# Algebraic generation of Orthogonal Fractional Factorial Designs

Roberto Fontana and Giovanni Pistone

**Abstract** The joint use of counting functions, Hilbert basis and Markov basis allows to define a procedure to generate all the fractional factorial designs that satisfy a given set of constraints in terms of orthogonality (Fontana, Pistone and Rogantin (JSPI,2000), Pistone and Rogantin (JSPI, 2008)). The general case of mixed level designs, without restrictions on the number of levels of each factor (such as power of prime number) is studied. The generation problem is reduced to finding positive integer solutions of a linear system of equations (e.g. Carlini and Pistone (JSTP, 2007)). This new methodology has been experimented on some significant classes of fractional factorial designs, including mixed level orthogonal arrays and sudoku designs (Fontana and Rogantin in Algebraic and Geometric Methods in Statistics, CUP (2009)). For smaller cases the complete generating set of all the solutions can be computed. For larger cases we resort to the random generation of a sample solution.

**Key words:** Design of Experiments, Hilbert basis, Markov basis, Algebraic statistics, Indicator polynomial, Counting function.

## 1 Introduction

The main result of this paper is discussed in Section 3 where the problem of finding fractional factorial designs that satisfy a set of orthogonality conditions is translated into the problem of finding non-negative integer solutions to a system of linear equa-

Roberto Fontana
DIMAT Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino Italy, e-mail: roberto.fontana@polito.it

Giovanni Pistone
DIMAT Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino Italy and Statistical Initiative, Collegio Carlo Alberto, Moncalieri Italy, e-mail: giovanni.pistone@polito.it

tions, so avoiding computations with complex numbers. Fractional factorial designs that satisfy a set of conditions in terms of orthogonality between factors have been described as the zero-set of a system of polynomial equations whose indeterminates are the complex coefficients of the counting polynomial functions, [7] and [4], see [5] for a short review. In section 4, we use the software [10] to find the generators of classes of orthogonal arrays. Finally, in section 5 we consider the problem of randomly sampling one fraction from a given class of orthogonality. Two simulation methods are considered, Simulated Annealing and Markov Chain Monte Carlo.

## 2 Full factorial design and fractions of a full factorial design

We recall notations and results from [7]:

- $\mathcal{D}_j$ is a *factor* with $n_j$ levels coded with the $n_j$-th roots of the unity, $\mathcal{D}_j = \{\omega_0, \ldots, \omega_{n_j-1}\}$, $\omega_k = \exp\left(\sqrt{-1}\,\frac{2\pi}{n_j}\,k\right)$; $\mathcal{D} = \mathcal{D}_1 \times \cdots \mathcal{D}_j \cdots \times \mathcal{D}_m$ is the *full factorial design* with *complex coding* and $\#\mathcal{D}$ is its cardinality;
- $X_j$ is the $j$-th component function, which maps a point 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*. The *interaction term* is $X^\alpha = X_1^{\alpha_1} \cdots X_m^{\alpha_m}$, $\alpha \in L = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_m}$, i.e. the monomial function $X^\alpha : \mathcal{D} \ni (\zeta_1, \ldots, \zeta_m) \mapsto \zeta_1^{\alpha_1} \cdots \zeta_m^{\alpha_m}$.

We underline that $L$ is both the full factorial design with integer coding and *the exponent set of all the simple factors and interaction terms* and $\alpha$ is both a treatment combination in the integer coding and a multi-exponent of an interaction term. These identifications make the complex coding especially simple.

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 $\mathcal{F}$.

**Definition 1.** If $f$ is a $\mathbb{C}$-valued polynomial function defined on $\mathcal{F}$, briefly a response, then its *mean value on* $\mathcal{F}$ is $E_{\mathcal{F}}(f) = \frac{1}{\#\mathcal{F}} \sum_{\zeta \in \mathcal{F}} f(\zeta)$, where $\#\mathcal{F}$ is the total number of treatment combinations of the fraction. A response $f$ is *centered* if $E_{\mathcal{F}}(f) = 0$. Two responses $f$ and $g$ are *orthogonal on* $\mathcal{F}$ if $E_{\mathcal{F}}(f\,\overline{g}) = 0$.

*Remark 1.* It should be noted that $\sum_{\zeta \in \mathcal{F}} f(\zeta)$ means $\sum_{\zeta \in \mathcal{F}_*} f_*(\zeta) f(\zeta)$.

With the complex coding the vector orthogonality of two interaction terms $X^\alpha$ and $X^\beta$, with respect to the Hermitian product $f \cdot g = E_{\mathcal{F}}(f\,\overline{g})$, corresponds to the combinatorial orthogonality as specified in Proposition 6. We consider the general case in which fractions can contain points that are replicated.

**Definition 2.** The *counting function* $R$ of a fraction $\mathcal{F}$ is a complex polynomial defined on $\mathcal{D}$ so that for each $\zeta \in \mathcal{D}$, $R(\zeta)$ equals the number of appearances of $\zeta$ in the fraction. A $0-1$ valued counting function is called *indicator function* of a single replicate fraction $\mathcal{F}$. We denote by $c_\alpha$ the coefficients of the representation of $R$ on $\mathcal{D}$ using the monomial basis $\{X^\alpha,\ \alpha \in L\}$: $R(\zeta) = \sum_{\alpha \in L} c_\alpha X^\alpha(\zeta)$, $\zeta \in \mathcal{D}$, $c_\alpha \in \mathbb{C}$.

**Proposition 1.** *If $\mathscr{F}$ is a fraction of a full factorial design $\mathscr{D}$, $R = \sum_{\alpha \in L} c_\alpha X^\alpha$ is its counting function and $[\alpha - \beta]$ is the m-tuple made by the componentwise difference in the ring $\mathbb{Z}_{n_j}$, $\left([\alpha_1 - \beta_1]_{n_1}, \ldots, [\alpha_j - \beta_j]_{n_j}, \ldots, [\alpha_m - \beta_m]_{n_m}\right)$, then*

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

We now define projectivity and its relation with orthogonal arrays.

**Definition 3.** A fraction $\mathscr{F}$ *factorially projects* onto the $I$-factors, $I \subset \{1, \ldots, m\}$, if the projection is a multiple full factorial design, i.e. a full factorial design where each point appears equally often. A fraction $\mathscr{F}$ is a *mixed orthogonal array* of strength $t$ if it factorially projects onto any $I$-factors with $\#I = t$.

**Proposition 2.** *A fraction is an* orthogonal array of strength $t$ *if, and only if, all the coefficients $c_\alpha$ of the counting function up to the order $t$ are zero*

## 3 Counting functions and strata

It follows from Proposition 1 and Proposition 2 that the problem of finding fractional factorial designs that satisfy a set of conditions in terms of orthogonality between factors can be written as a polynomial system in which the indeterminates are the complex coefficients $c_\alpha$ of the counting polynomial fraction.

Let us now introduce a different way to describe the full factorial design $\mathscr{D}$ and all its subsets. We consider the indicator functions $1_\zeta$ of all the single points of $\mathscr{D}$. The counting function $R$ of a fraction $\mathscr{F}$ can be written as $\sum_{\zeta \in \mathscr{D}} y_\zeta 1_\zeta$ 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 1 we obtain that the values of the counting function over $\mathscr{D}$, $y_\zeta$, are related to the coefficients $c_\alpha$ by $c_\alpha = \frac{1}{\#\mathscr{D}} \sum_{\zeta \in \mathscr{D}} y_\zeta \overline{X^\alpha(\zeta)}$. As described in Section 2, we consider $m$ factors, $\mathscr{D}_1, \ldots, \mathscr{D}_m$ where $\mathscr{D}_j \equiv \Omega_{n_j} = \{\omega_0, \ldots, \omega_{n_j - 1}\}$, for $j = 1, \ldots, m$. From [7], we recall two basic properties which hold true for the full design $\mathscr{D}$.

**Proposition 3.** *Let $X_j$ be the simple term with level set $\mathscr{D}_j = \Omega_{n_j} = \{\omega_0, \ldots, \omega_{n_j - 1}\}$. Over $\mathscr{D}$, the term $X_j^r$ takes all the values of $\Omega_{s_j}$ equally often, where $s_j = 1$ if $r = 0$ and $s_j = n_j / \gcd(r, n_j)$ if $r > 0$.*

**Proposition 4.** *Let $X^\alpha = X_1^{\alpha_1} \cdots X_m^{\alpha_m}$ be an interaction. $X_i^{\alpha_i}$ takes values in $\Omega_{s_i}$ where $s_i$ is determined according to the previous Proposition 3. Over $\mathscr{D}$, the term $X^\alpha$ takes all the values of $\Omega_s$ equally often, where $s = \operatorname{lcm}(s_1, \ldots, s_m)$.*

Let us now define the strata that are associated to simple and interaction terms.

**Definition 4.** Given a term $X^\alpha, \alpha \in L = \mathbb{Z}_{n_1} \times \ldots \times \mathbb{Z}_{n_m}$, the full design $\mathscr{D}$ is partitioned into the strata $D_h^\alpha = \left\{ \zeta \in \mathscr{D} : \overline{X^\alpha(\zeta)} = \omega_h \right\}$, where $\omega_h \in \Omega_s$ and $s$ is determined according to the previous Propositions 3 and 4.

We use $n_{\alpha,h}$ to denote the number of points of the fraction $\mathscr{F}$ that are in the stratum $D_h^\alpha$, $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, $h = 0, \ldots, s-1$, . The following Proposition 5 (see [3] for proof) links the coefficients $c_\alpha$ with $n_{\alpha,h}$.

**Proposition 5.** *Let $\mathscr{F}$ be a fraction of $\mathscr{D}$ with counting fraction $R = \sum_{\alpha \in L} c_\alpha X^\alpha$. Each $c_\alpha, \alpha \in L$, depends on $n_{\alpha,h}, h = 0, \ldots, s-1$, as $c_\alpha = \frac{1}{\#\mathscr{D}} \sum_{h=0}^{s-1} n_{\alpha,h} \omega_h$, where $s$ is determined by $X^\alpha$ (see Proposition 4).*

We now use a part of Proposition 3 of [7] to get conditions on $n_{\alpha,h}$ that makes $X^\alpha$ centered on the fraction $\mathscr{F}$.

**Proposition 6.** *Let $X^\alpha$ be a term with level set $\Omega_s$ on full design $\mathscr{D}$. Let $P(\zeta)$ be the complex polynomial associated to the sequence $(n_{\alpha,h})_{h=0,\ldots,s-1}$ so that $P(\zeta) = \sum_{h=0}^{s-1} n_{\alpha,h} \zeta^h$ and $\Phi_s$ the cyclotomic polynomial of the $s$-roots of the unity.*

1. *Let $s$ be prime. The term $X^\alpha$ is centered on the fraction $\mathscr{F}$ if, and only if, its $s$ levels appear equally often $n_{\alpha,0} = n_{\alpha,1} = \ldots = n_{\alpha,s-1} = \lambda_\alpha$;*
2. *Let $s = p_1^{h_1} \ldots p_d^{h_d}$, $p_i$ prime, $i = 1, \ldots, d$. The term $X^\alpha$ is centered on the fraction $\mathscr{F}$ if, and only if, the remainder $H(\zeta) = P(\zeta)$ mod $\Phi_s(\zeta)$, whose coefficients are integer linear combinations of $n_{\alpha,h}, h = 0, \ldots, s-1$, is identically zero.*

We observe that, being $D_h^\alpha$ a partition of $\mathscr{D}$, if $s$ is prime, we get $\lambda_\alpha = \frac{\#\mathscr{F}}{s}$.

If we remind that $n_{\alpha,h}$ are related to the values of the counting function $R$ of a fraction $\mathscr{F}$ by $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, this Proposition 6 allows to express the condition $X^\alpha$ *is centered on* $\mathscr{F}$ as integer linear combinations of the values $R(\zeta)$ of the counting function over the full design $\mathscr{D}$. In the Section 4, we will show the use of this property to generate fractional factorial designs.

## 4 Generation of fractions

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

**Definition 5.** Given $\mathscr{C} \subseteq \mathbb{Z}_{n_1} \times \ldots \mathbb{Z}_{n_m}$, a counting function $R = \sum_\alpha c_\alpha X^\alpha$ associated to $\mathscr{F}$ is a $\mathscr{C}$-compatible counting function if $c_\alpha = 0, \ \forall \alpha \in \mathscr{C}$.

We will denote by $OF(n_1 \ldots n_m, \mathscr{C})$ the set of all the fractions of $\mathscr{D}$ whose counting functions are $\mathscr{C}$-compatible. In the next sections, we will show our methodology on Orthogonal Arrays (other examples are in [3]). Let us consider $OA(n, s^m, t)$, i.e. orthogonal arrays with $n$ rows and $m$ columns where each columns has $s$ symbols, $s$ prime and with strength $t$. Using Proposition 2 we have that the coefficients of the corresponding counting functions must satisfy the conditions $c_\alpha = 0$ for all $\alpha \in \mathscr{C}$ where $\mathscr{C} \subseteq L = \{\alpha : 0 < \|\alpha\| \leq t\}$ and $\|\alpha\|$ is the number of non null elements of $\alpha$. It follows that $OF(s^m, \mathscr{C}) = \bigcup_n OA(n, s^m, t)$. Now using Proposition 6, we can express these conditions using strata. If we consider $\alpha \in \mathscr{C}$ we write the condition $c_\alpha = 0$ as $\sum_{\zeta \in D_h^\alpha} y_\zeta = \lambda, h = 0, \ldots, s-1$. To obtain all the conditions it is enough to vary $\alpha \in \mathscr{C}$. We therefore get the system of linear equations $AY = \lambda \underline{1}$ where $A$ is the $(\#\mathscr{C} \times s^m)$ matrix whose rows contains the values, over $\mathscr{D}$, of the indicator function of the strata, $1_{D_h^\alpha}$, $Y$ is the $s^m$ column vector whose entries are the values of the counting function over $\mathscr{D}$, $\lambda$ will be equal to $\frac{\#\mathscr{F}}{s}$ and $\underline{1}$ is the $s^m$ column vector whose entries are all equal to 1. We can write an equivalent homogeneous system if we consider $\lambda$ as a new variable. We obtain $\tilde{A}\tilde{Y} = 0$ where

$$\tilde{A} = \left[ A \left| \begin{matrix} -1 \\ \ldots \\ -1 \end{matrix} \right. \right] = [A, -\underline{1}] \text{ and } \tilde{Y} = \left[ \frac{Y}{\lambda} \right] = (Y, \lambda)$$

It is now immediate to verify that the sum of two Orthogonal Arrays, $Y_1 \in OA(n_1, s^m, t)$ and $Y_2 \in OA(n_2, s^m, t)$ is an Orthogonal Array $Y_1 + Y_2 \in OA(n_1 + n_2, s^m, t)$. The Hilbert Basis [9] is a minimal set of generators such that any $OA(n, s^m, t)$ becomes a linear combination of the generators with positive or null integer coefficients. This approach extends that of [1] where the conditions $c_\alpha = \frac{1}{\#\mathscr{D}} \sum_{\zeta \in \mathscr{F}} \overline{X^\alpha(\zeta)} = 0$ were used. The advantage of using strata is that we avoid computations with complex numbers ($\overline{X^\alpha(\zeta)}$). We explain this point in a couple of examples. For the computation we use 4ti2 [10].

- $OA(n, 2^5, 2)$ were investigated in [1]. We build the matrix $\tilde{A}$ that has 30 rows and 33 columns. We find the same 26,142 solutions as in the cited paper.
- For $OA(n, 3^3, 2)$ we build the matrix $\tilde{A}$ that has 54 rows and 28 columns. We find 66 solutions, 12 have 9 points, all different and 54 have 18 points, one replicated, i.e. support equal to 17.

Let us now consider the general case in which we do not put restrictions on the number of levels. We 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 we have that the coefficients of the corresponding counting functions must satisfy the conditions $c_\alpha = 0$ for all $\alpha \in \mathscr{C}$ where $\mathscr{C} \subseteq L = \{\alpha : \|\alpha\| = 1\}$. Let us consider $c_{1,0}$. From Proposition 3 we have that $X_1$ takes the values in $\Omega_s$ where $s = 4$. From Proposition 6, $X_1$ will be centered on $\mathscr{F}$ if, and only if, the remainder $H(\zeta) = P(\zeta) \bmod \Phi_4(\zeta)$ is identically zero. We have $\Phi_4(\zeta) = 1 + \zeta^2$ (see [6]) and so we can compute the remainder $H(\zeta) = n_{(1,0),0} - n_{(1,0),2} + (n_{(1,0),1} - n_{(1,0),3})\zeta$. The condition that $H(\zeta)$ must be identically zero translates into $n_{(1,0),0} - n_{(1,0),2} = 0$ and $n_{(1,0),1} - n_{(1,0),3} = 0$. Let us now consider $c_{2,0}$. From Proposition 3 we have that $X_1^2$ takes the values in $\Omega_s$ where $s = 2$. From Proposition 6, $X_1^2$ will be centered on $\mathscr{F}$ if, and only if, the remainder $H(\zeta) = P(\zeta) \bmod \Phi_2(\zeta)$ is identically zero. We have $\Phi_2(\zeta) = 1 + \zeta$ (see [6]) and so we can compute the remainder $H(\zeta) = n_{(2,0),0} - n_{(2,0),1}$.

If we repeat the same procedure for all the $\alpha$ such that $\|\alpha\| = 1$ and we recall that $n_{\alpha,h} = \sum_{\zeta \in D_h^\alpha} y_\zeta$, the orthogonal arrays $OA(n, 4^2, 1)$ become the positive integer solutions of the following integer linear homogeneous system:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 & 1 & -1 \\ 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 \\ 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & -1 & -1 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} y_{00} \\ y_{10} \\ y_{20} \\ y_{30} \\ y_{01} \\ y_{11} \\ y_{21} \\ y_{31} \\ y_{02} \\ y_{12} \\ y_{22} \\ y_{32} \\ y_{03} \\ y_{13} \\ y_{23} \\ y_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

It should be noted that the matrix of the coefficients is not full rank, e.g. the first and the fourth rows are equal. This aspect is discussed in [3]. Anyhow the solution method used here does not require a reduction to a full rank matrix. Using 4ti2 we find 24 solutions that correspond to all the Latin Hypercupe Designs (LHD). Analogously for $OA(n, 6^2, 1)$ we find 620 solutions that correspond to all the LHD.

## 5 Sampling

Sometimes, given a set of conditions $\mathscr{C}$ we are interested in picking up a solution more than in finding all the generators. The basic idea is to generate somehow a starting solution and then to randomly walk in the set of all the solutions for a certain number of steps, taking the arrival point as a new but still $\mathscr{C}$-compatible counting function. We can combine the previous results on strata with Markov Chain Monte Carlo Methods to sample one solution. We show the methods on some examples with indicator functions but it can be extended to counting functions.

Let us consider $OA(n, 3^3, 2)$ and let us suppose that we are searching for an orthogonal array with 9 design points and no replications. It means that we are interested in an indicator function whose values $y_\zeta, \zeta \in \mathscr{D}$ satisfy the system of linear equation $AY = \underline{1}$ where $A$, $Y$ and $\underline{1}$ have been defined as in see Section 4.

We now use standard simulated annealing to find one solution of our system [8]. We define the objective function to be maximised as the function that, for every indicator function defined over the design $\mathscr{D}$, counts the number of equations of the linear system $AY = \underline{1}$ that are satisfied. We have implemented this algorithm using SAS/IML. We found a solution in $2,702$ iterations (a couple of seconds on a common laptop). We have also tried the algorithm on $OA(9, 3^4; 3)$ and $4 \times 4$ sudoku (we found one solution in $2,895$ and $2,852$ iterations respectively) and $9 \times 9$ sudoku where we did not find any solution in $100,000$ iterations.

We conclude this section observing that the algorithm can also be used to explore the set of solutions simply replacing *stop when an optimal solution is found* with *store the optimal solution and continue until the maximum number of iterations is reached*. Let us now use the previous results on strata to get a suitable set of *moves*. We will show this procedure in the case in which all the factors have the same number of levels $s$, $s$ prime, but it can also be applied to the general case. In Section 4 we have shown that counting functions must satisfy the system of linear equations $AY = \lambda \underline{1}$, where $A$ corresponds to the set of conditions $\mathscr{C}$ written in terms of strata.

It follows that if, given a $\mathscr{C}$-compatible solution $Y$, such that $AY = \lambda \underline{1}$, we search for an additive move $X$ such that $A(Y+X)$ is still equal to $\lambda \underline{1}$, we have to solve the linear homogenous system $AX = 0$, with $X = (x_\zeta), \zeta \in \mathscr{D}$, $x_\zeta \in \mathbb{Z}$ and $y_\zeta + x_\zeta \geq 0$ for all $\zeta \in \mathscr{D}$. We observe that this set of conditions allows to determine new $\mathscr{C}$-compatible solutions *that give the same* $\lambda$. We know that $\lambda = \frac{\#\mathscr{F}}{s}$ so this homogenous system determines moves that *do not change the dimension of the solutions*.

Let us now consider the extended homogeneous system, where $\tilde{A}$ has already been defined in Section 4, $\tilde{A}\tilde{X} = 0$ with $\tilde{X} = (\tilde{x}_\zeta), \zeta \in \mathscr{D}$, $\tilde{x}_\zeta \in \mathbb{Z}$ and $\tilde{y}_\zeta + \tilde{x}_\zeta \geq 0$ for all $\zeta \in \mathscr{D}$. Given $\tilde{Y} = (Y, \lambda_Y)$, where $Y$ is $\mathscr{C}$-compatible counting function and $\lambda_Y = \frac{\sum_\zeta y_\zeta}{s}$, the solutions of $\tilde{A}\tilde{X} = 0$ determine all the other $\tilde{Y} + \tilde{X} = (Y+X, \lambda_{Y+X})$ such that $\tilde{A}(\tilde{Y} + \tilde{X}) = 0$. $Y+X$ are $\mathscr{C}$-compatible counting functions whose sizes, $s\lambda_{Y+X}$, are, in general, *different from that of* $Y$. We use the theory of Markov basis (see for example [2] where it is also available a rich bibliography on this subject) to determine a set of generators of the moves. We use the following procedure in order to randomly select a $\mathscr{C}$-compatible counting function. We compute a Markov basis of $\ker(A)$ using 4ti2 [10]. Once we have determined the Markov basis of $\ker(A)$, we make a random walk on the *fiber* of $Y$, where $Y$, as usual, contains the values of the counting function of an initial design $\mathscr{F}$. The fiber is made by all the $\mathscr{C}$-compatible counting functions that have the same size of $\mathscr{F}$. The randow walk is done randomly choosing one move among the feasible ones, i.e. among the moves for which we do not get negative values for the new counting function. In the next paragraphs we consider moves for the cases that we have already studied in Section 4.

We consider $OA(8, 2^5, 2)$. We use the matrix $A$, already built in Section 4 and give it as input to 4ti2 to obtain the Markov Basis, that we denote by $\mathscr{M}$. It contains 5.538 different moves. As an initial fraction $\mathscr{F}_0$, we consider the eight-run regular fraction whose indicator function is $R_0 = \frac{1}{4}(1 + X_1X_2X_3)(1 + X_1X_4X_5)$. We obtain the set $\mathscr{M}^f_{R_0}$ of the feasible moves from $R_0$, selecting from $\mathscr{M}$ the moves $M$ such that $R_0(\zeta) + M(\zeta) \geq 0\ \forall \zeta \in \mathscr{D}$ or $R_0(\zeta) - M(\zeta) \geq 0\ \forall \zeta \in \mathscr{D}$. We find 12 moves. We randomly choose one move, $M_{R_0}$, out of the 12 available ones and move to $R_1 = R_0 + \varepsilon_{M_{R_0}} M_{R_0}$ where $\varepsilon_{M_{R_0}}$ is the proper plus or minus sign. We run 1.000 simulations repeating the same loop, generating $R_i$ as $R_i = R_{i-1} + \varepsilon_{M_{R_{i-1}}} M_{R_{i-1}}$. We obtain all the 60 different 8-run fractions, each one with 8 different points as in [1].

We now consider $OA(9, 3^3, 2)$. As before, we use 4ti2 to generate the Markov basis $\mathscr{M}$. It contains 81 different moves. As an initial fraction we consider the nine-run regular fraction $\mathscr{F}_0$ whose indicator function is $R_0 = \frac{1}{3}(1 + X_1X_2X_3 + X_1^2X_2^2X_3^2)$.

Running 1.000 simulations we obtain all the 12 different 9-run fractions, each one with 9 different points as known in the literature and as found in Section 4.

## 6 Conclusions

We considered mixed level fractional factorial designs. Given the counting function $R$ of a fraction $\mathscr{F}$, we translated the constraint $c_\alpha = 0$, where $c_\alpha$ is a generic coefficient of its polynomial representation $R = \sum_\alpha c_\alpha X^\alpha$, into a set of linear constraints with integer coefficients on the values $y_\zeta$ that $R$ takes on all the points $\zeta \in \mathscr{D}$. We obtained the set of generators of the solutions of some problems using Hilbert bases. We also studied moves between fractions. We characterized these moves as the solution of a homogeneous linear system. We defined a procedure to randomly walk among the solutions that is based on the Markov basis of this system. We showed the procedure on some examples. Computations have been made using 4ti2 [10]. Main advantages of the method are that we do not put restrictions on the number of levels of factors and it is not necessary to use software that deals with complex polynomials. Main limit is in the high computational effort that is required. In particular, only a small part of the Markov basis is used because of the requirement that counting functions can only take values greater than or equal to zero. The possibility to generate only the moves that are feasible could make the entire process more efficient and is object of current research. The authors thank the referee for his suggestions.

## References

1. Carlini, E., Pistone, G.: Hibert bases for orthogonal arrays. Journal of Statistical Theory and practice pp. 299–309 (2007)
2. Drton, M., Sturmfels, B., Sullivant, S.: Lectures on Algebraic Statistics. No. 39 in Oberwolfach Seminars. Birkhäuser (2009)
3. Fontana, R., Pistone, G.: Algebraic strata for non symmetrical orthogonal fractional factorial designs and application. La matematica e le sue applicazioni 2010/1, Politecnico di Torino DIMAT (2010)
4. Fontana, R., Pistone, G., Rogantin, M.: Classification of two-level factorial fractions. Journal of Statistical Planning and Inference **87**(1), 149–172 (2000)
5. Fontana, R., Rogantin, M.P.: Indicator function and sudoku designs. In: P. Gibilisco, E. Riccomagno, M.P. Rogantin, H.P. Wynn (eds.) Algebraic and Geometric Methods in Statistics, pp. 203–224. Cambridge University Press (2010)
6. Lang, S.: Algebra. Addison Wesley, Reading, Mass. (1965)
7. Pistone, G., Rogantin, M.: Indicator function and complex coding for mixed fractional factorial designs. J. Statist. Plann. Inference **138**(3), 787–802 (2008)
8. Ross, S.M.: Simulation, third edn. Statistical Modeling and Decision Science. Academic Press Inc., San Diego, CA (2002)
9. Schrijver, A.: Theory of linear and integer programming. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons Ltd., Chichester (1986). A Wiley-Interscience Publication
10. 4ti2 team: 4ti2—a software package for algebraic, geometric and combinatorial problems on linear spaces. Available at www.4ti2.de