Algebraic statistics in mixed factorial design

Maria Piera Rogantin
DIMA – Università di Genova – rogantin@dima.unige.it

Giovanni Pistone
DIMAT – Politecnico di Torino – giovanni.pistone@polito.it

ICODOE, Memphis TN, May 12-15, 2005

Polynomial algebra and design of experiments

The application of computational commutative algebra to the study of estimability, confounding on the fractions of factorial designs has been proposed by Pistone & Wynn (Biometrika 1996).

1**st** **idea** Each set of points $\mathcal{D} \subseteq \mathbb{Q}^m$ is the set of the solutions of a system of polynomial equations.

2**nd** **idea** Each real valued function defined on $\mathcal{D}$ is a *polynomial function* with coefficients into the field of real number $\mathbb{R}$.

Polynomial representation of the full factorial design

- $A_i = \{a_{ij} \ : \ j = 1, \ldots, n_i\}$ **factors**

  a_{ij} **levels** coded by rational numbers or by complex numbers

- $\mathcal{D} = A_1 \times \ldots \times A_m \subset \mathbb{Q}^m$ (or $\mathcal{D} \subset \mathbb{C}^m$) **full factorial design**

$\mathcal{D}$ is the solution set of the system of polynomial equations

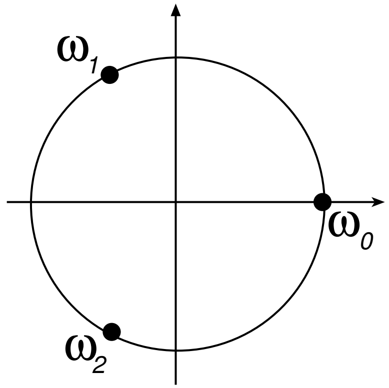
$$\left\{ \begin{array}{rcl} (X_1 - a_{11}) \cdots (X_1 - a_{1n_1}) & = & 0 \\ (X_2 - a_{21}) \cdots (X_2 - a_{2n_2}) & = & 0 \\ & \vdots & \\ (X_m - a_{m1}) \cdots (X_m - a_{mn_m}) & = & 0 \end{array} \right. \quad \text{or} \quad \left\{ \begin{array}{rcl} X_1^{n_1} & = & \sum_{k=0}^{n_1-1} \psi_{1k} \, X_1^k \\ & \vdots & \\ & \vdots & \\ X_m^{n_m} & = & \sum_{k=0}^{n_m-1} \psi_{mk} \, X_m^k \end{array} \right. \quad \text{rewriting rules}$$

A **fraction** is a subset of a full factorial design, $\mathcal{F} \subset \mathcal{D}$.

It is obtained by adding equations (*generating equations*) to restrict the set of solutions.

Complex coding for levels

We code the n levels of a factor A with the n-th roots of the unity:



$$\omega_k = \exp\left(\mathrm{i}\ \frac{2\pi}{n}\ k\right) \qquad \text{for } k = 0, \ldots, n-1$$

The mapping

$$\mathbb{Z}_n \longleftrightarrow \Omega_n \subset \mathbb{C}$$
$$k \longleftrightarrow \omega_k$$

is a group isomorphism of the additive group of $\mathbb{Z}_n$ on the multiplicative group $\Omega_n \subset \mathbb{C}$.

The **full factorial design** $\mathcal{D}$, as a subset of $\mathbb{C}^m$, is defined by the system of equations

$$\zeta_j^{n_j} - 1 = 0 \qquad \text{for } j = 1, \ldots, m$$

Responses on a design (functions defined on $\mathcal{D}$)

- $X_i : \mathcal{D} \ni (d_1, \ldots, d_m) \mapsto d_i$ **projection**, frequently called **factor**
- $X^\alpha = X_1^{\alpha_1} \cdots X_m^{\alpha_m}$ **monomial responses or terms or interactions**
  $\alpha = (\alpha_1, \ldots, \alpha_m) \qquad 0 \leq \alpha_i \leq n_i - 1,\ i = 1, \ldots, m$
- $L = \{(\alpha_1, \ldots, \alpha_m) \ :\ 0 \leq \alpha_i \leq n_i - 1,\ i = 1, \ldots, m\}$ **exponents of all the interactions**

Definitions:

- **Mean value of f** on $\mathcal{D}$: $\qquad E_{\mathcal{D}}(f) = \frac{1}{\#\mathcal{D}} \sum_{d \in \mathcal{D}} f(d)$
- A response f is **centered** if $\quad E_{\mathcal{D}}(f) = 0$
- Two responses f and g are **orthogonal on $\mathcal{D}$** if $\qquad < f, g >= 0$

$$< f, g >= E_{\mathcal{D}}(f\ \overline{g}) = \frac{1}{\#\mathcal{D}} \sum_{d \in \mathcal{D}} f(d)\ \overline{g}(d)$$

Space of the functions on a full or fractional design

• It is a **vector space** (classical results derive from this structure) whit Hermitian product defined before

• It is a **ring** (algebraic statistical approach)
The products are reduced with the rules derived by the polynomial representation of the full factorial design:

$$X_i^{n_i} = \sum_{k=0}^{n_i-1} \psi_{ik} \, X_i^k \; , \quad \psi_{ik} \in \mathbb{C} \qquad \text{for } i = 1, \ldots, m$$

Using the **complex coding**, the set of all the **monomial** responses on $\mathcal{D}$:
$\{X^\alpha, \alpha \in L\}$ is an **orthonormal monomial basis** of the set of all the complex functions defined on the full factorial design $\mathcal{C}(\mathcal{D})$

Each function defined on **full factorial design** is represented in a unique way by an identified complete regression model (i.e. as a linear combination of constant, simple terms and interactions):

$$\mathcal{C}(\mathcal{D}) = \left\{ \sum_{\alpha \in L} \theta_\alpha \, X^\alpha \; , \; \theta_\alpha \in \mathbb{C} \right\}$$

Indicator function of a fraction

The description of fractional factorial designs using the polynomial representations of their indicator functions has been

- introduced for **binary designs** in *Fontana R., Pistone G. and Rogantin M. P. (1997) and (2000)*

- introduced independently with the name of generalized word length patterns in *Tang B. and Deng L. Y. (1999)*

- generalized to replicates in *Ye, K. Q. (2003)*

- extended to not binary factors using orthogonal polynomials with an integer coding of levels in *Cheng S.-W. and Ye K. Q. (2004)*

Here we generalize to **multilevel factorial designs with replicates** using the complex coding.

The **indicator function** F **of a fraction** $\mathcal{F}$ is a response defined on the full factorial design $\mathcal{D}$ such that

$$F(\zeta) = \begin{cases} 1 & \text{if } \zeta \in \mathcal{F} \\ 0 & \text{if } \zeta \in \mathcal{D} \setminus \mathcal{F} \end{cases}$$

In a **fraction with replicates** $\mathcal{F}_{\text{rep}}$ the **counting function** R is a response on the full factorial design showing the number of replicates of a point ζ.

They are represented as polynomials:

$$F(\zeta) = \sum_{\alpha \in L} b_\alpha \, X^\alpha(\zeta) \qquad R(\zeta) = \sum_{\alpha \in L} c_\alpha \, X^\alpha(\zeta)$$

The coefficients b_α and c_α satisfy the following properties:

- $b_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}} \overline{X^\alpha(\zeta)}$ and $c_\alpha = \frac{1}{\#\mathcal{D}} \sum_{\zeta \in \mathcal{F}_{\text{rep}}} \overline{X^\alpha(\zeta)}$
- $\bar{b}_\alpha = b_{[-\alpha]}$ and $\bar{c}_\alpha = c_{[-\alpha]}$ because F is real valued.

Important statistical features of the fraction can be read out from the form of the polynomial representation of the indicator function.

Orthogonality

of responses in a vector space, based on a scalar or Hermitian product:

$$< f, g >= E_{\mathcal{D}}(f \ \overline{g}) = 0$$

Two orthogonal responses are not confounded and the estimators of their coefficients in a model are not correlated.

of factors : "all level combinations appear equally often"

Vector orthogonality is affected by the coding of the levels, while factor orthogonality is not.

If the levels are coded with the complex roots of the unity the two notion of orthogonality are essentially equivalent

Indicator function and orthogonality

1. A simple term or an interaction term X^α is **centered** on $\mathcal{F}$ if and only if $c_\alpha = c_{[-\alpha]} = 0$.

2. Two simple or interaction terms X^α and X^β **are orthogonal** on $\mathcal{F}$ if and only if $c_{[\alpha-\beta]} = c_{[\beta-\alpha]} = 0$;

3. If X^α is centered then,
   for any β and γ such that $\alpha = [\beta - \gamma]$ or $\alpha = [\gamma - \beta]$,
   X^β **is orthogonal to** X^γ.

Some other results about orthogonality

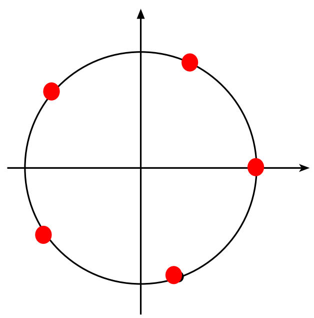
(following from the structure of the roots of the unity as cyclical group)

Let X^α be a term with level set Ω_s on the full factorial design $\mathcal{D}$.

The s levels of X^α appear equally often

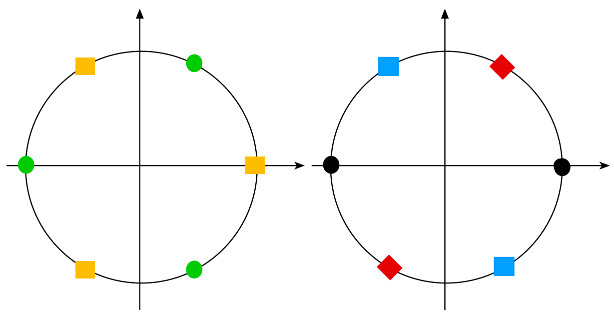
if and only if

s *prime*



the coefficient $c_\alpha = 0$
or, the term X^α is centered

s *not prime*



the coefficients $c_\alpha = 0$ and $c_{\alpha^r} = 0$
or, the terms X^α, $(X^\alpha)^r$ are centered

for any possible r

The two orthogonalities and the indicator functions

We split the factors into two blocks: $I \subset \{1, \ldots, m\}$ $J = I^c$

$$\mathcal{D} = \mathcal{D}_I \times \mathcal{D}_J$$

1. All **level combinations of the I-factors appear equally often**

if and only if

all the coefficients of the counting function involving only the I-factors are 0, that is $c_{\alpha_I} = 0$ with $\alpha_I \in L_I,\ \alpha_I \neq (0, 0, \ldots, 0)$

Then, for any β_I and γ_I in L_I such that $\alpha_I = [\beta_I - \gamma_I]$ or $\alpha_I = [\gamma_I - \beta_I]$,

X^{β_I} **is orthogonal to** X^{γ_I}

and, in particular, for simple terms:

$$X_k^{r_k} \perp X_h^{r_h} \qquad k, h \in I$$

2. A fraction is an **orthogonal array of strength t**

if and only if

all the coefficients of the counting function up to the order t are zero:

$$c_\alpha = 0 \quad \forall\, \alpha \text{ of order up to } t,\ \alpha \neq (0, 0, \ldots, 0)\ .$$

Then, for any β and γ of order up to t such that $\alpha = [\beta - \gamma]$ or $\alpha = [\gamma - \beta]$, **$X^\beta$ is orthogonal to X^γ**

3. If there exists a subset J of $\{1, \ldots, m\}$ such that the J-factors appear in all the non null elements of the counting function,

then

all **level combinations of the I-factors appear equally often** $(I = J^c)$

Regular fractions

- $\mathcal{F}$ a fraction without replicates where all factors have n levels
- Ω_n the set of the n-th roots of the unity, $\Omega_n = \{\omega_0, \ldots, \omega_{n-1}\}$
- $\mathcal{L}$ a subset of exponents, $\mathcal{L} \subset L = (\mathbb{Z}_n)^m$ containing $(0, \ldots, 0)$, $l = \#\mathcal{L}$
- e a map from $\mathcal{L}$ to Ω_n, $e : \mathcal{L} \to \Omega_n$

A fraction $\mathcal{F}$ is **regular** if:

1. $\mathcal{L}$ is a sub-group of L,

2. e is a homomorphism, $e([\alpha + \beta]) = e(\alpha)\ e(\beta)$ for each $\alpha, \beta \in \mathcal{L}$,

3. the *defining equations* are of the form

$$X^\alpha = e(\alpha) \quad , \qquad \alpha \in \mathcal{L}$$

   If $\mathcal{H}$ is a minimal generator of the group $\mathcal{L}$, then the equations $X^\alpha = e(\alpha)$ with $\alpha \in \mathcal{H} \subset \mathcal{L}$ are called minimal *generating* equations.

Notice that:

- we consider the general case where $e(\alpha)$ can be different from 1

- we have no restriction on the number of levels

- from items (1) and (2) it follows that a necessary condition is the $e(\alpha)$'s must belong to the subgroup spanned by the values X^{α}.
  For example for $n = 6$ an equation like $X_1^3 X_2^3 = \omega_2$ can not be a defining equation

Indicator function and regular fractions

(Pistone, Rogantin, 2005)

The following statements are equivalent:

1. The fraction $\mathcal{F}$ is **regular** according to previous definition with *defining equations* $X^\alpha = e(\alpha)$, $\alpha \in \mathcal{L}$

2. The **indicator function** of the fraction has the form
$$F(\zeta) = \frac{1}{l} \sum_{\alpha \in \mathcal{L}} \overline{e(\alpha)} \, X^\alpha(\zeta) \qquad \zeta \in \mathcal{D}$$
where $\mathcal{L}$ is a given subset of L and $e : \mathcal{L} \to \Omega_n$ is a given mapping.

3. For each $\alpha, \beta \in L$ the parametric functions represented on $\mathcal{F}$ by the terms X^α and X^β are **either orthogonal or totally confounded**

Example 1: a regular fraction of a 3^4 design

The generating equations of the fraction are

$$X_1X_2X_3^2 = 1 \quad \text{and} \quad X_1X_2^2X_4 = 1 \ .$$

Then: $\mathcal{H} = \{(1,1,2,0),(1,2,0,1)\}$

$e(1,1,2,0) = e(1,2,0,1) = \omega_0 = 1$

$\mathcal{L} = \{(0,0,0,0),(0,1,1,2),(0,2,2,1),(1,1,2,0),$
$(2,2,1,0),(1,2,0,1),(2,1,0,2),(1,0,1,1),(2,0,2,2)\}.$

The indicator function is:

$$F = \frac{1}{9}\Big(1 + X_2X_3X_4 + X_2^2X_3^2X_4^2 + X_1X_2X_3^2 + X_1^2X_2^2X_3 + X_1X_2^2X_4 + X_1^2X_2X_4^2 + X_1X_3X_4^2 + X_1^2X_3^2X_4\Big)$$

Example 2: a regular fraction of a 6^3 design

The terms X^α take values in:

Ω_6 or in one of the two subgroups $\{1, \omega_3\}$ and $\{1, \omega_2, \omega_4\}$.

The generating equations of the fraction are

$$X_1^3 X_3^3 = \omega_3 \qquad \text{and} \qquad X_2^4 X_2^4 X_3^2 = \omega_2$$

Then: $\mathcal{H} = \{(3,0,3),(4,4,2)\}$

$e(3,0,3) = \omega_3$, $e(4,2,2) = \omega_2$

$\mathcal{L} = \{(0,0,0),(3,0,3),(4,4,2),(2,4,4),(1,4,5),(5,2,1)\}$.

The full factorial design has 216 points and the fraction has 36 points

$$
\begin{array}{ccc}
X_1 & X_2 & X_3 \\
\end{array}
\begin{pmatrix}
\omega_0 & \omega_0 & \omega_1 \\
\omega_0 & \omega_1 & \omega_5 \\
\omega_0 & \omega_2 & \omega_3 \\
\omega_0 & \omega_3 & \omega_1 \\
\omega_0 & \omega_4 & \omega_5 \\
\omega_0 & \omega_5 & \omega_3 \\
\omega_1 & \omega_0 & \omega_2 \\
\omega_1 & \omega_1 & \omega_0 \\
\omega_1 & \omega_2 & \omega_4 \\
\omega_1 & \omega_3 & \omega_2 \\
\omega_1 & \omega_4 & \omega_0 \\
\omega_1 & \omega_5 & \omega_4
\end{pmatrix}
\quad
\begin{array}{ccc}
X_1 & X_2 & X_3 \\
\end{array}
\begin{pmatrix}
\omega_2 & \omega_0 & \omega_3 \\
\omega_2 & \omega_1 & \omega_1 \\
\omega_2 & \omega_2 & \omega_5 \\
\omega_2 & \omega_3 & \omega_3 \\
\omega_2 & \omega_4 & \omega_1 \\
\omega_2 & \omega_5 & \omega_5 \\
\omega_3 & \omega_0 & \omega_4 \\
\omega_3 & \omega_1 & \omega_2 \\
\omega_3 & \omega_2 & \omega_0 \\
\omega_3 & \omega_3 & \omega_4 \\
\omega_3 & \omega_4 & \omega_2 \\
\omega_3 & \omega_5 & \omega_0
\end{pmatrix}
\quad
\begin{array}{ccc}
X_1 & X_2 & X_3 \\
\end{array}
\begin{pmatrix}
\omega_4 & \omega_0 & \omega_5 \\
\omega_4 & \omega_1 & \omega_3 \\
\omega_4 & \omega_2 & \omega_1 \\
\omega_4 & \omega_3 & \omega_5 \\
\omega_4 & \omega_4 & \omega_3 \\
\omega_4 & \omega_5 & \omega_1 \\
\omega_5 & \omega_0 & \omega_0 \\
\omega_5 & \omega_1 & \omega_4 \\
\omega_5 & \omega_2 & \omega_2 \\
\omega_5 & \omega_3 & \omega_0 \\
\omega_5 & \omega_4 & \omega_4 \\
\omega_5 & \omega_5 & \omega_2
\end{pmatrix}
$$

The indicator function is:

$$F = \frac{1}{6}\left(1 + \omega_3 X_1^3 X_3^3 + \omega_4 X_1^4 X_2^4 X_3^2 + \omega_2 X_1^2 X_2^2 X_3^4 + \omega_1 X_1 X_2^4 X_3^5 + \omega_5 X_1^5 X_2^2 X_3\right)$$

Example 3: all the regular fractions of a 4^2 design

1. Using previous definition.

   All the inequivalent fractions with generating equations $X^\alpha = 1$

$$
\begin{array}{cc} X_1 & X_2 \end{array}
\begin{pmatrix} \omega_0 & \omega_0 \\ \omega_1 & \omega_1 \\ \omega_2 & \omega_2 \\ \omega_3 & \omega_3 \end{pmatrix}
\quad
\begin{array}{cc} X_1 & X_2 \end{array}
\begin{pmatrix} \omega_0 & \omega_0 \\ \omega_1 & \omega_3 \\ \omega_2 & \omega_2 \\ \omega_3 & \omega_1 \end{pmatrix}
\quad
\begin{array}{cc} X_1 & X_2 \end{array}
\begin{pmatrix} \omega_0 & \omega_0 \\ \omega_0 & \omega_2 \\ \omega_1 & \omega_1 \\ \omega_1 & \omega_3 \\ \omega_2 & \omega_0 \\ \omega_2 & \omega_2 \\ \omega_3 & \omega_1 \\ \omega_3 & \omega_3 \end{pmatrix}
\quad
\begin{array}{cc} X_1 & X_2 \end{array}
\begin{pmatrix} \omega_0 & \omega_0 \\ \omega_0 & \omega_2 \\ \omega_2 & \omega_1 \\ \omega_2 & \omega_3 \end{pmatrix}
\quad
\begin{array}{cc} X_1 & X_2 \end{array}
\begin{pmatrix} \omega_0 & \omega_0 \\ \omega_1 & \omega_2 \\ \omega_2 & \omega_0 \\ \omega_3 & \omega_2 \end{pmatrix}
$$

   Their indicator functions are respectively:

$$
\frac{1}{3}\left(1 + X_1X_2^3 + X_1^3X_2\right) \;, \qquad \tfrac{1}{3}\left(1 + X_1X_2 + X_1^3X_2^3\right) \;, \qquad \frac{1}{2}\left(1 + X_1^2X_2^2\right)
$$

$$
\frac{1}{3}\left(1 + X_1X_2^2 + X_1^3X_2^2\right) \;, \qquad \tfrac{1}{3}\left(1 + X_1X_2^3 + X_1^3X_2\right)
$$

   The last two fractions do not fully project on both factors

2. Using Galois Fields and pseudo-factors.

All the inequivalent fractions in polynomial-Galois notation and in pseudo-factor multiplicative notation

$$\begin{matrix} Z_1 & Z_2 \end{matrix} \qquad \begin{matrix} X_{10} & X_{11} & X_{20} & X_{21} \end{matrix}$$
$$\begin{pmatrix} 1+x & 1+x \\ 1 & 1 \\ x & x \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} -1 & -1 & -1 & -1 \\ -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

$$\begin{matrix} Z_1 & Z_2 \end{matrix} \qquad \begin{matrix} X_{10} & X_{11} & X_{20} & X_{21} \end{matrix}$$
$$\begin{pmatrix} 1+x & x \\ 1 & 1+x \\ x & 1 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} -1 & -1 & 1 & -1 \\ -1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$$

The first fraction corresponds to the first fraction in Item (1), but the latter is not equivalent to any fraction listed in Item (1).

References: Indicator function

- Fontana, R., Pistone, G. and Rogantin, M. P. (2000). Classification of two-level factorial fractions, *J. Statist. Plann. Inference* **87**(1), 149–172.
- Tang, B., Deng, L. Y., 1999. Minimum G_2-aberration for nonregular fractinal factorial designs. The Annals of Statistics 27 (6), 1914–1926.
- Ye, K. Q., 2003. Indicator function and its application in two-level factorial designs. The Annals of Statistics 31 (3), 984–994.
- Pistone, G., Rogantin, M.-P., 2003. Complex coding for multilevel factorial designs. Technical report n. 22 October 2003 Dipartimento di Matematica Politecnico di Torino.
- Ye, K. Q., 2004. A note on regular fractional factorial designs. Statistica sinica 14 (4), 1069–1074.
- Cheng, S.-W., Ye, K. Q., 2004. Geometric isomorphism and minimum aberration for factorial designs with quantitative factors. The Annals of Statistics 32 (5).
- Pistone, G., Rogantin, M.-P., 2005. Indicator function and complex coding for mixed fractional factorial designs. Technical report n. 13 April 2005 Dipartimento di Matematica Politecnico di Torino. Submitted.

References: Algebraic statistics in DOE

- Pistone, G., Wynn, H. P., Mar. 1996. Generalised confounding with Gröbner bases. Biometrika 83 (3), 653–666.

- Bates R., Giglio B., Riccomagno E. and Wynn H., 1998. Gröbner basis methods in polynomial modelling. Improceeding of COMPSTAT 98, ed. R. Payne p. 179–184

- Robbiano, L., 1998. Gröbner bases and statistics. In: Buchberger, B., Winkler, F. (Eds.), Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases). Vol. 251 of London Mathematical Society Lecture Notes. Cambridge University Press, pp. 179–204.

- Robbiano, L., Rogantin, M.-P., 1998. Full factorial designs and distracted fractions. In: Buchberger, B., Winkler, F. (Eds.), Gröbner Bases and Applications (Proc. of the Conf. 33 Years of Gröbner Bases). Vol. 251 of London Mathematical Society Lecture Notes Series. Cambridge University Press, pp. 473–482.

- Holliday T., Pistone G., Riccomagno E. and Wynn H., 1999. The application of computational algebraic geometry to the analysis of designed experiments: a case study, Comput. Statist., 14.2, p.213–231

- Pistone, G., Riccomagno, E., Wynn, H. P., 2001. Algebraic Statistics: Computational Commutative Algebra in Statistics. Chapman&Hall, Boca Raton.

- Galetto, F., Pistone, G., Rogantin, M. P., 2003. Confounding revisited with commutative computational algebra. J. Statist. Plann. Inference 117 (2), 345–363. and applications, With a foreword by C. R. Rao. Springer-Verlag, New York.

References: Complex coding

- Bailey, R. A., 1982. The decomposition of treatment degrees of freedom in quantitative factorial experiments. J. R. Statist. Soc., B 44 (1), 63–70.

- Kobilinsky, A., 1990. Complex linear model and cyclic designs. Linear Algebra and its Applications 127, 227–282.

- Kobilinsky, A., Monod, H., 1991. Experimental design generated by group morphism: An introduction. Scand. J. Statist. 18, 119–134.

- Edmondson, R. N., 1994. Fractional factorial designs for factors with a prime number of quantitative levels. J. R. Statist. Soc., B 56 (4), 611–622.

- Kobilinsky, A., Monod, H., 1995. Juxtaposition of regular factorial designs and the complex linear model. Scand. J. Statist. 22, 223–254.

- Collombier, D., 1996. Plans D'Expérience Factoriels. Construction et propriétés des fractions de plans. No. 21 in Mathématiques et Applications. Springer, Paris.