On the description and identifiability analysis of mixture designs

Hugo Maruri-Aguilar
Department of Statistics, The University of Warwick
Coventry CV4 7AL (UK), H.Maruri-Aguilar@warwick.ac.uk

Roberto Notari
Dipartimento di Matematica, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy

Eva Riccomagno
Dipartimento di Matematica, Politecnico di Torino
Corso Duca degli Abruzzi 24, 10129 Torino, Italy
and Department of Statistics, The University of Warwick
Eva.Riccomagno@polito.it

February 13, 2006

Abstract

Mixture designs are represented as sets of homogeneous polynomials thus allowing the use of computational commutative algebra to deduce generalised confounding relationships on monomials terms and to determine families of identifiable models.

Keywords: Mixture designs, cone of a mixture designs.

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 mixture experiment, ranging from the seminal work by H. Scheffé [30, 31] up to the work on optimal designs for second order mixture experiments by Zhang *et al.* [35]. An excellent textbook at the third edition is by J. Cornell [10] and we refer the reader to the bibliographical list therein.

We study mixture designs with tools from computational commutative algebra. Specifically we tailor the polynomial algebra approach to identifiability analysis introduced in [29] to mixture designs. In 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. Large part of the available literature is focused on obtaining suitable supports for models which are identifiable by the design and on determining "generalised confounding relations", see for example [6, 7, 18, 28, 21, 27]. Here we identify sets of polynomials suitable to represent mixture designs.

A characteristic of that approach is that it is computational. As we shall see later the algorithms for example in [28] and [29] apply to mixture experiments but the main results are in $k-1$ factors. In particular only supports for slack models are obtained and all but one of the basic "generalised confounding relations" exclude entirely a factor. The one that involves all factors translates the sum to one condition in the polynomial framework. In [18] the missing factor is reintroduced by homogenization with respect to it. This can be limiting as we show in Example 9. There is an intrinsic and unavoidable asymmetry in the computational technology behind the mentioned algorithms, as they depend on a technical tool called a term ordering which orders the factors, see Appendix 7.1. In [21] this has been used at the advantage of the statistical analysis of a complex data set. Despite this, we feel that for mixture experiments the mentioned technology resents too much of this asymmetry and we propose a modification.

The present paper is based on three observations, already present in the literature in different forms. First, a mixture design is a projective object of which we happen to take proportions summing to one. Thus each design point of the original mixture can be assimilated to a line through the point and the origin, excluding the origin itself. We call the set of all such lines the design cone. From an algebraic geometry perspective this leads naturally to consider homogeneous polynomials and thus homogeneous type regression models. A reference to mixture models based on homogeneous polynomials is [14], 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 cone and rational polynomial models play a relevant role. Cornell [10] collects and comments on many models for mixture experiments including ratios of polynomial models.

We shall make heavy use of computational commutative algebra. The material we use is collected in the appendix and in the main text only when necessary. There are many good books of computational commutative algebra, each with its peculiarities. We mainly use the undergraduate texts [11] and [12]. Good books are also [2] and [22]. We would like to put the reader in condition to perform the computations we present here for his/her own mixture designs. To this aim we show how to perform them in the freely available computer package CoCoA [9]. We could have used other excellent and free softwares for polynomial computation like *Singular* [20] or *Macaulay2* [19].

In this paper we use the terms "interaction" to mean a monomial of 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 models for mixture designs, we refer to the caveats, comments and solutions proposed in [10, 13, 8, 33, 26].

In Section 2 the mixture design is associated to the set of all homogeneous polynomials whose zeros include the design points. For completeness with respect to the current literature in algebraic statistics and design [17, 34] in Section 2.2 an indicator function of a mixture design is defined as a suitable ratio of polynomials.

In Section 3 we introduce a technology to retrieve supports for regression models of the same degree identified by the mixture. 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 [15][12, Ch.8§5]. 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 exemplifying analysis of two data sets is performed.

2 The cone of a mixture design

The design space of a mixture design in k factors with n distinct points, $\mathcal{D} \subset \mathbb{R}^k$, is a regular $(k-1)$-dimensional simplex. 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. This leads us naturally to identify uniquely $\mathcal{D}$ with the affine cone, $\mathcal{C}_{\mathcal{D}} \subset \mathbb{R}^k$, passing through the origin and $\mathcal{D}$:

$$\mathcal{C}_{\mathcal{D}} = \{\alpha d : d \in \mathcal{D} \text{ and } \alpha \in \mathbb{R}\} \subset \mathbb{R}^k$$

Recall that

i) two points $p_1 = (x_1, \dots, x_k)$ and $p_2 = (x'_1, \dots, x'_k)$ in $\mathbb{R}^k$ are equivalent if there exists $\lambda \in \mathbb{R}\backslash\{0\}$ such that $x_i = \lambda x'_i$ for $i = 1, \dots, k$, written as $p_1 \sim p_2$;

ii) the $(k-1)$-*dimensional projective space* over $\mathbb{R}$ is the set of equivalence classes on $\mathbb{R}^k\backslash\{(0, \dots, 0)\}$, written as $\mathbb{P}^{k-1}(\mathbb{R}) = (\mathbb{R}^k\backslash\{(0, \dots, 0)\})/\sim$;

iii) each $p = (x_1, \dots, x_k) \in \mathbb{R}^k\backslash\{(0, \dots, 0)\})$ defines a point in $\mathbb{P}^{k-1}(\mathbb{R})$ and the x_i are called the *homogeneous coordinates* of p.

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 the projective space $\mathcal{D}_2$ can be represented in $\mathbb{P}^2(\mathbb{R})$ by a $2^3\backslash\{(0,0,0)\}$ structure with levels $0, 1$: a fact we shall exploit in Section 4.

In this paper we study mixture designs with tools from computational commutative algebra specialising the theory developed in [28] and [29] for a general design. The first step in that theory is that any design $\mathcal{D} \subset \mathbb{R}^k$ is associated to the set of all polynomials whose zeros include the design points, called the *design ideal* and written as $\text{Ideal}(\mathcal{D})$.

In Appendix 7 we collect definitions and results from computational commutative algebra we use. Here we report only few essential ones. $\mathbb{R}[x_1, \dots, x_k]$ is the set of all polynomials in $x_1, \dots, x_k$ indeterminates and with real coefficients. A subset $I \subset \mathbb{R}[x_1, \dots, x_k]$ is a *(polynomial) ideal* if $f + g \in I$ and $hf \in I$ for all $f, g \in I$ and $h \in \mathbb{R}[x_1, \dots, x_k]$. The Hilbert basis theorem states that every polynomial ideal is finitely generated, where $G = \{g_1, \dots, g_q\} \in \mathbb{R}[x_1, \dots, x_k]$ generates I if for all $f \in I$ there exist $s_1, \dots, s_q \in \mathbb{R}[x_1, \dots, x_k]$ such that $f = \sum_{i=1}^q s_i g_i$. We write $I = \langle g_1, \dots, g_q \rangle$. There exist special generating sets called Gröbner bases which depend on a term-ordering (see Appendix 7.1). The computation of a Gröbner basis from a generating set is considered here an "elementary" operation. The design ideal is an ideal [28, 29].

Example 2

$$\text{Ideal}(\mathcal{D}_1) = \{s_1(x_1 + x_2 - 1) + s_2 x_1(x_1 - 1/2)(x_1 - 1) : s_1, s_2 \in \mathbb{R}[x_1, x_2]\}$$

and $x_1 + x_2 - 1$ and $x_1(x_1 - 1/2)(x_1 - 1)$ form a generator set of Ideal($\mathcal{D}_1$).

If $\mathcal{D}$ is a mixture design, then the polynomial $x_1 + \ldots + x_k - 1$ always vanishes on the design points and thus belongs to Ideal($\mathcal{D}$) [18, 27]. If the design lies on a face of the simplex then there will be a set $A \subseteq \{1, \ldots, k\}$ for which $\sum_{i \in A} x_i - 1 \in$ Ideal($\mathcal{D}$). As we shall show in Section 3, this restricts unduly the class of regression models for $\mathcal{D}$ retrieved with the algebraic statistics methodology and 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 particularly useful for non-regular fractions of a design as shown for example in [28, 29].

The notion of a polynomial vanishing on 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, $x_1 x_2 - x_3^2$ is a homogeneous polynomial of degree 2 which vanishes on $(\lambda, 4\lambda, 2\lambda)$ for all $\lambda \in \mathbb{R}$.

Definition 1 *The cone ideal of a mixture design is*

$$\text{Ideal}(\mathcal{C}_\mathcal{D}) = \{f \in \mathbb{R}[x_1 \ldots, x_k] : f(d) = 0 \textit{ for all } d \in \mathcal{C}_\mathcal{D}\}$$

that is the ideal of polynomials vanishing on every point of the cone of the design.

It is easy to show that Ideal($\mathcal{C}_\mathcal{D}$) is an ideal. Let $I, J \subset \mathbb{R}[x_1 \ldots, x_k]$ 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 generator 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 generator set formed by homogeneous polynomials [11, page 371].

Theorem 1 *For a mixture design* $\mathcal{D}$

1. $\text{Ideal}(\mathcal{C}_\mathcal{D}) = \langle f \in \mathbb{R}[x_1, \ldots, x_k] : f$ *is homogeneous and* $f(d) = 0$ *for all* $d \in \mathcal{D}\rangle$*, that is the largest homogeneous ideal in* $\mathbb{R}[x_1, \ldots, x_k]$ *vanishing on all the points of* $\mathcal{D}$*.*

2. $\text{Ideal}(\mathcal{D}) = \text{Ideal}(\mathcal{C}_\mathcal{D}) + \langle \sum_{i=1}^k x_i - 1\rangle$*, that is a polynomial vanishing on* $\mathcal{D}$ *can be written as combination of homogeneous components vanishing on* $\mathcal{D}$ *and the sum to one condition. If* G *is a generator set of* $\text{Ideal}(\mathcal{C}_\mathcal{D})$ *then* G *and* $\sum_{i=1}^k x_i - 1$ *form a generator set of* $\text{Ideal}(\mathcal{D})$*.*

Proof. *1.* Let $f \in I = \{f \in \mathbb{R}[x_1, \ldots, x_k] : f$ is homogeneous and $f(d) = 0$ for all $d \in \mathcal{D}\}$. As f is homogeneous then $f(\alpha d) = 0$ for all $\alpha \in \mathbb{R}$ and thus $f(d) = 0$ for $d \in \mathcal{C}_\mathcal{D}$. Hence $I \subseteq \text{Ideal}(\mathcal{C}_\mathcal{D})$.

Now we show that $\text{Ideal}(\mathcal{C}_\mathcal{D})$ is homogeneous. If $f \in \mathbb{R}[x_1, \ldots, x_k]$ and $f(d) = 0$ on the cone then as $\mathcal{D} \subset \mathcal{C}_\mathcal{D}$ $f(d) = 0$ on $\mathcal{D}$. For any polynomial $f = f_s + f_{s-1} + \cdots + f_0$ with f_i homogeneous polynomials of degree i. For $\alpha \in \mathbb{R}$ and $d \in \mathbb{R}^k$

$$\begin{aligned} f(\alpha d) &= f_s(\alpha d) + f_{s-1}(\alpha d) + \cdots + f_0(\alpha d) \\ &= \alpha^s f_s(d) + \alpha^{s-1} f_{s-1}(d) + \cdots + \alpha^0 f_0(d) \end{aligned} \quad (1)$$

If we take f vanishing on $\mathcal{C}_\mathcal{D}$ then we have $f(\alpha d) = 0$ for all $\alpha \in \mathbb{R}$. Equation (1) is a polynomial of degree s in α. As it is zero for infinitely many α's then its coefficients are zero that is $f_s(d) = \ldots = f_0(d)$ in particular for all $d \in \mathcal{D}$. As by construction f_i is homogeneous, $f_j(d) = 0$ for all d in the cone. Clearly $\text{Ideal}(\mathcal{C}_\mathcal{D}) \subseteq I$.

2. Clearly $\text{Ideal}(\mathcal{C}_D) \subsetneq \text{Ideal}(\mathcal{D}) \subset \mathbb{R}[x_1, \ldots, x_k]$ and $\langle \sum_{i=1}^k x_i - 1\rangle \subsetneq \text{Ideal}(\mathcal{D}) \subset \mathbb{R}[x_1, \ldots, x_k]$. Also $I, J \subset I + J$.

Let $g \in \text{Ideal}(\mathcal{D})$. Then there exists $s \in \mathbb{Z}_{\geq 0}$ such that $g = \sum_{i=0}^s f_i$ and the f_i's are homogeneous polynomials of total degree i. As $\sum_{i=1}^k x_i - 1 \in \text{Ideal}(\mathcal{D})$ we set $\sum_{i=0}^s f_i (x_1 + \cdots + x_k)^{s-i} = g_{homo}$ over $\mathcal{D}$. Then, for $l = x_1 + \cdots + x_k$

$$\begin{aligned} g - g_{homo} &= g - \textstyle\sum_{i=0}^s f_i (x_1 + \cdots + x_k)^{s-i} = g - \sum_{i=0}^s f_i l^{s-i} \\ &= (1 - l)\big(f_{s-1} + (1 + l) f_{s-2} + \cdots + (1 + l + \cdots + l^{s-1}) f_0\big) \\ &= (1 - l)\bar{f} \end{aligned}$$

and we have $g - g_{homo} = \bar{f}(1 - l)$. But both g and $(1-l)\bar{f}$ are in $\text{Ideal}(\mathcal{D})$, thus $g_{homo} \in \text{Ideal}(\mathcal{D})$. By 1. $g_{homo} \in \text{Ideal}(\mathcal{C}_\mathcal{D})$ and thus $g_{homo} \in \text{Ideal}(\mathcal{C}_\mathcal{D}) + \langle l - 1\rangle$ and the the proof is concluded. ∎

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

Theorem 1 states explicitly a constructive method to obtain a generating set of Ideal($\mathcal{D}$) from a generating set of Ideal($\mathcal{C}_{\mathcal{D}}$) by just adjoining the sum-to-one condition. Moreover we have the following theorem. A term ordering is graded if $x^\alpha < x^\beta$ whenever $\sum_{i=1}^k \alpha_i < \sum_{i=1}^k \beta_i$.

Theorem 2 *Let $\mathcal{D}$ be a mixture design and $\mathcal{C}_{\mathcal{D}}$ its cone. Let $G = \{l - 1, g_1, \ldots, g_r\}$ be a Gröbner basis of* Ideal($\mathcal{D}$) *with respect to a graded term ordering τ. Then $\{h(g_1), \ldots, h(g_r)\}$ is a generating set of* Ideal($\mathcal{C}_{\mathcal{D}}$)*, where $h(g)$ is the homogeneization of g with respect to $l = \sum_{i=1}^k x_i$.*

Proof. Let $f \in \text{Ideal}(\mathcal{D})$ be a homogeneous polynomial of degree s. From the defining property of a Gröbner basis, there exist $q, q_1, \ldots, q_r \in \mathbb{R}[x_1, \ldots, x_k]$ such that $f = q(l-1)+q_1g_1+\ldots+q_rg_r$ with $\deg q \leq s-1$ and $\delta_i = \deg(q_ig_i) \leq s$. Homogeneousing we obtain $h(f) = h(q)h(l-1) + l^{s-\delta_1}h(q_1)h(g_1) + \ldots + l^{s-\delta_r}h(q_r)h(g_r)$ and of course $h(l-1) = l - l = 0$. Thus $h(f) = \sum_{i=1}^r l^{s-\delta_i}h(q_i)h(g_i)$. But f is homogeneous and so $f = h(f)$ and $f = \sum_{i=1}^r l^{s-\delta_i}h(q_i)h(g_i)$. The claim now follows from Theorem 1. ■

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 $h(g_i)$. The next example shows that if G is not a Gröbner basis the thesis of Theorem 2 might not hold.

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 generator 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 Ideal($\mathcal{D}$) and to Ideal($\mathcal{C}_{\mathcal{D}}$). For a simple test to check ideal membership see [11] and [28].

In some computer algebra packages macros are implemented to compute generator sets of Ideal($\mathcal{D}$) and 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 [1]. For $\alpha_i \in \mathbb{R}_{>0}$, $i = 0, \ldots, k$, $\text{Ideal}(\mathcal{C}_{\mathcal{D}})+\langle\sum_{i=1}^k \alpha_i x_i - \alpha_0\rangle$ cuts the design cone not at the standard simplex. It returns another affine representative of the projective representation of the mixture design.

2.1 Notes on confounding for mixture designs

In [29] the authors use polynomials in 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 $x_1^2 x_2 = x_1 x_2^2$ over $\mathcal{D}_1$ as $x_1^2 x_2 - x_1 x_2^2 \in \text{Ideal}(\mathcal{D}_1)$. Indeed for all $d \in \mathcal{D}_1$, $(x_1^2 x_2)(d) = (x_1 x_2^2)(d) = 0$. In particular a Gröbner basis of Ideal($\mathcal{D}_1$) with respect to some term ordering (see Appendix 7.1) gives a 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 exists.

The confounding relationships $\sum_i x_i - 1$ belongs to Ideal($\mathcal{D}$) for every mixture design $\mathcal{D}$ [18, 27], thus confounding linear terms with the intercept. In particular the classical algebraic approach [28, 29] leads to the study of confounding relationships in a smaller set of factors and only when the sum-to-one condition is considered the remaining factors are reintroduced in the analysis.

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

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

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. We use mainly symmetric polynomials [11, page 311], that are invariant under permutations of the indeterminates. They lead to symmetric mixture designs which have interesting statistical properties like equal variance estimates for main factors and for interaction terms where reasonable [23]. They are considered to be particularly useful in the first stage of an experiment when the design region needs to be fairly screened.

2.2 Indicator function

The authors of [16, 17, 34] identify a fraction $\mathcal{F}$ of a larger design $\mathcal{D}$, typically taken to be a full factorial design, with a polynomial indicator function $S_{\mathcal{F}}$. They show how to read properties of the fraction from the coefficients of $S_{\mathcal{F}}$. Thus $S_{\mathcal{F}} - 1 \in \text{Ideal}(\mathcal{F})$, $S_{\mathcal{F}}^2 - S_{\mathcal{F}} \in \text{Ideal}(\mathcal{F})$ and $S_{\mathcal{F}}$ can be computed from the CoCoA function `SeparatorsOfPoints` [1].

Example 6 For $\mathcal{D}_3 = \{(0,0,1), (0,1,0), (1,0,0), (1/3,1/3,1/3)\}$, `SeparatorsOfPoints` returns four polynomials

$$3/2x_3^2 - 1/2x_3,\ \ 3/2x_3^2 + x_2 - 3/2x_3,\ \ 3/2x_3^2 - x_2 - 5/2x_3 + 1,\ \ -9/2x_3^2 + 9/2x_3$$

each vanishing on three of the four points in $\mathcal{D}_3$. The sum of the last two polynomials, $3x_3^2+x_2-2x_3$, is an indicator function of $\mathcal{F} = \{(0,0,1),(0,1,0)\} \subsetneq \mathcal{D}_3$.

Analogously for a mixture design $\mathcal{D}$ the CoCoA macro `SeparatorsOfProjectivePoints` returns a set of homogeneous polynomials each one vanishing on all points of $\mathcal{D}$ except one.

Example 7 Ideal$(\mathcal{C}_\mathcal{F}) = \langle x_1x_3 - x_2x_3, x_1x_2 - x_2x_3, x_2^2x_3 - x_2x_3^2\rangle$ and the short CoCoA script

```
Use T::=Q[x[1..3]];
D:=[[0,0,1],[0,1,0],[1,0,0],[1/3,1/3,1/3]];
S:=SeparatorsOfProjectivePoints(D);
Foreach Element In S Do PrintLn Element; EndForeach;
S[1]+S[2];
```

gives

$$-x_2x_3 + x_3^2,\ x_2^2 - x_2x_3,\ x_1^2 - x_2x_3,\ 9x_2x_3$$

Note again that separators of single points are not functions in the projective space, for example the value of $(9x_2x_3)(d)$ for $d \in \mathbb{P}^2(\mathbb{R})$ depends on the representative of d as projective point: for $(1/3,1/3,1/3)$ $(9x_2x_3)(d) = 1$ and for $(1,1,1)$ $(9x_2x_3)(d) = 9$. Thus we cannot sum separators of single points as in the affine case and obtain an indicator polynomial *function*. We resort to ratio of polynomials, a typical thing to do when working in projective spaces.

Theorem 3 *An indicator function for $\mathcal{F} \subset \mathcal{D}$, where $\mathcal{D}$ is a mixture design, is the rational polynomial function*

$$S_\mathcal{F} = \sum_{p\in\mathcal{F}} \frac{S_p}{(\sum_i x_i)^{s_p}}$$

where S_p is a homogeneous polynomial of total degree s_p such that $S_p(d) = 1$ if $d = p$ and 0 if $d \in \mathcal{D}\setminus\{p\}$.

Proof. In [1] it is shown that S_p exists for all p. For $\alpha \in \mathbb{R}$ and $d \in \mathcal{D}$,

$$\begin{aligned} S_\mathcal{F}(\alpha d) &= \textstyle\sum_{p\in\mathcal{F}} \frac{S_p(\alpha d)}{(\sum_i x_i)^{s_p}(\alpha d)} \\ &= \textstyle\sum_{p\in\mathcal{F}} \frac{\alpha^{s_p} S_p(d)}{\alpha^{s_p}(\sum_i x_i)^{s_p}(d)} \\ &= \textstyle\sum_{p\in\mathcal{F}} \frac{S_p(d)}{(\sum_i x_i)^{s_p}(d)} = S_\mathcal{F}(d) = \begin{cases} 1 & \text{if } d \in \mathcal{F} \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

■

Non significant differences occur if S_p takes any non zero value a other than 1 and the normalization is by $a(\sum_i x_i)^{s_p}$ or if we normalise each S_p by $(\sum_i \alpha_i x_i / \sum_i \alpha_i p_i)^{s_p}$, with $\alpha_i \in \mathbb{R}_{>0}$ and $p = (p_1, \ldots, p_k) \in \mathcal{F}$.

3 Supports for regression models

In [28] and [29] it is noted that for any design $\mathcal{D}$ the set of real functions over $\mathcal{D}$ is a $\mathbb{R}$-vector space and it is isomorphic to the coordinate ring $R[\mathcal{D}]$. In turn, $R[\mathcal{D}]$ is isomorphic to the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/\operatorname{Ideal}(\mathcal{D})$. 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.

For definition and properties of a coordinate ring over a variety see [11, Ch.5], for $R[\mathcal{D}]$ see [28, Ch.2§10,Ch.5] and [12]. See also Appendix 7.1. Here we report only the essential ones. The quotient ring $\mathbb{R}[x_1, \ldots, x_k]/\operatorname{Ideal}(\mathcal{D})$ is the set of equivalence classes for the equivalence relationship $f \sim g$ if $f - g \in \operatorname{Ideal}(\mathcal{D})$. 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 depend on a term ordering (see Appendix 7). 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 $x_2 > x_1$;

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 the example 1, x_1 and x_1^2 (see Figure 1a).

Note that any standard monomial set includes the intercept. Thus for a mixture design $\mathcal{D}$, this procedure returns supports for slack models [10, page 334]. These can be homogenized to return the support for a homogeneous regression model [18]. Models returned in Step 3. above have a hierarchical structure in that if they include the monomial x^α then by Step 3. above they also include x^β for all $\beta \leq \alpha$ component wise. A set of monomials with this property is called an order ideal. Standard monomials can be used as support for hierarchical polynomial models [25], [24]. The CoCoA macros `QuotientBasis` performs the algorithm above.

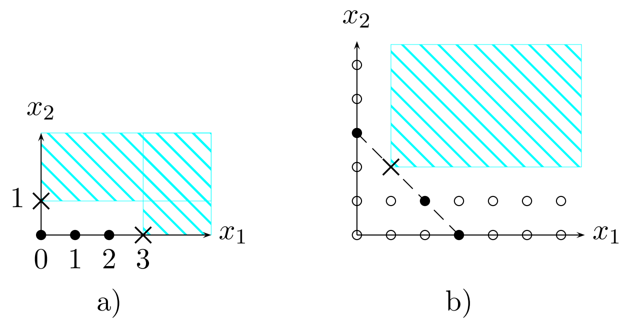

Figure 1: Standard monomials for Ideal($\mathcal{D}_1$) and Ideal($\mathcal{C}_{\mathcal{D}_1}$). Both cases were computed with a term order in which $x_2 > x_1$.

To overcome the lack of symmetry in the factors for mixture designs at the core of the procedure above, we propose to adapt it to the homogeneous component of the design ideal, that is to work directly over the cone ideal. There are two difficulties. First, $\mathbb{R}[x_1, \ldots, x_k]/\operatorname{Ideal}(\mathcal{C}_\mathcal{D})$ has infinite dimension. Figure 1b) shows this for Ideal($\mathcal{C}_{\mathcal{D}_1}$). Indeed $x_1^2x_2 - x_1x_2^2$ is a universal Gröbner bases with leading term $x_1x_2^2$ for any any term ordering for which $x_2 > x_1$. 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 1). One classical computational commutative algebra remedy to address the first problem considers only monomials of a certain degree say $s \in \mathbb{Z}_{\geq 0}$. The basic algebraic definitions and results are reported in Appendices 7.2 and 7.4. Below we just apply them. For a mixture design $\mathcal{D}$

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 $\{x_1x_2^2 - x_1^2x_2\}$;

2. compute the leading terms of each element of the Gröbner basis, for the example $x_1x_2^2$ for term orderings for which $x_2 > x_1$;

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 $x_1^3, x_1^2x_2, x_2^3$.

These monomials form a vector space basis for the quotient space $\mathbb{R}[x_1, \ldots, x_k]/\operatorname{Ideal}(\mathcal{D})$ and it is a subset of the set of standard monomials for the cone ideal. We refer to it with the term *degree s standard monomials*. 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 5. Appendix 8 includes CoCoA macros to perform this procedure.

Lemma 4 *Let $\mathcal{D}$ be a mixture design and $s \in \mathbb{Z}_{\geq 0}$ large enough. The $\mathbb{R}$ vector space $\mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$ has a basis $[g_1],\ldots,[g_n]$ where representatives of the equivalence classes can be chosen to be homogeneous of degree s.*

Proof. Let $[f]$ be an element in $\mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$. We want to prove that there exists $g \in [f]$ such that g is homogeneous of degree s. Let $l = x_1 + \ldots + x_k$ and let $f = f_t + \ldots + f_0$ where f_j is homogeneous of degree j and $t \leq s$. Let

$$g = l^{s-t}\left(f_t + l f_{t-1} + \ldots + l^t f_0\right)$$

Then,

$$\begin{aligned} g - f &= l^{s-t}\left(f_t + l f_{t-1} + \ldots + l^t f_0\right) - \left(f_t + l f_{t-1} + \ldots + l^t f_0\right) \\ &\quad + \left(f_t + l f_{t-1} + \ldots + l^t f_0\right) - \left(f_t + \ldots + f_0\right) \\ &= \left(l^{s-t} - 1\right)\left(f_t + l f_{t-1} + \ldots + l^t f_0\right) \\ &\quad + (l-1) f_{t-1} + (l^2 - 1) f_{t-2} + \ldots + (l^t - 1) f_0 \\ &= (l-1)\left[(l^{s-t-1} + \ldots + 1)\left(f_t + l f_{t-1} + \ldots + l^t f_0\right)\right. \\ &\quad \left. + f_{t-1} + (l+1) f_{t-2} + \ldots + (l^{t-1} + \ldots + 1) f_0\right] \end{aligned}$$

But $l - 1 \in \operatorname{Ideal}(\mathcal{D})$ and so $g \in [f]$. An upper bound for s is the number of points in $\mathcal{D}$. ∎

Theorem 5 *Let $\mathcal{D}$ be a mixture design. Then*

$$\dim \mathbb{R}[x_1,\ldots,x_k]_s/\operatorname{Ideal}(\mathcal{C}_{\mathcal{D}})_s = \dim \mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$$

If moreover $\mathcal{D}$ has n distinct points and s is sufficiently large then the dimensions equal n.

Proof. Let $[f_1],\ldots,[f_p]$ be a basis of the $\mathbb{R}$-vector space $\mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$ and let $g_1,\ldots,g_p$ be the degree s homogeneous polynomials constructed in Lemma 4. We want to prove that $[g_1],\ldots,[g_p]$ is a basis of $\mathbb{R}[x_1,\ldots,x_k]_s/\operatorname{Ideal}(\mathcal{C}_{\mathcal{D}})_s$. They are linearly independent: assume there exist $\lambda_1,\ldots,\lambda_p \in \mathbb{R}$ such that

$$\lambda_1[g_1] + \ldots + \lambda_p[g_p] = 0$$

Then $\lambda_1 g_1 + \ldots + \lambda_p g_p \in \operatorname{Ideal}(\mathcal{C}_{\mathcal{D}}) \subseteq \operatorname{Ideal}(\mathcal{D})$ and so $\lambda_1[g_1] + \ldots + \lambda_p[g_p] = 0$ in $\mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$. Hence, $\lambda_1[f_1] + \ldots + \lambda_p[f_p] = 0$ and so $\lambda_1 = \ldots = \lambda_p = 0$ because $[f_1],\ldots,[f_p]$ is a basis of $\mathbb{R}[x_1,\ldots,x_k]_{\leq s}/\operatorname{Ideal}(\mathcal{D})_{\leq s}$.

Let $g \in \mathbb{R}[x_1, \ldots, x_k]_s$. Thus, there exist $\lambda_1, \ldots, \lambda_p \in \mathbb{R}$ such that

$$[g] = \lambda_1[f_1] + \ldots + \lambda_p[f_p] = \lambda_1[g_1] + \ldots + \lambda_p[g_p]$$

and so $[g_1], \ldots, [g_p]$ are generators of $\mathbb{R}[x_1, \ldots, x_k]_s/\mathrm{Ideal}(\mathcal{C}_{\mathcal{D}})_s$. As a consequence, we get the claim. If s is sufficiently large then $\dim \mathbb{R}[x_1, \ldots, x_k]_{\leq s}/\mathrm{Ideal}(\mathcal{D})_{\leq s} = n$ (see e.g. [29]) and thus $p = n$. ■

A monomial basis of degree s can be computed with the Singular macro `kbase`. The corresponding CoCoA macro is in Appendix 8.

Example 8 The Gröbner basis of the homogeneous ideal for $\mathcal{D}_3 = \{(0,0,1), (0,1,0), (1,0,0), (1/3,1/3,1/3)\}$ and for any ordering for which $x_1 > x_2 > x_3$ 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 $x_1^3, x_2^3, x_3^3, x_3^2x_2$: the largest possible number of terms we can identify with a four point design. For $s = 1$ we obtained the support for a non saturated model x_1, x_2, x_3. Below we list the degree s standard monomials for all 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 9 The slack model obtained for $\mathcal{D}_3$ with respect to any ordering with $x_1 > x_2 > x_3$ has support $1, x_3, x_3^2, x_2$. By homogenising it following [18] 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 8.

Note the following things. *i)* For $s \geq n$ the procedure returns a degree s saturated support model. Example 8 shows that smaller values of s are possible, namely $s = 2, 3$ but the returned model support may not be saturated. *ii)* Equivalently for s large enough, the design/model matrix for $\mathcal{D}$ and the 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 as Example 9 shows (cf. [28]). *iv)* Different identifiable models can be obtained for a degree s standard monomial set as shown in Section 3.1.

3.1 Changing model

Often we want to substitute standard monomials obtained with the methodology of Section 3 or 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. That is, the new set should be a basis of the quotient space by $\text{Ideal}(\mathcal{D})$ expressed with some representatives taken from δ. Next we present an algorithm to perform such substitution.

For a mixture design $\mathcal{D}$ let $\text{SM}_{\tau,s}(\mathcal{C}_\mathcal{D})$ be the set of standard monomials of degree s with respect to a term ordering τ computed e.g. with the procedure of Section 3. We simplify the notation $\text{SM}_{\tau,s}(\mathcal{C}_\mathcal{D})$ to SM_s. It seems reasonable to start with a monomial set of the same size as the design, thus we take s sufficiently large. Set $l = \sum_{i=1}^k x_i$ and let G be a Gröbner basis of $\text{Ideal}(\mathcal{C}_\mathcal{D})$ with respect to τ.

Example 10 Our running example has $\mathcal{D} = \{(1/4, 1/4, 1/2),(1/8, 1/8, 3/4), (1/3, 1/3, 1/3) ,(1/5, 1/5, 3/5),(0, 0, 1)\}$, $s = 4$, τ is the default term ordering in CoCoA and $\delta = \{x_1, x_2, x_3, x_1x_2, x_1x_3, x_2x_3, x_1x_2x_3\}$ a Scheffé type model [31, page 237], [30], [10, page 334]. Thus

$$\text{SM}_s = \{x_2^4, x_2^3x_3, x_2^2x_3^2, x_2x_3^3, x_3^4\}.$$

Step 0. $\eta := \text{SM}_s$ is the current monomial basis of $\mathbb{R}[x_1, \ldots, x_k]/\text{Ideal}(\mathcal{D})$, $W := \emptyset$ set of rewriting rules, $\delta' := \delta$.

Step 1. Chose a monomial $w \in \delta'$ and let $\deg(w)$ be its total degree and update $\delta' := \delta' \setminus \{w\}$. Compute the normal form of $wl^{s-\deg(w)}$ with respect to G

$$\begin{aligned} \text{NF}(wl^{s-\deg(w)}) &= \textstyle\sum_{x^\alpha \in \text{SM}_s} \theta_\alpha x^\alpha \qquad \text{for } \theta_\alpha \in \mathbb{R} \\ &= \textstyle\sum_{x^\alpha \in \eta} \theta'_\alpha x^\alpha \end{aligned}$$

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 $x^\beta \in \text{SM}_s$. If there is not such β then repeat **Step 2.**

Step 3. Update $\eta := \eta \setminus \{x^\beta\} \cup \{w\}$. In each $g \in W$ substitute x^β with $\frac{1}{\theta'_\beta}(w - \sum_{x^\alpha\in\eta\setminus\{x^\beta\}} \theta'_\alpha x^\alpha)$ and get g'. Update $W = \{x^\beta \equiv \frac{1}{\theta'_\beta}(w - \sum_{x^\alpha\in\eta\setminus\{x^\beta\}} \theta'_\alpha x^\alpha), g' : g \in W\}$.

Step 4. Repeat from **Step 2.** until $\delta' = \emptyset$.

This is a variation of the algorithm in [4] where the set δ is the union of all the stairs and their border sets. A monomial set is a stair if it contains the divisors of any of its monomials. That is stair is another name for order ideal. The border of a monomial set is computed by multiplying any monomial in the set by x_i for any i and excluding monomials already in the set. The starting monomial set used in [4], what we call η, is a stair as well. The correctness of the above algorithm is proved as in [4]. Its termination is guaranteed by the updating of δ' in Step 1. and the finiteness of δ. While in [4] 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 [4] returns a support for a saturated hierarchical model. In the introduction we already mentioned the similarity with the algorithms in [15] and [12, Ch.8§5].

Example 11 For Example 10 the basic steps of the algorithm are as follows.

Step 1. We chose terms in δ in the order they are presented left-to-right in Example 10. Thus $w = x_1$ of degree 1 and for $(x_1 + x_2 + x_3)^3 x_1$

$$\mathrm{NF}(x_1 l^3) = 8x_2^4 + 12x_2^3x_3 + 6x_2^2x_3^2 + x_2x_3^3$$

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

$$\mathrm{NF}(x_2 l^3) = 8x_2^4 + 12x_2^3x_3 + 6x_2^2x_3^2 + x_2x_3^3 = x_1$$

There is no element to select as, over $\mathcal{D}$, $x_1 = x_2$ which is already included in η.

Steps 1. to 3. We try the next monomial in δ, $w = x_3$ which turns out it 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 which is 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 monomials which are linearly independent over $\mathcal{D}$. McConkey *et al.* (2000) [23] 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. The set δ could then 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 12 For $\mathcal{D}_3$ a brother algorithm of the above can be summarised in the following table, which expresses the inverse of the rewriting rules in W, for $\delta = \{x_i, x_i(1 - x_i) : i = 1, 2, 3\}$, $\mathrm{SM}_\tau = \{1, x_2, x_3, x_3^2\}$ and any τ for which $x_1 > x_2 > x_3$

$B =$

	1	x_2	x_3	x_3^2
x_1	1	−1	−1	0
x_2	0	1	0	0
x_3	0	0	1	0
$x_1(1-x_1)$	0	0	1	−1
$x_2(1-x_2)$	0	0	1	−1
$x_3(1-x_3)$	0	0	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 $\mathbb{R}[x_1, \ldots, x_k]/\operatorname{Ideal}(\mathcal{D})$ and considering ratios of homogeneous polynomials of the same degree as in Theorem 3.

Example 13 To $\mathcal{D}_1$ and $\{x_1, x_2, x_1x_2\}$ we associate the real valued rational functions $f_1 = \frac{x_1}{x_1+x_2}, f_2 = \frac{x_2}{x_1+x_2}, f_3 = \frac{x_1x_2}{(x_1+x_2)^2}$ where for example

$$\begin{array}{rccc} \frac{x_1}{(x_1+x_2)}: & \mathcal{C}_{\mathcal{D}_1} & \longrightarrow & \mathbb{R} \\ & (0,1) & \longmapsto & 0 \\ & (1,0) & \longmapsto & 1 \\ & (1,1) & \longmapsto & 1/2 \end{array}$$

The design matrix of $\mathcal{D}_1$ and f_1, f_2, f_3 is the same 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 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 [10, §7.9] or rational form can be considered. The natural rational model which includes terms like $\frac{x_1}{a}$ 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 $\theta_1 \frac{x_1}{(x_1+x_2)} + \theta_2 \frac{x_1}{(x_1+x_2)} + \theta_{11} \frac{x_1x_2}{(x_1+x_2)^2}$ which in turns 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 $x_i/(1 - x_i)$. These are defined over $\mathcal{D}$ and not over $\mathcal{C}_{\mathcal{D}}$. Models with support $\{x_i/(1 - x_i), i \in A\}$, with $A \subseteq \{1, \ldots, k\}$, are used as screening models [10]. Those are possible if the corner points with component 1 at the coordinates in A are not in the design. That is, if the normal form of the polynomials $1 - x_i$, $i \in A$, are not in zero. Moreover there is not automatic guarantee that the linear independence of a set $\{x^\alpha\}$ implies the linear independence of the "normalised" $\{x^\alpha / \prod_{i=1}^k (1 - x_i)^{\alpha_i}\}$ with $\alpha = (\alpha_1, \ldots, \alpha_k)$. See Section 5.2 for an example.

We could cut the design cone not with the standard simplex but with another hyperplane $\sum_{i=1}^k \alpha_i x_i = 1$ with all $\alpha_i > 0$. In this case there is no immediate interpretation of the points on the hyperplane as mixture design. The only obvious explanation is as a fraction of a bigger experiment with a linear generator.

Example 14 For $\mathcal{D}_1$ and x_1, x_2, x_1x_2 and the hyperplane $\alpha x_1 + \beta x_2$, $\alpha, \beta > 0$ the design matrix is

	$x_1/(\alpha x_1 + \beta x_2)$	$x_2/(\alpha x_1 + \beta x_2)$	$x_1x_2/(\alpha x_1 + \beta x_2)^2$
$(0, y)$	0	$1/\beta$	0
$(x, 0)$	$1/\alpha$	0	0
(x, x)	$1/(\alpha + \beta)$	$1/(\alpha + \beta)$	$1/(\alpha + \beta)^2$

Some mixture model forms include inverse terms to model extreme changes in the response behaviour (see [10, Ch.6]) for example

$$\sum_{i=1}^{k} \theta_i x_i + \sum_{i=1}^{k} \theta_{-i} x_i^{-1} \tag{2}$$

when no design point has a zero coordinate. A standard trick in algebra allows us to transform the above in a polynomial model in two ways at least. 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 [11, page 72]. Alternatively, rewrite Model (2) as

$$y \sum_{i=1}^{k} \theta_i x_i + \sum_{i=1}^{k} \theta_{-i} \prod_{j \neq i, j=1}^{k} x_j$$

and add the polynomial $y \prod_{i=1}^{k} x_i - 1$. See Section 4.4 for other transformations.

4 Some symmetric mixture designs

We state a simple characteristic of mixture designs including corner points and slack models, which is the algebraic representation of the well known fact that contrasts of all linear effects with the intercept are identifiable by such designs.

Lemma 6 *Let $\mathcal{D} \subset \mathbb{R}^k$ be the mixture design formed by the k corner points of the simplex. Let τ be a term order for which $x_k > x_i$ for all i. 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*

$$\mathrm{NF}(x^\alpha) = \begin{cases} 1 - \sum_{i=1}^{k-1} x_i & \text{if } x^\alpha = x_k^{\alpha_k} \\ x_i & \text{if } x^\alpha = x_i, i = 1, \ldots, k-1 \\ 0 & \text{if } \alpha \text{ has at least two non zero components} \\ 1 & \text{if } \alpha = (0, \ldots, 0). \end{cases} \tag{3}$$

Proof. A Gröbner basis for $I(\mathcal{D})$ for any term ordering τ is given by $\{x_1 + \ldots + x_k - 1, x_1^2 - x_1, \ldots, x_k^2 - x_k, x_1 \cdots x_k\}$. Then for τ such that $x_k > x_i$ for all i, the standard monomial set is $\{1, x_1, \ldots, x_{k-1}\}$. The result now follows from Example 5. ■

Theorem 7 *Let $\mathcal{D}$ be a mixture that contains the corner points. Let τ be a graded term ordering for which $x_k > x_i$ for all i. Then*

1. $1, x_1, \ldots, x_{k-1}$ *are linearly independent monomials over* $\mathcal{D}$,
2. *the coefficient of the term* 1 *in* $\mathrm{NF}(x_k^{\alpha_k})$ *is* 1,
3. *the coefficient of the term* 1 *in* $\mathrm{NF}(x^\alpha)$, *with* $x^\alpha \neq x_k^{\alpha_k}$ *is* 0.

Proof. For 1. observe that as the term ordering is graded than lower order terms are favoured over higher order terms and then included in the support for a slack model. It follows directly from the structure of the design/model matrix involved

	x_1	$\ldots$	x_{k-1}	1	$\cdots$
$(1,0,\ldots,0)$	1	$0\ldots$	0	1	$\cdots$
$\vdots$					0
$(0,\ldots,1,0)$	0	$0\ldots$	1	1	$\cdots$
$(0,\ldots,0,1)$	0	$0\ldots$	0	1	$\cdots$
$\vdots$					

For 2. let $\mathrm{NF}(x_k^{\alpha_k}) = \sum_{x^\alpha} \theta_\alpha x^\alpha$ where for a slack support no x^α involves x_k and evaluate it at the corner point $c_k = (0, \ldots, 0, 1)$. Deduce $\theta_0 = 1$. Similarly 3. is proved. ■

4.1 Simplex lattice designs

In [30] Scheffé 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 and with the $m+1$ uniformly spaced levels $\{0, 1/m, \ldots, 1\}$. The number of points in a $\{k, m\}$ simplex lattice design is $\binom{m+k-1}{m}$. Directly from the definition we deduce that for the $\{k, m\}$ simplex lattice design, $\mathcal{D}$,

$$\mathrm{Ideal}(\mathcal{D}) = \langle \prod_{j=0}^{m} (x_1 - j/m), \ldots, \prod_{j=0}^{m} (x_k - j/m), \sum_{i=1}^{k} x_i - 1 \rangle$$

where the first k polynomials are a simple generating set of the full factorial design and the last one is the simplex condition. The simplex lattice design is the set of points lying on the triangulation of the simplex obtained by drawing k sets of $m-1$ equidistant lines parallel to each edge of the simplex. In this sense its correspondence to the full factorial design is clearest.

The set of $\mathbb{R}$-vector space bases of the quotient space which correspond to hierarchical regression model support are well classified and they are k as Theorem 9 shows via Lemma 8. In [6] 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.

Lemma 8 *Let $\mathcal{D}$ be a $\{k, m\}$ simplex lattice design. Then a basis of the $\mathbb{R}$-vector space $\mathbb{R}[x_1, \ldots, x_k]_{\leq s}/\,\mathrm{Ideal}(\mathcal{D})_{\leq s}$ is*

$$\{1, x_2, \ldots, x_k, x_2^2, x_2x_3, \ldots, x_k^2, \ldots, x_2^{s'}, x_2^{s'}x_3, \ldots, x_k^{s'}\}$$

where $s' = \min\{s, m\}$.

Proof. The claim is equivalent to the following: $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})_s = 0$ for $s \leq m$. Indeed, $x_1 + \ldots + x_k - 1 \in \mathrm{Ideal}(\mathcal{D})$ and we can choose the other generators of $\mathrm{Ideal}(\mathcal{D})$ as homogenous polynomials in $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ by Theorem 1. Thus, let $f \in \mathrm{Ideal}(\mathcal{C}_\mathcal{D})_m$. We want to prove that $f = 0$ and we use induction on k and m. The base of the induction is as follows. First, we analyse the case $\{2, m\}$, for which $\mathcal{D} = \{P_0, \ldots, P_m\}$ with $P_i = (i/m, (m-i)/m)$ for $i = 0, \ldots, m$. But no homogeneous polynomial of degree m can have $m+1$ distinct zeros, unless it is the null polynomial. Second, we consider hte case $\{k, 1\}$. But this design was studied in Lemma 6.

Now, we consider the general case $\{k, m\}$ and we assume that no polynomial of degree $m-1$ belongs to a $\{k, m-1\}$ design and that no polynomial of degree m belongs to a $\{k-1, m\}$ design. Let $f \in \mathrm{Ideal}(\mathcal{C}_\mathcal{D})_m$. We need to show that $f = 0$. If we set $x_k = 0$ then we obtain a $\{k-1, m\}$ design $\mathcal{D}'$ and $f(x_1, \ldots, x_{k-1}, 0) \in \mathrm{Ideal}(\mathcal{C}'_\mathcal{D})_m$. By inductive hypothesis, $f(x_1, \ldots, x_{k-1}, 0)$ is the zero polynomial. Hence, $f = x_k f'$ for some f' suitable homogeneous polynomial of degree $m-1$. The affine transformation

$$\begin{aligned} X_i &= m/(m-1)x_i & i = 1, \ldots, k-1 \\ X_k &= -1/(m-1) + m/(m_1)x_k & \end{aligned}$$

takes $\mathcal{D} \setminus \mathcal{D}'$ into a $\{k, m-1\}$ simplex lattice design. Call it $\mathcal{D}''$ and f' into $\left(\frac{m}{m-1}\right)^{m-1} f'(X_1, \ldots, X_{k-1}) \in \mathrm{Ideal}(\mathcal{D}'')$. By inductive hypothesis, we have $f' = 0$ and so $f - 0$. As a consequence the Hilbert function of $\mathcal{D}$ is

$$\dim \mathbb{R}[x_1, \ldots, x_k]_{\leq s}/\,\mathrm{Ideal}(\mathcal{D})_{\leq 0} = 1 + \binom{k}{k-1} + \ldots + \binom{s'+k-1}{k-1} = \binom{s'+k}{k}$$

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

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

where $s' - \min\{s, m\}$ and the claim follows becasue $\binom{m+k}{k}$ is the number of points in $\mathcal{D}$. ■

Theorem 9 The algebraic fan of a $\{k, m\}$ simplex lattice design has size k. Each one of its elements is the set of all monomials up to degree m in $k-1$ factors.

Proof. In order to respect the order ideal property, not all factors can be included in the presence of the intercept. Moreover no higher degree power in any factor can be included as shown in Lemma 8. ■

Corollary 10 For no other polynomial, saturated and hierarchical model structure and for the $\{k, m\}$ simplex lattice design, $\mathcal{D}$, the design/model matrix is invertible.

Proof. Any other candidate model support would include the terms $1, x_1, , \ldots, x_k$, but they all cannot be identified as $x_1 + \ldots + x_k = 1$ over $\mathcal{D}$. ■

By Theorem 1 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 $l = \sum_{i=1}^{k}$. Table 4.1 reports a Gröbner basis for Ideal($\mathcal{C}_\mathcal{D}$) for various combinations of k and m. It uses the following functions

$$g(x_1, x_2, w) = \prod_{j=1}^{w} (x_1 - \frac{j x_2}{m-j})(x_1 - x_2 \frac{m-j}{j}) \text{ for } w \in \mathbb{Z}_{>0}$$

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

Fractions of a $\{k, m\}$ design, or of any other design, can be built by confusing identifiable terms [28]. A systematic use of the Hilbert function computes how many terms will be in any corresponding saturated model support and, in the homogeneous case, how many terms of each degree can be at most included. The relevant theory on Hilbert functions is in Appendix 7.4. Methods for identifying properties of the design and other information relevant to statistical analysis directly from the ideal of a design, without knowing the actual values of the design point, are under study (G. Pistone, personal communication). In some cases the generator set of the fraction is easy enough to allow the determination of the actual design points by direct investigation.

Example 15 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 generator set of the ideal of either the design or its cone, select the four corner points and the centroid point. 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 20. 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 15 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 e.g. CoCoA with the command `Saturation`. We do not study it here any further.

4.2 Simplex centroid design

Simplex centroid design introduced in [31] 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 $(\frac{1}{2}, \frac{1}{2}, 0, \ldots, 0)$, $\binom{k}{3}$ of the form $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}, 0, \ldots, 0)$, ..., and the point $(\frac{1}{k}, \ldots, \frac{1}{k})$: a total of $\sum_{j=1}^{k} \binom{k}{j} = 2^k - 1$ points. The simplex centroid lattice design in k-variables is the projection of the full factorial design with levels 0 and 1, on the simplex in $\mathbb{R}^k$ with respect to the origin. 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.

The ideal of the cone of $\mathcal{D}$ is easier to built than the affine ideal and it is

$$\text{Ideal}(\mathcal{C}_{\mathcal{D}}) = \langle x_i^2 x_j - x_i x_j^2 : i, j = 1, \ldots, k; i \neq j \rangle$$

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 generator set given for $\text{Ideal}(\mathcal{C}_{\mathcal{D}})$ is a Gröbner basis with respect to any term ordering (a universal Gröbner basis). The proof is rather technical and a straightforward application of the S-polynomial test [11, Ch.2§6Th.6].

Also the construction of $\text{Ideal}(\mathcal{D})$ can 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 [11, Ch.3]. We may want to do this when for some reasons we may not want to list the mixture point coordinates. The steps of the constructions 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 [11, Ch.7§2]. The elementary symmetric polynomials in $\mathbb{R}[x_1, \ldots, x_k]$ are

$$\begin{aligned}
\sigma_1 &= x_1 + \ldots + x_k \\
&\vdots \\
\sigma_r &= \textstyle\sum_{i_1 < i_2 < \cdots < i_r} x_{i_1} \ldots x_{i_r} \\
&\vdots \\
\sigma_k &= x_1 \ldots x_k
\end{aligned}$$

3. The simplicial projection is performed in two steps [5]. Extend the polynomial ring with the variables $y_1, \ldots, y_k$ and adjoin to the ideal above the polynomials $y_i(\sum_{j=1}^{k} x_j) - x_i$.

4. Eliminate the indeterminates $x_i, i = 1, \ldots, k$ from the ideal obtained in 3. above [11, Ch.3] to get Ideal$(\mathcal{D})$ which is now expressed in the y_i indeterminates.

Example 16 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 in 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$(\mathcal{D})$

$$\{y_1 + y_2 + y_3 - 1,\ y_3(y_3 - 1)(2y_3 - 1)(3y_3 - 1),\ y_2y_3(y_2 - y_3),\\ y_3(2y_3 - 1)(2y_2 + y_3 - 1),\ y_2(2y_2 - 1)(y_2 + 2y_3 - 1)\}$$

In [31] Scheffé considers two types of fractions of a simplex centroid. A fraction $\mathcal{D}$ of the type in [31, §Appendix B] is built from a fraction of the 2^k design, $\mathcal{F}$ not including the origin. In this case Ideal$(\mathcal{D})$ is computed starting the above algorithm with $\mathcal{F}$ and by homogenization as in Theorem 2 Ideal$(\mathcal{C}_\mathcal{D})$ can be obtained. The ideal of a fraction of the other type [31, §5] is built starting the algorithm from an echelon fraction of the 2^k design excluding the origin. For echelon designs see [28, §3.4]. Some of the difficulties met by Scheffé [31, §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 17 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, where $\mathcal{F}_k$ is the full simplex centroid. Clearly, $\mathcal{F}_m$ satisfies the description in [31, §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

$$\langle x_i^2 x_j - x_i x_j^2, x_{i_1} \cdots x_{i_{m+1}} : i \neq j \text{ and } i_1 \neq \cdots \neq i_{m+1} \rangle \text{ if } m > 1$$

which for $m = 1$ simplifies to $\langle x_i x_j : i \neq j \rangle$. Differently from Example 15 the given generators are those of a saturated ideal.

Example 18 We compute the algebraic fan of $\mathcal{D} = \mathcal{F}_m$ of Example 17 as an example of the application of the techniques in Subsection 4.2. First note that the given generator set is a universal Gröbner basis. 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. $x_1 < x_2 < x_3$, the leading term of $x_i^2 x_j - x_i x_j^2$ is $x_i^2 x_j$ if $x_i > x_j$ 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 $x_i > x_j$ and $x_{i_1} x_{i_2} \cdots x_{i_{m+1}}$, namely for $m = 3$

$$\{x_i^s, x_i^{s-1} x_j, x_i^{s-2} x_j x_l : i, j, l = 1, \ldots, k, i < j < l\}$$

4.3 Snee-Marquardt designs

In [32] simplex screening designs which are axial designs are presented and now they are known as Snee-Marquardt designs. The Snee-Marquardt design in k factors, $\mathcal{M}$, has 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	$(\frac{k+1}{2k}, \frac{1}{k}, \ldots, \frac{1}{k}), \ldots, (\frac{1}{k}, \ldots, \frac{1}{k}, \frac{k+1}{2k})$
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 $\text{Ideal}(\mathcal{M})$ observe that each point in $\mathcal{M}$ lies on one axis between a vertex ith and its opposite end effect point. Call $\mathcal{A}_i$ these axes, $i = 1, \ldots, k$. Note

$$\text{Ideal}(\mathcal{M} \cap \mathcal{A}_i) = \langle g, f_i, x_j - x_l : \text{ for } j \neq i, l, j = 1, \ldots, k \rangle$$

where $g = \sum_{i=1}^{k} x_i - 1$ is the simplex condition, $f_i = x_i(x_i - \frac{k+1}{2k})(x_i - 1)$ selects points with ith coordinate in $\left\{0, \frac{k+1}{2k}, 1\right\}$, and the other polynomials describe the ith axis. Note that the centroid point, c, is not included and $\text{Ideal}(\{c\}) = \langle x_i - \frac{1}{k} : i = 1, \ldots, k \rangle$.

$\text{Ideal}(\mathcal{M})$ is obtained as the product of the $\text{Ideal}(\mathcal{M} \cap \mathcal{A}_i)$ and of $\text{Ideal}(\{c\})$ [11, page 210], as $\mathcal{M}$ is the union of the points on the $\mathcal{A}_i$ axis and the centroid point. A generator set is obtained as the set of all products of the generators of all factor ideals [11, page 183]. $\text{Ideal}(\mathcal{C}_\mathcal{M})$ is obtained by homogenising as in previous examples.

The ideals of other types of axial designs are obtained by changing the f_i polynomials.

4.4 Logistic transformations

Mixture designs in $\mathbb{R}^{k+1}$ with no points on the boundary are obtained from a full factorial designs in $\mathbb{R}^k$ 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 i. Then

$$\text{Ideal}(\mathcal{F}) = \langle \prod_{j=1}^{n_i} (z_i - l_{ij}), \qquad i = 1, \ldots, k \rangle \subset \mathbb{R}[z_1, \ldots, z_k] \tag{4}$$

with the unique standard monomial set

$$\left\{ z^\alpha : \alpha \in \prod_{i=1}^{k} \{0, 1, \ldots, n_i - 1\} \right\} \tag{5}$$

The additive logistic transformation

$$\begin{aligned} x_i &= \frac{e^{z_i}}{1 + \sum_{j=1}^k e^{z_j}} \quad \text{for } i = 1, \ldots, k \\ x_{k+1} &= \frac{1}{1 + \sum_{j=1}^k e^{z_j}} \end{aligned}$$

with inverse transformation

$$z_i = \ln \frac{x_i}{x_{k+1}} \quad i = 1, \ldots, k \tag{6}$$

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 (5) returns the support for a generalisation of the model (12.6) in [3].

Substitution of (6) in (4) and inclusion of the sum to one condition in the x_i space gives

$$\text{Ideal}(\mathcal{G}) = \langle \sum_{i=1}^{k+1} x_i - 1, \prod_{j=1}^{n_i} (x_i - x_{k+1} e^{l_{ij}}), \qquad i = 1, \ldots, k \rangle \subset \mathbb{R}[x_1, \ldots, x_{k+1}]$$

Direct application of the Buchberger algorithm [11, Ch.2§7] shows that the polynomials above form a Gröbner basis for any term ordering for which $x_{k+1} > x_i$ for all $i = 1, \ldots, k$. The corresponding standard monomial set is directly linked with that of the full factorial in (5) and it gives the support for a slack model identified by $\mathcal{G}$

$$\left\{x_1^{\alpha_1}\cdots x_k^{\alpha_k} : \alpha_i \in \{0,1,\ldots,n_i-1\}, i=1,\ldots,k\right\} \tag{7}$$

As another example of the simplicity and elegance of the algebraic statistics note that the recursive structure of the multiplicative logistic transformation

$$\begin{aligned} x_i &= \frac{e^{z_i}}{\prod_{j=1}^{i}(1+e^{z_j})} \quad \text{for } i=1,\ldots,k \\ x_{k+1} &= \frac{1}{\prod_{j=1}^{k}(1+e^{z_j})} \end{aligned}$$

with inverse

$$z_i = \ln\frac{x_i}{1-x_1-\ldots-x_i} \quad i=1,\ldots,k$$

sending $\mathcal{F}$ into $\mathcal{H}$ is reflected in the recursive structure of the polynomials in

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

There exists at least a term ordering for which the leading terms of the polynomials above are $x_i^{n_i}$ and for the sum to one condition it is x_{k+1}. The corresponding standard basis is again (7) while the substitution of the inverse relationship in (5) returns the support for a generalisation of the model (12.7) in [3].

5 Notes on the analysis of two data sets

5.1 A non regular mixture design

In [18] 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. For the same initial ordering the support for a homogeneous saturated model of degree 2 is

$$\{df, ef, f^2, ag, bg, cg, dg, eg, fg, g^2, ah, bh, ch, dh, eh, fh, gh, h^2\}$$

Call it M_1. Some of the terms in M_1 are replaced by terms of different degree using the algorithm in Subsection 3.1. In particular we may want to check

Initial model	Final terms	R^2	R_A^2	$\hat{\sigma} \times 10^2$
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

Table 2: Results of model selection

if we can replace the quadratic terms of f^2, g^2, h^2 by the linear terms f, g, h. Indeed that is the case and we have a (more) Scheffé (like) model, named M_2. We could as well have replaced some interactions terms with linear terms, for example building models degree by degree using a suitable δ set in the algorithm in Subsection 3.1. But we do not pursue this here. Finally, following [10] 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.

For practical purposes, often a reduced model which fits reasonably well to the data, is preferred to the saturated one. Table 5.1 shows the values of the determination coefficient R^2, the adjusted one R_A^2 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.

5.2 A fraction of the simplex centroid design

A particular fraction of the simplex centroid with k factors is proposed in [23] 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 relabelling of the factors is clearest from the structure of the polynomial representation below. The fraction obtained is of the echelon type described in [31, §5], and it is labeled $\{k|p\}$ in [23]. In [23] 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 $\{9|3\}$ analysed in [23].

To construct the cone ideal consider the polynomials

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

where the second set of polynomials give 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 triplets in our design. The centroid point $(1,\ldots,1)$ still satisfies that set of equations. The algebraic operation to remove it is the *colon* of ideals [11, Ch.4§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 other degree three monomial with exponents not in A, for example $x_4x_8x_9$, where the saturation is with respect to the usual ideal $\mathrm{Ideal}(x_1,\ldots,x_9)$. The Hilbert function (Appendix 7.4) of the cone ideal is

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

and thus we can construct a saturated homogeneous model of degree two. For the default term ordering in CoCoA with $x_1 > \ldots > x_9$ the support for such a model is

$$\{x_1^2, x_2^2, x_2x_3, x_3^2, x_4^2, x_4x_7, x_4x_8, x_4x_9, x_5^2, x_5x_6, x_5x_7, x_5x_8, x_5x_9, x_6^2, x_6x_7, x_6x_8, x_6x_9, x_7^2, x_8^2, x_8x_9, x_9^2\} \tag{8}$$

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 $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$, 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 the analysis in [23, Eqn.(3)] includes the sum $x_1x_2 + x_1x_3 + x_2x_3$.

The terms x_i can replace the terms x_i^2 in Equation (8), 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 diagonal 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 Acknowledgments

The authors thank Prof. H. P. Wynn and A. di Bucchianico for providing initial bibliographical references on mixture designs. The first author acknowledge financial support from CONACYT.

7 Appendix

Reference texts for this section include [2, 11, 22].

7.1 Basic concepts

With $\mathbb{R}[x_1, \ldots, x_k]$ we indicated the set of polynomials in $x_1, \ldots, x_k$ and with real coefficients. The theory holds for whatever field K instead of $\mathbb{R}$. For us T^k indicates the set of power products or monomials in $\mathbb{R}[x_1, \ldots, x_k]$: $x^\alpha = x_1^{\alpha_1} \ldots x_k^{\alpha_k}$ for $\alpha_i \in \mathbb{Z}_{\geq 0}$ and a polynomial $f \in \mathbb{R}[x_1, \ldots, x_k]$ is a finite sum $f = \sum_{\alpha \in A} a_\alpha x^\alpha$ with $x^\alpha \in T^k$, $a_\alpha \in \mathbb{R}$ and for a finite subset $A \subset \mathbb{Z}^k_{\geq 0}$.

Definition 2 *A set $I \subset \mathbb{R}[x_1, \ldots, x_k]$ is a polynomial ideal if*

1. *$f + g \in I$ for all $f, g \in I$*
2. *$hf \in I$ for all $h \in \mathbb{R}[x_1, \ldots, x_k]$ and $f \in I$.*

We state the very deep property of polynomial ideals known as Hilbert Basis Theorem [11, Ch.2§5]

Theorem 11 *Every ideal $I \subseteq \mathbb{R}[x_1, \ldots, x_k]$ is finitely generated, i.e. there exist $g_1, \ldots, g_t \in I$ such that for every $f \in I$ there exist $h_1, \ldots, h_t \in \mathbb{R}[x_1, \ldots, x_k]$ that satisfy $f = h_1 g_1 + \cdots + h_t g_t$.*

The polynomials $g_1, \ldots, g_t$ in the previous theorem form a set of generators of I and we write $I = \langle g_1, \ldots, g_t \rangle$. There are special sets of generators called Gröbner bases. To introduce them we need the notion of term ordering. A term ordering τ is a total order relation on T^k that satisfies *i)* $x^\alpha > 1$ for all non zero $\alpha \in \mathbb{Z}^k_{\geq 0}$ and *ii)* if $x^\alpha > x^\beta$ then $x^\alpha x^\gamma > x^\beta x^\gamma$ for all $\alpha, \beta, \gamma \in \mathbb{Z}^k_{\geq 0}$.

Definition 3 *Given a term ordering τ, the leading term of a polynomial $f \in \mathbb{R}[x_1, \ldots, x_k]$ is its largest term with respect to τ, and we write it as $\mathrm{LT}_\tau(f)$.*

Given a term ordering τ and an ideal I, we consider the set of leading terms of all polynomials in I: $\mathrm{LT}_\tau(I) = \langle \mathrm{LT}_\tau(f) : f \in I \rangle$. If $g_1, \ldots, g_t$ is a generator set of an ideal I, in general $\mathrm{LT}_\tau(g_1), \ldots,, \mathrm{LT}_\tau(g_t)$ is not a set of generators of $\mathrm{LT}_\tau(I)$. This remark justifies the following definition.

Definition 4 *Let I be an ideal, τ a term ordering and $G = \{g_1, \ldots, g_t\} \subseteq I$. G is a Gröbner basis (sometimes called a standard basis) of I if $\mathrm{LT}_\tau(I)$ is generated by $\langle \mathrm{LT}_\tau(g) : g \in G \rangle$.*

Theorem 12 *For every ideal I and term ordering τ there exist finite Gröbner bases of I.*

Definition 5 *Let $r = \sum_{\alpha \in A} a_\alpha x^\alpha$ be a polynomial, τ a term ordering and I be an ideal. r is in normal form w.r.t. I and τ if $x^\alpha \notin \mathrm{LT}_\tau(I)$ for all $\alpha\ inA$.*

The following result holds.

Proposition 1 *Let τ be a term ordering, I an ideal and let $G = \{g_1, \ldots, g_t\}$ be a Gröbner basis of I w.r.t. τ. For every polynomial $f \in \mathbb{R}[x_1, \ldots, x_k]$ there exists a unique $r \in \mathbb{R}[x_1, \ldots, x_k]$ in normal form and $h_1, \ldots, h_t \in \mathbb{R}[x_1, \ldots, x_k]$ such that $f = h_1 g_1 + \cdots + h_t g_t + r$. Furthermore, $r = 0$ if and only if $f \in I$.*

Given an ideal I, we can consider the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I$ whose elements are the equivalence classes $[f]$ of the relation $f \sim g$ if $f - g \in I$. It is easy to prove that if r is the normal form of f w.r.t. I and τ, then $[f] = [r]$ and so the elements of $\mathbb{R}[x_1, \ldots, x_k]/I$ (are represented) by polynomials obtained as combination of terms not in $\mathrm{LT}_\tau(I)$. The set $\mathrm{SM}_\tau(I) = T^k \setminus \mathrm{LT}_\tau(I)$ is called the set of the standard monomials of I w.r.t. τ. As $\mathbb{R}$-vector spaces, $\mathbb{R}[x_1, \ldots, x_k]/I$ is isomorphic to $\mathbb{R}[x_1, \ldots, x_k]/\mathrm{LT}_\tau(I)$ and so it is isomorphic to the vector space spanned by $\mathrm{SM}_\tau(I)$ over $\mathbb{R}$. The Singular macro *kbasis* returns $\mathrm{SM}_\tau(I)$ for an ideal of points.

7.2 Affine Hilbert function for ideals

For $s \in \mathbb{Z}_{\geq 0}$ let $\mathbb{R}[x_1, \ldots, x_k]_{\leq s} = \mathrm{Span}(x^\alpha \in T^k : \sum_{i=1}^k \alpha_i \leq s)$. For an ideal $I \subset \mathbb{R}[x_1, \ldots, x_k]$, let $I_{\leq s} = I \cap \mathbb{R}[x_1, \ldots, x_k]_{\leq s}$. As $\mathbb{R}[x_1, \ldots, x_k]_{\leq s}$ is a $\mathbb{R}$-vector space of dimension $\binom{k+s}{s}$ and $I_{\leq s}$ is a subvector space of $\mathbb{R}[x_1, \ldots, x_k]_{\leq s}$, we can define the *affine Hilbert function* of I as

$$^{\mathrm{a}}\mathrm{HF}_I(s) = \dim \mathbb{R}[x_1, \ldots, x_k]_{\leq s}/I_{\leq s} = \dim \mathbb{R}[x_1, \ldots, x_k]_{\leq s} - \dim I_{\leq s}.$$

There exists s_0 called the *index of regularity of* I such that for all $s \geq s_0$ ${}^{\mathrm{a}}\mathrm{HF}_I(s)$ is a polynomial with integer coefficients. It is called the *affine Hilbert polynomial* of I and denoted as ${}^{\mathrm{a}}\mathrm{HP}_I(s)$. That is

$$ {}^{\mathrm{a}}\mathrm{HP}_I(s) = \sum_{i=0}^{k} b_i \binom{s}{k-i} $$

with $b_i \in \mathbb{Z}_{\geq 0}$ and $b_i > 0$. The following theorem gives the affine Hilbert function for the design ideal $I(\mathcal{D})$.

Theorem 13 *Let $I(\mathcal{D})$ be the ideal generated by a design $\mathcal{D}$ with n distinct points. Then for $s \geq n$, ${}^{\mathrm{a}}\mathrm{HF}_{I(\mathcal{D})}(s) = {}^{\mathrm{a}}\mathrm{HP}_{I(\mathcal{D})}(s) = n$.*

Proof. This is in [11, Ex.10,Ch.9§4]. ■

The Hilbert function counts the monomials that are not in $I(\mathcal{D})$; this set of monomials is precisely the set of standard monomials as described in Subsection 7.1. As ${}^{\mathrm{a}}\mathrm{HF}_{I(\mathcal{D})}(s)$ is a constant, we retrieve the standard result $\dim \mathbb{R}[x_1, \ldots, x_k]/I = n$.

A term ordering τ is graded if x^α is larger than x^β whenever $\sum_{i=1}^k \alpha_i > \sum_{i=1}^k \beta_i$. Let τ be a graded term ordering, then for all $s \in \mathbb{Z}_{\geq 0}$

$$ {}^{\mathrm{a}}\mathrm{HF}_I(s) = \# \left(\mathrm{SM}_\tau(I) \cap \mathbb{R}[x_1, \ldots, x_k]_{\leq s}\right) $$

where $\#A$ is the size of the set A.

7.3 Homogenising a mixture ideal

A key point in this paper is the study of mixture designs through cone ideals, namely $\mathrm{Ideal}(\mathcal{C}_\mathcal{D}) \subset \mathbb{R}[x_1, \ldots, x_k]$ for a mixture design $\mathcal{D}$. As mentioned in the main text, there are macros e.g. `IdealOfProjectivePoints` which construct a generator set for $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ from the coordinates of $\mathcal{D}$. Next we outline the basic construction of $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ which can be performed in any software for ideal computation. Let $\mathcal{D} = \{P_1, \ldots, P_n\}$ be the design and assume that $P_i = (a_{i1}, \ldots, a_{ik})$ with $\sum_{j=1}^k a_{ij} = 1$. Then, P_i belongs to the hyperplane H defined by the single equation $x_1 + \ldots + x_k = 1$ for $i = 1, \ldots, n$. Moreover P_i is the intersection of H with the line L_i containing P_i and the origin $0 = (0, \ldots, 0)$. In particular, we have $\mathrm{Ideal}(\{P_i\}) = \langle \mathrm{Ideal}(\{L_i\}), x_1 + \ldots + x_k - 1\rangle$. But $\mathrm{Ideal}(\mathcal{D}) = \bigcap_{i=1}^n \mathrm{Ideal}(\{P_i\}) = \langle \bigcap_{i=1}^n \mathrm{Ideal}(\{L_i\}), x_1 + \ldots + x_k - 1\rangle$. We set $\mathrm{Ideal}(\mathcal{C}_\mathcal{D}) = \bigcap_{i=1}^n \mathrm{Ideal}(\{L_i\})$, and so

$$ \mathrm{Ideal}(\mathcal{D}) = \langle \mathrm{Ideal}(\mathcal{C}_\mathcal{D}), x_1 + \ldots + x_k - 1\rangle $$

Now, we describe some properties of $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$.

Theorem 14 Ideal($\mathcal{C}_\mathcal{D}$) *is generated by homogeneous polynomials.*

Proof. The ideal defining the lines L_i is generated by the 2×2 minors of the matrix

$$\begin{pmatrix} x_1 & \dots & x_k \\ a_{i1} & \dots & a_{ik} \end{pmatrix}$$

and so it is generated by homogeneous linear polynomials. The intersection of ideals generated by homogeneous polynomials is again generated by homogeneous polynomials. So the claim follows. ■

Ideal($\mathcal{C}_\mathcal{D}$) can be characterized as follows.

Theorem 15 Ideal($\mathcal{C}_\mathcal{D}$) *is the largest homogeneous ideal in* Ideal($\mathcal{D}$)*.*

Proof. Let $f \in$ Ideal($\mathcal{D}$), f homogeneous. Then

$$f(ta_{i1}, \dots, ts_{ik}) = t^{\deg f} f(a_{i1}, \dots, f_{ik}) = 0$$

for every $i = 1, \dots, n$ and for all $t \in \mathbb{R}$. Hence, $f \in$ Ideal($\{L_i\}$) for all $i = 1, \dots, n$ and so $f \in$ Ideal($\mathcal{C}_\mathcal{D}$). That is every homogeneous polynomial in Ideal($\mathcal{D}$) is in Ideal($\mathcal{C}_\mathcal{D}$) and the claim follows. ■

7.4 Hilbert function

An ideal $I^h \subset \mathbb{R}[x_1, \dots, x_k]$ is homogeneous if it is generated by a set of homogeneous polynomials.

For $s \in \mathbb{Z}_{\geq 0}$ let $\mathbb{R}[x_1, \dots, x_k]_s = \text{Span}(x^\alpha \in T^k : \sum_{i=1}^k \alpha_i = s) \cup \{0\}$ and for a homogeneous ideal $I^h \subset \mathbb{R}[x_1, \dots, x_k]$, let $I^h_s = I^h \cap \mathbb{R}[x_1, \dots, x_k]_s$. $\mathbb{R}[x_1, \dots, x_k]_s$ is a $\mathbb{R}$-vector space of dimension $\binom{k+s-1}{s}$ and I^h_s is a subvector space. The Hilbert function of the homogeneous ideal I is

$$\text{HF}_I(s) = \dim \mathbb{R}[x_1, \dots, x_k]_s / I^h_s$$

Theorem 16 *Let* $I^h \subset \mathbb{R}[x_1, \dots, x_k]$ *be a homogeneous ideal.*

1. *For s sufficiently large* $\text{HF}_{I^h}(s)$ *is a polynomial with rational coefficients and integer values.*

2. *For* $s \geq 1$
$$\text{HF}_{I^h}(s) = {}^{\text{a}}\text{HF}_{I^h}(s) - {}^{\text{a}}\text{HF}_{I^h}(s-1) \tag{9}$$

3. *If I^h is a monomial ideal and thus trivially homogeneous, then* $\text{HF}_{I^h}(s)$ *is the number of monomials not in I^h and in* $\mathbb{R}[x_1, \dots, x_k]_s$*.*

4. *If τ is a term ordering and I^h a homogeneous ideal, then*

$$\mathrm{HF}_{I^h}(s) = \mathrm{HF}_{\langle \mathrm{LT}(I^h)\rangle}(s)$$

5. *(The dimension theorem) Let*

$$V = V(I) = \left\{ a \in \mathbb{P}^{k-1}(\mathbb{C}) : f(a) = 0 \text{ for all } f \in I \right\}$$

be non empty. Then

$$\dim(V) = \deg \mathrm{HP}_I(s)$$

where $\dim(V)$*, for V a projective variety, is defined as the degree of the Hilbert polynomial of I. Furthermore,*

$$\dim(V) = \deg \mathrm{HP}_{\langle \mathrm{LT}(I)\rangle}(s)$$

and it equals the maximum dimension of a projective coordinate subspace in $V(\langle \mathrm{LT}(I)\rangle)$. If $I = \mathrm{Ideal}(V)$ the last statements hold over $\mathbb{R}$.

6. *The previous statement holds for I an ideal, not necessarily homogeneous, $V = V(I)$ and $\mathrm{HP}_I(s)$ is substituted by ${}^{\mathrm{a}}\mathrm{HP}_I(s)$*

For the proof we refer to any classical text such as [11]. Here we just need to observe that as we deal with a regular structure as $V = \mathcal{C}_{\mathcal{D}}$ then $I = \mathrm{Ideal}(V)$.

The CoCoA macro `Hilbert` applied to a homogeneous ideal computes the Hilbert function of the ideal. In Singular we use `hilb` and `vdim`. The affine Hilbert function of the homogeneous ideal can be retrieved by Equation (9) together with the initial condition ${}^{\mathrm{a}}\mathrm{HF}_{I^h}(0) = 1$. If the ideal is not homogeneous then `Hilbert` returns the Hilbert function of the corresponding leading term ideal w.r.t. whatever term ordering is running in the open computer session.

Example 19 For $\mathcal{D} = \{(1/2, 1/2), (1/4, 3/4), (0, 1)\}$ and a term order in which $x_1 > x_2$, $\mathrm{Ideal}(\mathcal{C}_{\mathcal{D}}) = \langle x_1^3 - 4/3x_1^2x_2 + 1/3x_1x_2^2\rangle$. Then

s	$\mathrm{HF}_{\mathrm{Ideal}(\mathcal{C}_{\mathcal{D}})}(s)$	${}^a\,\mathrm{HF}_{\mathrm{Ideal}(\mathcal{C}_{\mathcal{D}})}(s)$
0	1	1
1	2	3
2	3	6
3	3	9
4	3	12
$\vdots$	3	$3 + {}^a\,\mathrm{HF}_{\mathrm{Ideal}(\mathcal{C}_{\mathcal{D}})}(s-1)$

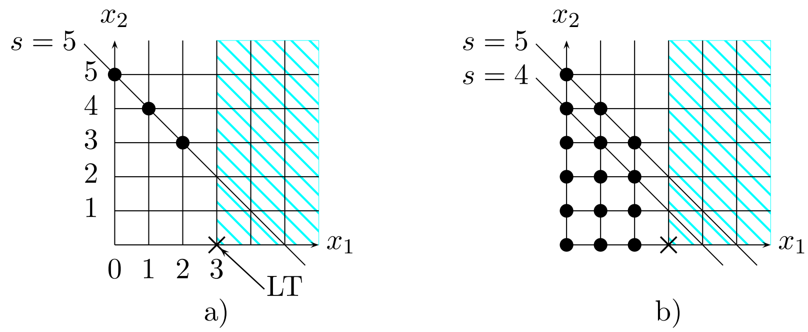

Figure 2: Standard monomials counted by a) the Hilbert function with $s = 5$ and b) the affine Hilbert function for $s = 4, 5$. Both cases refer to $\mathrm{I}(\mathcal{C}_\mathcal{D})$ of Example 19.

See Figure 2.

Theorem 17 *Let $\mathcal{D}$ be a mixture design with n distinct points and let $\mathcal{C}_\mathcal{D}$ be its cone; let* $\mathrm{Ideal}(\mathcal{D})$ *and* $\mathrm{Ideal}(\mathcal{C}_\mathcal{D})$ *be their corresponding ideals. Then for s large enough,*

$$\mathrm{HF}_{I(\mathcal{C}_\mathcal{D})}(s) = {}^{\mathrm{a}}\mathrm{HF}_{I(\mathcal{D})}(s)$$

Proof. This is Theorem 5. ■

Example 20 The Hilbert function of the cone ideal of the $\{4, 4\}$ design in Example 15 is

$$\mathrm{HF}_{\mathrm{Ideal}(\mathcal{C}_\mathcal{D})}(s) = \begin{cases} 1 & \text{if } s = 0 \\ 4 & \text{if } s = 1 \\ 10 & \text{if } s = 2 \\ 20 & \text{if } s = 3 \\ 35 & \text{if } s \geq 4 \end{cases}$$

For the fraction cut by $(x_1 - x_2)(x_3 - x_4)$ it is

$$\mathrm{HF}_{\mathrm{Ideal}(\mathcal{C}_\mathcal{F})}(s) = \begin{cases} 1 & \text{if } s = 0 \\ 4 & \text{if } s = 1 \\ 9 & \text{if } s = 2 \\ 13 & \text{if } s = 3 \\ 15 & \text{if } s \geq 4 \end{cases}$$

We use the CoCoA macro `Hilbert`.

8 CoCoA programs

8.1 Function for homogeneous model support

```
// Compares two monomials, the first argument is "leadterm".
// Returns 1 if Compa is divisible by Led
Define Divides(Led,Compa);
Suma:=0; Difference:=Log(Compa)-Log(Led);
For I:=1 To NumIndets() Do
    If Difference[I]<0 Then Suma:=Suma+1;
EndIf; EndFor;
If IsZero(Suma) Then R:=1; Else R:=0; EndIf;
Return R; EndDefine;

// Output is a homogeneous basis for R[x]s/Is.
// Inputs are homogeneous Ideal and degree Gr
Define HomogeneousBasis(Id,Gr);
Le:=Gens(LT(Id)); C:=Support(DensePoly(Gr)); L:=[];
ForEach I In C Do
   Total:=0; J:=1;
   While Total=0 And J<=Len(Le) Do
       Total:=Total+Divides(Le[J],I); J:=J+1;  EndWhile;
   If IsZero(Total) Then Append(L,I); EndIf;
EndForEach;
Return L; EndDefine;
```

8.2 Function for the point coordinates of some classic mixture designs

```
// Simplex lattice design in K factors M+1 levels
Define SimplexLattice(K,M);
L:=(0..M); For I:=1 To (M+1) Do L[I]:=L[I]/M EndFor;
L:=Tuples(L,K);
ForEach I In L Do If Sum(I)<>1 Then L:=Diff(L,[I]); EndIf;  EndForEach;
Return L; EndDefine;

// 2^K design minus the origin, i.e. simplex centroid design in K factors
Define SimplexCentroidProjective(K);
R:=Tuples([0,1],K); Remove(R,1);
Return R; EndDefine;
```

```
// Snee Marquardt axial design: K factors, if E=1 then add end points
Define Snee(K,E);
R:=[];PV:=[]; PI:=[]; PE:=[];
ForEach I In 1..K Do
   Zero:=NewList(K,0); One:=NewList(K,1); Zero[I]:=1; One[I]:=0;
   PV:=Concat(PV,[Zero]); PE:=Concat(PE,[One]);  One[I]:=K+1;
   PI:=Concat(PI,[One]);
EndForEach;
   R:=Concat(PV,[NewList(K,1)],PI); If E=1 Then R:=Concat(R,PE); EndIf;
Return R; EndDefine;
```

8.3 Function for the cone ideal of some classic mixture designs

```
// Generators for the design ideal of the simplex lattice
Define GSimplexLattice(K,M);
G:=-1;  L:=[];
ForEach I In 1..K Do;
  P:=1;  ForEach J In 0..M Do  P:=P*(x[I]-J/M); EndForEach;
  G:=G+x[I]; L:=Concat(L,[P]);
EndForEach; L:=Concat(L,[G]);
Return L; EndDefine;

// Generators for the cone ideal of the simplex centroid design
Define GSimplexCentroidProjective(K);
L:=[];
ForEach I In 1..(K-1) Do ForEach J In (I+1)..K Do
     L:=Concat(L,[ x[I]^2*x[J]-x[I]*x[J]^2]);
EndForEach; EndForEach;
Return L; EndDefine;

// Generators for the I-th axis of a Snee Marquardt with K factors
Define GAxisSnee(K,I);
G:=Sum(Indets())-1; L:=[];
ForEach M In 1..(K-1) Do ForEach J In (M+1)..K Do
     If (J<>I And M<>I) Then L:=Concat(L,[ x[M]-x[J] ]); EndIf;
EndForEach; EndForEach;
L:=Concat(L,[G],[ x[I]*(x[I]-(K+1)/(2*K))*(x[I]-1) ]);
Return L; EndDefine;
```

```
// Generators for the ideal of centroid with K variables
Define GCentroid(K);
L:=[]; ForEach I In 1..K Do L:=Concat(L,[x[I]-1/K]); EndForEach;
Return L; EndDefine;
```

References

[1] J. Abbott, A. Bigatti, M. Kreuzer, and L. Robbiano. Computing ideals of points. *J. Symbolic Comput.*, 30(4):341–356, 2000.

[2] W. W. Adams and P. Loustaunau. *An introduction to Gröbner bases*, volume 3 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1994.

[3] J. Aitchison. *The statistical analysis of compositional data*. Monographs on Statistics and Applied Probability. Chapman & Hall, London, 1986.

[4] E. Babson, S. Onn, and R. Thomas. The Hilbert zonotope and a polynomial time algorithm for universal Gröbner bases. *Adv. in Appl. Math.*, 30(3):529–544, 2003.

[5] C. Bocci, G. Dalzotto, R. Notari, and M. L. Spreafico. An iterative construction of Gorenstein ideals. *Trans. Amer. Math. Soc.*, 357(4):1417–1444 (electronic), 2005.

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

[7] M. Caboara and E. Riccomagno. An algebraic computational approach to the identifiability of Fourier models. *JSC*, 26(2):245–260, 1998.

[8] P. J. Claringbold. Use of the simplex design in the study of joint action of related hormones. *Biometrics*, 1955.

[9] CoCoATeam. CoCoA: a system for doing Computations in Commutative Algebra. Available at `http://cocoa.dima.unige.it`.

[10] J. A. Cornell. *Experiments with mixtures*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], New York, third edition, 2002. Designs, models, and the analysis of mixture data.

[11] D. Cox, J. Little, and D. O'Shea. *Ideal, Varieties, and Algorithms*. Springer-Verlag, New York, 1997. Second Edition.

[12] D. Cox, J. Little, and D. O'Shea. *Using algebraic geometry.* Springer-Verlag, New York, 2004. Second Edition.

[13] D. R. Cox. A note on polynomial response functions for mixtures. *Biometrika*, 1971.

[14] N. R. Draper and F. Pukelsheim. Mixture models based on homogeneous polynomials. *J. Statist. Plann. Inference*, 71(1-2):303–311, 1998.

[15] J. C. Faugère, P. Gianni, D. Lazard, and T. Mora. Efficient computation of zero-dimensional Gröbner bases by change of ordering. *J. Symbolic Comput.*, 16(4):329–344, 1993.

[16] R. Fontana, G. Pistone, and M.-P. Rogantin. Algebraic analysis and generation of two level designs. *Statistica Applicata*, 9(1):15–29, 1997.

[17] R. Fontana, G. Pistone, and M.-P. Rogantin. Classification of two-level factorial fractions. *J. Statist. Plann. Inference*, 87(1):149–172, 2000.

[18] B. Giglio, H. P. Wynn, and E. Riccomagno. Gröbner basis methods in mixture experiments and generalisations. In *Optimum design 2000 (Cardiff)*, volume 51 of *Nonconvex Optim. Appl.*, pages 33–44. Kluwer Acad. Publ., Dordrecht, 2001.

[19] D. R. Grayson and M. E. Stillman. Macaulay 2, a software system for research in algebraic geometry. Available at http://www.math.uiuc.edu/Macaulay2/.

[20] G.-M. Greuel, G. Pfister, and H. Schönemann. SINGULAR 3.0. A Computer Algebra System for Polynomial Computations, Centre for Computer Algebra, University of Kaiserslautern, 2005. `http://www.singular.uni-kl.de`.

[21] T. Holliday, G. Pistone, E. Riccomagno, and H.P. Wynn. The application of algebraic geometry to analysis of designed experiments: A case study. *Computational Statistics*, 14(2):213–231, 1999.

[22] M. Kreuzer and L. Robbiano. *Computational commutative algebra. 1.* Springer-Verlag, Berlin, 2000.

[23] B.J. McConkey, P.G. Mezey, D.G. Dixon, and B.M. Grenberg. Fractional simplex designs for interaction screening in complex mixtures. *Biometrics*, 56:824–832, 2000.

[24] P. McCullagh and J. A. Nelder. *Generalised linear model*. Chapman and Hall, London, 1989.

[25] J. L. Peixoto and J. Díaz. On the construction of hierarchical polynomial regression models. *Estadística*, 48(150-151):175–210 (1998), 1996.

[26] G. F. Piepel, R. D. Hicks, J. M. Szychowski, and J. L. Loeppky. Methods for assessing curvature and interaction in mixture experiments. *Technometrics*, 44(2):161–172, 2002.

[27] G. Pistone, E. Riccomagno, and H. P Wynn. Gröbner basis methods for structuring and analysing complex industrial experiments. *International Journal of Reliability, Quality and Safety Engineering*, 7(4):285–300, 2000.

[28] G. Pistone, E. Riccomagno, and H. P. Wynn. *Algebraic Statistics*, volume 89 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, 2001.

[29] G. Pistone and H. P. Wynn. Generalised confounding with Gröbner bases. *Biometrika*, 83(3):653–666, 1996.

[30] H. Scheffé. Experiments with mixtures. *J. Roy. Statist. Soc. Ser. B*, 20:344–360, 1958.

[31] H. Scheffé. The simplex-centroid design for experiments with mixtures. *J. Roy. Statist. Soc. Ser. B*, 25:235–263, 1963.

[32] R.D. Snee and D.W. Marquardt. Screening concepts and designs for experiments with mixtures. *Technometrics*, 18:19–29, 1976.

[33] J. N. Darroch; J. Waller. Additivity and interaction in three-component experiments with mixtures. *Biometrika*, 1985.

[34] K. Q. Ye. Indicator function and its application in two level factorial designs. *The Annals of Statistics*, 31(3):984–994, 2003.

[35] C. Q. Zhang, L. Y. Chan, Y. N. Guan, K. H. Li, and T. S. Lau. Optimal designs for an additive quadratic mixture model involving the amount of mixture. *Statist. Sinica*, 15(1):165–176, 2005.