Generalised design: interpolation and statistical modelling over varieties

(Generalised design)

Hugo Maruri-Aguilar and Henry P Wynn

Contents

1

Generalised design: Interpolation and statistical modeling over varieties

In the classical formulation an experimental design is a set of sites at each of which an observation is taken on a response Y. The algebraic method treats the design as giving an "ideal of points" from which potential monomial bases for a polynomial regression can be derived. If the Gröbner basis method is used then the monomial basis depends on the monomial term ordering. The full basis has the same number of terms as the number of design points and gives an exact interpolator for the Y-values over the design points. Here the notation of design point is generalized to a variety. Observation means, in theory, that one observes the value of the response on the variety. A design is a union of varieties and the assumption is, then, that on each variety we observe the response. The task is to construct an interpolator for the function between the varieties. Motivation is provided by transect sampling in a number of fields. Much of the algebraic theory extends to the general case. But special issues arise including the consistency of interpolation at the intersection of the varieties and the consequences of taking a design of points restricted to the varieties.

1.1 Introduction

Experimental design is defined simply as the choice of sites, or observation points, at which to observe a response, or output. A set of such points is the experimental design. Terminology varies according to the field. Thus, sites may be called "treatment combinations", "input configurations", "runs", "data points" and so on. For example in interpolation theory "observation point" is common. Whatever the terminology or field we can nearly always code up the notion of an observation point as a single point in k dimensions which represents a single combination of levels of k independent variables.

The purpose of this paper is to extend the notation of an observation point to a whole algebraic variety. An experimental design is then a union of such varieties. An observation would be the acquired knowledge of the restriction of the response to the variety. This is an idealization, but one with considerable utility. It may be, for example that one models the restriction of the response to each variety by a separate polynomial.

An important example of sampling via a variety is *transect sampling*. This is a method used in the estimation of species abundance in ecology and geophysics. A key text is Buckland et al. (1993) and the methods are developed further in Mack & Quang (1998). There one collects information about the distance of objects from the transects and tries to estimate the average density of the objects in the region of interest, namely to say something about a feature connected with the whole region. A useful idea is that of "reconstruction"; one tries to reconstruct a function given the value on the transects. This reconstruction we interpret here as "interpolation", or perhaps we should say "generalized" interpolation. Other examples are geophysics, tomography, computer vision and imaging.

Our task is to extend the algebraic methods used for observation points to this generalized type of experimental design and interpolation. Within this, the main issue is to create monomial bases to interpolate between the varieties on which we observe. At one level this is a straightforward extension, but there are a number of special constructions and issues the discussion of which should provide an initial guide to the area.

(i) The most natural generalization is to the case where the varieties are hyperplanes, and therefore we shall be interested in hyperplane arrangements. This covers the case of lines in two dimensions, the traditional transects mentioned above.

(ii) There are consistency issues when the varieties intersect: the observation on the varieties must agree on the intersection.

(iii) Since observing a whole function on a variety may be unrealistic one can consider traditional point designs restricted to the varieties. That is, we may use standard polynomial interpolation on the varieties and then combine the results to interpolate between varieties, but having in mind the consistency issue just mentioned.

(iv) It is also natural to use power series expansions on each variety:

is it possible to extend the algebraic interpolation methods to power series? We are here only able to touch on the answer.

We now recall some basic ideas. Interpolation is the construction of a function $f(x)$ that coincides with observed data at n given observation points. That is, for a finite set of distinct points $\mathcal{D} = \{d_1, \ldots, d_n\}$, $d_1, \ldots, d_n \in \mathbb{R}^k$ and observation values $y_1, \ldots, y_n \in \mathbb{R}$, we build a function such that $f(d_i) = y_i$, $i = 1, \ldots, n$. We set our paper within design of experiments theory where the design is a set of points $\mathcal{D}$, n is the design (sample) size and k is the number of factors. Approaches to interpolation range from statistically oriented techniques such as kriging, see Stein (1999), to more algebraic techniques involving polynomials, splines or operator theory, see Phillips (2003) and Sakhnovich (1997).

Pistone & Wynn (1996) build polynomial interpolators using an isomorphism between the following real vector spaces: the set of real valued polynomial functions defined over the design, $\phi : \mathcal{D} \longrightarrow \mathbb{R}$, and the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{D})$. To construct the quotient ring they first consider the design $\mathcal{D}$ as the set of solutions to a system of polynomial equations. Then this design corresponds to the *design ideal* $I(\mathcal{D})$, that is the set of all polynomials in $\mathbb{R}[x_1, \ldots, x_k]$ that vanish over the points in $\mathcal{D}$. The polynomial interpolator has n terms and is constructed using a basis for $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{D})$ called *standard monomials*.

This algebraic method of constructing polynomial interpolators can be applied to, essentially, any finite set of points, see for example Holliday et al. (1999) and Pistone et al. (2006). In fractional factorial designs it has lead to the use of indicator functions, see Fontana et al. (1997), Pistone & Rogantin (2008). Another example arises when the design is a mixture, i.e. the coordinate values of each point in $\mathcal{D}$ add up to one. In such a case the equation $\sum_{i=1}^{k} x_i = 1$ is incorporated into the design ideal, namely the polynomial $\sum_{i=1}^{k} x_i - 1 \in I(\mathcal{D})$, see Giglio et al. (2001). More recently, Maruri-Aguilar et al. (2007) used projective algebraic geometry and considered the projective coordinates of the mixture points. Their technique allows the identification of the support for a homogeneous polynomial model.

If, instead of a set of points, we consider the design as an affine variety, then the algebraic techniques discussed are still valid. As a motivating example, consider the circle in two dimensions with radius two and center at the origin. Take the *radical ideal* generated by the circle as its design ideal, i.e. the ideal generated by $x_1^2 + x_2^2 - 4$. The set of standard monomials is infinite in this case. For a monomial order with initial order $x_2 \prec x_1$, the set of standard monomials is $\{x_2^j, x_1 x_2^j : j \in \mathbb{Z}_{\geq 0}\}$, and can be used to interpolate over the circle. However, a number of questions arise: What is the interpretation of observation on such a variety? What method of statistical analysis should be used?

In this paper, then, we are concerned with extending interpolation to when the design no longer comprises a finite set of points, but it is defined as the union of a finite number of affine varieties, see Definition 1. Only real affine varieties (without repetition) and the radical ideals generated by them are considered. Real affine varieties can be linked to complex varieties, see Whitney (1957) for an early discussion on properties of real varieties. In Section 1.2.2 we study the case when the design $\mathcal{V}$ comprises the union of $(k-1)$-dimensional hyperplanes. In Section 1.2.3 we present the case, when every affine variety is an intersection of hyperplanes. The following is a motivating example of such linear varieties.

Example 1 Consider a general bivariate Normal distribution $(X_1, X_2)^T \sim N\left((\mu_1, \mu_2)^T, \Sigma\right)$ with

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_1\sigma_2\rho & \sigma_2^2 \end{pmatrix},$$

where σ_1, σ_2 are real positive numbers, and $\rho \in [-1, 1] \subset \mathbb{R}$. Now when Σ is fixed, $\log p(x_1, x_2)$ is a quadratic form in μ_1, μ_2, where $p(x_1, x_2)$ is the normal bivariate density function. Imagine that, instead of observing at a design point, we are able to observe $\log p(x_1, x_2)$ over a set of lines $\mathcal{V}_i$, $i = 1, \ldots, n$. That is, the design $\mathcal{V}$ is a union of lines (transects), and suppose we have perfect transect sampling on every line on the design. This means that we know the value of $\log p(x_1, x_2)$ on every line.

The question is: how do we reconstruct the entire distribution? Are there any conditions on the transect location?

We do not attempt to resolve these issues here. Rather we present the ideas as a guide to experimentation on varieties in the following sense. If $I(\mathcal{V})$ is the design ideal, then the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$ is no longer of finite dimension, but we can still obtain a basis for it and use it to construct statistical models for data observed on $\mathcal{V}$.

Even though we can create a theory of interpolation by specifying, or "observing" polynomial functions on a fixed variety $\mathcal{V}$, we may wish to observe a point set design $\mathcal{D}$ which is a subset of $\mathcal{V}$. In Section 1.3 we present this alternative, that is, to subsample a set of points $\mathcal{D}$ from a general design $\mathcal{V}$.

If instead, a polynomial function is given at every point on the algebraic variety, it is often possible to obtain a general interpolator which in turn coincides with the individual given functions. In Section 1.4 we give a simple technique for building an interpolator over a design and in Section 1.5 we survey the interpolation algorithm due to Becker & Weispfenning (1991). A related approach is to obtain a reduced expression for an analytic function defined over a design, which is discussed in Section 1.6. In Section 1.7 we discuss further extensions.

1.2 Definitions

In this Section we restrict to only the essential concepts for the development of the theory, referring the reader to the books in algebraic geometry by Cox et al. (1997), Kreuzer & Robbiano (2000) and Kreuzer & Robbiano (2005); we also refer the reader to the monograph in algebraic statistics by Pistone et al. (2001).

An affine algebraic set is the solution in $\mathbb{R}^k$ of a finite set of polynomials. The affine algebraic set of a polynomial ideal J is $Z(J)$. The set of polynomials which vanish on a set of points W in $\mathbb{R}^k$ is the polynomial ideal $I(W)$, which is radical. Over an algebraically closed field, such as $\mathbb{C}$, the ideal $I(Z(J))$ coincides with the radical ideal $\sqrt{J}$. However, when working on $\mathbb{R}$, which is not algebraically closed, the above does not necessarily hold.

Example 2 Take $J = \langle x^3 - 1\rangle \subset \mathbb{R}[x]$, i.e. the ideal generated by $x^3 - 1$. Therefore $Z(J) = \{1\}$ and $I(Z(J)) = \langle x - 1\rangle$. However J is a radical ideal and yet $I(Z(J)) \neq J$.

Recall that for $W \subset \mathbb{R}^k$, the set $Z(I(W))$ is the closure of W with respect to the Zariski topology on $\mathbb{R}^k$. There is a one to one correspondence between closed algebraic sets in $\mathbb{R}^k$ and radical ideals in $\mathbb{R}[x_1, \ldots, x_k]$ such that $I(Z(J)) = J$.

Example 3 Consider $I = \langle x^2\rangle \subset \mathbb{R}[x]$. Clearly I is not a radical ideal. However, its affine algebraic set is $Z(I) = \{0\}$, which is irreducible.

A real affine variety $\mathcal{V}$ is the affine algebraic set associated to a prime ideal. Remind that an algebraic variety $\mathcal{V}$ is irreducible, whenever $\mathcal{V}$ is written as the union of two affine varieties $\mathcal{V}_1$ and $\mathcal{V}_2$ then either $\mathcal{V} = \mathcal{V}_1$ or $\mathcal{V} = \mathcal{V}_2$.

Definition 1 *A design variety $\mathcal{V}$ is affine variety in $\mathbb{R}^k$ which is the union of irreducible varieties, i.e. for $\mathcal{V}_1, \ldots, \mathcal{V}_n$ irreducible varieties, $\mathcal{V} = \bigcup_{i=1}^n \mathcal{V}_i$.*

We next review quotient rings and normal forms computable with the variety ideal $I(\mathcal{V})$.

Two polynomials $f, g \in \mathbb{R}[x_1, \ldots, x_k]$ are congruent modulo $I(\mathcal{V})$ if $f - g \in I(\mathcal{V})$. The quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$ is the set of equivalence classes for congruence modulo $I(\mathcal{V})$. The ideal of leading terms of $I(\mathcal{V})$ is the monomial ideal generated by the leading terms of polynomials in $I(\mathcal{V})$, which is written as $\langle \mathrm{LT}(I(\mathcal{V}))\rangle = \langle \mathrm{LT}(f) : f \in I(\mathcal{V})\rangle$.

Two isomorphisms are considered. Firstly, as real vector space, the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/\langle \mathrm{LT}(I(\mathcal{V}))\rangle$ is isomorphic to $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$. Secondly, the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$ is isomorphic (as real vector space) to $\mathbb{R}[\mathcal{V}]$, the set of polynomial functions defined on $\mathcal{V}$.

For a fixed monomial ordering $\prec$, let G be a Gröbner basis for $I(\mathcal{V})$ and let $L_\prec(I(\mathcal{V}))$ be the set of all monomials that cannot be divided by the leading terms of the Gröbner basis G, that is

$$L_\prec(I(\mathcal{V})) := \{x^\alpha \in T^k : x^\alpha \text{ is not divisible by } \mathrm{LT}_\prec(g), g \in G\} \quad (1.1)$$

where T^k is the set of all monomials in $x_1, \ldots, x_k$. This set of monomials is known as the set of standard monomials, and when there is no ambiguity, we refer to it simply as $L(\mathcal{V})$. We reformulate in the setting of interest of this paper the following proposition (Cox et al. 1997, Section 5§3, Proposition 4).

Proposition 1 *Let $I(\mathcal{V}) \subset \mathbb{R}[x_1, \ldots, x_k]$ be a radical ideal. Then $\mathbb{R}[x_1, \ldots, x_k]/\langle LT(I(\mathcal{V}))\rangle$ is isomorphic as a $\mathbb{R}$-vector space to the polynomials which are real linear combinations of monomials in $L(\mathcal{V})$.*

In other words, the monomials in $L(\mathcal{V})$ are linearly independent modulo $\langle \mathrm{LT}(I(\mathcal{V}))\rangle$. By the two isomorphisms above, monomials in $L(\mathcal{V})$ form a basis for the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$ and for polynomial functions on $\mathcal{V}$. The division of a polynomial f by the elements of a Gröbner basis for $I(\mathcal{V})$ leads to a remainder r which is a linear combinations of monomials in $L(\mathcal{V})$, which is called the normal form of f.

Theorem 1 *(Cox et al. 1997, Section 2§3, Theorem 3) Let $I(\mathcal{V})$ be the ideal of a design variety $\mathcal{V}$; let $\prec$ be a fixed monomial order on $\mathbb{R}[x_1, \ldots, x_k]$ and let $G = \{g_1, \ldots, g_m\}$ be a Gröbner basis for $I(\mathcal{V})$ with* *respect to $\prec$. Then every polynomial $f \in \mathbb{R}[x_1, \ldots, x_k]$ can be expressed as $f = \sum_{i=1}^{m} g_i h_i + r$, where $h_1, \ldots h_m \in \mathbb{R}[x]$ and r is a linear combination of monomials in $L(\mathcal{V})$.*

We have that $f - r \in I(\mathcal{V})$ and, in the spirit of this paper, we say that the normal form r *interpolates* f on $\mathcal{V}$. That is, f and r coincide over $\mathcal{V}$. We may write $r = \mathrm{NF}_{\prec}(f, \mathcal{V})$ to denote the normal form of f with respect to the ideal $I(\mathcal{V})$ and the monomial ordering $\prec$.

1.2.1 Designs of points

The most elementary experimental point design has a single point $d_1 = (d_{11}, \ldots, d_{1k}) \in \mathbb{R}^k$, whose ideal is $I(d_1) = \langle x_1 - d_{11}, \ldots, x_k - d_{1k} \rangle$. An experimental design in statistics is the set of distinct points $\mathcal{D} = \{d_1, \ldots, d_n\}$, whose corresponding ideal is the following intersection:

$$I(\mathcal{D}) = \bigcap_{i=1}^{n} I(d_i). \tag{1.2}$$

Example 4 For $\mathcal{D} = \{(0,0), (1,0), (1,1), (2,1)\} \subset \mathbb{R}^2$, the set $G = \{x_1^3 - 3x_1^2 + 2x_1, x_1^2 - 2x_1x_2 - x_1 + 2x_2, x_2^2 - x_2\}$ is a Gröbner basis for $I(\mathcal{D})$. If we set a monomial order for which $x_2 \prec x_1$ then the leading terms of G are x_1^3, x_2^2 and x_1^2 and thus $L(\mathcal{D}) = \{1, x_1, x_2, x_1x_2\}$. Any real-valued polynomial function defined over $\mathcal{D}$ can be expressed as a linear combination of monomials in $L(\mathcal{D})$.

That is, for any function $f : \mathcal{D} \longrightarrow \mathbb{R}$, there is a unique polynomial $r(x_1, x_2) = c_0 + c_1x_1 + c_2x_2 + c_{12}x_1x_2$ where the constants c_0, c_1, c_2, c_{12} are real numbers whose coefficients can be determined by solving the linear system of equations $r(d_i) = f(d_i)$ for $d_i \in \mathcal{D}$. In particular if we observe real values y_i at $d_i \in \mathcal{D}$, in statistical terms, r is a saturated model. For example, if we observe the data $2, 1, 3, -1$ at the points in $\mathcal{D}$ then $r = 2 - x_1 + 5x_2 - 3x_1x_2$ is the saturated model for the data.

1.2.2 Designs of hyperplane arrangements

Let $H(a, c)$ be the ($(k-1)$-dimensional) affine hyperplane directed by a non-zero vector $a \in \mathbb{R}^k$ and with intercept $c \in \mathbb{R}$, i.e.

$$H(a, c) = \left\{ x = (x_1, \ldots, x_k) \in \mathbb{R}^k : l_a(x) - c = 0 \right\}$$

with $l_a(x) := \sum_{i=1}^{n} a_i x_i$. Now for a set of vectors $a_1, \ldots, a_n \in \mathbb{R}^k$, and real scalars $c_1, \ldots, c_n$, the hyperplane arrangement $\mathcal{A}$ is the union of the affine hyperplanes $H(a_i, c_i)$, that is

$$\mathcal{A} = \bigcup_{i=1}^{n} H(a_i, c_i). \tag{1.3}$$

We restrict the hyperplane arrangement to consist of distinct hyperplanes, i.e. no repetitions. The polynomial $Q_{\mathcal{A}}(x) := \prod_{i=1}^{n} (l_{a_i}(x) - c_i)$ is called the defining polynomial of $\mathcal{A}$. Combinatorial properties of hyperplane arrangements have been studied extensively in the mathematical literature, see (Grünbaum 2003, Chapter 18).

Clearly $\mathcal{A}$ is a variety as in Definition 1, $I(\mathcal{A})$ is a radical ideal and it is generated by $Q_{\mathcal{A}}(x)$. Furthermore for any monomial ordering $\prec$, $\{Q_{\mathcal{A}}(x)\}$ is a Gröbner basis for $I(\mathcal{A})$.

Example 5 Let a_i be the i-th unit vector and $c_i = 0$ for $i = 1, \ldots, k$, then $Q_{\mathcal{A}}(x) = x_1 \cdots x_k$ and $\mathcal{A}$ comprises the k coordinate hyperplanes.

Example 6 The *braid arrangement* plays an important role in combinatorial studies of arrangements. It has defining polynomial $Q_{\mathcal{A}}(x) = \prod(x_i - x_j - 1)$, where the product is carried on $i, j : 1 \leq i < j \leq k$, see Stanley (1996).

In the arrangement generated by the k coordinate hyperplanes of Example 5 and for any monomial order, the set of standard monomials comprises all monomials which miss at least one indeterminate, and this set does not depend on the term ordering used. For other hyperplane arrangements, the leading term of $Q_{\mathcal{A}}(x)$ may depend on the actual monomial order used. We have the following elementary result, which we state without proof.

Lemma 2 *Let $\mathcal{A}$ be an arrangement of n hyperplanes as in Equation (1.3). Then for any monomial ordering, the total degree of $LT_{\prec}(Q_{\mathcal{A}}(x))$ is n.*

Lemma 2 implies that the set of standard monomials for $\mathcal{A}$ always contains all monomials up to a total degree $n-1$. This result can be used in conjunction with the methodology of Section 1.3: an arrangement of n hyperplanes has the potential to identify a full model of total degree $n - 1$.

1.2.3 Generalised linear designs (GLDs)

The design variety in Section 1.2.2 can be generalised to include unions of intersections of distinct hyperplanes. Namely, $\mathcal{V} = \bigcup_{i=1}^{n} \mathcal{V}_i$ where $\mathcal{V}_i = \bigcap_{j=1}^{n_i} H(a_j^i, c_j^i)$ where a_j^i are non-zero vectors in $\mathbb{R}^k$ and $c_j^i \in \mathbb{R}^k$ for $j = 1, \ldots, n_i$ $i = 1, \ldots, n$ and n and $n_1, \ldots, n_n$ are positive integers. Consequently, the design ideal is the intersection of sums of ideals

$$I(\mathcal{V}) = \bigcap_{i=1}^{n} \sum_{j=1}^{n_i} I(H(a_j^i, c_j^i)).$$

Example 7 Let $\mathcal{V} \subset \mathbb{R}^3$ be constructed by the union of the following eleven affine sets: $\mathcal{V}_1, \ldots, \mathcal{V}_8$ are the eight hyperplanes $\pm x_1 \pm x_2 \pm x_3 - 1 = 0$, and $\mathcal{V}_9, \mathcal{V}_{10}, \mathcal{V}_{11}$ are the three lines in direction of the every coordinate axis. The varieties $\mathcal{V}_1, \ldots, \mathcal{V}_8$ form a hyperplane arrangement $\mathcal{A}'$. The variety $\mathcal{V}_9$ is the axis x_1 and thus is the intersection of the hyperplanes $x_2 = 0$ and $x_3 = 0$, i.e $I(\mathcal{V}_9) = \langle x_2, x_3 \rangle$. Similarly $I(\mathcal{V}_{10}) = \langle x_1, x_3 \rangle$ and $I(\mathcal{V}_{11}) = \langle x_1, x_2 \rangle$. The design is $\mathcal{V} = \mathcal{A}' \cup \mathcal{V}_9 \cup \mathcal{V}_{10} \cup \mathcal{V}_{11}$ and the design ideal is $I(\mathcal{V}) = I(\mathcal{A}') \cap I(\mathcal{V}_9) \cap I(\mathcal{V}_{10}) \cap I(\mathcal{V}_{11})$. For the lexicographic monomial ordering in which $x_3 \prec x_2 \prec x_1$, the Gröbner basis of $I(\mathcal{V})$ has three polynomials whose leading terms have total degree ten and are $x_1^9 x_2, x_1^9 x_3, x_1^8 x_2 x_3$ and thus

$$\begin{aligned} L(\mathcal{V}) \quad = \quad & \left\{1, x_1, x_1^2, x_1^3, x_1^4, x_1^5, x_1^6, x_1^7\right\} \otimes \left\{x_2^i x_3^j : (i,j) \in \mathbb{Z}_{\geq 0}^2\right\} \\ & \bigcup \left\{x_1^{j+9} : j \in \mathbb{Z}_{\geq 0}\right\} \bigcup \left\{x_1^8 x_2^{j+1} : j \in \mathbb{Z}_{\geq 0}\right\} \\ & \bigcup \left\{x_1^8 x_3^{j+1} : j \in \mathbb{Z}_{\geq 0}\right\} \bigcup \left\{x_1^8\right\}, \end{aligned}$$

where $\otimes$ denotes the Kronecker product of sets. That is, the set of exponents of monomials in $L(\mathcal{V})$ comprises the union of eight shifted copies of $\mathbb{Z}_{\geq 0}^2$, three shifted copies of $\mathbb{Z}_{\geq 0}$ and a finite set of monomials. This finite union of disjoint sets is an example of the Stanley decomposition of an $L(\mathcal{V})$, see Stanley (1978) and Sturmfels & White (1991).

1.3 Subsampling from a variety: "fill-up"

Varieties give a taxonomy which informs experimentation. Indeed, suppose that, for fixed $\mathcal{V}$, we take a finite sample of design points $\mathcal{D}$ from $\mathcal{V}$, i.e. $\mathcal{D} \subset \mathcal{V}$. We have the following inclusion between the quotient rings as real vector spaces

$$\mathbb{R}[x_1, \ldots, x_k]/\langle LT(I(\mathcal{D}))\rangle \subset \mathbb{R}[x_1, \ldots, x_k]/\langle LT(I(\mathcal{V}))\rangle. \qquad (1.4)$$

That is, the basis for the quotient ring $\mathbb{R}[x_1, \ldots, x_k]/I(\mathcal{V})$ provides an indication of the capability of models we can fit over $\mathcal{D}$ by setting the design $\mathcal{D}$ to lie on the affine variety $\mathcal{V}$. In particular, the sets of standard monomials for interpolating over $\mathcal{D}$ and over $\mathcal{V}$ satisfy $L_{\prec}(\mathcal{D}) \subset L_{\prec}(\mathcal{V})$. A question of interest is: given any finite subset $L' \subset L_{\prec}(\mathcal{V})$, can we find a set of points $\mathcal{D} \subset \mathcal{V}$ so that $L' \subseteq L_{\prec}(\mathcal{D})$?

An interesting case is the circle. Can we "achieve" a given L' from some finite design of points on the circle? The authors are able, in fact, to answer affirmatively with a sufficiently large equally spaced design around the circle, and a little help from discrete Fourier analysis. For instance set $\mathrm{LT}(x_1^2 + x_2^2 - 1) = x_2^2$ and thus $L = \{1, x_2\} \otimes \{x_1^j : j \in \mathbb{Z}_{\geq 0}\}$ and let $L' \subset L$ be the finite sub-basis. For $i = 0, \ldots, n-1$ let $(x_i, y_i) = \left(\cos\left(\frac{2\pi i}{n}\right), \sin\left(\frac{2\pi i}{n}\right)\right)$. For n sufficiently large, the design matrix $X = [x_i^u y_i^v]_{(u,v)\in L', i=0,\ldots,n-1}$ has full rank $|L'|$. Indeed we can explicitly compute the non zero determinant of $X^T X$ using Fourier formulæ.

The general case is stated as a conjecture.

Conjecture 3 *Let $\mathcal{V}$ be a design variety with set of standard monomials $L_{\prec}(\mathcal{V})$. Then, for any model with finite support on $L' \subset L_{\prec}(\mathcal{V})$, there is a finite design with points on the real part of $\mathcal{V}$ such that the model is identifiable.*

This conjecture can be proven when the design $\mathcal{V}$ is in the class of generalised linear designs (GLD) of Section 1.2.3. We believe that the construction may be of some use in the important inverse problem: finding a design which allows identification of a given model.

Proof Let $\mathcal{V} = \bigcup_{i=1}^n \mathcal{V}_i$ be a GLD, where the irreducible components are the $\mathcal{V}_i = \bigcap_{j=1}^{n_i} H(a_j^i, c_j^i)$. Take a finite set of monomials $L' \subset L(\mathcal{V})$ and consider a polynomial in this basis:

$$p(x) = \sum_{\alpha \in L'} \theta_\alpha x^\alpha,$$

i.e. $p(x)$ is a polynomial with monomials in L' and real coefficients. Select a $\mathcal{V}_i$ and consider the values of $p(x)$ on this variety. Suppose $\dim(\mathcal{V}_i) = k_i$, then by a linear coordinatisation of the variety we can reduce the design problem on the variety to the identification of a model of a particular order on $\mathbb{R}^{k_i}$. But using the "design of points" theory and because L' is finite, with a sufficiently large design $\mathcal{D}_i \subset \mathcal{V}_i$ we can carry out this identification and therefore can completely determine the value of $p(x)$ on the variety $\mathcal{V}_i$. Carrying out such a construction for each variety gives the design $\mathcal{D} = \bigcup_{i=1}^n \mathcal{D}_i$. Then the values of $p(x)$ are then completely known on each variety and the normal form over $\mathcal{V}$ recaptures $p(x)$, which completes the proof. A shorthand version is: fix a polynomial model on each $\mathcal{V}_i$ and the normal form (remainder) is fixed. The normal form of $p(x)$ with respect to $I(\mathcal{D})$ must agree with the normal forms of $p(x)$ with respect to $I(\mathcal{D}_i)$, for all i, otherwise a contradiction can be shown. This is enough to shown that $p(x)$ can be reconstructed on $\mathcal{V}$ from $\mathcal{D}$. □

This points to a sequential algorithms in which we "fix" the values on $\mathcal{V}_1$, reduce the dimension of the model as a result, fix the reduced model on $\mathcal{V}_2$ and so on. Further research is needed to turn such algorithms into a characterization of designs satisfying Conjecture 3 and minimal sample size for the existence of such designs. The following example shows heuristically how such an algorithm might work.

Example 8 Take $k = 2$ and the design $\mathcal{V}$ to be the GLD of four lines $x_1 = \pm 1, x_2 = \pm 1$. A Gröbner basis for $I(\mathcal{V})$ is $\left\{(x_1^2 - 1)(x_2^2 - 1)\right\}$ with leading term $x_1^2 x_2^2$ and

$$\begin{aligned} L(\mathcal{V}) &= \{x_2^2, x_1 x_2^2\} \otimes \{x_2^j : j \in \mathbb{Z}_{\geq 0}\} \bigcup \{x_1^2, x_1^2 x_2\} \otimes \{x_1^j : j \in \mathbb{Z}_{\geq 0}\} \\ &\quad \bigcup \{1, x_1, x_2, x_1 x_2\} \end{aligned}$$

Take the model with all terms of degree three or less, which has ten terms, see the dashed triangle on the right hand in Figure 1.1. On $x_1 = 1$ the model is cubic in x_2 so that four distinct points are enough to fix it. Thus any design with four distinct points on each line is enough. The design $\mathcal{D} = \{(\pm 1, \pm 1), (\pm 1, \pm 2), (\pm 2, \pm 1)\}$ in Figure 1.1 satisfies our needs.

1.4 Interpolation over varieties

Let $\mathcal{V} = \cup_{i=1}^n \mathcal{V}_i$ with $\mathcal{V}_i$ irreducible real affine variety and assume that the $\mathcal{V}_i$'s do not intersect i.e. $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$ for $1 \leq i < j \leq n$. Then the polynomial ideal driving an interpolation on $\mathcal{V}$ can be constructed as the intersection of the n polynomial ideals, each one driving interpolation on a separate $\mathcal{V}_i$. We discuss this approach with an example.

Let $z_1, \ldots, z_4$ be real values observed at design points $(\pm 1, \pm 1) \in \mathbb{R}^2$. Suppose we are able to observe a function over the variety defined by a circle with radius $\sqrt{3}$ and center at the origin and for simplicity, suppose

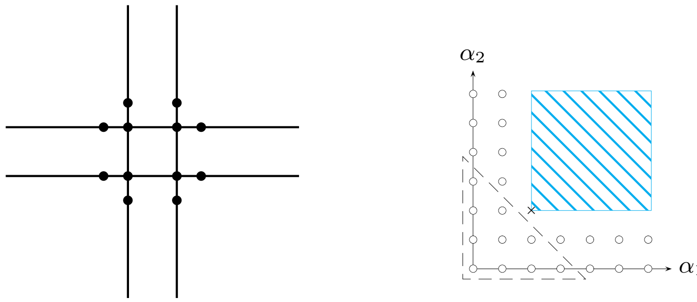

Fig. 1.1. GLDs $\mathcal{V}$ and $\mathcal{D}$ of Example 8 (left) and exponents $\alpha = (\alpha_1, \alpha_2)$ for monomials in $L(\mathcal{V})$ (right). The symbol $\times$ corresponds to the leading term $x_1^2 x_2^2$, while the shaded area contains monomials not in $L(\mathcal{V})$.

that we observe the zero function on the circle. We want a polynomial function that interpolates both the values z_i over the factorial points and takes the value zero over the circle. Note that the design $\mathcal{V}$ is the union of five varieties: one for each point, plus the circle. Start by constructing an ideal $I_i \subset \mathbb{R}[x_1, x_2, y]$ for every point d_i, e.g. $I_1 = \langle y - z_1, x_1 - 1, x_2 - 1 \rangle$. A similar approach for the circle gives: $I_C = \langle y, x_1^2 + x_2^2 - 3 \rangle$. Then intersect all the ideals $I^* = I_1 \cap \cdots \cap I_4 \cap I_C$. The ideal I^* contains all the restrictions imposed by all the varieties as well as the restrictions imposed by the observed functions. Then, for a monomial order $x^\alpha \prec y^\beta$, the desired interpolator is $\mathrm{NF}(y, I^*) \in \mathbb{R}[x_1, \ldots, x_k]$. In our current example we have $\mathrm{NF}(y, I^*) = g(x_1, x_2)(x_1^2 + x_2^2 - 3)/4$, where

$$\begin{aligned} g(x_1, x_2) \quad = \quad & -(z_1 + z_2 + z_3 + z_4) + (z_2 + z_4 - z_1 - z_3)x_1 \\ & +(z_3 + z_4 - z_1 - z_2)x_2 + (z_2 + z_3 - z_1 - z_4)x_1 x_2 \end{aligned}$$

is the interpolator for the four points, adjusted with a negative sign to compensate for the inclusion of $x_1^2 + x_2^2 - 3$. This is the standard formula appearing in books of design of experiments.

The monomial ordering used above is called a blocked ordering; for an application of such type of orders in algebraic statistics see Pistone et al. (2000). This method works well in a number of cases for which the varieties do not intersect, and when the functions defined on each variety are polynomial functions. If the varieties that compose the design intersect, then the methodology needs to ensure compatibility between the observed functions at the intersections. For example, consider again observing the zero function over the circle with radius $\sqrt{3}$; and the function $f(x_1, x_2) = 1$ over the line $x_1 + x_2 - 1 = 0$. The observed functions are not compatible at the two intersection points between the circle and the line, which is reflected on the fact that $\mathrm{NF}(y, I^*) = y \notin \mathbb{R}[x_1, x_2]$.

1.5 Becker-Weispfenning interpolation

Becker & Weispfenning (1991) define a technique for interpolation on varieties. It develops a polynomial interpolator for a set of pre-specified polynomial functions defined on a set of varieties in $\mathbb{R}^k$.

For a design variety $\mathcal{V} = \bigcup_{i=1}^n \mathcal{V}_i$ with $\mathcal{V}_i$ irreducible, the ideal of $\mathcal{V}_i$ is generated in parametric form and a pre-specified polynomial function is determined for each variety. For every variety $\mathcal{V}_i$, let $g_{i1}, \ldots, g_{ik} \in \mathbb{R}[z_1, \ldots, z_m]$ be the set of parametric generators for the ideal of the variety $I(\mathcal{V}_i)$ so that $I(\mathcal{V}_i) = \langle x_1 - g_{i1}, \ldots, x_k - g_{ik} \rangle \subset \mathbb{R}[x_1, \ldots, x_k, z_1, \ldots, z_m]$. Also, for every variety $\mathcal{V}_i$, a polynomial function $f_i(z) \in \mathbb{R}[z_1, \ldots, z_m]$ is pre-specified. Now for indeterminates $w_1, \ldots, w_n$, let I^* be the ideal generated by the set of polynomials

$$\bigcup_{i=1}^{n} \{w_i (x_1 - g_{i1}), \ldots, w_i (x_k - g_{ik})\} \bigcup \left\{ \sum_{i=1}^{n} w_i - 1 \right\} \tag{1.5}$$

We have $I^* \subset \mathbb{R}[x_1, \ldots, x_k, w_1, \ldots, w_n, z_1, \ldots, z_m]$. The technique of introducing dummy variables w_i is familiar from the specification of point ideals: when any $w_i \neq 0$ we must have $x_j - g_{ij} = 0$ for $j = 1, \ldots, k$, that is, we automatically select the i-th variety ideal. The statement $\sum_{i=1}^n w_i - 1 = 0$ prevents all the w_i being zero at the same time. If several w_i are non-zero, the corresponding intersection of $\mathcal{V}_i$ is active. Consistency of the parametrization is, as Becker and Weispfenning (1991) point out, a necessary, but not sufficient, condition for the method to work.

Let $\prec$ be a block monomial order for which $x^\alpha \prec w^\beta z^\gamma$. Set $f^* = \sum_{i=1}^m w_i f_i(z)$ and let $f' = \mathrm{NF}(f^*, I^*)$. The interpolation problem has a solution if the normal form of f^* depends only on x, that is if $f' \in \mathbb{R}[x_1, \ldots, x_k]$. Although the solution does not always exist, an advantage of the approach is the freedom to parametrise each variety separately from a functional point of view, but using a common parameter z.

Example 9 (Becker & Weispfenning 1991, Example 3.1) We consider interpolation over $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \mathcal{V}_3 \subset \mathbb{R}^2$ The first variety is the parabola $x_2 = x_1^2 + 1$, defined through the parameter z by $g_{11} = z, g_{12} = z^2 + 1$.

The second and third varieties are the axes x_1 and x_2 and therefore $g_{21} = z, g_{22} = 0$ and $g_{31} = 0, g_{32} = z$. The prescribed functions over the varieties are $f_1 = z^2, f_2 = 1$ and $f_3 = z + 1$. The ideal I^* is constructed using the set in Equation (1.5) and we set $f^* = w_1 f_1 + w_2 f_2 + w_3 f_3$. For a block lexicographic monomial order $\prec$ in which $x^\alpha \prec w^\beta z^\gamma$, we compute the normal form of f^* with respect to I^* and obtain $f' = x_2 + 1$.

1.6 Reduction of power series by ideals

Let us revisit the basic theory. Here $x = (x_1, \ldots, x_k)$. A polynomial $f \in \mathbb{R}[x]$ can be reduced by the ideal $I(\mathcal{V}) \subset \mathbb{R}[x]$ to an equivalent polynomial f' such that $f = f'$ on the affine variety $\mathcal{V}$. By Theorem 1, the reduced expression is $f' = \mathrm{NF}(f, \mathcal{V})$ and clearly $f - f' \in I(\mathcal{V})$.

Example 10 Consider the hyperplane arrangement $\mathcal{V}$ given by the lines $x_1 = x_2$ and $x_1 = -x_2$. We have $I(\mathcal{V}) = \langle x_1^2 - x_2^2 \rangle$. Now for $i = 1, 2, \ldots$, consider the polynomial $f_i = (x_1 + x_2)^i$. For a monomial ordering in which $x_2 \prec x_1$, we have that $\mathrm{NF}(f_i, \mathcal{V}) = 2^{i-1}(x_1 + x_2)x_2^{i-1}$, for instance $\mathrm{NF}((x_1 + x_2)^5, \mathcal{V}) = 16(x_1 + x_2)x_2^4 = 16x_1x_2^4 + 16x_2^5$.

A convergent series of the form

$$f(x) = \sum_{i=0}^{\infty} \alpha_i x^{\alpha_i},$$

can be written on the variety $\mathcal{V}$ as

$$\mathrm{NF}(f, \mathcal{V}) = \sum_{i=0}^{\infty} \alpha_i \mathrm{NF}(x^{\alpha_i}, \mathcal{V}). \tag{1.6}$$

See Apel et al. (1996) for a discussion of conditions for the validity of Equation (1.6).

We may also take the normal form of convergent power series with respect to the ideal of an affine variety in $\mathbb{C}$. For example by substituting $x^3 = 1$ in the expansion for e^x we obtain

$$\begin{aligned} \mathrm{NF}(e^x, \langle x^3 - 1 \rangle) &= 1 + \frac{1}{3!} + \frac{1}{6!} + \frac{1}{9!} + \ldots + x\left(1 + \frac{1}{4!} + \frac{1}{7!} + \frac{1}{10!} + \ldots\right) \\ &\quad + x^2\left(\frac{1}{2!} + \frac{1}{5!} + \frac{1}{8!} + \ldots\right) \\ &= \frac{1}{3}e + \frac{2}{3}e^{-\frac{1}{2}} \cos\left(\frac{\sqrt{3}}{2}\right) \end{aligned}$$

$$+ x\left(\frac{1}{3}e - \frac{1}{3}e^{-\frac{1}{2}}\cos\left(\frac{\sqrt{3}}{2}\right) + \frac{1}{3}e^{\frac{1}{2}}\sin\left(\frac{\sqrt{3}}{2}\right)\right)$$
$$+ x^2\left(\frac{1}{3}e - \frac{1}{3}e^{-\frac{1}{2}}\cos\left(\frac{\sqrt{3}}{2}\right) - \frac{1}{3}e^{\frac{1}{2}}\sin\left(\frac{\sqrt{3}}{2}\right)\right)$$

The relation $\mathrm{NF}(e^x, \langle x^3-1\rangle) = e^x$ holds at the roots d_1, d_2, d_3 of $x^3-1=0$, with d_1 the only real root. Note that the above series is not the same as the Taylor expansion at, say, 0.

Example 11 Consider the ideal $I = \langle x_1^3 + x_2^3 - 3x_1x_2\rangle$. The variety $\mathcal{V}$ that corresponds to I is the Descartes' *folium*. For a monomial ordering in which $x_2 \prec x_1$, the leading term of the ideal is x_1^3. Now consider the function $f(x) = \sin(x_1+x_2)$, whose Taylor expansion is

$$f(x) \quad = \quad (x_1+x_2) - \frac{1}{3!}(x_1+x_2)^3 + \frac{1}{5!}(x_1+x_2)^5 + \ldots \qquad (1.7)$$

The coefficients for every term of Equation (1.7) which is divisible by x_1^3 is absorbed into the coefficient of some of the monomials in $L(\mathcal{V})$. For the second term in the summation we have the following remainder

$$\mathrm{NF}\left(-\frac{(x_1+x_2)^3}{3!}, \mathcal{V}\right) = -\frac{1}{2}\left(x_1^2x_2 + x_1x_2^2 + x_1x_2\right).$$

Note that different terms of the Taylor series may have normal forms with common terms. For instance the normal form for the third term in the summation is

$$\mathrm{NF}\left(\frac{(x_1+x_2)^5}{5!}, \mathcal{V}\right) = \frac{3}{40}x_1^2x_2^3 - \frac{3}{40}x_2^5 + \frac{1}{8}x_1^2x_2^2 + \frac{1}{4}x_1x_2^3 - \frac{1}{40}x_2^4 + \frac{3}{40}x_1x_2^2.$$

The sum of the normal forms for first ten terms of Equation (1.7) is

$$\begin{aligned}
\tilde{f}(x) \quad = \quad & x_2 + x_1 - \frac{1}{2}x_1x_2 - \frac{17}{40}x_1x_2^2 - \frac{1}{2}x_1^2x_2 - \frac{1}{40}x_2^4 + \frac{137}{560}x_1x_2^3 \\
& + \frac{1}{8}x_1^2x_2^2 - \frac{41}{560}x_2^5 - \frac{167}{4480}x_1x_2^4 + \frac{1}{16}x_1^2x_2^3 + \frac{167}{13440}x_2^6 \\
& - \frac{4843}{492800}x_1x_2^5 - \frac{17}{896}x_1^2x_2^4 + \frac{2201}{492800}x_2^7 + \frac{197343}{25625600}x_1x_2^6 \\
& + \frac{89}{44800}x_1^2x_2^5 - \frac{65783}{76876800}x_2^8 - \frac{4628269}{5381376000}x_1x_2^7 + \frac{1999}{5913600}x_1^2x_2^6 \\
& + \frac{118301}{1793792000}x_2^9 - \frac{305525333}{1463734272000}x_1x_2^8 - \frac{308387}{1076275200}x_1^2x_2^7 + \ldots
\end{aligned}$$

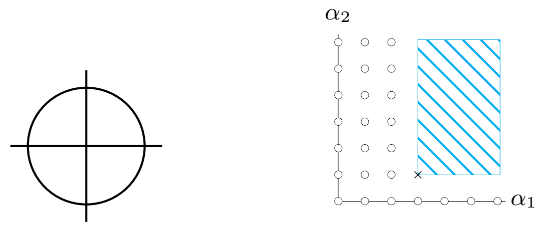

Fig. 1.2. Variety for the ideal $\langle x_1x_2(x_1^2+x_2^2-2)\rangle$ (left) and exponents $\alpha=(\alpha_1,\alpha_2)$ for monomials in $L(\mathcal{V})$ (right). The symbol $\times$ in the right diagram corresponds to the leading term $x_1^3x_2$, while the shaded area contains monomials not in $L(\mathcal{V})$.

The equality $\tilde{f}(x)=\sin(x_1+x_2)$ is achieved over $\mathcal{V}$ by summing the normal forms for all terms in Equation (1.7): $\tilde{f}(x)$ interpolates $\sin(x_1+x_2)$ over $\mathcal{V}$.

1.7 Discussion and further work

In this paper we consider the extension of the theory of interpolation over points to interpolation over varieties with in mind applications to design of experiments in statistics. We associate to the design variety a radical ideal and the quotient ring induced by this variety ideal is a useful source of terms which can be used to form the basis for a (regression) model. In particular, knowledge of the quotient ring for the whole variety can be a useful guide to models which can be identified with a set of points selected from the variety.

If the design variety is not a GLD, the technique still can be applied. As an example consider the structure $\mathcal{V}$ consisting of a circle with a cross, see Figure 1.2. For any monomial ordering, the polynomial $g=x_1x_2(x_1^2+x_2^2-2)=x_1^3x_2+x_1x_2^3-2x_1x_2$ is a Gröbner basis for $I(\mathcal{V})$. Now, for a monomial order in which $x_2\prec x_1$, we have $\mathrm{LT}_{\prec}(g)=x_1^3x_2$ and $L(\mathcal{D})=\{x_2,x_1x_2,x_1^2x_2\}\otimes\{x_2^j : j\in\mathbb{Z}_{\geq 0}\}\bigcup\{x_1^{3+j} : j\in\mathbb{Z}_{\geq 0}\}\bigcup\{1,x_1,x_1^2\}$ see Figure 1.2. If we are interested in $L'=\{1,x_1,x_2,x_1^2,x_1x_2,x_2^2\}$ then a good subset of $\mathcal{V}$ which estimates L' is $\mathcal{D}=\{(\pm1,\pm1)\}\cup\{(0,\pm\sqrt{2}),(\pm\sqrt{2},0)\}\cup\{(0,0)\}$. This is the classic central composite design of response surface methodology.

We have not discussed the issue of statistical variation in interpolation, that is, when observations come with error. In the case of selecting points from $\mathcal{V}$ of Section 1.3, standard models can be used, but when an observation is a whole function as in Sections 1.4 and 1.5, a full statistical theory awaits development. It is likely that such a theory would involve random functions, that is stochastic processes on each variety $\mathcal{V}_i$.

Finally, we note that elsewhere in this volume there is emphasis on probability models defined on discrete sets. Typically the set may be a product set which allows independence and conditional independence statements. A simple approach but with deep consequences is to consider not interpolation of data (y-values) in a variety, but $\log p$ where p is a probability. It is a challenge, therefore, to consider $\log p$ models on varieties, that is, distributions on varieties. One may count occurrences rather than observe real continuous y-values. With counts we may be able to reconstruct a distribution on the transect as in Example 1. Again the issue would be to reconstruct the full distribution both on and off the transect. This points to a theory of exponential families anchored by prescribing the value in varieties. We trust that the development of such a theory would be in the spirit of this volume and the very valuable work of its dedicatee.

Acknowledgements

The authors acknowledge the EPSRC grant EP/D048893/1, considerable help from referees and an early conversation with Professor V. Weispfenning.

Bibliography

Apel, J., Stückrad, J., Tworzewski, P. & Winiarski, T. (1996), 'Reduction of everywhere convergent power series with respect to Gröbner bases', *J. Pure Appl. Algebra* **110**(2), 113–129.

Becker, T. & Weispfenning, V. (1991), The chinese remainder problem, multivariate interpolation, and Gröbner bases, *in* 'ISSAC '91: Proceedings of the 1991 international symposium on Symbolic and algebraic computation', pp. 64–69.

Buckland, S. T., Anderson, D. R., Burnham, K. P. & Laake, J. L. (1993), *Distance sampling*, Chapman & Hall, London. Estimating abundance of biological populations, With a foreword by G. A. F. Seber.

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

Fontana, R., Pistone, G. & Rogantin, M. (1997), 'Algebraic analysis and generation of two level designs', *Statist. Appl.* **9**(1), 15–29.

Giglio, B., Wynn, H. & Riccomagno, E. (2001), Gröbner basis methods in mixture experiments and generalisations, *in* 'Optimum design 2000 (Cardiff)', Vol. 51 of *Nonconvex Optim. Appl.*, Kluwer Acad. Publ., Dordrecht, pp. 33–44.

Grünbaum, B. (2003), *Convex polytopes*, Vol. 221 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York. Prepared and with a preface by V. Kaibel, V. Klee and G. M. Ziegler.

Holliday, T., Pistone, G., Riccomagno, E. & Wynn, H. (1999), 'The application of computational algebraic geometry to the analysis of designed experiments: a case study', *Comput. Statist.* **14**(2), 213–231.

Kreuzer, M. & Robbiano, L. (2000), *Computational commutative algebra. 1*, Springer-Verlag, Berlin.

Kreuzer, M. & Robbiano, L. (2005), *Computational commutative algebra. 2*, Springer-Verlag, Berlin.

Mack, Y. P. & Quang, P. X. (1998), 'Kernel methods in line and point transect sampling', *Biometrics* **54**(2), 606–619.

Maruri-Aguilar, H., Notari, R. & Riccomagno, E. (2007), 'On the description and identifiability analysis of mixture designs', *Statist. Sinica* **17**(4), 1417–1440.

Phillips, G. M. (2003), *Interpolation and approximation by polynomials*, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, 14, Springer-Verlag, New York.

Pistone, G., Riccomagno, E. & Rogantin, M. (2006), 'Algebraic statistics methods in DOE (with a contribution by Maruri-Aguilar, H.)'. (Forthcoming).

Pistone, G., Riccomagno, E. & Wynn, H. (2000), 'Gröbner basis methods for structuring and analysing complex industrial experiments', *Int. J. Rel. Qual. Saf. Eng.* **7**(4), 285–300.

Pistone, G., Riccomagno, E. & Wynn, H. P. (2001), *Algebraic Statistics*, Vol. 89 of *Monographs on Statistics and Applied Probability*, Chapman & Hall/CRC, Boca Raton.

Pistone, G. & Rogantin, M. (2008), 'Indicator function and complex coding for mixed fractional factorial designs', *JSPI* **138**, 787–802.

Pistone, G. & Wynn, H. (1996), 'Generalised confounding with Gröbner bases', *Biometrika* **83**(3), 653–666.

Sakhnovich, L. A. (1997), *Interpolation theory and its applications*, Vol. 428 of *Mathematics and its Applications*, Kluwer Academic Publishers, Dordrecht.

Stanley, R. P. (1978), 'Hilbert functions of graded algebras', *Adv. Math.* **28**(1), 57–83.

Stanley, R. P. (1996), 'Hyperplane arrangements, interval orders, and trees', *Proc. Nat. Acad. Sci. U.S.A.* **93**(6), 2620–2625.

Stein, M. L. (1999), *Interpolation of spatial data*, Springer Series in Statistics, Springer-Verlag, New York.

Sturmfels, B. & White, N. (1991), 'Computing combinatorial decompositions of rings', *Combinatorica* **11**(3), 275–293.

Whitney, H. (1957), 'Elementary structure of real algebraic varieties', *Ann. Math. (2)* **66**, 545–556.