in \{2, \dots, s_{\max}\}$, where $s_{\max}$ is the maximum of all the s_α, $\alpha \in \mathcal{C}$, we compute the matrix B_s.

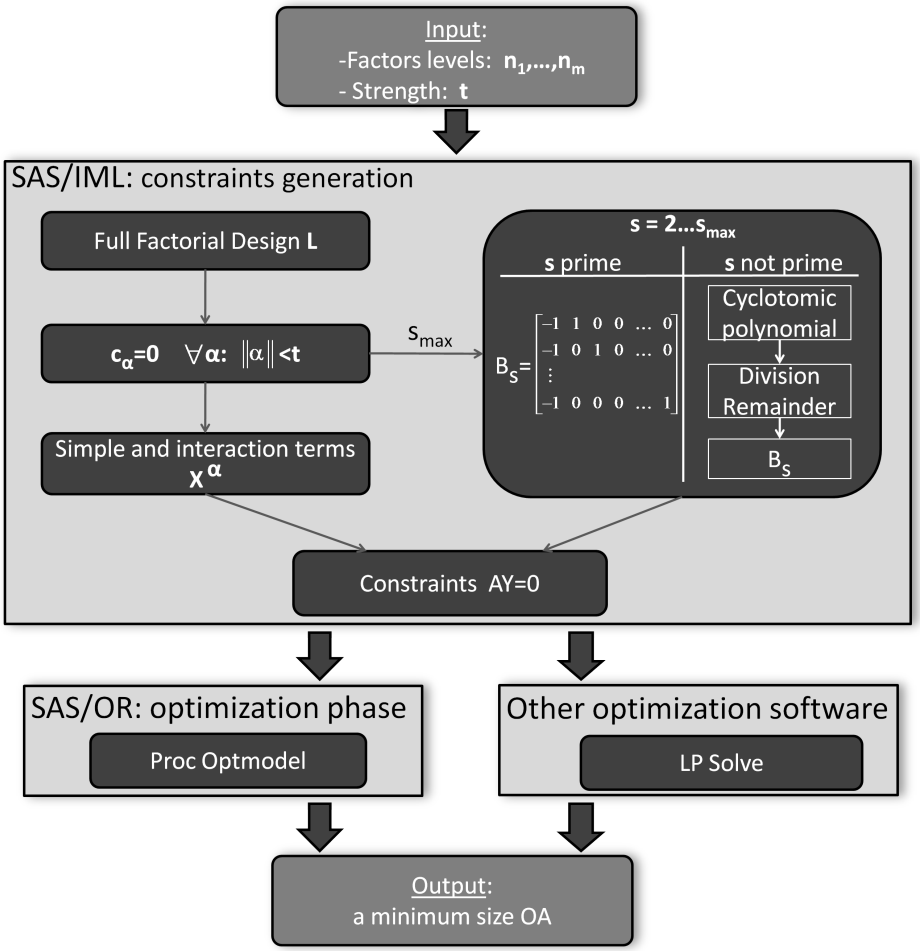

Figure 1: The main modules of the algorithm.

(a) If s is prime, then $B_s = \begin{bmatrix} -1 & 1 & 0 & 0 & \cdots & 0 \\ -1 & 0 & 1 & 0 & \cdots & 0 \\ \vdots & & & & & \vdots \\ -1 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$.

(b) If s is not prime, then B_s is determined taking the remainder H_s of the division between the polynomial P_s and the cyclotomic polynomial Φ_s.

4. Using all the $\left[\overline{X^{\alpha}(\zeta)} : \zeta \in \mathcal{D}\right]$, $\alpha \in \mathcal{C}$ and the corresponding B_{s_α} and recalling that $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, we build the constraints matrix A.

Once the constraints $AY = \underline{0}$ are obtained, the optimization phase can be carried out in SAS or in another software tool like **lp_solve** (Berkelaar, Eikland, and Peter 2004).

In the next section we explain in detail all the steps of the algorithm.

3.1. Full factorial design generation

We must build up the full factorial design $L = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_m}$, which represents the set of

multi-indexes α. The total number of points in L is $\#L = n_1 \cdot \ldots \cdot n_m$. The set of multi-indexes L can be represented by a matrix L with m columns and $\#L$ rows.

To write the rows of L, we generalize the common conversion base algorithm (Cox, Little, and O'Shea 1997). We make a loop over the $\#L$ rows: the row index (from 0 to $\#L - 1$) is the number k to be converted and $r_1 \ldots r_m$ is the corresponding row of L. This is the algorithm:

Input: $k, n_1, \ldots, n_m$
Output: $r_1, \ldots, r_m$
  for $j = 1$ **to** m **do**
    $r_{m-j+1} = [k]_{n_{m-j+1}}$
    $k = \mathrm{int}(k/n_{m-j+1})$
  end for

The notation $[k]_n$ indicates the residue of k modulo n, and $\mathrm{int}(k/n)$ is the integer part of the division $\frac{k}{n}$.

3.2. Simple and interaction terms

From now on we use ω_{k,n_j} in place of $\omega_k^{(n_j)}$ to simplify the notation. We observe that the mapping τ_{n_j}

$$\mathcal{D}_j = \left\{\omega_{0,n_j}, \ldots, \omega_{n_j-1,n_j}\right\} \ni \omega_{k,n_j} \overset{\tau_{n_j}}{\longleftrightarrow} k \in \mathbb{Z}_{n_j} = \{0, \ldots, n_j - 1\}$$

is a group isomorphism between Ω_{n_j} with the complex product and $\mathbb{Z}_{n_j}$ with the sum modulo n_j. It follows that $\tau_{n_1} \times \ldots \times \tau_{n_m}$ is a group isomorphism between $\mathcal{D}$ and L. Let us consider the term X_j^r and its values $X_j^r(\omega_{k,n_j}) = \omega_{k,n_j}^r$ over $\mathcal{D}_j$, $j = 1, \ldots, m, r \in \mathbb{Z}_+$. Using the isomorphism τ_{n_j} we can represent the vector $[X_j^r(\omega_{k,n_j}) : \omega_{k,n_j}^r \in \mathcal{D}_j]$ with the vector

$$[\tau_{n_j}(\omega_{k,n_j}^r) = [kr]_{n_j} : k = 0, \ldots, n_j - 1]$$

Now we consider the interaction term X^α and its values

$$X^\alpha(\omega_{k_1,n_1}, \ldots, \omega_{k_m,n_m}) = \omega_{k_1,n_1}^{\alpha_1} \cdot \ldots \cdot \omega_{k_m,n_m}^{\alpha_m}$$

over $\mathcal{D}$, $\alpha \in L$. From Proposition 2.4, X^α takes all the values of Ω_{s_α}, where $s_\alpha = \mathrm{lcm}(s_1, \ldots, s_m)$ and $s_j, j = 1, \ldots, m$ are determined as in Item 1 of the same Proposition. It follows that we can represent the vector $[X^\alpha(\omega_{k_1,n_1}, \ldots, \omega_{k_m,n_m}) : \omega_{k,n_j} \in \mathcal{D}_j, j = 1, \ldots, m]$ with the vector

$$\left[\tau_{s_\alpha}\left(\omega_{k_1,n_1}^{\alpha_1} \cdot \ldots \cdot \omega_{k_m,n_m}^{\alpha_m}\right) = \left[k_1\left(\frac{s_\alpha}{n_1}\right) + \ldots + k_m\left(\frac{s_\alpha}{n_m}\right)\right]_{s_\alpha} : \right.$$
$$\left. k_j \in \{0, \ldots, n_j - 1\}, j = 1, \ldots, m\right]$$

Finally, by observing that the complex conjugate $\overline{\omega}_{h,s}$ of $\omega_{h,s} \in \Omega_s$ is $\omega_{[s-h]_s,s}$ and that, using τ_s, can be represented as $[s-h]_s$, we build a matrix whose columns contain the integer representation of $\overline{X^\alpha(\zeta)}$ for all $\alpha \in \mathcal{C}$ and for all $\zeta \in \mathcal{D}$.

Furthermore we store all the values $s_\alpha, \alpha \in \mathcal{C}$ in a vector and we calculate the maximum value $s_{\max}$ of it, which will be used in the next steps of the algorithm.

3.3. Remainders

For each s, where $s = 2, \ldots, s_{\max}$, there is one or more linear constraint over the coefficients $n_{\alpha,h}$. If s is prime we know from Proposition 2.6 that the conditions can be easily determined. For example for $s = 5$ the conditions correspond to the matrix B_5:

$$B_5 = \begin{array}{c} \begin{array}{ccccc} n_{\alpha,0} & n_{\alpha,1} & n_{\alpha,2} & n_{\alpha,3} & n_{\alpha,4} \end{array} \\ \begin{bmatrix} -1 & 1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 & 1 \end{bmatrix} \end{array}$$

If s is not prime, then we have to generate the cyclotomic polynomial Φ_s, the polynomial $P_s(\zeta)$ and then to calculate the remainder H_s of the division P_s/Φ_s.

From the fact that the roots of a cyclotomic polynomial are the primitive roots of unity (Lang 2002), we can calculate Φ_s by dividing $X^s - 1$ by the product of the cyclotomic polynomials Φ_{d_i} corresponding to the proper divisors $d_1, \ldots, d_s$ of s :

$$\Phi_s(X) = \frac{X^s - 1}{\prod_{d|s,\ d<s} \Phi_d(X)}.$$

We create a function that implements the polynomials division (Cox *et al.* 1997). Let f be the dividend polynomial and g the divisor, the pseudocode of the algorithm for finding q and r such that $f = qg + r$ is:

Input: g, f
Output: q, r
  $q = 0$; $r = f$
  while $r \neq 0$ and $LT(g)$ divides $LT(r)$ **do**
    $q = q + LT(r)/LT(g)$
    $r = r - (LT(r)/LT(g))g$
  end while

LT means *leading term*, i.e., the term of highest degree. To construct the cyclotomic polynomial Φ_s we need three steps:

1. Find all proper divisors of s: $d_1, \ldots, d_k$.
2. Analyze each d_i: if it is prime then we can calculate the cyclotomic polynomial Φ_{d_i} dividing $X^{d_i} - 1$ by $\Phi_1 = X - 1$; if it is not prime then we have to return to step 1 setting $s = d_i$.
3. When $\Phi_{d_1}, \ldots, \Phi_{d_k}$ are built, we obtain Φ_s by dividing $X^s - 1$ by the product $\Phi_{d_1} \cdot \ldots \cdot \Phi_{d_k}$.

From Proposition 2.6 we know that we are interested in calculating the remainder of the division of P_s by the cyclotomic polynomial Φ_s. In this case the dividend polynomial has symbolic coefficients $n_{\alpha,h},\ h = 0, \ldots, s-1$:

$$P_s(\zeta) = n_{\alpha,0}\zeta^0 + \ldots + n_{\alpha,s-1}\zeta^{s-1}.$$

The division algorithm remains the same but at each step we have to store also these coefficients. We create an $s \times s$ matrix, where the column index is related to the degree and the row

index to the polynomial coefficient. We explain this point considering $s = 4$. By definition $P_4(\zeta)$ is

$$P_4(\zeta) = n_{\alpha,0} + n_{\alpha,1}\zeta + n_{\alpha,2}\zeta^2 + n_{\alpha,3}\zeta^3,$$

and the corresponding matrix becomes

	ζ^3	ζ^2	ζ^1	ζ^0
$n_{\alpha,3}$	1	0	0	0
$n_{\alpha,2}$	0	1	0	0
$n_{\alpha,1}$	0	0	1	0
$n_{\alpha,0}$	0	0	0	1

When we divide $P_4(\zeta)$ by $\Phi_4(\zeta) = 1 + \zeta^2$, we obtain the remainder $H_4(\zeta) = n_{\alpha,0} - n_{\alpha,2} + (n_{\alpha,1} - n_{\alpha,3})\zeta$ and its related matrix

	ζ^3	ζ^2	ζ^1	ζ^0
$n_{\alpha,3}$	0	0	-1	0
$n_{\alpha,2}$	0	0	0	-1
$n_{\alpha,1}$	0	0	1	0
$n_{\alpha,0}$	0	0	0	1

Finally, by simply deleting the zero columns, we obtain the matrix B_4:

$$B_4 = \begin{array}{c} \begin{array}{cccc} n_{\alpha,0} & n_{\alpha,1} & n_{\alpha,2} & n_{\alpha,3} \end{array} \\ \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \end{array}.$$

3.4. Strata and constraints

The last step is to transform each matrix $B_s, s = 2, \ldots, s_{\max}$, that refers to $n_{\alpha,h}, h = 0, \ldots, s-1$ into a matrix A_s that refers to $y_\zeta, \zeta \in \mathcal{D}$ and then to stack all the A_s into the matrix A. In this way, it is possible to set up the optimization problem as stated in Equation 1 of Section 2.4.

It is enough to recall that

$$n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta, h = 0, \ldots, s-1.$$

We explain this point by using an example. Let us consider, as in Section 2.3, $OA(n, 4^2, 1)$. We obtain that $\mathcal{C} = \{\alpha \in \mathbb{Z}_4 \times \mathbb{Z}_4 : \|\alpha\| \leq 1, \alpha \neq (0,0)\} = \{(0,k), (k,0) : k = 1,2,3\}$. Let us consider $\alpha = (0,1)$. We represent $\overline{X^{(0,1)}(\zeta)}, \zeta \in \mathcal{D}$, whose level set is Ω_4, using τ_4 and obtaining its integer representation.

We obtain

$$\begin{array}{ccccccc}
\mathcal{D}=(\zeta_1\ ,\zeta_2) & & X^{(0,1)}(\zeta_1,\zeta_2) & & \overline{X^{(0,1)}(\zeta_1,\zeta_2)} & & \tau_4(\overline{X^{(0,1)}(\zeta_1,\zeta_2)}) \\
\begin{bmatrix} 1 & 1 \\ 1 & i \\ 1 & -1 \\ 1 & -i \\ i & 1 \\ \vdots & \vdots \\ -i & 1 \\ -i & i \\ -i & -1 \\ -1 & -i \end{bmatrix} & \Rightarrow & \begin{bmatrix} 1 \\ i \\ -1 \\ -i \\ 1 \\ \vdots \\ 1 \\ i \\ -1 \\ -i \end{bmatrix} & \Rightarrow & \begin{bmatrix} 1 \\ -i \\ -1 \\ i \\ 1 \\ \vdots \\ 1 \\ -i \\ -1 \\ i \end{bmatrix} & \Rightarrow & \begin{bmatrix} 0 \\ 3 \\ 2 \\ 1 \\ 0 \\ \vdots \\ 0 \\ 3 \\ 2 \\ 1 \end{bmatrix}
\end{array}$$

The remainder matrix for $s = 4$ is

$$B_4 = \begin{array}{c} \begin{array}{cccc} n_{\alpha,0} & n_{\alpha,1} & n_{\alpha,2} & n_{\alpha,3} \end{array} \\ \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix} \end{array},$$

so that we obtain the following A_4 matrix

$$A_4 = \begin{bmatrix} 1 & 0 & -1 & 0 & 1 & \dots & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & \dots & 0 & -1 & 0 & 1 \end{bmatrix}.$$

By repeating the same procedure for all $\alpha \in \mathcal{C}$ and stacking all the A_s matrices, we obtain the A matrix, as reported at the end of Section 2.3.

4. The optimization problem

At this point our problem is to find the fraction that satisfies all the constraints while minimizing the number of points. We can use any optimization software that solves mixed-integer linear problems (MILP). In particular, we have constructed the set of constraints in such a way that can be directly used in SAS/OR and can also be exported for use with **lp_solve** (Berkelaar *et al.* 2004).

SAS/OR procedures are used to optimize a linear function subject to linear and integer constraints. Specifically, we use the `Optmodel` procedure, which solves the general MILP of the form:

$$\begin{aligned} \text{minimize } & c^\top x \\ \text{subject to } & Ax\{\geq, =, \leq\}b \\ & l \leq x \leq u \\ & x_i \in \mathbb{Z}, \forall i \in \mathscr{S}, \end{aligned}$$

where:

- A is an $r \times v$ matrix of technological coefficients;

- b is an $r \times 1$ matrix of right-hand side constants;
- c is a $v \times 1$ matrix of objective function coefficients;
- x is a $v \times 1$ matrix of structural variables;
- l is a $v \times 1$ matrix of lower bounds on x;
- u is a $v \times 1$ matrix of upper bounds on x;
- $\mathscr{S}$ is a subset of the set of indices $1, \ldots, v$ and identifies those variables that must take only integer values.

In our case:

- A is the $r \times v$ constraints matrix, with $v = n_1 \cdot \ldots \cdot n_m$;
- b is the $r \times 1$ matrix whose entries are all equal to 0;
- $c = \underline{1}$ is the $n_1 \cdot \ldots \cdot n_m \times 1$ matrix whose entries are all equal to 1;
- x is an $n_1 \cdot \ldots \cdot n_m \times 1$ matrix that contains the unknown number of appearances of each point of $\mathcal{D}$ in the fraction $\mathcal{F}$;
- l is the $n_1 \cdot \ldots \cdot n_m \times 1$ matrix whose entries are all equal to 0; we do not need to set any value for the upper bounds u;
- $\mathscr{S}$ contains all $n_1 \cdot \ldots \cdot n_m$ variables, because all of them must take integer values.

Finally, in order to exclude the empty fraction from the solutions, we add the following constraint:

$$\underline{1}^\top x \geq 1$$

For further details about the `Optmodel` procedure see SAS Institute, Inc. (2008b).

As mentioned before, the constraints matrix A is stored in a dataset and can be exported to other optimization programs. Our algorithm exports the matrix A into a file that can be used by the well-known and widely-used open-source code **lp_solve** (Berkelaar *et al.* 2004).

5. Results

We tested the algorithm by comparing our results with those reported in Schoen *et al.* (2010). We also ran other examples. The results are summarized in Table 1. The first column reports *the factor set*, i.e., the number of levels of each factor, $\#\mathcal{D}$ is the cardinality of the full factorial design $\mathcal{D}$, t and n are respectively the strength and the cardinality of the OA as generated by the algorithm.

Factor set	$\#\mathcal{D}$	t	n
$2^6 16$	1024	2	32
$4^3 8$	512	2	32
$3^4 12$	972	2	36
$4^3 12$	768	2	48
3 6 9	162	2	54
$4^3 16$	1024	2	64
$3^4 9$	729	3	81

Table 1: The results of some examples.

6. Conclusion

6.1. OFFD

Given the levels of each factor, $n_1, \ldots, n_m$, and the strength t as input, the algorithm provides one of the minimum size mixed level OAs $OA(n, n_1 \cdot \ldots \cdot n_m, t)$ as output.

To deal with generic OFFD, it is enough, using Proposition 2.1, to modify $\mathcal{C}$, the subset of the multi-indexes $\alpha \in L$ for which the coefficients of the counting function must be equal to zero and to leave the remaining part of the code unchanged. We illustrate this point with an example. Let us suppose that we consider 4 factors with 3 levels each and that we search for a minimum size OFFD such that all the main effects of the factors, $\mathcal{D}_1, \ldots, \mathcal{D}_4$, and the interaction between factor $\mathcal{D}_2$ and $\mathcal{D}_3$ are estimable and orthogonal.

From Item 1 of Proposition 2.1,

- the main effects of the factors $\mathcal{D}_1, \ldots, \mathcal{D}_4$ will be centered, if and only if, $c_\alpha = 0$ for $\alpha \in \{(e,0,0,0),(0,e,0,0),(0,0,e,0),(0,0,0,e)\}$ with $e = 1,2$;
- the interaction between factor $\mathcal{D}_2$ and $\mathcal{D}_3$ will be centered, if and only if, $c_\alpha = 0$ for $\alpha \in \{(0,e,f,0)\}$ with $e, f = 1,2$.

From Item 2 of Proposition 2.1,

- the main effect of the factor $\mathcal{D}_i$ will be orthogonal to the main effect of the factor $\mathcal{D}_j$, $i < j$, if and only if, $c_\alpha = 0$ for $\alpha \in \left\{e(\delta_1^i, \delta_2^i, \delta_3^i, \delta_4^i) + f(\delta_1^j, \delta_2^j, \delta_3^j, \delta_4^j)\right\}$ with $e, f = 1,2$ and $\delta_h^k = 0$ if $h \neq k$ and $\delta_h^k = 1$ if $h = k$;
- the main effect of the factor $\mathcal{D}_1$ (resp. $\mathcal{D}_4$) will be orthogonal to the interaction between factor $\mathcal{D}_2$ and $\mathcal{D}_3$, if and only if, $c_\alpha = 0$ for $\alpha \in \{(e,f,g,0)\}$ (resp. $\alpha \in \{(0,e,f,g)\}$) with $e, f, g = 1,2$;
- the main effect of the factor $\mathcal{D}_2$ or of the factor $\mathcal{D}_3$ will be orthogonal to the interaction between factor $\mathcal{D}_2$ and $\mathcal{D}_3$, if and only if, $c_\alpha = 0$ for $\alpha \in \{(0,e,f,0)\}$ with $e, f = 1,2$.

It follows that $\mathcal{C}$ can be written as:

$$\mathcal{C} = \mathcal{C}_0 \cup \mathcal{C}_1 \cup \mathcal{C}_2,$$

where $\mathcal{C}_0 = \left\{\alpha \in \mathbb{Z}_2^4 : 0 < \|\alpha\| \leq 2\right\}$, $\mathcal{C}_1 = \{(k_1, k_2, k_3, 0) : k_1, k_2, k_3 = 1, 2\}$ and $\mathcal{C}_2 = \{(0, k_2, k_3, k_4) : k_2, k_3, k_4 = 1, 2\}$. We obtain a solution $\mathcal{F}$ with $\#\mathcal{F} = 27$.

6.2. The FACTEX procedure

The algorithm presented is an improvement which completes existing procedures in the SAS/QC software (SAS Institute, Inc. 2010). In particular the FACTEX procedure constructs an orthogonal fractional design for q-level factors by using the Galois field (or finite field) of size q. For this reason there is a restriction on the number of levels: all the factors must have the same number q of levels and q must be either a prime number or an integer power of a prime number. Finally the **FACTEX** procedure constructs *regular* fractional factorial design: these fractions have no replicates and any two factorial effects are either estimated independently of each other or are fully aliased.

The algorithm puts no restriction on the number of levels of each factor, thereby including mixed level designs. It also manages non-regular fractions.

6.3. Optimization

Finally, this algorithm has the advantage of keeping the algebraic construction of the constraints distinct from the optimization phase. We can use the output of the first part of the algorithm either with SAS/OR procedures or with other optimization programs.

The range of application is limited only by the amount of computational effort required, which depends on the size of the full design $\mathcal{D}$.

References

Bailey RA (2008). *Design of Comparative Experiments*. Cambridge University Press, Cambridge.

Berkelaar M, Eikland K, Peter (2004). ***lp_solve***. Version 5.1.0.0, Licence terms: GNU LGPL, URL http://lpsolve.sourceforge.net/5.5/.

Bose RC (1947). "Mathematical Theory of the Symmetrical Factorial Design." *Sankhyā*, **8**, 107–166.

Collombier D (1996). *Plans D'Expérience Factoriels. Construction et propriétés des fractions de plans*. 21. Springer-Verlag.

Cox D, Little J, O'Shea D (1997). *Ideals, Varieties and Algorithms. An Introduction to Computational Algebraic Geometric and Commutative Algebra*. Springer-Verlag, New York.

Dey A, Mukerjee R (1999). *Fractional Factorial Plans*. John Wiley & Sons, New York.

Fontana R (2013). "Algebraic Generation of Minimum Size Orthogonal Fractional Factorial Designs: An Approach Based on Integer Linear Programming." *Computational Statistics*, **28**(1), 241–253.

Fontana R, Pistone G (2010a). "Algebraic Generation of Orthogonal Fractional Factorial Designs." In *Proceedings of the 45th Scientific Meeting of the Italian Statistical Society, Padova, 16–18 June 2010*. URL http://homes.stat.unipd.it/mgri/SIS2010/Program/contributedpaper/654-1297-3-DR.pdf.

Fontana R, Pistone G (2010b). "Algebraic Strata for Non Symmetrical Orthogonal Fractional Factorial Designs and Application." *La matematica e le sue applicazioni 2010/1*, Politecnico di Torino DIMAT.

Fontana R, Pistone G, Rogantin M (2000). "Classification of Two-Level Factorial Fractions." *Journal of Statistical Planning and Inference*, **87**(1), 149–172.

Hedayat AS, Sloane NJA, Stufken J (1999). *Orthogonal Arrays. Theory and Applications*. Springer-Verlag, New York.

Lang S (1965). *Algebra*. Addison Wesley, Reading.

Lang S (2002). *Algebra, Graduate Texts in Mathematics*. Springer-Verlag, New York.

Mukerjee R, Wu C (2006). *A Modern Theory of Factorial Designs*. Springer-Verlag.

Pistone G, Rogantin M (2008). "Indicator Function and Complex Coding for Mixed Fractional Factorial Designs." *Journal of Statistical Planning and Inference*, **138**(3), 787–802.

Pistone G, Wynn HP (1996). "Generalised Confounding with Gröbner Bases." *Biometrika*, **83**(3), 653–666.

Raktoe BL, Hedayat A, Federer WT (1981). *Factorial Designs*. John Wiley & Sons, New York.

Sampò S (2011). *Orthogonal Fractional Factorial Designs Generation: A SAS-Based Algorithm*. Master's thesis, Politecnico di Torino. Supervisor: Fontana R., winner of the SAS University Challenge 2011.

SAS Institute, Inc (2008a). *SAS/IML 9.2 User's Guide*. Cary.

SAS Institute, Inc (2008b). *SAS/OR 9.2 User's Guide: Mathematical Programming*. Cary.

SAS Institute, Inc (2010). *SAS/QC 9.2 User's Guide, Second Edition*. Cary.

Schoen ED, Eendebak PT, Nguyen MVM (2010). "Complete Enumeration of Pure-Level and Mixed-Level Orthogonal Arrays." *Journal of Combinatorial Designs*, **18**(2), 123–140.

Wu CFJ, Hamada M (2000). *Experiments. Planning, Analysis, and Parameter Design Optimization*. John Wiley & Sons, New York.

Affiliation:

Roberto Fontana
Dipartimento di Scienze Matematiche
Politecnico di Torino
10129 Torino, Italy
E-mail: roberto.fontana@polito.it
URL: http://www.swas.polito.it/rubrica/scheda_pers.asp?matricola=002454

Journal of Statistical Software http://www.jstatsoft.org/
published by the American Statistical Association http://www.amstat.org/
Volume 53, Issue 10 *Submitted:* 2012-03-02
May 2013 *Accepted:* 2012-12-26