Statistica Sinica **17**(2007), 1417-1440

ON THE DESCRIPTION AND IDENTIFIABILITY ANALYSIS OF EXPERIMENTS WITH MIXTURES

Hugo Maruri-Aguilar[1], Roberto Notari[2] and Eva Riccomagno[2,3]

[1]*London School of Economics,* [2]*Politecnico di Torino and* [3]*The University of Warwick*

Abstract: In a mixture experiment the collinearity problems, implied by the sum to one functional relationship among the factors, have strong consequences on the identification and analysis of regression models for such designs. Here to address these problems, mixture designs are represented as sets of homogeneous polynomials. Techniques from computational commutative algebra are employed to deduce generalized confounding relationships on power products, and to determine families of identifiable models.

Key words and phrases: Algebraic statistics, cone of a mixture design, experiments with mixtures, fan of a design, regression analysis.

1. Introduction

In a mixture experiment the response variables depend on the proportion of the components or factors, but not on the absolute amount of the mixture. There is a vast literature on experiments with mixtures, including the seminal work by Scheffé (1958, 1963) and the highly cited textbooks by Cornell (2002) and Aitchison (1986). We refer the reader to the bibliographical list therein.

We study mixture designs with tools from computational commutative algebra (CCA). Specifically, we tailor to mixture designs the polynomial algebra approach to identifiability analysis introduced in Pistone and Wynn (1996). In a few words, that approach consists of representing a design with a set of polynomials in k indeterminates, where k is the total number of factors in the design. Relevant statistical information and objects are retrieved by analysis of that polynomial set. From a practical view point, it is particularly useful in the analysis of non-regular designs for describing the set of polynomials which take the same values over the design points, for determining a finite generating set, called *generalised confounding relations*, and for determining classes of saturated hierarchical models identified by the design. A technical advantage of the algebraic statistic framework is the avoidance of the computation of the rank of the design/model matrix, which can be numerically ill-conditioned. Ill-conditioned problems and statistical consequences of multi-collinearity are well known in regression analysis and statistical inference, see e.g., Thisted (1988, Sec. 3.5), Miller (2002) and Björck (1996). Example 17 in the on-line supplement illustrates the instability of a usually adopted procedure for identification and least square estimates when a small perturbation of a coordinate of a design point of a mixture data set occurs.

The CCA approach is computational and the algorithms, provided in e.g., Pistone, Riccomagno and Wynn (2001), apply to mixture experiments. But the main results are in $k-1$ factors. In particular only slack models are obtained and all but one of the basic generalised confounding relations entirely exclude a factor. The polynomial that involves all factors corresponds to the sum to one condition. In Giglio, Riccomagno and Wynn (2001), the missing factor is reintroduced by homogenization. This might not be fully satisfactory, see Example 7. This asymmetry is intrinsic to the computational technology behind the mentioned algorithms, as they depend on a technical algebraic tool called a term ordering, see Appendix 7.1 in the online supplement. In Holliday, Pistone, Riccomagno and Wynn (1999), term orderings have been used to advantage in the statistical analysis of a complex data set. Here we suggest representing a mixture design not as *the set of all polynomials whose zeros include the design points*, but as the subset of all *homogeneous* polynomials whose zeros include the design points. The first set is called the *design ideal* in the algebraic statistics literature, and we call the second one the *cone ideal*. The use of the cone ideal reduces the effect of the aforementioned asymmetry, gives a natural representation of a compositional data set as a set of polynomials, and retains the advantages, both computational and mathematical, of the use of algebraic statistics. The needed algorithms are suitably modified.

Our argument is based on three observations, already present in the literature in different forms. First, a mixture design is a projective object. Each point of the original mixture can be assimilated to a line through the point and the origin, excluding the origin itself. The design cone is the set of all such lines. From an algebro-geometrical perspective this leads naturally to consideration of homogeneous polynomials, and thus to homogeneous type regression models. A reference to mixture models based on homogeneous polynomials is Draper and Pukelsheim (1998), where the mathematical tool employed is the Kronecker product. So homogeneous polynomials are at the base of our second observation. The third one is that no non-trivial polynomial function can be defined over a projective variety, e.g., our cone, and rational polynomial models play a relevant role. Cornell (2002) collects and comments on many models for mixture experiments, including ratios of polynomial models.

We make heavy use of CCA. In Appendices 7.1-7.3 in the online supplement, we collect definitions and results from CCA that we use, while in the main text we report only few essential ones. For an algebraic statistics neophyte, it might be useful to read Appendix 7.1 first. There are many good books on computational commutative algebra, each with its peculiarities. We mainly use the undergraduate texts by Cox, Little and O'Shea (1997, 2004) and Kreuzer and Robbiano (2000, 2005). We would like the reader to be able to perform the computations we present here for his/her own mixture designs. To this aim we specify the name of the commands and macros required in the syntax of CoCoA, which is a freely available system for computing with multivariate polynomials at the webpage http://cocoa.dima.unige.it/. We could have used other excellent and free softwares like *Singular*, see http://www.singular.uni-kl.de/, or *Macaulay*2 at http://www.math.uiuc.edu/Macaulay2/. The proofs of the results we present are collected in Appendix 7.4 in the online supplement, exemplifying the way geometric properties of the experimental plan are used.

In this paper we use the terms "interaction" to mean a monomial of total degree larger than one, and "main effect" for monomials of degree one. For proper use of the terminology, statistical interpretation and analysis of the presence or absence of an interaction in the obtained model when dealing with mixture experiments, we refer to the caveats, comments, and solutions proposed in Claringbold (1955), Cornell (2002), Cox (1971), Piepel, Hicks, Szychowski and Loeppky (2002) and Darroch and Waller (1985).

In Section 2 we study the cone ideal and its link with the design ideal. We choose mixture experiments with n distinct points for simplicity. In Section 3 we discuss a method to retrieve supports for homogeneous regression models identified by a mixture experiment. The algorithm in Section 3.1, which allows us to substitute some terms of the obtained model support retaining identifiability, strongly resembles the algebraic FGLM and Gröbner walk algorithms in Faugère, Gianni, Lazard and Mora (1993). It proved to be very useful in practice. Some typical model structures from the literature are considered in Section 3.2. Practical examples are collected in Section 4, where the theoretical results of the paper are applied to simplex lattice designs, simplex centroid designs, and axial designs. A brief analysis of two data sets follows.

2. The Cone of a Mixture Design

The design space of a mixture design in k factors, $\mathcal{D} \subset \mathbb{R}^k$, is a regular $(k-1)$-dimensional simplex, namely $\{x = (x_1, \ldots, x_k) \in \mathbb{R}^k : \sum_{i=1}^k x_i = 1 \text{ and } 0 \le x_i \le 1\}$. For this reason we can see $\mathcal{D}$ alternatively in the affine space $\mathbb{R}^k$ or in the projective space $\mathbb{P}^{k-1}(\mathbb{R})$, where every point is associated to a line through the origin. We recall that $\mathbb{P}^{k-1}(\mathbb{R})$ is defined as the set of equivalence classes of points in $\mathbb{R}^k$ where p_1 and p_2 are equivalent if p_1, p_2, and $0 = (0, \ldots, 0) \in \mathbb{R}^k$ lie on the same line. Moreover, if $p = (x_1, \ldots, x_k) \in \mathbb{R}^k$, then a representative of the equivalence class of p in $\mathbb{P}^{k-1}(\mathbb{R})$ is $(x_1 : \ldots : x_k)$, called the homogeneous coordinates of p. By definition of the equivalence relationship, they are defined up to a multiple scalar. This leads us to identify, naturally and uniquely, $\mathcal{D}$ with the affine cone $\mathcal{C}_{\mathcal{D}} \subset \mathbb{R}^k$ passing through the origin and $\mathcal{D}$, namely $\mathcal{C}_{\mathcal{D}} = \{\alpha d : d \in \mathcal{D} \text{ and } \alpha \in \mathbb{R}\} \subset \mathbb{R}^k$.

Example 1. The cone of $\mathcal{D}_1 = \{(0,1), (1,0), (1/2,1/2)\} \subset \mathbb{R}^2$ is $\mathcal{C}_{\mathcal{D}_1} = \{(0,a), (b,0), (c,c) : a,b,c, \in \mathbb{R}\} \subset \mathbb{R}^2$, to which we can associate three projective points. For example $(0:1), (1:0), (1:1) \in \mathbb{P}^1(\mathbb{R})$ are representative of the points in $\mathcal{D}_1$ as well. An analogous construction of $\mathcal{C}_{\mathcal{D}_2}$ for $\mathcal{D}_2 = \{(0,0,1), (0,1,0), (1,0,0), (0,1/2,1/2), (1/2,0,1/2), (1/2,1/2,0), (1/3,1/3,1/3)\} \subset \mathbb{R}^3$ shows that in $\mathbb{P}^2(\mathbb{R})$, $\mathcal{D}_2$ can be represented by a $2^3\backslash\{(0,0,0)\}$ structure with levels 0 and 1, a fact we shall exploit in Section 4.

In order to define the design ideal and the cone ideal, let $R = \mathbb{R}[x_1, \ldots, x_k]$ be the set of all polynomials in $x_1, \ldots, x_k$, indeterminates with real coefficients, and let $I \subset R$ be a *(polynomial) ideal.* See Definition 3 in the online supplement for its definition and main properties. We work with particular types of ideals defined in Definitions 1 and 2 below. A set $G = \{g_1, \ldots, g_q\} \subseteq I$ generates I if for all $f \in I$, there exist $s_1, \ldots, s_q \in R$ such that $f = \sum s_i g_i$, and we write $I = \langle g_1, \ldots, g_q \rangle$. There exist special generating sets called Gröbner bases which depend on a term-ordering (see Appendix 7.1 and in particular Definition 6, in the online supplement). The computation of a Gröbner basis from a generating set is considered here as an "elementary" operation. The CoCoA command is `GBasis`.

Definition 1. For $\mathcal{D} \subset \mathbb{R}^k$ with n distinct points, define $\text{Ideal}(\mathcal{D}) = \{f \in \mathbb{R}[x_1, \ldots, x_k]$ such that $f(d) = 0$ for all $d \in \mathcal{D}\}$.

$\text{Ideal}(\mathcal{D})$ is a polynomial ideal studied in Pistone, Riccomagno and Wynn (2001).

Example 2. (cont. Example 1). $\text{Ideal}(\mathcal{D}_1) = \{s_1(x_1 + x_2 - 1) + s_2 x_1(x_1 - 1/2)(x_1 - 1) : s_1, s_2 \in \mathbb{R}[x_1, x_2]\}$, and the polynomials $x_1 + x_2 - 1$ and $x_1(x_1 - 1/2)(x_1 - 1)$ form a generating set of $\text{Ideal}(\mathcal{D}_1)$.

If $\mathcal{D}$ is a mixture experiment, then the polynomial $x_1 + \ldots + x_k - 1$ always vanishes on the design points and thus belongs to $\text{Ideal}(\mathcal{D})$. If the design lies on a face of the $(k-1)$-simplex then there will be a set $A \subseteq \{1, \ldots, k\}$ for which $\sum_{i \in A} x_i - 1 \in \text{Ideal}(\mathcal{D})$. As we show in Section 3, this unduly restricts the class of regression models for $\mathcal{D}$ retrieved with the algebraic statistics methodology; we need a more general theory. The idea is to exploit the representation of a mixture design as a cone. This will have consequences on the structure of the regression models we can associate to $\mathcal{D}$, thus extending the general theory of modelling and confounding. This has been proved to be particularly useful in case studies for the analysis of non-regular fractions of a design in e.g., Holliday, Pistone, Riccomagno and Wynn (1999) and Giglio, Riccomagno and Wynn (2001).

The notion of a polynomial vanishing at a projective point is rather delicate. Indeed, the polynomial $x_2 - x_3^2$ vanishes on $p = (1, 4, 2)$. The points p and $q = (2, 8, 4) = 2p$ are the same point of $\mathbb{P}^2(\mathbb{R})$, but $x_2 - x_3^2$ does not vanish in q. A way to overcome this problem is to use only homogeneous polynomials. A polynomial is *homogeneous* if the total degree (sum of exponents) of each one of its terms (or power products) is the same. For example, $x_1x_2 - x_3^2$ is a homogeneous polynomial of degree 2 which vanishes on $(\lambda, 4\lambda, 2\lambda)$ for all $\lambda \in \mathbb{R}$.

Definition 2. The cone ideal of a mixture design is $\text{Ideal}(\mathcal{C}_\mathcal{D}) = \{f \in \mathbb{R}[x_1 \ldots, x_k]$ such that $f(d) = 0$ for all $d \in \mathcal{C}_\mathcal{D}\}$; that is, the ideal of polynomials vanishing at every point of the cone of the design.

It is easy to show that $\text{Ideal}(\mathcal{C}_\mathcal{D})$ is an ideal. Let $I, J \subset R$ be two ideals generated by the sets G_I and G_J, respectively. Then $I+J = \{f + g : f \in I \text{ and } g \in J\}$ is an ideal and $G_I \cup G_J$ is a generating set of $I+J$. A polynomial ideal is said to be *homogeneous* if, for each $f \in I$, the homogeneous components of f are in I as well, equivalently if I admits a generating set formed by homogeneous polynomials. In some computer algebra packages macros are implemented to compute generating sets of $\text{Ideal}(\mathcal{D})$ and $\text{Ideal}(\mathcal{C}_\mathcal{D})$ directly from the coordinates of the points in $\mathcal{D}$. In CoCoA they are called `IdealOfPoints` and `IdealOfProjectivePoints`, respectively. See Abbott, Bigatti, Kreuzer and Robbiano (2000).

Theorem 1. *For a mixture design* $\mathcal{D}$*,* $\text{Ideal}(\mathcal{C}_\mathcal{D}) = \langle f \in R : f$ *is homogeneous and* $f(d) = 0$ *for all* $d \in \mathcal{D}\rangle$ *and* $\text{Ideal}(\mathcal{D}) = \text{Ideal}(\mathcal{C}_\mathcal{D}) + \langle \sum x_i - 1\rangle$.

Thus, $\text{Ideal}(\mathcal{C}_\mathcal{D})$ is the largest homogeneous ideal in R vanishing at all the points of $\mathcal{D}$. Moreover, a polynomial vanishing on $\mathcal{D}$ can be written as a combination of homogeneous components vanishing on $\mathcal{D}$ and the sum to one condition. If G is a generating set of $\text{Ideal}(\mathcal{C}_\mathcal{D})$ then G and $\sum x_i - 1$ form a generating set of $\text{Ideal}(\mathcal{D})$.

Example 3. (cont. Example 2). $\text{Ideal}(\mathcal{C}_{\mathcal{D}_1}) = \langle x_1^2x_2 - x_1x_2^2\rangle$ and $\text{Ideal}(\mathcal{C}_{\mathcal{D}_2}) = \langle x_1^2x_2 - x_1x_2^2, x_1^2x_3 - x_1x_3^2, x_3^2x_2 - x_3x_2^2\rangle$. For $\mathcal{D}_3 = \{(1, 0, 0), (0, 1, 0), (0, 0, 1), (1/3, 1/3, 1/3)\}$, $\text{Ideal}(\mathcal{C}_{\mathcal{D}_3}) = \langle x_1x_3 - x_2x_3, x_1x_2 - x_2x_3\rangle$.

Theorem 1 states explicitly a method to construct a generating set of $\text{Ideal}(\mathcal{D})$ from a generating set of $\text{Ideal}(\mathcal{C}_\mathcal{D})$ by just adjoining the sum to one condition. Theorem 2 provides the converse. A term order is graded if $x^\beta \succ x^\alpha$ whenever $\sum \alpha_i < \sum \beta_i$.

Theorem 2. *Let* $\mathcal{D}$ *be a mixture design and* $\mathcal{C}_\mathcal{D}$ *its cone. Let* $G = \{\sum_{i=1}^k x_i - 1, g_1, \ldots, g_r\}$ *be a Gröbner basis of* $\text{Ideal}(\mathcal{D})$ *with respect to a graded term or-*

der. Then $\{g_1^{hom}, \ldots, g_r^{hom}\}$ *is a generating set of* $\text{Ideal}(\mathcal{C}_\mathcal{D})$*, where* g^{hom} *is the homogenization of* g *with respect to* $\sum x_i$.

See Kreuzer and Robbiano (2005, Sec. 4.3) for generalities on homogenization. The generating set of the cone ideal obtained in Theorem 2 might not be a Gröbner basis because we do not control the leading term of g_i^{hom} (see Definition 5 in the online supplement for the leading term). The next example shows that if G is not a Gröbner basis, the conclusion of Theorem 2 might not hold. See also Kreuzer and Robbiano (2005, Tutorial 53 and Sec. 4.5)

Example 4. For $\mathcal{D} = \{(0,0,1), (0,1,0), (1,0,0), (1/2,1/2,0), (1/2,0,1/2), (0,1/2,1/2)\}$, $\text{Ideal}(\mathcal{D}) = \langle x_1 + x_2 + x_3 - 1, x_i(x_i - 1/2)(x_i - 1) : i = 1,2,3\rangle$ and the four listed polynomials form a generating set. For $l = x_1 + x_2 + x_3$, the ideal $I = \langle x_i(x_i - 1/2l)(x_i - l) : i = 1,2,3\rangle \subsetneq \text{Ideal}(\mathcal{D})$ does not contain the polynomial $x_2^2x_3 - x_2x_3^2$, which instead belongs to $\text{Ideal}(\mathcal{D})$ and to $\text{Ideal}(\mathcal{C}_\mathcal{D})$. For a simple test to check ideal membership see Cox, Little and O'Shea (1997, p.93), Kreuzer and Robbiano (2000, p.114), or Pistone, Riccomagno and Wynn (2001). See Kreuzer and Robbiano (2005, Corollary 4.4.16) for a homogeneous membership test.

If $\alpha_i \in \mathbb{R}_{>0}$, $i = 0, \ldots, k$, and the hyperplane corresponding to the equation $\sum \alpha_i x_i$ does not contain any point in $\mathcal{C}_\mathcal{D}$, then $\text{Ideal}(\mathcal{C}_\mathcal{D}) + \langle \sum \alpha_i x_i - \alpha_0\rangle$ corresponds to a cutting of the design cone not at the standard simplex. This returns another affine representative of the projective representation of the mixture experiment. In this case there is no immediate interpretation of the points on the hyperplane as a mixture experiment. An obvious interpretation is as a fraction of a bigger experiment with a linear generating constraint.

2.1. Notes on confounding for mixture experiments

In Pistone, Riccomagno and Wynn (2001), the authors use polynomials in $\text{Ideal}(\mathcal{D})$ to deduce (generalised) confounding relations between functions defined over a design $\mathcal{D}$. For example, $x_1+x_2-1 \in \text{Ideal}(\mathcal{D}_1)$ testifies that the polynomial functions x_1 and $1-x_2$ take the same values over $\mathcal{D}_1$, likewise $x_1^2x_2 = x_1x_2^2$ over $\mathcal{D}_1$ because $x_1^2x_2 - x_1x_2^2 \in \text{Ideal}(\mathcal{D}_1)$. Indeed for all $d \in \mathcal{D}_1$, $(x_1^2x_2)(d) = (x_1x_2^2)(d) = 0$. Here with abuse of notation we do not distinguish between the polynomial and its associated polynomial function. In particular, a Gröbner basis of $\text{Ideal}(\mathcal{D}_1)$ with respect to some term ordering gives a finite set of confounding relations which is sufficient to deduce all the others. Usually in classical experimental design theory this information is encoded in the alias table for the design, if it is defined.

As already mentioned, the polynomial $\sum x_i - 1$ belongs to $\text{Ideal}(\mathcal{D})$ for every mixture design $\mathcal{D}$, thus confounding linear terms with the intercept. Thus the classical algebraic approach leads to the study of confounding relationships in a smaller set of factors and the remaining factors are reintroduced in the analysis only when the sum to one condition is considered.

Example 5. For the design $\mathcal{D}$ containing the corner points of the simplex in $\mathbb{R}^k$, for any corner point d and $\alpha \in \mathbb{Z}^k_{\geq 0}$,

$$(x^\alpha)(d) = \begin{cases} 1 & \text{if } \alpha = (0, \ldots, 0) \\ (x_i)(d) & \text{if } \alpha = (0, \ldots, \alpha_i, 0, \ldots, 0) \\ 0 & \text{if at least two components of } \alpha \text{ are not zero.} \end{cases}$$

Ideal($\mathcal{D}$) represents all generalised confounding relations over $\mathcal{D}$. Likewise a polynomial in Ideal($\mathcal{C}_\mathcal{D}$) expresses confounding among homogeneous components. In Section 4 we study some classes of mixture designs and discuss methods to construct classes of fractions by describing the generating polynomials of the cone of the fraction, that is, by confounding some power products. In Section 5.2 we consider some mixture designs which exhibit some geometrical symmetries, and which have interesting statistical properties like equal variance estimates for main factors and for interaction terms where reasonable. They are considered to be particularly useful in the first stage of an experiment when the design region needs to be fairly screened.

3. Supports for Regression Models

In Pistone and Wynn (1996) and Pistone, Riccomagno and Wynn (2001) it is noted that, for any design $\mathcal{D}$, the set of real functions over $\mathcal{D}$ is a $\mathbb{R}$-vector space isomorphic to the coordinate ring $R[\mathcal{D}]$. In turn, $R[\mathcal{D}]$ is isomorphic to the quotient ring $R/\text{Ideal}(\mathcal{D})$. The quotient space is a "computable algebraic object", for example using Gröbner bases. This makes it an important tool to discuss functions over a design, in particular model functions.

For the definition and properties of a coordinate ring over a variety see Cox, Little and O'Shea (1997, Chap. 5), for $R[\mathcal{D}]$ see Pistone, Riccomagno and Wynn (2001, Chap. 2, Sec. 10, Chap. 5), and Cox, Little and O'Shea (2004). See also Appendix 7.1 in the online supplement. Here we only recall that the quotient ring $R/\text{Ideal}(\mathcal{D})$ is the set of equivalence classes for the equivalence relationship $f \sim g$ if $f - g \in \text{Ideal}(\mathcal{D})$. Special monomial $\mathbb{R}$-vector space bases of the quotient ring, called *standard monomials*, can be obtained from particular generating sets of Ideal($\mathcal{D}$), namely Gröbner bases, and thus depend on a term ordering. The main steps of the computation are as follows:

1. determine a Gröbner basis of Ideal($\mathcal{D}$) with respect to a term ordering, for example a Gröbner basis of Ideal($\mathcal{D}_1$) is $\{x_1^3 - 3/2x_1^2 + 1/2x_1, x_1 + x_2 - 1\}$ with respect to any term ordering for which $x_2 \succ x_1$;

2. compute the leading term of each element of the Gröbner basis, for the example, x_1^3 and x_2;
3. determine all monomials which are not divisible by the leading terms, for Example, 1, x_1 and x_1^2 (see Figure 3.1a).

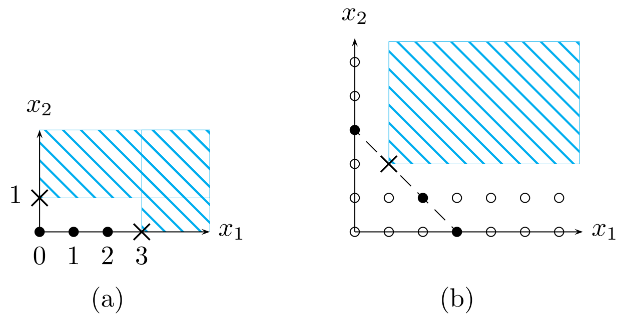

Figure 3.1. Standard monomials for Ideal($\mathcal{D}_1$) and Ideal($\mathcal{C}_{\mathcal{D}_1}$). Both cases were computed with a term order in which $x_2 \succ x_1$.

The CoCoA macro `QuotientBasis` performs the algorithm above. Models returned in Step 3 have a hierarchical structure in that if they include the monomial x^α, then they also must include x^β for all $\beta \leq \alpha$ component-wise. A set of monomials with this property is called an order ideal. Order ideals can be used as support for saturated hierarchical polynomial models. McCullagh and Nelder (1989) and Peixoto and Díaz (1996), among others, strongly argue in favour of hierarchical regression models. Note that any standard monomial set includes the intercept. This might not be good when analysing a mixture experiment. Indeed, for a mixture experiment $\mathcal{D}$, the procedure above returns supports for slack models. See Cornell (2002, p.334) and Cox (1971) for comments on the difficulties in interpreting model parameters. Slack models can be homogenized to return the support for a homogeneous regression model. We proceed differently and propose to adapt the above procedure to the homogeneous component of the design ideal; that is, to work with the cone ideal instead of the ideal. The resulting homogeneous models can be different from those obtained by homogenization of a slack model, as shown in Example 7.

There are two difficulties. First, $R/\text{Ideal}(\mathcal{C}_\mathcal{D})$ is infinite dimensional. Figure 3.1b) shows this for Ideal($\mathcal{C}_{\mathcal{D}_1}$). Second, usually a polynomial does not define a polynomial function on $\mathbb{P}^k(\mathbb{R})$, equivalently on $\mathcal{C}_\mathcal{D}$ (see the comment before Definition 2). One classical CCA remedy to address the first problem considers only monomials of a certain degree $s \in \mathbb{Z}_{\geq 0}$. The basic algebraic definitions and results are in Appendix 7.3 in the online supplement. Below we just apply them. For a mixture design $\mathcal{D}$, the above algorithm is modified as follows:

1. determine a Gröbner basis of Ideal($\mathcal{C}_{\mathcal{D}}$) with respect to a term ordering, for Ideal($\mathcal{C}_{\mathcal{D}_1}$) it is $\{x_1x_2^2 - x_1^2x_2\}$ for any term ordering;
2. compute the leading terms of each element of the Gröbner basis, for the example, $x_1x_2^2$ for term orderings for which $x_2 \succ x_1$;
3. consider all monomials of a sufficiently large total degree, for example in $\mathbb{R}[x_1, x_2]$, there are four monomials of degree $s = 3$, namely $x_1^3, x_1^2x_2, x_1x_2^2, x_2^3$;
4. determine all monomials of degree s not divisible by the leading terms of the Gröbner basis, in the example, $x_1^3, x_1^2x_2, x_2^3$.

Let the symbol A_s represent the set of polynomials in A of degree s and analogously define $A_{\leq s}$. The monomials obtained in Step 4 above form a $\mathbb{R}$-vector space basis of the quotient space $R_s/\text{Ideal}(\mathcal{D})_s$, and form a subset of the set of standard monomials for the cone ideal. We call it the *degree s standard monomial set*. As in the affine case it can be used to construct the support for regression models for $\mathcal{D}$. The correctness of this statement follows directly from Theorem 4 below.

Lemma 3. *Let $\mathcal{D}$ be a mixture design and $s \in \mathbb{Z}_{\geq 0}$ be large enough. The $\mathbb{R}$-vector space $R_{\leq s}/\text{Ideal}(\mathcal{D})_{\leq s}$ has a basis $[g_1], \ldots, [g_n]$, where representatives of the equivalence classes can be chosen to be homogeneous of degree s.*

Theorem 4. *For a mixture design $\mathcal{D}$, $\dim R_s/\text{Ideal}(\mathcal{C}_{\mathcal{D}})_s = \dim R_{\leq s}/\text{Ideal}(\mathcal{D})_{\leq s}$. If, moreover, $\mathcal{D}$ has n distinct points and s is sufficiently large, then the dimensions equal n.*

A monomial basis of degree s can be computed with the Singular macro `kbase`.

Example 6. The Gröbner basis of the homogeneous ideal of $\mathcal{D}_3 = \{(0,0,1), (0,1,0), (1,0,0), (1/3,1/3,1/3)\}$, and for any ordering for which $x_1 \succ x_2 \succ x_3$, is $\{x_1x_3 - x_2x_3, x_1x_2 - x_2x_3, x_2^2x_3 - x_2x_3^2\}$. The leading terms are x_1x_3, x_1x_2, $x_2^2x_3$, respectively. For $s = 3$ the standard monomials are $x_1^3, x_2^3, x_3^3, x_3^2x_2$, giving the largest possible number of terms we can identify with a four point design. For $s = 1$ we obtain the support for a non saturated model: x_1, x_2, x_3. Below we list the degree s standard monomials for all possible values of s.

s	list of monomials of degree s	degree s standard monomials
0	1	1
1	x_1, x_2, x_3	x_1, x_2, x_3
2	x_1^2, x_1x_2, x_2^2, x_1x_3, x_2x_3, x_3^2	x_1^2, x_2^2, x_2x_3, x_3^2
3	x_1^3, $x_1^2x_2$, $x_1x_2^2$, x_2^3, $x_1^2x_3$, $x_1x_2x_3$, $x_2^2x_3$, $x_1x_3^2$, $x_2x_3^2$, x_3^3	x_1^3, x_2^3, $x_2x_3^2$, x_3^3
$s > 3$	x_1^s, $x_1^{s-1}x_2$, $x_1^{s-2}x_2^2$, $\ldots$, x_3^s	x_1^s, x_2^s, $x_2x_3^{s-1}$, x_3^s

Example 7. The slack model obtained for $\mathcal{D}_3$, with respect to any ordering with $x_1 \succ x_2 \succ x_3$, has support $1, x_3, x_3^2, x_2$. By homogenizing it, following Giglio, Riccomagno and Wynn (2001), we obtain $x_1^3, x_3x_1^2, x_3^2x_1, x_2x_1^2$, which is the support of a saturated homogeneous model of total degree 3, but different from the degree 3 model in Example 6. The slack model is the "orthogonal" projection over the subspace of $\mathbb{Z}_{\geq 0}$ defined by $x_k = 0$ of a degree s model support.

Note the following. (i) For $s \geq n$, the procedure returns the support for a saturated model of degree s. Example 6 shows that smaller values of s are possible, but the returned model support may not be saturated. (ii) Equivalently for s large enough, the design/model matrix for $\mathcal{D}$ and degree s standard monomials is invertible, and for any s it is full rank. (iii) These standard monomials are not usually retrieved with the homogenization of a slack model. (iv) Different identifiable models can be obtained by varying the term ordering, as in the affine case. (v) The degree s standard monomial set can be used as a starting set to obtain other types of identifiable sets, as shown in Section 3.1.

3.1. Changing models

Often we want to substitute standard monomials in the set obtained with the methodology of Section 3, or in any other monomial basis of the quotient space, with monomials from a set δ that for some reason we would prefer to consider for the construction of the final regression model. The new set should still be a basis of the quotient space by Ideal($\mathcal{D}$). We present an algorithm to perform such a substitution.

For a mixture design $\mathcal{D}$, let $\mathrm{SM}_{\tau,s}$ or SM_s be the set of standard monomials of degree s with respect to a term ordering τ. It seems reasonable to start with a monomial set of the same size as the design, thus we take s sufficiently large. Set $l = \sum x_i$, and let G be a Gröbner basis of Ideal($\mathcal{C}_\mathcal{D}$) with respect to τ.

Example 8. We consider $\mathcal{D} = \{(1/4, 1/4, 1/2), (1/8, 1/8, 3/4), (1/3, 1/3, 1/3), (1/5, 1/5, 3/5), (0, 0, 1)\}$, with $s = 4$; τ is the default term ordering in CoCoA and $\delta = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}$ is a Scheffé type model, see Scheffé (1963, p.237), Scheffé (1958) and Cornell (2002, p.334). Thus $\mathrm{SM}_4 = \{x_2^4, x_2^3x_3, x_2^2x_3^2, x_2x_3^3, x_3^4\}$.

Step 0. $\eta := \mathrm{SM}_s$ is the current monomial basis of $R/\mathrm{Ideal}(\mathcal{D})$, $W := \emptyset$ is the set of rewriting rules, $\delta' := \delta$.

Step 1. Chose a monomial $w \in \delta'$, let $\deg(w)$ be its total degree, and update $\delta' := \delta' \setminus \{w\}$. Compute the normal form (see Definition 7 in the online supplement) of $wl^{s-\deg(w)}$ with respect to G, obtaining $\mathrm{NF}(wl^{s-\deg(w)}) = \sum_{x^\alpha \in \mathrm{SM}_s} \theta_\alpha x^\alpha = \sum_{x^\alpha \in \eta} \theta'_\alpha x^\alpha$ for $\theta_\alpha, \theta'_\alpha \in \mathbb{R}$. These equalities are valid over

$\mathcal{D}$. The second one follows by substituting the rules in W where necessary (this can be cumbersome in practice).

Step 2. Chose a term x^β in $\sum_{x^\alpha\in\eta}\theta'_\alpha x^\alpha$ for which $\theta'_\beta\neq 0$ and $x^\beta\notin\delta$, equivalently $x^\beta\in \mathrm{SM}_s$. If there is not such β, repeat *Step* 1.

Step 3. Update $\eta := \eta\setminus\{x^\beta\}\cup\{w\}$. In each $g\in W$ substitute x^β with $(w-\sum_{x^\alpha\in\eta\setminus\{x^\beta\}}\theta'_\alpha x^\alpha)/\theta'_\beta$ and get g'. Update $W=\{x^\beta\equiv(w-\sum_{x^\alpha\in\eta\setminus\{x^\beta\}}\theta'_\alpha x^\alpha)/\theta'_\beta,\ g' : g\in W\}$.

Step 4. Repeat from *Step* 1 until $\delta'=\emptyset$.

This is a variation of the algorithm in Babson, Onn and Thomas (2003) where the set δ is the union of all the stairs of a given size and their border sets. Stair is another name for an order ideal. The border of a monomial set is computed by multiplying any monomial in the set by x_i, for $i=1,\dots,k$, and excluding monomials already in the set. The starting monomial set used in Babson, Onn and Thomas (2003), what we call η, is a stair as well. The correctness of our algorithm is proved as in Babson, Onn and Thomas (2003). Its termination is guaranteed by the updating of δ' in *Step* 1 and the finiteness of δ; in Babson, Onn and Thomas (2003), the algorithm terminates when η contains n monomials which are linearly independent and form an order ideal according to the chosen term ordering. In particular, the algorithm in Babson, Onn and Thomas (2003) returns a support for a saturated hierarchical model. Different final monomial sets, of possibly different sizes, might be obtained by choosing different monomials in *Step* 1. In the introduction we mentioned the similarity with the algorithms in Faugère, Gianni, Lazard and Mora (1993), see also Cox, Little and O'Shea (2004, Chap. 8, Sec. 5).

Example 9. (cont. Example 8). *Step* 1. We chose terms in δ in the order they are presented left-to-right in Example 8. Thus $w=x_1$ of degree 1 and $\mathrm{NF}(x_1l^3)=8x_2^4+12x_2^3x_3+6x_2^2x_3^2+x_2x_3^3$. We update $\delta'=\delta'\setminus\{x_1\}$. *Steps* 2 and 3. We select $x^\beta=x_2^4$ and update $\eta=\{x_1,x_2^3x_3,x_2^2x_3^2,x_2x_3^3,x_3^4\}$ and $W=\{\ x_2^4\equiv 1/8x_1-12/8x_2^3x_3-3/4x_2^2x_3^2-1/8x_2x_3^3\}$. *Steps* 1 and 2. Next $w=x_2$, update $\delta'=\delta'\setminus\{x_2\}$ and $\mathrm{NF}(x_2l^3)=8x_2^4+12x_2^3x_3+6x_2^2x_3^2+x_2x_3^3=x_1$. There is no element to select as, over $\mathcal{D}$, $x_1=x_2$ is already included in η. *Steps* 1 to 3. We try the next monomial in δ, $w=x_3$, which can replace $x_2^3x_3$. We update $\eta=\{x_1,x_3,x_2^2x_3^2,x_2x_3^3,x_3^4\}$, $W=W\cup\{x_2^3x_3\equiv 1/8x_3-12/8x_2^2x_3^2-3/4x_2x_3^3-x_3^4\}$, and δ'. *Steps* 1 to 3. We update η, substituting $x_2^2x_3^2$ with x_1x_2 and add the rule $x_2^2x_3^2\equiv x_1x_2-x_2x_3^3-1/4x_3^4-1/2x_1+1/4x_3$ to W. *Steps* 1 to 3. Now we substitute in η the monomial $x_2x_3^3$ with x_1x_3 and add the rule $x_2x_3^3\equiv -1/16x_3^4+4/9x_1x_2+2/9x_1x_2-2/9x_1+4/243x_3$ to W. The current η is $\{x_1,x_3,x_1x_2,x_1x_3,x_3^4\}$. *Steps* 1 and 2. The next candidate in δ is x_2x_3. However, there is no interchange possible, as over $\mathcal{D}$, $x_2x_3=x_1x_3$ and $x_1x_3\in\eta$. At this step $\delta'=\{x_1x_2x_3\}$. *Steps* 1 to

3. The final monomial to be removed from η is x_3^4, substituted with $x_1x_2x_3$. We add the rule $x_3^4 \equiv 6x_1x_2x_3 + 14/3x_1x_2 - 11/3x_1x_3 - 7/3x_1 + 235/162x_3$. *Step* 4. As now $\delta' = \emptyset$, the algorithm ends with the new model/representatives of classes of the quotient space $\eta = \{x_1, x_3, x_1x_2, x_1x_3, x_1x_2x_3\}$, and with the updated set of rules W to express polynomials in terms of monomials in η.

The starting monomial set does not need to be a SM_s set, but could be any other set of polynomials which are linearly independent over $\mathcal{D}$. McConkey, Mezey, Dixon and Grenberg (2000) describe the confounding relationship between the parameters of the Scheffé quadratic model and the model with support x_i and $x_i(1-x_i)$, $i = 1, \ldots, k$, used to describe the average deviation from linearity caused by an individual component on mixing with the other components. Indeed the set δ could be this support, and for $w = x_i(1-x_i)$ the normal form of $x_i \sum_{j \neq i} x_j$ is computed.

Example 10. For $\mathcal{D}_3$ a brother algorithm of the above can be summarised in the following table. It expresses the inverse of the rewriting rules in W, for $\delta = \{x_i, x_i(1-x_i) : i = 1, 2, 3\}$, $\text{SM}_\tau = \{1, x_2, x_3, x_3^2\}$, and any τ for which $x_1 \succ x_2 \succ x_3$.

$B =$

	x_1	x_2	x_3	$x_1(1-x_1)$	$x_2(1-x_2)$	$x_3(1-x_3)$
1	1	0	0	0	0	0
x_2	-1	1	0	0	0	0
x_3	-1	0	1	1	1	1
x_3^2	0	0	0	-1	-1	-1

3.2. Rational models

Sets of linearly independent functions over $\mathcal{D}$ can be defined starting from a $\mathbb{R}$-vector space basis of $R/\text{Ideal}(\mathcal{D})$ and considering ratios of homogeneous polynomials of the same degree.

Example 11. To $\mathcal{D}_1$ and $\{x_1, x_2, x_1x_2\}$ we associate the real-valued rational functions $f_1 = x_1/(x_1+x_2)$, $f_2 = x_2/(x_1+x_2)$ and $f_3 = x_1x_2/(x_1+x_2)^2$ where, for example, the function $x_1/(x_1+x_2) : \mathcal{C}_{\mathcal{D}_1} \longrightarrow \mathbb{R}$ is defined by $(0,1) \longmapsto 0$, $(1,0) \longmapsto 1$ and $(1,1) \longmapsto 1/2$. The design matrix of $\mathcal{D}_1$ and f_1, f_2, f_3 is the same one as that of $\mathcal{D}_1$ and x_1, x_2, x_1x_2. As over $\mathcal{D}_1$ $x_1 + x_2 = 1$, there is no issue in considering a polynomial regression model as usually done. If $x_1 + x_2 = a$ for some $a \in \mathbb{R} \setminus \{0\}$, then a mixture-amount model either in polynomial form (see Cornell (2002, Sec. 7.9)) or rational form can be considered. The natural rational model, which includes terms like x_1/a, can be written as a polynomial model by introducing two extra indeterminates, say $t = 1/a$ and the extra polynomial

$ta-1$. Namely, for $\theta_1, \theta_2, \theta_{11}$ parameters, $\theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1 x_2$ becomes the rational model $\theta_1 x_1/(x_1+x_2) + \theta_2 x_1/(x_1+x_2) + \theta_{11} x_1 x_2/(x_1+x_2)^2$, which in turn translates into the pair of polynomials $at-1$ and $\theta_1 x_1 + \theta_2 x_2 + \theta_{11} x_1 x_2 a$.

Sometimes in the literature x_i is substituted with $x_i/(1-x_i)$ for $i \in A \subseteq \{1,\ldots,k\}$. These functions are defined over $\mathcal{D}$, and not over $\mathcal{C}_{\mathcal{D}}$, and are used as screening models, see Cornell (2002). As the corner points with component 1 at the coordinates in A should not be in the design, the normal forms of the polynomials $1-x_i$, $i \in A$, are not zero. The authors have not been able to prove or disprove the assertion that the linear independence of a set $\{x^\alpha\}$ implies the linear independence of the "normalised" set $\{x^\alpha/\prod_{i=1}^k (1-x_i)^{\alpha_i}\}$ with $\alpha = (\alpha_1,\ldots,\alpha_k)$. An example is analysed in Section 5.2,

Some mixture model forms include inverse terms to model extreme changes in the response behaviour. For example, Cornell (2002, Chap. 6) suggests the model $\sum \theta_i x_i + \sum \theta_{-i} x_i^{-1}$ when no design point has a zero coordinate. Rather than checking that the design/model matrix is full rank we could employ a standard trick in algebra which allows us to transform the above model to a polynomial model in at least two ways. Set $y_i = x_i^{-1}$, to Ideal($\mathcal{D}$) add the polynomials $y_i x_i - 1$, $i = 1\ldots,k$ and work in $\mathbb{R}[y_1,\ldots,y_k,x_1,\ldots,x_k]$ with a term ordering which eliminates the y_i indeterminates. For elimination theory, see Cox, Little and O'Shea (1997, p.72) and Kreuzer and Robbiano (2000, Sec. 3.4). Alternatively, rewrite the suggested model as $y\sum \theta_i x_i + \sum_i \theta_{-i} \prod_{j\neq i, j=1}^k x_j$ and add the polynomial $y\prod_i x_i - 1$.

3.3. Logistic transformations

Mixture designs in $\mathbb{R}^{k+1}$ with no point on the boundary are obtained from a full factorial design in $\mathbb{R}^k$ by applying the additive logistic transformation, or any other transformation that maps $\mathbb{R}^k$ into the interior of the simplex in one higher dimension. Let $\mathcal{F} \subset \mathbb{R}^k$ be a full factorial design with $l_{i1},\ldots,l_{in_i} \in \mathbb{R}$ levels for factor z_i. Then

$$\text{Ideal}(\mathcal{F}) = \left\langle \prod_{j=1}^{n_i} (z_i - l_{ij}), \quad i = 1,\ldots,k \right\rangle \subset \mathbb{R}[z_1,\ldots,z_k] \tag{3.1}$$

with the unique standard monomial set

$$\left\{ z^\alpha : \alpha \in \prod_{i=1}^{k} \{0,1,\ldots,n_i-1\} \right\}. \tag{3.2}$$

The additive logistic transformation $x_i = e^{z_i}/(1+\sum e^{z_j})$, for $i=1,\ldots,k$, and $x_{k+1} = (1+\sum e^{z_j})^{-1}$, with inverse transformation $z_i = \ln x_i/x_{k+1}$, $i=1,\ldots,k$, maps $z = (z_1, \ldots, z_k) \in \mathcal{F}$ into a mixture point. Call $\mathcal{G}$ the collection of such mixture points. Note that substitution of the inverse relationship in (3.2) returns the support for a generalisation of the model (12.6) in Aitchison (1986).

Substitution of the inverse transformation in (3.1), and inclusion of the sum to one condition in the x_i space, gives $\text{Ideal}(\mathcal{G}) = \langle \sum_{i=1}^{k+1} x_i - 1, \prod_{j=1}^{n_i}(x_i - x_{k+1}e^{l_{ij}}),\ i = 1, \ldots, k\rangle \subset \mathbb{R}[x_1, \ldots, x_{k+1}]$.

Direct application of the Buchberger algorithm (see Cox, Little and O'Shea (1997, Chap. 2, Sec. 7) or Kreuzer and Robbiano (2000, Sec.2.5)) shows that the polynomials above form a Gröbner basis for any term ordering for which $x_{k+1} \succ x_i$ for all $i = 1, \ldots, k$. The corresponding standard monomial set is directly linked with the one of the full factorial in (3.2) and it gives the support for a slack model identified by $\mathcal{G}$, call it $V = \{x_1^{\alpha_1} \cdots x_k^{\alpha_k} : \alpha_i \in \{0, 1, \ldots, n_i - 1\}, i = 1, \ldots, k\}$.

As another example of the simplicity and elegance of algebraic statistics, note that the recursive structure of the multiplicative logistic transformation $x_i = e^{z_i}/\prod_{j=1}^{i}(1 + e^{z_j})$ for $i = 1, \ldots, k$, and $x_{k+1} = \prod_{j=1}^{k}(1 + e^{z_j})^{-1}$, with inverse $z_i = \ln(x_i/(1 - x_1 - \ldots - x_i))$, $i = 1, \ldots, k$, sending $\mathcal{F}$ into $\mathcal{H}$ is reflected in the recursive structure of the polynomials in

$$\text{Ideal}(\mathcal{H}) = \langle \sum_{i=1}^{k+1} x_i - 1, \prod_{j=1}^{n_i} \Big(x_i(1 + e^{l_{ij}}) - (1 - x_1 - \ldots - x_{i-1})e^{l_{ij}}\Big) : i = 1, \ldots, k\rangle,$$

which is a Gröbner basis if we chose a term ordering for which $x_1 \prec \ldots \prec x_{k+1}$. In fact, the leading terms of the above polynomials are $x_{k+1}, x_1^{n_1}, \ldots, x_k^{n_k}$, respectively, and so we can apply standard techniques and complete the argument. The corresponding standard basis is again V, while the substitution of the inverse relationship in (3.2) returns the support for a generalisation of the model (12.7) in Aitchison (1986).

4. Some Symmetric Mixture Designs

We start by stating a simple fact valid for mixture designs including corner points. It is the algebraic representation of the well known fact that contrasts of all linear effects with the intercept are identifiable by such an experiment.

Lemma 5. *Let $\mathcal{D} \subset \mathbb{R}^k$ be the mixture design formed by the k corner points of the simplex and τ be a term order. If $x_k \succ x_i$ for all $i \in \{1, \ldots, k\}$, then the (generalised) confounding relationship for a general interaction $x^\alpha = x_1^{\alpha_1} \ldots x_k^{\alpha_k}$, $\alpha \in \mathbb{Z}_{\geq 0}^k$, is*

$$\text{NF}(x^\alpha) = \begin{cases} 1 - \sum_{i=1}^{k-1} x_i & \text{if } x^\alpha = x_k^{\alpha_k} \\ x_i & \text{if } x^\alpha = x_i, i = 1, \ldots, k-1 \\ 0 & \text{if } \alpha \text{ has at least two non-zero components} \\ 1 & \text{if } \alpha = (0, \ldots, 0). \end{cases} \tag{4.1}$$

Theorem 6. *Let $\mathcal{D}$ be a mixture design that contains the corner points, and τ a graded term ordering for which $x_k \succ x_i$ for all i. Then* (i) $1, x_1, \ldots, x_{k-1}$ *are linearly independent over $\mathcal{D}$,* (ii) *the coefficient of the term* 1 *in* $\mathrm{NF}(x_k^{\alpha_k})$ *is* 1, *and* (iii) *the coefficient of the term* 1 *in* $\mathrm{NF}(x^\alpha)$, *with* $x^\alpha \neq x_k^{\alpha_k}$, *is* 0.

4.1. Simplex lattice designs

Scheffé (1958) discusses uniformly spaced distributions of points on the simplex to explore the whole factor space, and calls them *simplex lattice designs.* A $\{k, m\}$ simplex lattice design is the intersection of the simplex in $\mathbb{R}^k$ and the full factorial design in k factors, with the $m + 1$ uniformly spaced levels $\{0, 1/m, \ldots, 1\}$. It has $\binom{m+k-1}{m}$ points. Directly from that description we deduce that for the $\{k, m\}$ simplex lattice design, $\mathcal{D}$, $\mathrm{Ideal}(\mathcal{D}) = \langle \prod_{j=0}^m (x_1 - j/m), \ldots, \prod_{j=0}^m (x_k - j/m), \sum x_i - 1\rangle$, where the first k polynomials are a simple generating set of the full factorial design and the last one is the simplex condition.

The set of slack models identified by $\mathcal{D}$ are well classified and they are in number k as Theorem 7 shows. In Caboara, Pistone, Riccomagno and Wynn (1999) the set of order ideals identified by a design, and obtained via the procedure in Section 3, is called the algebraic fan of the design.

Theorem 7. *The algebraic fan of a $\{k, m\}$ simplex lattice design has size k. Each of its elements is the set of all monomials up to degree m in $k - 1$ factors.*

Corollary 8. *There are no other saturated hierarchical polynomial models identified by the $\{k, m\}$ simplex lattice design, apart from those of Theorem* 7.

By Theorem 1, $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ is the radical of the ideal generated by the homogeneous polynomials $\prod_{j=0}^m (x_i - lj/m)$ for $i = 1, \ldots, k$, and $l = \sum x_i$. Table 1 reports a Gröbner basis for $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ for various combinations of k and m. It uses the following functions: $g(x_1, x_2, w) = \prod_{j=1}^w (x_1 - jx_2/m - j)(x_1 - x_2((m-j)/j))$ and, for $w \in \mathbb{Z}_{>0}$,

$$f(x_1, x_2) = \begin{cases} 1 & \text{if } m = 1 \\ g(x_1, x_2, w) & \text{if } m \text{ odd, } m \neq 1 \text{ and } w = \lfloor \frac{m}{2} \rfloor \\ (x_1 - x_2) g(x_1, x_2, w) & \text{for } m \text{ even and } w = \frac{m}{2} - 1. \end{cases}$$

Table 1. $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ for some simplex lattice designs.

$\mathcal{D}$	$\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$	Number of terms
$\{k, 1\}$	$\mathrm{Ideal}(\mathcal{C}_\mathcal{D}) = \langle x_i x_j : i \neq j \rangle$	$\binom{k}{2}$
$\{k, 2\}$	$\mathrm{Ideal}(\mathcal{C}_\mathcal{D}) = \langle x_i^2 x_j - x_i x_j^2, x_i x_j x_l : i \neq j \neq l \rangle$	$\binom{k}{2} + \binom{k}{3}$
$\{2, m\}$	$\mathrm{Ideal}(\mathcal{C}_\mathcal{D}) = \langle x_1 x_2 f(x_1, x_2) \rangle$	

Fractions of a $\{k, m\}$ design, or of any other design, can be built by confounding identifiable terms. A systematic use of the Hilbert function computes how many terms will be in any corresponding saturated model support and, in the homogeneous case, the maximum number of terms of each degree that can be included. The relevant theory on Hilbert functions is in Appendix 7.3 in the online supplement. In some cases the generating set of the fraction is simple enough to allow the determination of the actual design points by direct investigation.

Example 12. For the $\{4, 4\}$ design, the binomials $x_1x_2 - x_3x_4$, $x_1x_3 - x_2x_4$ and $x_1x_4 - x_2x_3$ added to the generating set of the ideal of either the design or its cone, select the four corner points and the centroid point. They also establish that the terms in each binomial are confounded; that is, take the same values over the selected fraction.

The polynomial $(x_1 - x_2)(x_3 - x_4)$ selects the 15 points for which $x_1 = x_2$ or $x_3 = x_4$, see Example 19 in the on-line supplement. With respect to the default term ordering in CoCoA we obtain the support for a slack model $1, x_4, x_4^2, x_4^3, x_4^4, x_3, x_3^2, x_2, x_2^2, x_2^3, x_2^4, x_3x_4,\ x_3x_4^2, x_2x_4, x_2^2x_4$. For the same fraction and term ordering, the support for a homogeneous model of total degree $s = 0, \ldots, 4$, is

s	SM_s
0	1
1	x_1, x_2, x_3, x_4
2	$x_1^2, x_1x_2, x_2^2, x_2x_3, x_3^2, x_1x_4, x_2x_4, x_3x_4, x_4^2$
3	$x_1^3, x_1^2x_2, x_1x_2^2, x_2^3, x_2x_3^2, x_3^3, x_2^2x_4, x_2x_3x_4, x_3^2x_4, x_1x_4^2, x_2x_4^2, x_3x_4^2, x_4^3$
4	$x_1^4, x_1^3x_2, x_1^2x_2^2, x_1x_2^3, x_2^4, x_2x_3^3, x_3^4, x_3^3x_4, x_2^2x_4^2, x_2x_3x_4^2, x_3^2x_4^2, x_1x_4^3,$ $x_2x_4^3, x_3x_4^3, x_4^4$

In Example 12 we had to take the saturation of the ideal generated by the homogeneous polynomials $\prod_{j=0}^{4}(x_i - lj/4)$, $i = 1, 2, 3, 4$, and $(x_1 - x_2)(x_3 - x_4)$ with respect to x_1, x_2, x_3, x_4. The saturation is an algebraic operation which allows us to take the largest homogeneous ideal defined over a variety, namely the ideal of the variety. It can be performed in CoCoA with the command `Saturation`. We do not study it here any further and refer to Hartshorne (1977), but we add another example and some comments in order to clarify the algebraic motivation.

Example 13. In $\mathbb{P}^3$ with coordinates x, y, z, w, consider the two skew lines $L_1 = \mathrm{V}(x, y)$ and $L_2 = \mathrm{V}(z, w)$, and the curve $C = L_1 \cup L_2$ whose ideal is $\mathrm{Ideal}(C) = \mathrm{Ideal}(L_1) \cap \mathrm{Ideal}(L_2) = \langle xz, xw, yz, yw\rangle$. If we cut C with the plane $H = \mathrm{V}(y + z)$, we obtain the points $A_1 = (0{:}0{:}0{:}1)$ and $A_2 = (1{:}0{:}0{:}0)$ whose ideal is $\mathrm{Ideal}(A_1, A_2) = \langle y, z, xw\rangle$. Of course it is more natural to compute the ideal $J = \text{Ideal}(C) + \text{Ideal}(y+z)$ than the coordinates of the intersection points, and we have $J = \langle y+z, xy, xw, y^2, yw\rangle$.

Clearly $J \neq \text{Ideal}(A_1, A_2)$ and it is easy to verify that $J_s = \text{Ideal}(A_1, A_2)_s$ for $s \geq 2$. So, we can say that the sum of the two ideals I and J is asymptotically equal to the ideal of the intersection of the varieties $\text{V}(I)$ and $\text{V}(J)$. In fact, when we compute combinations of homogeneous polynomials we always get polynomials of degree larger than or equal to the degree of the operands.

The algebraic operation that allows us to compute the ideal of $\text{V}(I) \cap \text{V}(J)$ from $I + J$ is *saturation* with respect to the ideal generated by all the indeterminates, and it consists in looking for homogeneous polynomials f with the property that $fx_i^{m_i} \in I + J$ for some $m_i \in \mathbb{Z}_{>0}$ and for every $i = 1, \ldots, k$.

In the affine space this phenomenon does not show up because when computing combinations of non-homogeneous polynomials, we can obtain polynomials of degree strictly smaller than the degree of the operands.

4.2. Simplex centroid designs

Simplex centroid designs, introduced in Scheffé (1963), are mixture designs in which coordinates are zero or equal to each other. Thus in the k dimensional simple centroid design, there are k points of the form $(1, 0, \ldots, 0)$, $\binom{k}{2}$ of the form $(1/2, 1/2, 0, \ldots, 0)$, $\binom{k}{3}$ of the form $(1/3, 1/3, 1/3, 0, \ldots, 0)$, ..., and the point $(1/k, \ldots, 1/k)$: a total of $\sum \binom{k}{j} = 2^k - 1$ points. This design is the projection with respect to the origin of the full factorial design with levels 0 and 1, on the simplex in $\mathbb{R}^k$. Again, easily, we see that there are $2^k - 1$ points. We rename "2^k design" the full factorial design with levels 0 and 1 in k factors.

If $\mathcal{D}$ is the simplex centroid design in $\mathbb{R}^k$ then $\text{Ideal}(\mathcal{C}_\mathcal{D}) = \langle x_i^2 x_j - x_i x_j^2 : i, j = 1, \ldots, k; i \neq j\rangle$. The geometry of the design is easily deduced by inspection of the factorised generators $x_i x_j (x_i - x_j)$: coordinates of a point in $\mathcal{D}$ are either 0 or equal to each other. The generating set given for $\text{Ideal}(\mathcal{C}_\mathcal{D})$ is a Gröbner basis with respect to any term ordering. The proof is a straightforward application of the S-polynomial test; see e.g., Cox, Little and O'Shea (1997, Chap. 2, Sec. 6, Thm. 6).

The construction of $\text{Ideal}(\mathcal{D})$ can also be based on the derivation of the simplex centroid design from the 2^k design, but it is more complicated and involves techniques from elimination theory. We may want to do this when, for some reasons, we do not want to list the mixture point coordinates. The steps of the construction are as follows.

1. The ideal of the 2^k design is $\langle x_i^2 - x_i : i = 1, \ldots, k\rangle$.
2. The origin can be removed by adjoining the polynomial given by the sum of the elementary symmetric polynomials and 1 with alternate signs (see

Cox, Little and O'Shea (1997, Chap. 7, Sec. 2). The elementary symmetric polynomials in R are $\sigma_1 = (x_1 + \ldots + x_k)$, …, $\sigma_r = (\sum_{i_1<i_2<\cdots<i_r} x_{i_1} \ldots x_{i_r})$, …, $\sigma_k = (x_1 \ldots x_k)$.

3. The simplicial projection is performed in two steps: extend the polynomial ring with the variables $y_1, \ldots, y_k$, and adjoin to the above ideal the polynomials $y_i(\sum x_j) - x_i$.
4. Eliminate the indeterminates $x_i, i = 1, \ldots, k$, from the ideal obtained in the previous step to get Ideal($\mathcal{D}$), which is now expressed in the y_i indeterminates.

Example 14. For $k = 3$ the affine ideal of a 2^3 design is $\langle x_1^2 - x_1, x_2^2 - x_2, x_3^2 - x_3\rangle$. The origin is removed with the ideal operation $\text{Ideal}(2^3 \setminus \{(0,0,0)\}) = \text{Ideal}(2^3) + \langle \sigma_3 - \sigma_2 + \sigma_1 - 1\rangle$, where $\sigma_3 - \sigma_2 + \sigma_1 - 1 = x_1x_2x_3 - x_1x_2 - x_1x_3 - x_2x_3 + x_1 + x_2 + x_3 - 1$. Extend the polynomial ring with y_1, y_2, y_3 and create the following ideal: $\text{Ideal}(2^3 \setminus \{(0,0,0)\}) + \langle y_1 l - x_1, y_2 l - x_2, y_3 l - x_3\rangle \subset \mathbb{R}[x_1, x_2, x_3, y_1, y_2, y_3]$, where $l = x_1 + x_2 + x_3$. Eliminate the variables x_1, x_2, x_3, for instance with the CoCoA macro `Elim`. This last step gives a set of generators for Ideal($\mathcal{D}$) $\{y_1 + y_2 + y_3 - 1,\ y_3(y_3 - 1)(2y_3 - 1)(3y_3 - 1),\ y_2y_3(y_2 - y_3),\ y_3(2y_3 - 1)(2y_2 + y_3 - 1),\ y_2(2y_2 - 1)(y_2 + 2y_3 - 1)\}$.

Scheffé (1963) considers two types of fractions of a simplex centroid. A fraction $\mathcal{D}$ of the type in Appendix B of Scheffé (1963) is built from a fraction of the 2^k design, excluding the origin. Call it $\mathcal{F}$. In this case Ideal($\mathcal{D}$) is computed starting the above algorithm with $\mathcal{F}$, and Ideal($\mathcal{C}_\mathcal{D}$) can be obtained by homogenization as in Theorem 2. The ideal of a fraction of the other type in Scheffé (1963, Sec. 5) is built starting the algorithm from an echelon fraction of the 2^k design excluding the origin. For echelon designs, see (Pistone, Riccomagno and Wynn, 2001, Sec. 3.4). Some of the difficulties lamented in Scheffé (1963, Appendix B), in determining identifiably models for these fractions, are then overcome by the algebraic approach to design, specifically the algorithms in Section 3.

Example 15. For $1 < m \leq k$, let $\mathcal{F}_m$ be the fraction of a simplex centroid design that includes all points with at most m non-zero components, and let $\mathcal{F}_k$ be the full simplex centroid. Clearly, $\mathcal{F}_m$ satisfies the description in Scheffé (1963, Sec. 5). The number of points in $\mathcal{F}_m$ is $\sum_{j=1}^{m} \binom{k}{j}$. The cone ideal for $\mathcal{F}_m$ is $\langle x_i^2 x_j - x_i x_j^2, x_{i_1} \cdots x_{i_{m+1}} : i \neq j \text{ and } i_1 \neq \cdots \neq i_{m+1}\rangle$ if $m > 1$, and for $m = 1$ simplifies to $\langle x_i x_j : i \neq j\rangle$. Differently from Example 12, the given generators are those of a saturated ideal.

Example 16. We compute the algebraic fan of $\mathcal{D} = \mathcal{F}_m$ in Example 15. First note that the given generating set is a Gröbner basis for any term ordering. For $m = 1$ and any term ordering, the leading term of $x_i x_j \in \text{Ideal}(\mathcal{C}_\mathcal{D})$ is the monomial itself. Thus the homogeneous model has support $\{x_1^s, x_2^s, \ldots, x_k^s\}$ for any $s \in \mathbb{Z}_{\geq 1}$. If $m > 1$, the leading term of $x_{i_1} x_{i_2} \cdots x_{i_{m+1}}$ is the monomial itself. For a given initial term ordering on $x_1, \ldots, x_k$, e.g., $x_3 \succ x_2 \succ x_1$, the leading term of $x_i^2 x_j - x_i x_j^2$ is $x_i^2 x_j$ if $x_i \succ x_j$, and $x_i x_j^2$ otherwise. For a given initial term ordering, there are $\sum_{j=1}^{m} \binom{k}{j}$ monomials of total degree s not divisible by $x_i^2 x_j$, with $x_i \succ x_j$ and $x_{i_1} x_{i_2} \cdots x_{i_{m+1}}$, namely for $m = 3$ we have $\{x_i^s, x_i^{s-1} x_j, x_i^{s-2} x_j x_l : i, j, l = 1, \ldots, k, i < j < l\}$.

4.3. Snee-Marquardt designs

In Snee and Marquardt (1976), simplex screening designs which are axial designs are introduced and now they are known as Snee-Marquardt designs. The Snee-Marquardt design in k factors, $\mathcal{M}$, is formed by the points

k vertices	$(1, 0, \ldots, 0), \ldots, (0, \ldots, 0, 1)$
1 centroid	$(\frac{1}{k}, \ldots, \frac{1}{k})$
k interior points	$(\frac{k+1}{2k}, \frac{1}{2k}, \ldots, \frac{1}{2k}), \ldots, (\frac{1}{2k}, \ldots, \frac{1}{2k}, \frac{k+1}{2k})$
k end effects	$(0, \frac{1}{k-1}, \ldots, \frac{1}{k-1}), \ldots, (\frac{1}{k-1}, \ldots, \frac{1}{k-1}, 0)$.

To construct Ideal($\mathcal{M}$) observe that each point in $\mathcal{M}$ lies on a line A_i through the ith vertex and its opposite end effect point, for $i = 1, \ldots, k$. Ideal($\mathcal{M} \cap \mathcal{A}_i$) is generated by $g = \sum x_i - 1$ and $f_i = x_i x_m (x_i - (k+1) x_m)(x_i - x_m)$, where $m \in \{1, \ldots, i-1, i+1, \ldots, k\}$ and $x_j - x_m$, $1 \leq j < m \leq k$, $j \neq i$, $l \neq i$. The ideals of other types of axial designs are obtained by changing the f_i polynomials. First we prove that if h and m are different from i, then $x_i x_m (x_i - (k+1) x_m)(x_i - x_m)$ and $x_i x_h (x_i - (k+1) x_h)(x_i - x_h)$ cut A_i on the same subset. This remark justifies the fact that, in our notation, f_i does not depend on x_m. In fact, $x_i x_m (x_i - (k+1) x_m)(x_i - x_m) - x_i x_h (x_i - (k+1) x_h)(x_i - x_h) = (x_m - x_h) x_i [x_i^2 - (k+2) x_i (x_m + x_h) + (k+1)(x_m^2 + x_m x_h + x_h^2)] \in$ Ideal(A_i). The ideal defining $\mathcal{M}$ is the intersection of the Ideal($\mathcal{M} \cap \mathcal{A}_i$). We compute Ideal($\mathcal{C}_{\mathcal{M}}$) as usual. If $k = 3$, a straightforward computation shows that Ideal($\mathcal{C}_{\mathcal{M}}$) $= \langle (x_1 - x_2)(x_1 - x_3)(x_2 - x_3), x_1 x_2 (x_1 - x_2)(x_1 + x_2 - 5x_3), x_1 x_3 (x_1 - x_3)(x_1 + x_3 - 5x_2), x_2 x_3 (x_2 - x_3)(x_2 + x_3 - 5x_1) \rangle$. Next, we want to compute a finite generating set of Ideal($\mathcal{C}_{\mathcal{M}}$) for $k \geq 4$.

Proposition 1. *For $k \geq 4$, Ideal($\mathcal{C}_{\mathcal{M}}$) is generated by $q_{ijkm} = (x_i - x_j)(x_h - x_m)$, where i, j, h, m are different from each other in $\{1, \ldots, k\}$, and by $f_{rs} = x_r x_s (x_r - x_s)(x_r + x_s - (k+1) x_t)$, where r, s, t are different from each other in $\{1, \ldots, k\}$.*

A corollary of Proposition 1 is that

$$\mathrm{HF}_{R/\mathrm{Ideal}(\mathcal{C}_{\mathcal{M}})}(s) = \begin{cases} 1 & \text{if } s = 0 \\ k & \text{if } s = 1 \\ 2k & \text{if } s = 2 \\ 3k & \text{if } s = 3 \\ 3k+1 & \text{if } s \geq 4. \end{cases}$$

5. Notes on the Analysis of Two Data Sets

5.1. A non-regular mixture design

In Giglio, Riccomagno and Wynn (2001), a non-regular mixture experiment with $k = 8$ and $n = 18$ is analyzed. For the initial term ordering $h \prec g \prec f \prec e \prec d \prec c \prec b \prec a$ on the factors, a hierarchical slack model for the response is obtained. Instead of homogenising that model support, we consider the cone ideal and, for the same initial ordering, we obtain the degree 2 homogeneous support $\mathrm{M}_1 = \{df, ef, f^2, ag, bg, cg, dg, eg, fg, g^2, ah, bh, ch, dh, eh, fh, gh, h^2\}$. Some of the terms in M_1 are replaced by terms of different degree using the algorithm in Subsection 3.1. In particular, we can replace the quadratic terms of f^2, g^2, h^2 by the linear terms f, g, h and obtain a (more) Scheffé (like) model, named M_2. We could as well have replaced some interaction terms with linear terms, for example building models degree by degree using a suitable δ set in the algorithm in Subsection 3.1. We do not pursue this here. Finally, following Cornell (2002), we can construct a support for a third model where $x_i x_j$ in M_1 are replaced by the rational terms $x_i x_j/((1 - x_i)(1 - x_j))$. We refer to this model as M_3. Such a substitution with rational terms is not always possible, but in this specific example it can be shown that the linear independence of the terms in M_3 over the design follows from the linear independence of the terms in M_1, because of the particular structure of the design.

Often for practical purposes, a reduced model which fits the data reasonably well is preferred to the saturated one. Table 2 shows the values of the determination coefficient R^2, the adjusted one, R^2_A, and the residual error $\hat{\sigma}$ for the submodels obtained with backward stepwise regression. We use the `leaps` function in the statistical software `R`; see http://cran.r-project.org.

Table 2. Results of submodel selection.

Initial model	Final terms	R^2	R^2_A	$10^2\hat{\sigma}$
M_1	h^2, bh, df, eh	0.977	0.958	6.1
M_2	f, h, bh, fh	0.983	0.978	4.4
M_3	$\frac{ef}{(1-e)(1-f)}, \frac{g^2}{(1-g)^2}, \frac{bh}{(1-b)(1-h)}, \frac{ch}{(1-c)(1-h)}, \frac{gh}{(1-g)(1-h)}$	0.974	0.964	5.7

5.2. A fraction of the simplex centroid design

A particular fraction of the simplex centroid with k factors is proposed in McConkey, Mezey, Dixon and Grenberg (2000) for screening for significant interactions. It exhibits some sort of symmetries. The fraction is constructed by considering the k corners of the simplex, and those combinations with p

non-zero factors such that any pair of non-zero factors appears in the design just once. The fact that there are many such fractions, obtained by relabeling of the factors, is clearest from the structure of the polynomial representation below. The fraction obtained is of the echelon type described in Scheffé (1963, Sec. 5), and it is labeled $\{k|p\}$ in McConkey, Mezey, Dixon and Grenberg (2000). There it is noted that there are some values of k for which a $\{k|p\}$ fraction cannot be constructed. We focus our attention on the $\{9|3\}$ case analysed in McConkey, Mezey, Dixon and Grenberg (2000). To construct the cone ideal consider the polynomials

$$x_i(x_j - x_k), x_j(x_i - x_k), x_k(x_j - x_i) : (i, j, k) \in A \quad \text{and} \quad x_ix_j(x_i - x_j) : i \neq j, i, j \in \{1, \ldots, 9\},$$

where the second set of polynomials gives the simplex centroid design in 9 factors, and the set $A = \{(1,2,3), (1,4,8), (2,5,9), (3,6,7), (4,5,6), (2,4,7), (3,5,8), (1,6,9), (7,8,9), (1,5,7), (2,6,8), (3,4,9)\}$ corresponds to the non-zero factor triplets in our design points. The centroid point $(1 : \ldots : 1)$ still satisfies that set of equations. The algebraic operation to remove it is the *colon* of ideals (see e.g., Cox, Little and O'Shea (1997, Chap. 4, Sec. 4) and can be achieved by taking the saturation of the ideal generated by the above polynomials and $x_1x_2x_3x_4x_5x_6x_7x_8x_9$, or any degree three monomial with exponents not in A, for example $x_4x_8x_9$, where the saturation is with respect to $\text{Ideal}(x_1, \ldots, x_9)$. The Hilbert function of the cone ideal is

$$\text{HF}_{R/\text{Ideal}(\mathcal{C}_\mathcal{D})}(s) = \begin{cases} 1 & \text{if } s = 0 \\ 9 & \text{if } s = 1 \\ 21 & \text{if } s \geq 2, \end{cases}$$

and thus we can construct a saturated homogeneous model of degree two. For the default term ordering in CoCoA with $x_1 \succ \ldots \succ x_9$, the support for such a model is $\{x_1^2, x_2^2, x_2x_3, x_3^2, x_4^2, x_4x_7, x_4x_8, x_4x_9, x_5^2, x_5x_6, x_5x_7, x_5x_8, x_5x_9, x_6^2, x_6x_7, x_6x_8, x_6x_9, x_7^2, x_8^2, x_8x_9, x_9^2\}$. A feature of a $\{k|p\}$ fraction is that double interactions are completely confounded over the design in sets of size p, e.g., for the $\{9|3\}$ fraction, the polynomials $x_1x_2 - x_1x_3, x_1x_2 - x_2x_3$ and $x_1x_3 - x_2x_3$ belong to $\text{Ideal}(\mathcal{C}_\mathcal{D})$; that is, the column of a design/model involving the polynomials x_1x_2, x_2x_3 and x_1x_3 are equal. For this reason, equation (3) in McConkey, Mezey, Dixon and Grenberg (2000) includes the sum $x_1x_2+x_1x_3+x_2x_3$ as a model term.

In the model support obtained above, the term x_i can be replaced by the terms x_i^2 for all $i = 1, \ldots, 9$, e.g., by application of the algorithm in Section 3.1. The design/model matrix for the obtained model and the fraction $\{9|3\}$ is a block matrix of the form

$$\left[\begin{array}{c|c} I_9 & 0 \\ \hline P & \frac{1}{9}I_{12} \end{array}\right],$$

where I_k is the identity matrix of size k, and P is the 12×9 matrix listing the coordinates of the mixture points.

6. Further Comments

If the points of $\mathcal{D}$ do not lie on a hyperplane, none of them is the origin, and each line through the origin and a design point does not contain any other design point, then the cone ideal is still well defined. The identifiability theory of the homogeneous model supports works exactly as for mixture designs. In particular Ideal($\mathcal{C}_{\mathcal{D}}$) is the largest homogeneous ideal in Ideal($\mathcal{D}$). Although mathematically sensible, this operation does not seem to be particularly interesting if the design points do not lie on a hyperplane, except perhaps for mixture amount experiments in the original scale, where the point coordinates retain an interpretation as proportions of a total amount which can vary from design point to design point.

For an experiment where the relative proportions of the components are significant rather than the total amount, few relevant facts are implied by considering the cone ideal. The design points are recovered as the variety obtained from intersecting the cone ideal with the simplex ideal as shown in Theorem 1. The generalised confounding relationships collected in Ideal($\mathcal{C}_{\mathcal{D}}$) are the same whatever the total amount of the mixture is. Likewise, the homogeneous model supports are the same independently of the total mixture amount.

Both the confounding relationships and the model support are easily computed even for fairly irregular designs, i.e., designs that do not manifest any geometric symmetry. An exact evaluation of the speed of the algorithms as a function of the sample size and number of factors has not been done. An estimation can be obtained from Abbott, Bigatti, Kreuzer and Robbiano (2000). Macros in the computational algebra package CoCoA to compute homogeneous model supports, the ideals, and the cone ideals of the designs in Section 4 are available from the first author.

A general remark on the algebraic statistics approach is that it allows a symbolic approach to identifiability. Thus numerical approximations are postponed to the estimation phase of an analysis. For example, rather than checking numerically if the rank of the design/model matrix for a candidate model is maximal, one computes a basis of the quotient space. This might be advantageous or disadvantageous according to situation. We find that the information embedded in the ideal of a design, or of its cone, are useful in visualising the constraints imposed on the power terms by the design.

Acknowledgement

The authors thank Prof. Henry Wynn and Alessandro di Bucchianico for providing initial bibliographical references on mixture designs. The first author acknowledges financial support from CONACYT.

References

Abbott, J., Bigatti, A., Kreuzer, M. and Robbiano, L. (2000). Computing ideals of points. *J. Symbolic Comput.* **30**, 341-356.

Aitchison, J. (1986). *The Statistical Analysis of Compositional Data*. Chapman & Hall, London.

Babson, E., Onn, S. and Thomas, R. (2003). The Hilbert zonotope and a polynomial time algorithm for universal Gröbner bases. *Adv. Appl. Math.* **30**, 529-544.

Björck, A. (1996). *Numerical Methods for Least Square Problems*. SIAM Press, Philadelphia.

Caboara, M., Pistone, G., Riccomagno, E. and Wynn, H. P. (1999). The fan of an experimental design. Eurandom Preprint n. 99-038.

Claringbold, P. J. (1955). Use of the simplex design in the study of joint action of related hormones. *Biometrics* **11**, 174-185.

Cornell, J. A. (2002). *Experiments With Mixtures*. Third edition. Wiley, New York.

Cox, D., Little, J. and O'Shea, D. (1997). *Ideal, Varieties, and Algorithms*. Second edition. Springer-Verlag, New York.

Cox, D., Little, J. and O'Shea, D. (2004). *Using Algebraic Geometry*. Springer-Verlag, New York. Second Edition.

Cox, D. R. (1971). A note on polynomial response functions for mixtures. *Biometrika* **58**, 155-159.

Darroch, J. N. and Waller, J. (1985). Additivity and interaction in three-component experiments with mixtures. *Biometrika* **72**, 153-163.

Draper, N. R. and Pukelsheim, F. (1998). Mixture models based on homogeneous polynomials. *J. Statist. Plann. Inference* **71**, 303-311.

Faugère, J. C., Gianni, P., Lazard, D. and Mora, T. (1993). Efficient computation of zero-dimensional Gröbner bases by change of ordering. *J. Symbolic Comput.* **16**, 329-344.

Giglio, B., Riccomagno, E. and Wynn, H. P. (2001). Gröbner basis methods in mixture experiments and generalisations. In *Optimum design* 2000 (*Cardiff*), 33-44. Kluwer Acad. Publ., Dordrecht.

Hartshorne, R. (1977). *Algebraic Geometry*. Springer-Verlag, New York.

Holliday, T., Pistone, G., Riccomagno, E. and Wynn, H. P. (1999). The application of algebraic geometry to analysis of designed experiments: A case study. *Computat. Statist.* **14**, 213-231.

Kreuzer, M. and Robbiano, L. (2000). *Computational Commutative Algebra*. 1. Springer-Verlag, Berlin.

Kreuzer, M. and Robbiano, L. (2005). *Computational Commutative Algebra*. 2. Springer-Verlag, Berlin.

McConkey, B., Mezey, P., Dixon, D. and Grenberg, B. (2000). Fractional simplex designs for interaction screening in complex mixtures. *Biometrics* **56**, 824-832.

McCullagh, P. and Nelder, J. A. (1989). *Generalised Linear Model*. Chapman and Hall, London.

Miller, A. (2002). *Subset Selection in Regression*. Chapman & Hall/CRC, Boca Raton.

Peixoto, J. L. and Díaz, J. (1996). On the construction of hierarchical polynomial regression models. *Estadistica* **48**, 175-210.

Piepel, G. F., Hicks, R. D., Szychowski, J. M. and Loeppky, J. L. (2002). Methods for assessing curvature and interaction in mixture experiments. *Technometrics* **44**, 161-172.

Pistone, G., Riccomagno, E. and Wynn, H. P. (2001). *Algebraic Statistics.* Chapman & Hall/CRC, Boca Raton.

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

Prescott, P., Dean, A. M., Draper, N. R. and Lewis, S. M. (2002). Mixture experiments: ill-conditioning and quadratic model specification. *Technometrics* **44**, 260-268.

Scheffé, H. (1958). Experiments with mixtures. *J. Roy. Statist. Soc. Ser. B* **20**, 344-360.

Scheffé, H. (1963). The simplex-centroid design for experiments with mixtures. *J. Roy. Statist. Soc. Ser. B* **25**, 235-263.

Snee, R. and Marquardt, D. (1976). Screening concepts and designs for experiments with mixtures. *Technometrics* **18**, 19-29.

Thisted, R. A. (1988). *Elements of Statistical Computing.* Chapman and Hall, New York.

Department of Statistics, London School of Economics, WC2A 2AE, UK.

E-mail: H.Maruri-Aguilar@lse.ac.uk

Dipartimento di Matematica, Politecnico di Torino, 10141, Italy.

E-mail: Roberto.Notari@polito.it

Department of Statistics, The University of Warwick, CV4 7AL, United Kingdom.
Dipartimento di Matematica, Politecnico di Torino, 10141, Italy.

E-mail: Eva.Riccomagno@polito.it

(Received March 2006; accepted October 2006)