A course on Algebraic Statistics

Hugo Maruri-Aguilar, Eva Riccomagno, Henry P Wynn

1 Algebraic geometry

1.1 Introduction

It is useful to start with some basic notation. Consider a set of d indeterminates:

$$x = (x_1, \ldots, x_k)$$

For a set of non-negative integers $\alpha = (\alpha_1, \ldots, \alpha_n)$ we define a *monomial*

$$x^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_k^{\alpha_k}$$

We should note at this early stage that any monomial x^α can be represented by its exponent α and it often useful to think in terms of the non-negative integer grid, in discussing monomials. For example, the *total degree* of a monomial x^α is $|\alpha| = \sum_{i=1}^d \alpha_i$. In statistics we are familiar with monomials as linear, quadratic, interaction etc terms in regression functions:

$$x_1, x_2^2, x_1 x_2, \ldots$$

By taking linear combinations of monmials with coefficients in a base field K we obtain a ring of polynomials, $R = k[x_1, \ldots, x_k]$ over K. We can write a polynomial compactly as

$$f(x) = \sum_{\alpha \in M} \theta_\alpha x^\alpha,$$

where M is a set of distinct exponents. For example the standard quadratic *response surface* in two variable is:

$$f(x_1, x_2) = \theta_{00} + \theta_{10} x_1 + \theta_{01} x_2 + \theta_{20} x_1^2 + \theta_{11} x_1 x_2 + \theta_{02} x_2^2,$$

and

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

1.2 Ideals

For a ring R we have special subsets called *ideals*.

Definition 1 *A subset $I \subset R$ is an ideal if for any $f, g \in I$ we have $f+g \in I$ and for any $f \in I$ and $g \in R$ we have $fg \in I$.*

Starting with a set of polynomials $\{f_1, \ldots, f_m\}$ we can form the *ideal generated by* $\{f_1, \ldots, f_m\}$ the set of all polynomial combination:

$$\langle f_1, \ldots, f_m \rangle = \{f_1 g_1 + \cdots + f_m g_m : \ g_1, \ldots, g_m \in R\}$$

The Hilbert basis theorem says that any ideal I is finitely generated: for any ideal we can find a finite collection $f_1, \ldots, f_k \in R$ such that $I = \langle f_1, \ldots, f_k \rangle$.

An important operation in ring theory is that of a *quotient*. For two polynomials $f, g \in K[x_1, \ldots, x_n]$ and an ideal I define the equivalence class $f \sim_I g$ if and only if $f - g \in I$. The objects in the quotient $K[x_1, \ldots x_n]/I$ are the equivalence classes. Since $f_1 \sim_I f_2$ and $g_1 \sim_I g_2$ imply $f_1 + g_1 \sim_I f_1 + g_2$ and $f_1 g_1 \sim_I f_2 g_2$, $K[x_1, \ldots x_n]/I$ is also a ring. Finding $K[x_1, \ldots x_n]/I$ in particular case requires a *division algorithm*. One of the breakthroughs was to find a division algorithm which was amenable to symbolic computation.

1.3 Varieties

In statistics we often work with real curves and surfaces. In this case we should begin to think of the indeterminates $x_1, \ldots, x_k$ as real variables, for which we have many uses. Two uses are important: (i) variables as *factors* as in the design of experiments (or statistical modeling more generally) and (ii) *probabilities*.

Much more, perhaps, than is usually recognised in statistics we often deal with polynomial equations. In ordinary regression we may have equations saying that the mean of the dependent variable is a polynomial function of the independent variable and we shall see below that an experimental design can be considered as a solution of a set of equations. In probability models conditions such as independence and conditional independence are expressed via polynomial equations.

Definition 2 *Let $f_1, \ldots, f_m \in K[x_1, \ldots, x_k]$ be a set of polynomials. The affine variety is the solution of a set of simultaneous equations:*

$$V(f_1, \ldots, f_n) = \{(a_1, \ldots, a_k) \in K^k : f_i(a_1, \ldots, a_d) = 0, \ i = 1, \ldots, m\}$$

Every affine variety has an associated ideal which we write $I(V)$. It is the set of *all* polynomial which are zero on the variety:

$$I(V) = \{f \in K[x_1, \ldots, x_k] : f(a_1, \ldots, a_k) = 0, \text{ for all } (a_1, \ldots, a_k) \in V\}.$$

A difficult question arises: if we (i) start with polynomials $f_1, \ldots, f_m$ and (ii) construct the corresponding variety V (iii) form the ideal $I(V)$, is it true that $I(V) = \langle f_1, \ldots, f_m \rangle$. We can claim that $I(V) = \langle f_1, \ldots, f_m \rangle \subset I(V)$, but that the converse may not be true will be covered in section (1.6). However, for one important case that is when the variety is collection of isolated single points the equality does hold, see section (2.3).

1.4 Division and Gröbner bases

One might recall elementary division for polynomials in one variable. In that case the usual algorithm benefits from the fact that monomials in one variable are totally ordered by degree:

$$1 \prec x \prec x^2 \prec x^3 \prec \ldots$$

This is generalised to a special total ordering on monomials $\{x^\alpha\}$.

Definition 3 *A* monomial term ordering τ *is a total ordering of monomials* $\prec$ *such that* $1 \prec x^\alpha$ *for all* $\alpha \geq 0$ *and for all* $\gamma \geq 0$

$$x^\alpha \prec_\tau x^\beta$$

implies

$$x^{\alpha+\gamma} \prec_\tau x^{\beta+\gamma}$$

We shall use the term *monomial ordering* for short and quietly drop τ unless we need to emphasize a special monomial ordering.

There are, indeed, a number of standard monomials orderings.

1. *Lexicographic ordering, lex.* $\alpha \prec_{lex} \beta$ when (i) $\beta - \alpha \geq 0$ and the leftmost entry of $\beta - \alpha$ is positive.

2. *Graded lexicographic ordering, grlex*: $\alpha \prec_{grlex} \beta$ if (i) the degree of α is less than that of β, $|\alpha| \langle |\beta|$ and (ii) $\alpha \prec_{lex} \beta$

3. *Reverse lexicographic ordering, grevlex*: (i) $|\alpha| \langle |\beta|$ and (ii) $\overline{\alpha} \prec_{lex} \overline{\beta}$, where the overline means: reverse the entries.

Given an monomial ordering, because it is by definition a total ordering, any set of monomials has a leading term. In particular, since a polynomial,

f, uses a finite set of monomials it has a unique leading term and we write $LT_{\prec}(f)$, or, if $\prec$ is assume, just $LT(f)$.

In any division algorithm in more than one dimension it seems a good guess that monomial orderings will play an important part, by analogy with their role in the one dimensional algorithm. This is indeed the case. In fact to quite a large extent monomials drive the whole theory.

Definition 4 *A monomial ideal I is one such that there is a collection of monomials $f_1, \ldots, f_m$ such that any $g \in I$ can be expressed as a sum*

$$g = \sum_{i=1}^{m} g_i(x) f_i(x).$$

Now, we can appeal to the representation a monomial x^α by its exponent α. If $\beta \geq 0$ is another exponent then

$$x^\alpha x^\beta = x^{\alpha+\beta},$$

and $\alpha + \beta$ is in the positive (shorthand for non-negative) "orthant" with corner at α. The set of all monomials in a monomial ideal is the union of all positive orthants whose corners are given by the exponent vectors of the generating monomial $f_1, \ldots, f_m$.

Dickson's Lemma states that, even if we define a monomial ideal with an infinite set of f_i, we can find a finite set $h_1, \ldots h_n$ such that $I = \langle h_1, \ldots, h_k \rangle$. The way to see this intuitively is that the lower boundary of the union of all positive orthants with corners at the original f_i, although possibly very complex, does not need to employ more than a finite number of f_i as generators.

There are, in general, many ways to express a given ideal I as being generated from a basis $I = \langle f_1, \ldots, f_m \rangle$. That is to say there are many choices of basis.

Definition 5 *Given an ideal I a set $\{g_1, \ldots g_m\}$ is called a Gröbner basis if:*

$$\langle LT(g_1), \ldots, LT(g_m) \rangle \ = \ \langle LT(I) \rangle,$$

where $\langle LT(I) \rangle$ is the ideal generated by all the monomials in I

Lemma 6 *Any ideal I has a basis which is a Gröbner basis and any Gröbner basis in the ideal is a basis of the ideal.*

Monomial orderings are critical in establishing that for any given monomial ordering $\prec$ any ideal I has a unique "reduced" associated Gröbner basis. The basic idea is that given a monomial ordering and an ideal expressed in terms of the G-basis, $I = \langle g_1, \ldots, g_m \rangle$, any polynomial f has a unique remainder with respect the quotient operation $K[x_1, \ldots, x_k]/I$. That is

$$f = \sum_{i=1}^{m} s_i(x) g_i(x) + r(x)$$

We call the remainder $r(x)$ the *normal form* of f with respect to I and write $NF(f)$. Here are some formal definitions.

Definition 7 *Given a monomial ordering $\prec$, a polynomial $f = \sum_{\alpha \in L} \theta_\alpha x^\alpha$ is a normal form with respect to $\prec$ if $x^\alpha \notin LT(f)$ for all $\alpha \in L$.*

Lemma 8 *Given an ideal I and a monomial ordering $\prec$, for every $f \in K[x_1, \ldots, x_k]$ there is a unique normal form $NF(f)$ such that $f - NF(f) \in I$.*

We now need to relate (i) the Gröbner basis (ii) a division algorithm and (iii) the nature of the normal form. We have partly covered this but let us collect the results together.

1. There are algorithms, which given an ideal, I, a monomial ordering $\prec$ and a polynomial f delivers the remainder r, in Lemma (8), by successively dividing by the G-basis terms $g_i,\ i = 1, \ldots, m$. The best know algorithm is the Buchberger algorithm.

2. Suppose the remainder $r(x) = NF(f) = \sum_{\alpha \in L} \theta_{\alpha \in L} x^\alpha$, then no monomial $x^\alpha, \alpha \in L$ is divisible by any of the leading terms of the G-basis of $I:\ \mathrm{LT}(g_i),\ i = 1, \ldots, m$.

3. The remainder $r(x) = NF(f)$ does not depend on which order the G-basis terms $g_i(x)$ are used in the division algorithm.

4. The largest possible set terms $\{x^\alpha, \alpha \in L\}$, which can appear in a remainder $r(x)$ is a basis of the quotient ring, considered as a vector space of function over $k[x_1, \ldots, x_k]/I$. Moreover the terms are linearly independent over I:
$$\sum_{\alpha} \theta_\alpha x^\alpha \sim_I 0$$
implies $\theta_\alpha = 0$ for all $\alpha \in L$.

1.5 Elimination

If there is one operation which demonstrates the usefulness of computer algebra based on Gröbner bases and related methods it is *elimination*. For linear algebra we are familiar with Gaussian elimination. In that case we start with d linear equations in d variables and try to reduce the set to (upper) triangular form. If we are able to do this then, for $1 \leq s \langle d$ the last $d-s$ variables will appear in the last $d-s$ equations and we have eliminated s variables. If $s = d-1$ we can solve for x_d and by successive back substitution up the triangular form obtain a full solution.

Remarkably, we have an analogue for polynomial equations. We work with ideals. The start is the notion of an *elimination ideal*. Let $I \subset K[x_1, \ldots, x_n]$ be an ideal.

Definition 9 *Let* $I = \langle f_1, \ldots, f_m \rangle \subset K[x_1, \ldots, x_d]$. *Select variables* $x_1, \ldots, x_s$. *Then the elimination ideal is*

$$I_s = I \cap K[x_{s+1}, \ldots, x_k]$$

We can eliminate using the lex monomial ordering. The rationale is that since the lex order "prefers" variable which are low in the "initial" order it will place these first in the G-basis.

Theorem 10 *Let* $I \in K[x_1, \ldots, x_k]$ *be an ideal and let* $G = \langle g_1, \ldots, g_m \rangle$ *be the G-basis with respect to the lex ordering with initial ordering*

$$x_k \prec \cdots \prec x_1.$$

then

$$G_s = G \cap k[x_{s+1}, \ldots, x_n]$$

is a G-basis of the elimination ideal for $x_s, \ldots, x_k$.

Consider the equation:

$$\begin{aligned} x_1x_2 + x_1x_3 + x_2x_3 &= 1 \\ 2x_1x_2 + x_1x_3 + 3x_2x_3 &= 5 \\ x_1x_2 + 2x_1x_3 - 3x_2x_3 &= 2 \end{aligned}$$

The elimination ideal is

$$\langle 3x_3^2 - 7, 7x_2 + 12x_3, x_1 + 3x_3 \rangle$$

Now by back substitution starting with the solutions to the first equation we obtain all the solutions: $\left(-\frac{1}{3}\sqrt{\frac{7}{3}}, -\sqrt{\frac{12}{7}}, \sqrt{\frac{1}{3}}\right)$, $\left(\frac{12}{7}\sqrt{\frac{7}{3}}, -\sqrt{\frac{7}{3}}, -\sqrt{\frac{7}{3}}\right)$.

1.6 More on ideals and varieties

The relationship between varieties and ideals depends subtly on what field we are operating over. It is not true for a general field that $I(V(f_1, \ldots, f_m)) = \langle f_1, \ldots, f_m \rangle$. For an algebraically closed field there are some fundamental results. The first shows that if the field is closed the ideal of any non-zero variety is the whole ring.

Theorem 11 *(Hilbert's Weak Nullstellensatz) For an algebraically closed field K and $I \subset K[x_1, \ldots, x_k]$ an ideal, then $V(I) =$ if and only if $I = K[x_1, \ldots, x_k]$.*

The next result, Hilbert's Strong Nullstellensatz, says that some power of a member of the variety ideal must be in the generated ideal.

Theorem 12 *Let k be an algebraically closed field. Then for polynomials f and $f_1, \ldots, f_m$ in $k[x_1, \ldots, x_d]$ with $f \in I(V(f_1, \ldots, f_m)$, there is exist an integer $r \geq 1$ such that $f^r \in \langle f_1, \ldots, f_d \rangle$.*

We list some key points about the relationship between ideals and varieties.

1. If $f^r \in I(V)$ then $f \in I(V)$.
2. An ideal is radical if $f^r \in I$ for any integer $f \geq 1$ we have $f \in I$. Hence $I(V)$ is radical.
3. The *radical* of I, written $\sqrt{I}$ is the set of f such that $f^r \in I$ for some integer $r \geq 1$.
4. In general: $I_1 \subset I_2$ implies $V(I_2) \subset V(I_1)$. Also $V_1 \subset V_2$ implies $I(V_1) \supset I(V_2)$. For any variety V, $V(I(V)) = V$.
5. If we restrict to algebraically closed fields k and *radical* ideals, then we have a bijection which send $\subset$ to $\supset$ and vice-versa.
6. There are two related ideas, *prime* ideals and *irreducible* varieties. Working over $K[x_1, \ldots, x_k]$ a ideal I is prime if $fg \in I$ implies $f \in I$ or $g \in I$. A variety is irreducible if $V = V_1 \cup V_2$ implies $V_1 = V$ or $V_2 = V$. A result is that if V is an affine variety it is irreducible if and only if $I(V)$ is prime. When the field is algebraically closed prime ideals and irreducible varieties are in $1 - 1$ correspondence.

Despite the difficulty of working over the real field one should get used to taking unions, sums and intersections.

1. The sum of two ideals: $I+J=\{f+g,\ f\in Ig\in J\}$. We simply adjoin the two bases
$$I+J=\langle f_1,\ldots,f_m,g_1,\ldots,g_n\rangle$$
The corresponding variety $V(I+J)$ is the full set of solutions to
$$f_1=\cdots=f_m=g_1=\cdots=g_n=0$$
and $V(I+J)=V(I)\cap V(J)$.
2. $V(f_1,\ldots,f_m)\cup V(g_1,\ldots,g_m)=V(f_ig_j,\ i=1,\ldots,m;\ j=1,\ldots,n)$.

2 Experimental design

2.1 Introduction

In these notes an experimental design will be a set of n points in Q^k at each of which a single observation is taken. In statistics this is called the *single replicate case*. We are interested in the relationship between the designs and the potential polynomial regression models may one fit to data at the design points.

The history of experimental design is rich in algebraic and combinatorial theory and a large number of special structures have been studied. Perhaps the most common is a *full factorial design* which is simply a product space $\mathcal{N} = N_{n_1} \times \cdots \times N_{n_d}$ where

$$N_j = \{0, \ldots, n_j - 1\},\ j = 1, \ldots, d.$$

A subset of a given design is sometimes called a *fraction*. But another type of design, sometimes called a *response surface design* has some features of a fraction but may have a more adventurous structure. Another area of design is so-called combinatorial design, more usual in the context of qualitative factors. There is a long list: *balance incomplete block design (BIBD), Latin Squares* and so on.

2.2 Regression and the X-matrix

Classical linear regression can be described by a model for an output $Y(x)$ as linear combination of a basis of functions $\{f_j(x)\}_{j=1,\ldots,m}$ of an independent variable x, typically in R^k:

$$Y_x = \sum_{j=1}^{m} \theta_j f_j(x) + \epsilon_x, \tag{1}$$

where ϵ_x is random error. An experimental design is a set of distinct points $D = \{x^{(1)}, \ldots, x^{(n)}\}$, of size n, the *sample size*. At each design point $x^{(i)}$ we take and observation Y_i. It should be noted that where we have qualitative factors, as opposed to quantitative or ordinal factors, we can still use real variables by converting to indicator variables eg $x_{ij} = 1$ if we are at the j-th level of factor i and $x_{ij} = 0$ otherwise.

We may write the statistical regression model which relates the response Y to the factors x, as

$$Y_i = \sum_{j=1}^{k} \theta_j f_j(x^{(i)}) + \epsilon_i,\ i = 1, \ldots, n$$

The X-matrix is the matrix with observations indexing the rows and functions indexing the columns:

$$X = \{f_j(x^{(i)})\}$$

In matrix terms with obvious notation:

$$Y = X\theta + \varepsilon \tag{2}$$

The standard second order assumptions are that the errors have mean zero $\mathrm{E}(\varepsilon) = 0$ and covariance $\mathrm{cov}(\varepsilon) = \sigma^2 I_{n\times n}$, where $\sigma^2\rangle 0$ is then error variance. The standard distributional assumption are that ε is multivariate normal (Gaussian).

In the algebraic theory the basis is typically monomial so that

$$f_j(x) = m_j(x)$$

and then

$$X = \{m_j(x^{(i)})\}$$

Another important matrix is the information matrix $X^T X$. The following is standard.

1. A least squares estimator of the parameter vector θ is

$$\hat{\theta} = \arg\min_{\theta \in R^k} ||Y - X\theta||^2, \tag{3}$$

   where Y is the vector of observations, and if X has full rank,

$$\hat{\theta} = (X^T X)^{-1} X^T Y$$

   We shall assume X is full rank from now on.

2. The covariance matrix of the estimator $\hat{\theta}$ is

$$\mathrm{cov}(\hat{\theta}) = \sigma^2 (X^T X)^{-1}$$

3. The matrix $P = X(X^TX)^{-1}X^T$ is the symmetric idempotent projector onto the column space (range) of the X-matrix and we have a partition of the identity

$$I = P + I - P$$

which give the basic decomposition of $||Y||^2$ arising out of the least squares operation:

$$||Y||^2 = Y^TPY + Y^T(I - P)Y$$

Thus $Y^TPY = ||\hat{Y}||^2$ and $Y^T(I - P)Y = ||Y - \hat{Y}||^2$ where

$$\hat{Y} = PY = X\hat{\theta}$$

is the vector of predicted values and $R = Y - \hat{Y} = (I - P)Y$ is the vector of residuals. This decomposition is a simple version of what is called the *analysis of variance (ANOVA)* in statistics and $||Y||^2, ||\hat{Y}||^2$ and $||R||^2 = ||Y - \hat{Y}||^2$ are respectively the total sum of squares, the regression sum of squares and the residual sum of squares.

If $n = k$ so that X is square we have exact interpolation and

$$\hat{\theta} = X^{-1}Y,$$

and trivially $\hat{Y} = Y$. Another case is where the number of *distinct* locations for the x_i is k. In that case we might re-index the Y_i as Y_{ij} where i runs over the distinct locations and j runs over the *replications* at each location. In that case, when at every design point we have the same number of replications, $\hat{\theta} = X^{-1}[\bar{Y}]$ where $[\bar{Y}] = (\bar{Y}_1, \ldots, \bar{Y}_n)^T$ and $\bar{Y}_i$ average of the Y_{ij} values at the design point $x^{(i)}$.

2.3 Ideals of points and saturated interpolators

A main contribution of the algebra is to obtain a saturated basis for an arbitrary design.

1. A design is a finite set of distinct points, D, in R^d (Q^d) and can be expressed as the solution of a set of equations and can be thought of as a zero dimensional variety. The set of all polynomials with zeros on a D is the ideal, $I(D)$.

2. There is a Gröbner basis for $I(D)$ for a given monomial ordering.

3. The quotient ring
$$K[x_1, \ldots, x_k]/I(D)$$
of the ring of polynomials $K[x_1, \ldots, x_k]$ in $x_1, \ldots, x_k$ forms is a vector space spanned by a special set of monomials: $x^\alpha, \alpha \in L$. These are all the monomials not divisible by the leading terms of the G-basis and $|L| = |D|$.

4. The set of multi-indices L has the "order ideal" property: $\alpha \in L$ implies $\beta \in L$ for any $0 \leq \beta \leq \alpha$. For example, if $x_1^2 x_2$ in the model so is $1, x_1, x_2, x_1 x_2$.

5. Any function $y(x)$ on D has a unique polynomial interpolator given by
$$f(x) = \sum_{\alpha \in L} \theta_\alpha x^\alpha$$
such that $y(x) = f(x),\ x \in D$.

6. The X-matrix is $n \times n$, has rank n and has rows indexed by the design points and columns indexed by the basis:
$$X = \{x^\alpha_{x \in D, \alpha \in L}\}$$

It is an exercise to construct the ideal of a design, before we computer the G-basis. Consider a single point $(z_1, \ldots, z_d)$. The ideal for this point is clearly
$$I_z = \langle x_1 - z_1, \ldots, x_d - z_d \rangle.$$
A design D is a union of points $z \in D$ so the design ideal is the intersection of the point ideals
$$I(D) = \cap_{z \in D} I_z.$$
So, if we have an algorithm to compute intersections of ideals, we are done.

There is a "by hand" way of the computing intersection of ideals using elimination.

Theorem 13 *Let I and J be ideals in $K[x_1, \ldots, x_k]$. Introduce and extra variable t. Then*
$$I \cap J = (tI + (1-t)J) \cap K[x_1, \ldots, x_k]$$

Here $tI = \{tf : f \in I\}$ and $\cap k[x_1, \ldots, x_n]$ means the ideal formed by eliminating t.

Here is a small example. Let $D = \{(1,1),(3,2)\}$. Then take the two point ideals

$$I = \langle x_1 - 1, x_2 - 1\rangle,\ J = \langle x_1 - 3, x_2 - 2\rangle,$$

and eliminate t from

$$tI + (1-t)J = \langle t(x_1-1), t(x_2-1), (1-t)(x_1-3), (1-t)(x_2-2)\rangle$$

We obtain the G-basis (using lex):

$$\langle x_2^2 - 3x_2 + 2, x_1 - 2x_2 + 1, x_2 - 2 + t\rangle.$$

The first two terms give the elimination ideal:

$$I \cap J = \langle x_2^2 - 3x_2 + 2, x_1 - 2x_2 + 1\rangle.$$

The leading terms are x_2^2 and x_1, giving as expected a two-term basis: $\{1, x_2\}$.

2.3.1 Aliasing

In the last section we defined the equivalance class with respect to an ideal I as $f \sim_I g$ if $f - g \in I$. If $I = I(D)$, a design ideal then, equivalently $f(x) \sim_I g(x)$ for all $x \in D$: we cannot distinguish f from g by considering the values on D. We might call this *algebraic aliasing*. But this is not quite as broad a meaning as in statistics.

To make a connection to a branch of experimental design we consider regular 2^k factorial fractions and use the classical capital letter notion to begin with. We start with a k-letter alphabet $A, B, \ldots$ etc which form an Abelian group $\mathcal{G}_k$ under the condition $A^2 = B^2 = \ldots = I$, I is the identity. Then we consider a subgroup $\mathcal{G}_r$ of order r generated by r algebraically independent words $(G_1, G_2, \ldots, G_r)$. We label the actual factors by lower case words $a, b, c, \ldots$ which take levels ± 1. The subgroup $\mathcal{G}_r$ splits the full factorial design

$$\{\pm 1, \pm 1, \ldots, \pm 1\}$$

into 2^k blocks in the following way. Let $g = (g_1, \ldots, g_r)$ be the lower case version of $G_1, G_2, \ldots, G_r$. Then the blocks are given by the solutions of

$$\{a^2 = b^2 = \cdots = 1, g = e\}$$

where e ranges over all 2^r r-vectors $(\pm 1, \pm 1, \ldots, \pm 1)$.

One of several interpretations is that the capital letters refer to sign change reflections of the relevant coordinate of a k-vector $(a, b, \ldots)$ with entries ± 1. Then $\mathcal{G}$ is the full sign change group and $\mathcal{G}_r$ is a subgroup. The blocks are the invariant sets under the subgroup G_r.

Suppose we take one of the blocks as our design, which will then have sample size 2^{k-r} and is called a regular fraction. The attractive part of the theory is that we can read off the aliasing among monomials by looking at the cosets of $\mathcal{G}_r$. Each distinct coset $g\mathcal{G}_r$, for some $w \in G$ forms an equivalence class of aliased monomials. Each such equivalence class has 2^r members and can be thought of as a member of the quotient group $\mathcal{G}/\mathcal{G}_k$. Moreover every block gives this same alias structure.

Consider a very simple example for $d = 3, r = 1$ and $\mathcal{G}_r = \{I, ABC\}$. The design is one of two blocks, say.

$$\{(-1,-1,-1), (1,1,-1), (1,-1,1), (-1,1,1)\}$$

The alias table is

$$\begin{array}{rcl} I & = & ABC \\ A & = & BC \\ B & = & AC \\ C & = & AB \end{array}$$

The algebra (with $\prec_{lex}$) gives the ideal in G-basis form as

$$\langle x_1^2 - 1, x_2^2 - 1, x_3^2 - 1, x_1 + x_2 x_3 \rangle$$

This says that $x_2 x_3$ and $-x_1$ are algebraically aliased. Indeed, we can read this from the second row of the alias table (the alias table ignores the sign changes).

In a statistical regression model, before we fit the model, the coefficient of a term is unknown. A monomial appears in a regression term multiplied by a parameter: eg $\theta_{011} x_2 x_3$. This leads to the following more general definition

Definition 14 *Two sets of polynomials* $\{f\}$ *and* $\{g\}$ *statistically aliased over the design* D *if*

$$span(\{NF(f)\}) = span(\{NF(g)\}).$$

2.4 Use of indicator functions

When the design D is a subset of a full factorial $\mathcal{N}$ it is sometimes convenient to describe it via an indicator function: F_D. This is most usual when we start with some basic design, such as a full factorial, and consider a fraction. We saw such a fraction in the last subsection. A indicator function is a *single* additional function which we add the generators of the design ideal to form the ideal of the fraction. We can write the last example as

$$I(D) = \langle x_1^2 - 1, x_2^2 - 1, x_3^2 - 1, x_1x_2x_3 + 1\rangle.$$

The first three terms form the G-basis of the full factorial $\{(\pm 1, \pm 1, \pm 1)\}$.

The equation $x_1x_2x_3 + 1 = 0$ can be written equivalently in terms of an indicator function. Consider $g(x_1, x_2, x_3) = \frac{1}{2}(-x_1x_2x_3 + 1)$. This takes the value 1 on the design and 0 on the complementary fraction. Then, on D,:

$$x_1x_2x_3 + 1 = 0 \Leftrightarrow g(x_1, x_2, x_3) = 1$$

More generally let D be the basic, starting design, and $D' \subset D$ the fraction. Fix a monomial order and, via the $I(D)$, construct the basis for interpolation over D. Then the indicator function interpolates the $0, 1$ values as required:

$$y_{D'}(x) = \begin{cases} 1, & x \in D' \\ 0, & x \in D \backslash D' \end{cases}$$

It it sometimes convenient to use the indicator if we require that the fraction has a special property, such as orthogonality. Consider the 2^d case with coding $\{-1, 1\}$. Two (square-free) monomials x^α, x^β in the model with fraction D' are said to be are said to be *orthogonal* if the corresponding columns in the X-matrix are orthogonal:

$$\sum_{x \in D'} x^\alpha x^\beta = \sum_{x \in D'} x^{\alpha+\beta} = 0.$$

We can express this in terms of the indicator over $\{-1, 1\}^d$ and write

$$\sum_{x \in D} x^{\alpha+\beta} g(x) = 0.$$

Another use is that we can take union and intersections rather easily by using Boolean type operations over D:

$$g_{D_1 \cap D_2} = g_{D_1} g_{D_2},\ g_{D_1 \cup D_2} = g_{D_1} + g_{D_2} - g_{D_1} g_{D_2}.$$

2.5 Weighted orders and zonotopes

In design, by considering all terms orders one achieves the full *algebraic fan* of models achievable by the algebraic. It turns out that it is enough to consider a finite set of *weighted* term orderings.

The following material provides construction of weighted term orderings for finite subsets of the set of all monomials $T^d = \{x^\alpha : \alpha \in \mathbb{Z}^d_{\geq 0}\}$. This simplification is important as most of the time Gröbner basis computations are carried out over a (relatively small) finite set of monomials.

2.5.1 Weighted orders

Recall that term orders are total orders over monomials in T^d, but for some computations, only a total order over a subset of T^d is required. The importance of this is for applications in which we may wish to consider all, or a range, of term orders. A weighted term ordering which uses vector dot-product is presented here, and conditions for this ordering to be a total order are given.

Definition 15 *(w-**order**) Let $B \subseteq T^d$, $B \neq \emptyset$ and let $w \in \mathbb{Z}^k_{\geq 0}, w \neq (0, \ldots, 0)$ be fixed. For $x^\alpha, x^\beta \in B$, we say $x^\alpha \succeq_w x^\beta$ if $w \cdot \alpha \geq w \cdot \beta$, where $w \cdot \alpha$ is the usual dot product between w and α. The weighted ordering relation thus defined is referred to as w-order and represented by $\succeq_w$.*

If the set B is all of T^d, then for any vector w, $\succeq_w$ is a partial order, i.e. there are ties among monomials. Even for a finite set B, it is not difficult to find a vector w such that $\succeq_w$ is only a partial ordering. For example, if $B = \{x_1, x_2\} \subset T^2$ take $w = (1, 1)$. However, a careful selection of w with respect to the set B allows to use w-orders that define a total ordering over the set B. The following Lemma states a sufficient condition for a vector w to define a total ordering over B.

Lemma 16 *Let $B \subset T^d$ be finite; let w be such that w is not orthogonal to any of $\{\alpha - \beta : x^\alpha, x^\beta \in B\}$. Then w-order is total ordering in B, and we write $\succ_w$.*

Example 1 Consider $B = \{x_1, x_2, x_3\} \subset T^3$. Using Lemma **??**, any $w = (w_1, w_2, w_3) \in \mathbb{Z}^k_{\geq 0} \setminus \{(0, 0, 0)\}$ that does not satisfy the conditions $w_1 = w_2$, $w_1 = w_3$ and $w_2 = w_3$ creates a total w-ordering on B. For instance,

$w = (3, 2, 1)$ defines a total ordering in B for which $x_1 \succ_w x_2 \succ_w x_3$, while $w = (1, 3, 2)$ corresponds to the total ordering $x_2 \succ_w x_3 \succ_w x_1$. Both cases are examples of different initial orders.

In general, selection of different vectors w that satisfy Lemma **??** will create different total w-orderings on B. However, some of the choices may correspond to the same total ordering on B. In such case, it is natural to create equivalence classes of ordering vectors w. Such classes will depend on the set B.

Definition 17 *Let $B \subset T^d$ be finite; let $\succ_{w_1}, \succ_{w_2}$ be total orderings in B. We say $w_1 \sim w_2$ when $x^\alpha \succ_{w_1} x^\beta$ iff $x^\alpha \succ_{w_2} x^\beta$ for all $x^\alpha, x^\beta \in B$.*

The relation $\sim$ is a well posed equivalence relation, and we say that w_1, w_2 belong to the same equivalence class. From a practical point of view, we only need one representative vector w for each equivalence class, as use of other vectors in the same class does not yield a new total ordering on B.

For fixed finite B, the set of all weighing vectors that create $\succ_w$ is the equivalence class of w. Each equivalence class is a polyhedral cone (intersection of halfspaces), intersected with the positive integer lattice. All the equivalence classes of ordering vectors can be related to central hyperplane arrangements by the following result.

Theorem 18 *The central hyperplane arrangement constructed with all pairwise differences between exponents of monomials in B partitions the positive integer lattice into cones, the interior of which corresponds to equivalence classes $\sim$.*

The ultimate object of interest is a collection of representatives, one for each equivalence class of total w-orderings. This set is referred to a the universal set of ordering vectors W_+, and depends on the finite set of monomials B under consideration.

Example 2 For the same set $B = \{x_1, x_2, x_3\}$ of Example **??**, equivalence classes of ordering vectors are created in the first orthant by the planes $w_1 = w_2, w_1 = w_3$ and $w_2 = w_3$. The three hyperplanes split the first orthant into six polyhedral cones. Taking a representative in the interior of each polyhedral cone forms the desired set

$$W_+ = \{(1, 2, 3), (1, 3, 2), (2, 1, 3), (2, 3, 1), (3, 1, 2), (3, 2, 1)\}$$

.

2.5.2 Weighted orders and zonotopes

Theorem **??** can be explained in terms of the fan of called zonotopes. Zonotopes are convex polytopes which can be written as Minkowski sums. Using zonotopes not only provides an interesting link between geometry and algebra, but allows to obtain bounds for the number of equivalence classes. We present here a brief summary of results.

Definition 19 *Let B, C be two subsets of $\mathbb{R}^k$, the Minkowski sum of B and C is defined as follows: $B + C = \{b + c : b \in B, c \in C\}$.*

Zonotopes are polytopes constructed by the Minkowski sum of line segments.

Definition 20 *Let V be a finite set of vectors in $\mathbb{R}^d$. The **zonotope** of V is the Minkowski sum*

$$Z(V) = \sum_{v \in V} [0, v], \tag{4}$$

where $[0, v]$ is the line segment between 0 and v.

The following result encodes all equivalence classes of ordering vectors into the normal fan a single geometrical figure. Recall that the normal fan of a polytope is the collection of pointed polyhedral cones in $\mathbb{R}^d$, one for every vertex of the polytope such that its union is all of $\mathbb{R}^d$.

Theorem 21 *Let $B \subset T^d$ be finite, let $D = D(B) = \{\alpha - \beta : x^\alpha, x^\beta \in B\}$ and let $Z(D)$ be the zonotope of D. Then the restricted* normal fan *of $Z(D)$ partitions the positive integer lattice into the cones $\sim$.*

Example 3 Consider the set $B = \{x_1, \ldots, x_k\}$ of Examples and **??** and **??**. The zonotope of pairwise differences of elements in B is constructed with the set $D(B) = \pm\{(1, -1, 0), (1, 0, -1), (0, 1, -1)\}$. The zonotope $Z(D)$ is a permutahedron that has six vertexes $\pm\{(2, -2, 0), (2, 0, -2), (0, 2, -2)\}$. The normal fan of this polytope partitions precisely the first orthant in the same six classes as in Example **??**.