

Complex linear models for general fractional factorial designs

Modelli lineari complessi per piani fattoriali frazionari generali

Giovanni Pistone
DIMAT
Politecnico di Torino
pistone@calvino.polito.it

Maria Piera Rogantin¹
Dima
Università di Genova
rogantin@dima.unige.it

Riassunto: I livelli di un fattore possono essere rappresentati dalle radici n -esime dell'unità. In questo lavoro mostriamo come definire modelli lineari adatti a questa codifica e confrontiamo tali modelli con quelli relativi alla codifica tradizionale $0, 1, \dots, n$.

Keywords: factorial design, complex design, polynomial ring, Gröbner basis.

1. Introduction

Confounding designs is a field in factorial where algebraic methods are widely used. However, the use of finite field algebra, introduced in Bose (1947), has not received much consideration in recent books. The classical monograph by Raktov *et al.* (1981) and the recent one by Wu and Hamada (2000) rely more on linear algebra.

The use of a different algebraic method, called commutative algebra or polynomial ring algebra, has been firstly advocated in Pistone and Wynn (1996) and developed in Pistone *et al.* (2001). The specific application to factorial designs has been the object of a series of papers: Fontana *et al.* (2000) on the binary case with the coding $-1, +1$, and Robbiano (1998), Robbiano and Rogantin (1998) on general designs.

In the present paper, we use a complex coding of levels, where the k -th level of a n -level factor is represented by the complex number $\exp(i \frac{2\pi}{n} k)$, $i = \sqrt{-1}$. Complex linear models on group-generated designs were discussed in Kobilinsky (1990), while results on non-binary factorial designs are in Galetto *et al.* (2001).

A basic dictionary of commutative algebra is given below. Main references are Cox *et al.* (1997) and Kreuzer and Robbiano (2000).

Given a numerical coding of factor levels in a number field k , we identify a design D with the set of all solutions of a suitable system of polynomial equations. The set of all polynomials which are zero on the design points is called the *design ideal*, $\text{Ideal}(D)$. Any set of polynomial equations defining (with multiplicity one) the same design D is called a *set of generators* of $\text{Ideal}(D)$. A *monomial ordering* is a total ordering of monomials which is compatible with the product. Many such orderings exist. Given one of them τ , in each polynomial the maximal monomial is called the *leading term*.

There exist special sets of generators, called *Gröbner bases*, with the following property: *the set of monomials which are not divided by any of the leading terms form a linear basis for the k -vector space of functions on the design*. These bases are denoted by $\text{Est}_\tau(D)$ because depend on τ only. By construction such a monomial basis is *hierarchi-*

¹Il lavoro è stato svolto con il contributo del Politecnico di Torino (Ricerca di Base) e del CNR (PCS "Algoritmi algebrici per la pianificazione statistica degli esperimenti simulati")

cal, that is all divisors of each term in $\text{Est}_\tau(D)$ are in $\text{Est}_\tau(D)$. We will consider general hierarchical monomial bases besides those generated by Gröbner algorithms.

For a computable number field k , mainly the rationals, $k = \mathbb{Q}$, all the needed computations are implemented in largely available softwares. Among them we use CoCoA, Capani *et al.* (2000), and Maple, Char *et al.* (1991). However no of these softwares appear to be able to deal with the complex case, $k = \mathbb{Q} + i\mathbb{Q}$.

Our basic aim is to give a new mathematical background to the traditional “additive” and “multiplicative” notations. Our conclusion is the following: as most of the authors actually do, the additive notation can be avoided and the results on confounding are deduced from commutative algebra using a new interpretation of the multiplicative notation.

2. Fractional factorial designs in the complex plane

Let m be the number of factors and n_j be the number of levels of each factor, $j = 1, \dots, m$. We code the n levels of a factor A with the complex solutions of the equation $\zeta^n = 1$, e.g. $\omega_k = \exp(i \frac{2\pi}{n}k)$ for $k = 0, 1, \dots, n-1$. We do not consider here the natural cyclic ordering of the roots of unity, because we are concerned with *qualitative factors* without any natural ordering of factor levels.

The full factorial design D , as a subset of the m -dimensional complex vector space $\mathbb{C}^m$, is defined by the equations $X_j^{n_j} - 1 = 0$, for $j = 1, \dots, m$. A fraction is a subset $F \subseteq D$; all fractions are obtained by adding further equations, called *generating equations*, to restrict the number of solutions. This is discussed in Fontana *et al.* (2000), where they are improperly termed “defining” equations.

For example, a classical 3_{III}^{4-2} regular fraction (Wu and Hamada, 2000, Table 5A.1) is defined by $X_j^3 - 1 = 0$, $j = 1, \dots, 4$, together with the generating equations $X_1X_2X_3 - 1 = 0$ and $X_1X_2^2X_4 - 1 = 0$. Such a representation of the fraction is called “multiplicative” notation. The equations are actually defined on the complex field $\mathbb{C}$. The recoding mapping between the Galois field $\mathbb{Z}_3$ and the field $\mathbb{C}$

$$k \leftrightarrow \exp\left(i \frac{2\pi}{3}k\right)$$

is an homomorphism of the additive group on the multiplicative group. Then it maps generating equations which are additive in $\mathbb{Z}_3$ into binomial equations in $\mathbb{C}$. In this case, and in all the other regular cases, the generating equations are binomial and have real coefficients. In the following, we allow for general subsets of the full factorial design and, as a consequence, no special form for the generating equations is assumed.

A *complex response* f on the fraction F is a $\mathbb{C}$ -valued function defined on F . It can be considered as the restriction to F on a complex polynomial. The general algebraic theory applies and the Gröbner algorithms produce an hierarchical monomial basis for the $\mathbb{C}$ -vector space of responses:

$$\text{Est}_\tau(F) = \{X^\beta : \beta \in M\}$$

where β is a multi-index, $\beta = (\beta_1, \dots, \beta_m)$, and X^β is a monomial, $X^\beta = X_1^{\beta_1} \dots X_m^{\beta_m}$. Each response f is represented as an unique linear combination of elements of $\text{Est}_\tau(F)$:

$$f = \sum_{\beta \in M} \theta_\beta X^\beta \quad \text{with } \theta_\beta \in \mathbb{C} \quad . \quad (1)$$

In applications, we are interested in the vector space of *real* responses, and in its linear representation analogous to Eq. (1), which is usually called full linear model on F . In the real coding case, a real valued linear model is simply obtained by taking the model matrix of the monomial basis. In this complex coding case there are two basic options we describe below. Examples related to standard designs, either regular or non-regular, will appear elsewhere.

Over-parametrized complex model. The linear model is given by the Eq. (1) together with a linear restriction on the θ 's implying f is real valued, e.g. $f = \overline{f}$. One way of writing this condition is the following. Let C be the matrix $[C(\alpha, \beta)]_{\alpha, \beta \in M}$ defined by $\overline{X^\alpha} = \sum_{\beta \in M} C(\alpha, \beta) X^\beta$. Representing functions on F as column vectors and column vectors with square brackets, we have $\overline{[X]} = C [X]$. From $[f] = [\theta]^t [X]$ we get $\overline{[f]} = \overline{[\theta]}^t C [X]$, then

$$f = \overline{[f]} \quad \text{if} \quad [\theta]^t = \overline{[\theta]}^t C \quad .$$

This model is over-parametrized because the number of monomial in $\text{Est}_\tau(F)$ is equal to the number of points of F , while the number of real parameter in Eq. (1) is twice as many.

Real model. We derive a real linear model using an argument due to (Kobilinsky, 1990, Sec. 3). If we multiply the set of the real responses by a complex number we get a self-conjugate subspace of the complex responses: $\overline{af} = \overline{a}\overline{f} = \overline{a}f$. As such subspace has at least one real valued vector basis, we proceed as follows. First, we look for monomials X^α which are real valued on F ; they form the first block of the real basis. Then, we look for monomials which are imaginary valued on F and take their imaginary parts form the second block of the real basis. Finally, we chose a suitable subset of remaining monomials and for each of them we take the real and imaginary parts to complete the real basis.

3. Different codings of levels

Finally, we compare monomial bases under different codings of the levels.

Let $S \in \mathbb{C}$ and $T \in \mathbb{C}$ be two sets of levels and $\phi : S \rightarrow T$ a bijective affine transformation of level codings. Assume that similar transformations are applied to each factor independently, and let us call Φ the multivariate affine transformation. Let F_1 and F_2 , $F_2 = \Phi(F_1)$, be the same fraction with the different codings. We denote by X_j, Y_j the generic factors on F_1, F_2 .

Let $B = \{X^\alpha : \alpha \in M\}$ be a hierarchical monomial basis of the responses on F_1 . Then $B' = \{Y^\alpha : \alpha \in M\}$ is a hierarchical monomial basis of the responses on F_2 .

In fact, we have $Y_j = a_j X_j + b_j$, and for each $\alpha \in M$,

$$Y^\alpha = \prod_{j=1}^m (a_j X_j + b_j)^{\alpha_j} = A^\alpha X^\alpha + X^\beta \text{'s terms with } \beta \leq \alpha \text{ componentwise,}$$

where, because of the hierarchical structure of B , each β belongs to M . Then, as $\beta \leq \alpha$ implies that β precedes α in all monomial orders, the matrix transformation from X^β 's to Y^α 's is lower triangular, in particular non-singular.

More complex transformations between codings or designs, in general will change the monomial basis. A case of particular interest is the transformation of a factor X with n levels to $n-1$ pseudo-factors $Y_1, \dots, Y_{n-1}$ defined by $Y_j = X^j$. In this case each element of a monomial basis is transformed in a square free monomial.

Finally, in other cases the monomial basis of the recoded design can be computed by the general Gröbner algorithm.

We illustrate this with an example for an n -level factor with the complex coding $\omega_0, \omega_1, \dots, \omega_{n-1}$ or the coding $0, 1, \dots, n-1$. We denote by Z and A the factor values respectively in multiplicative and additive notation.

Let $E_0, E_1, \dots, E_{n-1}$ be 0, 1-valued factors, such that E_k represents k and ω_k :

$$A = \sum_{k=0}^{n-1} k E_k, \quad Z = \sum_{k=0}^{n-1} \omega_k E_k$$

From the previous formulae it is easy to show that there is a linear transformation between the pseudo-factors

$$A^h = \sum_{k=1}^{n-1} k^h E_k = \sum_{k=1}^{n-1} k^h \frac{\prod_{j=0, j \neq k}^{n-1} (Z - \omega_j)}{\prod_{j=0, j \neq k}^{n-1} (\omega_k - \omega_j)}$$

and a similar formula applies to the transformation from A to Z .

References

- Bose R. (1947) Mathematical theory of the symmetrical factorial designs, *Sankhyā*, 8, 107–166.
- Capani A., Niesi G. and Robbiano L. (2000) *CoCoA, a system for doing Computations in Commutative Algebra*, `cocoa.dima.unige.it`, 4.0 edition.
- Char B., Geddes K., Gonnet G., Leong B., Monogan M. and Watt S. (1991) *MAPLE V Library Reference Manual*, Springer-Verlag, New York.
- Cox D., Little J. and O’Shea D. (1997) *Ideal, Varieties, and Algorithms*, Springer-Verlag, New York, 2nd edition, 1st ed. 1992.
- Fontana R., Pistone G. and Rogantin M.P. (2000) Classification of two-level factorial fractions, *J. Statist. Plann. Inference*, 87, 1, 149–172.
- Galetto F., Pistone G. and Rogantin M.P. (2001) Confounding revisited with commutative computational algebra, submitted.
- Kobilinsky A. (1990) Complex linear model and cyclic designs, *Linear Algebra and its Applications*, 127, 227–282.
- Kreuzer M. and Robbiano L. (2000) *Computational Commutative Algebra 1*, Springer, Berlin-Heidelberg.
- Pistone G., Riccomagno E. and Wynn H.P. (2001) *Algebraic Statistics: Computational Commutative Algebra in Statistics*, Chapman&Hall.
- Pistone G. and Wynn H.P. (1996) Generalised confounding with Gröbner bases, *Biometrika*, 83, 3, 653–666.
- Raktoe B.L., Hedayat A. and Federer W.T. (1981) *Factorial designs*, John Wiley & Sons Inc., New York.
- Robbiano L. (1998) Gröbner bases and statistics, in: *Gröbner Bases and Applications*, Buchberger B. and Winkler F., eds., Cambridge University Press, 179–204.
- Robbiano L. and Rogantin M.P. (1998) Full factorial designs and distracted fractions, in: *Gröbner Bases and Applications*, Buchberger B. and Winkler F., eds., Cambridge University Press, 473–482.
- Wu C.F.J. and Hamada M. (2000) *Experiments*, John Wiley & Sons Inc., New York, planning, analysis, and parameter design optimization, A Wiley-Interscience Publication.