

Chapter 1: Polynomials and Ideals

1 Polynomials in n variables

1.1 Polynomials in one variable

A **polynomial** in the **variable** x with **coefficients** in the ring R is an expression of the form

$$\sum_{i=0}^n a_i x^i = a_0 + a_1 x + \dots + a_{n-1} x^{n-1} + a_n x^n, \tag{1}$$

where $n \geq 0$ and the a_i are elements of R . If the \sum notation is used, then $x^0 = 1$ by convention. Note that x itself is NOT an element of R , but is a formal symbol which helps to make calculations as natural as possible. If $n = 0$, the polynomial is simply an element $a_0 \in R$, and is called a **constant** polynomial.

The set of all polynomials (1) is a ring under the usual addition and multiplication operations. It is denoted by $R[x]$ and is called **the ring of polynomials in one variable over R** . We shall use letters such as f, g, h or p, q, r to denote polynomials; if it is necessary to draw attention to the variable being used, then we write e.g. $f(x), f(y)$ instead of f . We shall often assume that R is a **field** F . We shall **always** assume that R is a **commutative ring with 1**. You should be familiar with the meaning of these terms from MT2262.

The most important fact about divisibility in $F[x]$ is the **division theorem**. If $f \neq 0$ (i.e. f is not the zero polynomial, so that some coefficient $a_i \neq 0$), then we may assume in (1) that $a_n \neq 0$; n is then called the **degree** of the polynomial f , and we write $n = \deg f$. Note that the degree of the zero polynomial is not defined.

Theorem 1.1 (Division Theorem) *Let F be a field and let $f, g \in F[x]$, where $g \neq 0$. Then there are unique polynomials q , the **quotient**, and r , the **remainder**, in $F[x]$ such that (i) $f = qg + r$, and (ii) either $\deg r < \deg g$ or $r = 0$.*

We shall see that the division theorem fails for polynomials in two or more variables. This is one fundamental reason why the algebra of polynomials in more than one variable is more difficult. One of the objects of this course is to convince you that it is also more interesting and rewarding! However, before moving on, we note two more things we can do with one-variable polynomials.

(i) **Substitution** Given polynomials $f(x), g(x) \in R[x]$, we can replace every occurrence of x in $f(x)$ by $g(x)$ to obtain a new polynomial $f(g(x))$. For example if $f(x) = x^2 + 1$ and $g(x) = x - 2$, then $f(g(x)) = (x - 2)^2 + 1 = x^2 - 4x + 5$ and $g(f(x)) = (x^2 + 1) - 2 = x^2 - 1$.

(ii) **Evaluation** Given a polynomial $f(x) \in R[x]$ and an element $a \in R$, we can replace every occurrence of x in $f(x)$ by a to obtain a new element of R , which we denote by $f(a)$. In this way, the polynomial $f(x)$ defines a **function** from R to R . This function is called a **polynomial function**.

Although there is a logical distinction between polynomials and polynomial functions, it really does no harm to identify these two concepts. However, you should be aware that two polynomials are equal if and only if they have identical **coefficients**, whereas two functions are equal if and only if they have the same **graphs**. The difference can be seen by taking the field F to be finite; for example if p is a prime number and $F = \mathbf{F}_p$, the field with p elements, then the polynomials x^p and x define the **same** function. But if F is an infinite field such as \mathbf{Q} (the rational numbers), \mathbf{R} (the real numbers) or \mathbf{C} (the complex numbers), two polynomials are equal if and only if the corresponding polynomial functions are equal. Especially for $F = \mathbf{R}$, polynomial functions give a geometric way to visualise polynomials via their graphs.

1.2 Polynomials in two variables

When we try to write down a definition like (1) for a general polynomial in two variables x and y with coefficients in R , we have a difficulty: which order do we write the terms in? For example should the ‘standard form’ of a quadratic polynomial in x and y be $a_{0,0} + a_{1,0}x + a_{0,1}y + a_{2,0}x^2 + a_{1,1}xy + a_{0,2}y^2$, or should the term in x^2 come before the term in y ? Notice that we can collect terms in either variable, and write this as $(a_{0,0} + a_{1,0}x + a_{2,0}x^2) + (a_{0,1} + a_{1,1}x)y + a_{0,2}y^2$ or as $(a_{0,0} + a_{0,1}y + a_{0,2}y^2) + (a_{1,0} + a_{1,1}y)x + a_{2,0}x^2$.

We can hide this problem by writing the polynomial in the form

$$\sum_{i,j \geq 0} a_{i,j} x^i y^j \tag{2}$$

where $a_{i,j} \in R$ and the sum is **finite**, i.e. $a_{i,j} = 0$ for all but a finite number of indexing pairs (i, j) . Although the problem resurfaces when we try to do practical calculations, we adopt (2) as the formal definition. We call an expression $x^i y^j$ a **monomial** in x and y , and if $a_{i,j} \neq 0$ we call $a_{i,j} x^i y^j$ a **term** of the polynomial. The **(total) degree** of this monomial is $i + j$, and the **(total) degree** of a nonzero polynomial is the maximum of the degrees of its terms. A polynomial is called **homogeneous of degree d** if it has no term of degree $\neq d$. (Thus the zero polynomial is ‘homogeneous of degree d ’ for all $d \geq 0$, although its degree is not defined!)

The set of all polynomials (2) forms a ring $R[x, y]$ with the usual addition and multiplication operations, called the ring of **polynomials in two variables over R** . By collecting terms with the same power of x , we can regard a polynomial in $R[x, y]$ as a polynomial in x with coefficients which are polynomials in y , i.e. as

an element of $S[x]$ where $S = R[y]$. Thus we have a natural way to **identify** the rings $R[x, y]$ and $(R[x])[y]$. Similarly, we can identify $R[x, y]$ and $(R[y])[x]$.

As for the one variable case, we shall often use letters such as f, g, h or p, q, r to denote polynomials; if it is necessary to draw attention to the variables being used, then we write $f(x, y)$ etc instead of f . If $R = F$ is a **field**, then a polynomial f in $F[x, y]$ is always divisible by any nonzero constant polynomial, i.e. any polynomial of degree 0.

It is easy to see that the division theorem fails for polynomials in two variables. Suppose, for example, that we try to divide the polynomial x by the polynomial y . This means we are looking for polynomials q and r such that $x = qy + r$ and $\deg r < 1$ or $r = 0$. Thus r must be a constant polynomial, and clearly there is then no polynomial $q \in R[x, y]$ such that $x = qy + r$.

1.3 Polynomials in n variables

The definitions above are easily extended to the case of $n > 2$ variables. A **monomial** in n variables x_1, \dots, x_n is an expression of the form

$$m = x_1^{\alpha_1} \cdots x_n^{\alpha_n},$$

where $\alpha_1, \dots, \alpha_n$ are integers ≥ 0 . The **(total) degree** of the monomial m is $\deg m = \alpha_1 + \dots + \alpha_n$. It is convenient to adopt a ‘multi-index’ notation for writing monomials, i.e.

$$\mathbf{x}^\alpha := x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

where $\mathbf{x} = (x_1, \dots, x_n)$ is a ‘vector variable’ and $\alpha = (\alpha_1, \dots, \alpha_n)$ is a ‘vector exponent’. We denote the total degree of \mathbf{x}^α by $|\alpha| = \alpha_1 + \dots + \alpha_n$.

Let R be a commutative ring with 1. A **polynomial** over R in the variables x_1, \dots, x_n is a **finite** sum $f = f(x_1, \dots, x_n)$ of **terms** of the form $a_\alpha \mathbf{x}^\alpha$, where the **coefficient** $a_\alpha \in R$. If $a_\alpha \neq 0$, we say that the term $a_\alpha \mathbf{x}^\alpha$ **appears** in f . We identify the monomial $x_1^0 \cdots x_n^0$ with the identity element $1 \in R$. The **constant term** of f is the element $a_{(0, \dots, 0)} \in R$; it may of course be zero.

In multi-index notation, a polynomial can be written as follows.

$$f = a_{\alpha(1)} \mathbf{x}^{\alpha(1)} + \dots + a_{\alpha(r)} \mathbf{x}^{\alpha(r)} \quad (3)$$

where the coefficients $a_{\alpha(1)}, \dots, a_{\alpha(r)}$ are elements of R . The set of all polynomials of the form (3) forms a ring $R[x_1, \dots, x_n]$ with the usual addition and multiplication operations, called the ring of **polynomials in n variables over R** . Notice that by collecting powers of one of the variables, x_n say, we can identify the rings $R[x_1, \dots, x_n]$ and $R[x_1, \dots, x_{n-1}][x_n]$.

The **(total) degree** of a polynomial $f \neq 0$ is the maximum of the degrees of the terms appearing in f . If all terms appearing in f have the same degree d , f is called a **homogeneous polynomial of degree d** or a **form of degree d** . In the case $d = 1$ we speak of a **linear form**, in the case $d = 2$ of a **quadratic form**, in the case $d = 3$ of a **cubic form**, etc.

2 Ideals in polynomial rings

2.1 The ideal generated by a finite set

Let R be a commutative ring with 1 and let $a_1, \dots, a_n \in R$. The **ideal generated** by a_1, \dots, a_n is the set of all elements of R of the form $a = r_1 a_1 + \dots + r_n a_n$, where $r_1, \dots, r_n \in R$. We denote this ideal by $\langle a_1, \dots, a_n \rangle$.

In particular, the ideal $\langle a \rangle$ generated by a single element $a \in R$ is the set of all multiples of a . In the general case, elements of the ideal are formed analogously to linear combinations of vectors in a vector space.

Example 2.1 Let $R = F[x, y]$ where F is any field. Then $\langle x^2, y \rangle$ is the set of all polynomials $f(x, y)$ of the form $x^2 g + yh$ where g and h are polynomials in x and y . By thinking of f as a sum of monomials, we can see that $f \in \langle x^2, y \rangle$ if and only if every monomial which appears in f is divisible by either x^2 or y . Thus $\langle x^2, y \rangle$ is the set of all polynomials such that the constant term and the coefficient of x are both zero.

2.2 Principal ideals

In general, an **ideal** in a ring R is a non-empty subset of R which is closed under the operations of addition and multiplication by elements of R . More formally, $I \subseteq R$ is an ideal if and only if

- $0 \in I$;
- if $a, b \in I$ then $a + b \in I$;
- if $a \in I$ and $r \in R$ then $ra \in I$.

It is easy to check that if $a_1, \dots, a_n \in R$ then $\langle a_1, \dots, a_n \rangle$ is an ideal. Conversely, given an ideal $I \subseteq R$, we can ask if there exist elements $a_1, \dots, a_n \in R$ such that $I = \langle a_1, \dots, a_n \rangle$. Such a set of elements is called a **generating set** or a **basis** for I . If I is generated by a single element, i.e. if $I = \langle a \rangle$ for some $a \in R$, then I is called a **principal ideal** of R .

It is easy to see from Example 2.1 that the ideal $I = \langle x^2, y \rangle$ is **not** a principal ideal in $R = F[x, y]$. For if $I = \langle f \rangle$, then the polynomial f would have to divide both x^2 and y . But the only common divisors of x^2 and y are nonzero constant polynomials $a \in R$. However a is invertible in R , and hence $\langle a \rangle = R$.

Remarkably, the situation is quite different in the case of polynomials in **one** variable over a **field**.

Theorem 2.2 *Let F be a field. Then every ideal in the polynomial ring $F[x]$ is principal.*

Proof Let $I \subseteq F[x]$ be an ideal. If $I = \{0\}$ then it is principal, so we may assume that I contains a nonzero polynomial g of minimum degree. Clearly $\langle g \rangle \subseteq I$. We shall prove that $I \subseteq \langle g \rangle$, so that $I = \langle g \rangle$ and hence I is principal.

Thus let $f \in I$. By the Division Theorem 1.1, there exist polynomials $q, r \in F[x]$ such that $f = qg + r$ and $\deg r < \deg g$ or $r = 0$. Since $r = f - qg$ and $f, g \in I$, it follows from the definition of an ideal that $r \in I$. By our choice of g , it is impossible that $\deg r < \deg g$, and hence $r = 0$. But then $f = qg$, and so $f \in \langle g \rangle$. \square

2.3 Finitely generated ideals

We see from Theorem 2.2 that the failure of the Division Theorem in the polynomial ring $F[x, y]$ in two variables over the field F is closely related to the existence of ideals which are not principal. From Example 2.1, we might be tempted to guess that every ideal in $F[x, y]$ is generated by at most two elements. The following example shows that this is far too optimistic.

Example 2.3 Let F be a field, let $n \geq 1$ and let $I = \langle x^n, x^{n-1}y, \dots, xy^{n-1}, y^n \rangle$ be the ideal in $F[x, y]$ generated by the set of all monomials of degree n . We shall prove that I cannot be generated by fewer than $n + 1$ polynomials.

First notice that the polynomials in I are the polynomials in which no term of degree $< n$ appears. Let f_1, \dots, f_m be a generating set for I : we must prove that $m \geq n + 1$. Each of the $n + 1$ monomials $x^i y^{n-i}$ can be expressed in the form $x^i y^{n-i} = \sum_{j=1}^m g_{i,j} f_j$ where $g_{i,j} \in F[x, y]$. Since the f_j have no terms of degree $< n$, equating terms of degree n gives $x^i y^{n-i} = \sum_{j=1}^m a_{i,j} h_j$, where h_j is the degree n homogeneous part of f_j and $a_{i,j} \in F$ is the constant term of $g_{i,j}$. However, the homogeneous polynomials of degree n form a vector space of dimension $n + 1$ over F , with basis $\{x^n, x^{n-1}y, \dots, xy^{n-1}, y^n\}$, and the equations $x^i y^{n-i} = \sum_{j=1}^m a_{i,j} h_j$ express each of these basis elements as linear combinations of the h_1, \dots, h_m . Hence $m \geq n + 1$, by standard facts of linear algebra.

We see from this example that there is no integer N such that every ideal in $F[x, y]$ is generated by some set with at most N elements. Obviously, the same is true for $F[x_1, \dots, x_n]$ for every $n \geq 2$. Thus the algebraic structure of a polynomial ring in two or more variables is very different from the case of one variable. Here are some questions we might ask about polynomials in n variables.

- Do two polynomials have a greatest common divisor? (Recall that for $n = 1$ the answer is yes, and it can be computed by the Euclidean Algorithm.)
- Given an ideal $I = \langle f_1, \dots, f_n \rangle$ and a polynomial f , how can we determine whether f is in I or not? (For $n = 1$, $f \in I$ if and only if the generator of the ideal divides f .)

- Does every ideal in $F[x_1, \dots, x_n]$ have a finite generating set?

The last question is answered by

Theorem 2.4 (Hilbert's Basis Theorem) *Let F be a field and let I be an ideal in the polynomial ring $P = F[x_1, \dots, x_n]$. Then I is finitely generated, i.e. there exists a positive integer N and a set of polynomials f_1, \dots, f_N in I such that every polynomial $f \in I$ can be written in the form $f = \sum_{i=1}^N r_i f_i$ for some polynomials r_1, \dots, r_N in P .*

We shall return to Hilbert's theorem later.

The following example shows that ideals in polynomial rings need not be finitely generated in general. This example will not be needed in our future work, and so the details (including the precise definitions!) are left for you to think about.

Exercise 2.5 Let F be a field and let x_1, x_2, \dots be an infinite sequence of variables. Let P be the polynomial ring $F[x_1, x_2, \dots]$ and let I be the ideal consisting of all polynomials with zero constant term. Then I is not finitely generated.

An ideal which is not finitely generated can always be generated by some set, for example, the set of all its elements. Here we make precise the definition of a generating set for an ideal.

Let R be a commutative ring with 1, let I be an ideal in R , and let A be a subset of R . Then A is a **generating set** or **basis** for I (written $I = \langle A \rangle$) if and only if every element $a \in I$ can be written in the form

$$a = r_1 a_1 + \dots + r_n a_n, \tag{4}$$

where $a_1, \dots, a_n \in A$ and $r_1, \dots, r_n \in R$, for some positive integer n .

The point is that although the set A can be infinite, only finitely many elements of A are involved in (4) for a given $a \in I$.

2.4 Quotient Rings

Recall (MT2262) that when I is an ideal in the commutative ring R , then the **quotient ring** R/I is constructed as follows. An element of P/I is a **coset** $\overline{f} = f + I = \{f + g \mid g \in I\}$, and that the ring operations in P/I are defined by

$$\begin{aligned} (f_1 + I) + (f_2 + I) &= f_1 + f_2 + I, & \text{i.e. } \overline{f_1} + \overline{f_2} &= \overline{f_1 + f_2}, \\ (f_1 + I) \cdot (f_2 + I) &= f_1 \cdot f_2 + I, & \text{i.e. } \overline{f_1} \cdot \overline{f_2} &= \overline{f_1 \cdot f_2}. \end{aligned}$$

This is a good way to construct lots of interesting rings, as we shall see later.

3 Monomial Ideals and Dickson's Lemma

Monomials are much easier to handle than general polynomials. For example, we can write down the GCD and LCM of two given monomials at sight, e.g. $\text{GCD}(x^3y^2z, x^2z^2) = x^2z$, $\text{LCM}(x^3y^2z, x^2z^2) = x^3y^2z^2$.

Let F be a field and let $P = F[x_1, \dots, x_n]$ be the polynomial ring in n variables over F . A **monomial ideal** in P is an ideal I which is generated by a set of monomials M . We do **not** assume that M is finite. Examples 2.1 and 2.3 are examples of monomial ideals in $F[x, y]$.

Theorem 3.1 *Let I be a monomial ideal in $P = F[x_1, \dots, x_n]$, and let f be a polynomial in P . Then $f \in I$ if and only if every monomial which appears in f is in I .*

Proof The 'if' part is obvious; we prove the 'only if' part. Let $f \in I$, and let $a_\alpha \mathbf{x}^\alpha$ be a term in f . Thus $f = a_\alpha \mathbf{x}^\alpha + f_1$, where the monomial \mathbf{x}^α does not appear in f_1 . We must prove that $\mathbf{x}^\alpha \in I$.

Let S be a set of monomials which generates I . Then f is a finite sum

$$f = g_1 \mathbf{x}^{\alpha(1)} + \dots + g_r \mathbf{x}^{\alpha(r)} \quad (5)$$

for some polynomials $g_1, \dots, g_r \in P$ and some monomials $\mathbf{x}^{\alpha(1)}, \dots, \mathbf{x}^{\alpha(r)} \in S$. Equating terms in (3.1) with exponent vector α , we obtain

$$a_\alpha \mathbf{x}^\alpha = t_1 \mathbf{x}^{\alpha(1)} + \dots + t_r \mathbf{x}^{\alpha(r)} \quad (6)$$

where each t_i is either zero or is the term appearing in g_i with exponent vector $\beta = \alpha - \alpha(i)$. Since $a_\alpha \neq 0$, equation (6) expresses \mathbf{x}^α as an element of I . \square

This result reduces the problem of testing a polynomial for membership of a monomial ideal I to the problem of testing a monomial. The next result provides such a test.

Theorem 3.2 *Let S be a set of monomials in $P = F[x_1, \dots, x_n]$, and let \mathbf{x}^α be any monomial in P . Then $\mathbf{x}^\alpha \in \langle S \rangle$ if and only if \mathbf{x}^α is divisible by some element of S .*

Proof Let I be the set of polynomials f such that all monomials appearing in f are divisible by some element of S . Then I is an ideal in P . Since every element of S is in I , we have $\langle S \rangle \subseteq I$. But if $f \in I$, every monomial appearing in f is in $\langle S \rangle$ and so $f \in \langle S \rangle$. Hence $I \subseteq \langle S \rangle$, and so $I = \langle S \rangle$. Taking $f = \mathbf{x}^\alpha$, we have the required statement. \square

Theorems 3.1 and 3.2 allow us to describe the elements of a monomial ideal $\langle S \rangle$ precisely. For example, $\langle xy, y^3 \rangle \subset F[x, y]$ consists of all polynomials which have zero constant term and in which the monomials y, y^2 and all x^n ($n \geq 1$) do not appear.

3.1 Sums of ideals

Let I and J be ideals in R . Then their **sum** $I + J$ is defined by

$$I + J = \{a + b \mid a \in I, b \in J\}. \quad (7)$$

This is an ideal in R . If A and B are subsets of R such that A generates I and B generates J , then $I + J$ is generated by the union $A \cup B$ of these generating sets. It follows that the sum of two monomial ideals is again a monomial ideal.

Example 3.3 If $I = \langle x^2y, y^3 \rangle$ and $J = \langle x^4, xy^2 \rangle$, then $I + J = \langle x^2y, y^3, x^4, xy^2 \rangle$.

Notice that in general $I \cup J$ is **not** an ideal, but if $I \subseteq J$ then $I \cup J = J = I + J$. We can think of $I + J$ as the smallest ideal which contains both I and J .

3.2 Products of ideals

Let I and J be ideals in R . Then their **product** IJ is the set of elements $r \in R$ of the form

$$r = \sum_{i=1}^n a_i b_i, \text{ where } a_i \in I, b_i \in J. \quad (8)$$

This is an ideal in R . If A and B are subsets of R such that A generates I and B generates J , then IJ is generated by the product $AB = \{a_i b_i \mid a_i \in A, b_i \in B\}$ of these generating sets. It follows that the product of two monomial ideals is again a monomial ideal.

Example 3.4 If $I = \langle x^2y, y^3 \rangle$, $J = \langle x^4, xy^2 \rangle$, then $IJ = \langle x^6y, x^3y^3, x^4y^3, xy^5 \rangle$. Notice that we can throw out x^4y^3 from this list of generators, because it is divisible by another monomial in the list, x^3y^3 . The remaining list $\{x^6y, x^3y^3, xy^5\}$ is a **minimal** generating set for IJ .

3.3 Intersections of ideals

The **intersection** $I \cap J$ of ideals I and J in R is defined to be their intersection as sets, i.e. the set of all $r \in R$ such that $r \in I$ and $r \in J$. This is an ideal in R .

If \mathbf{x}^α and \mathbf{x}^β are monomials, then the intersection of the principal ideals $\langle \mathbf{x}^\alpha \rangle$ and $\langle \mathbf{x}^\beta \rangle$ is the principal ideal $\langle \mathbf{x}^\gamma \rangle$ where $\mathbf{x}^\gamma = \text{LCM}(\mathbf{x}^\alpha, \mathbf{x}^\beta)$. For example, $\langle x^2y \rangle \cap \langle xy^3 \rangle = \langle x^2y^3 \rangle$. This follows immediately from Theorem 3.2. More generally, let I and J be the monomial ideals generated by sets of monomials S_1 and S_2 respectively. Then by Theorem 3.2 a monomial \mathbf{x}^γ is in $I \cap J$ if and only if \mathbf{x}^γ is divisible by some element $\mathbf{x}^\alpha \in S_1$ and also by some element $\mathbf{x}^\beta \in S_2$. Hence the set $\{\text{LCM}(\mathbf{x}^\alpha, \mathbf{x}^\beta) \mid \mathbf{x}^\alpha \in S_1, \mathbf{x}^\beta \in S_2\}$ is a set of monomial generators for $I \cap J$.

Example 3.5 If $I = \langle x^2y, y^3 \rangle$ and $J = \langle x^4, xy^2 \rangle$, $I \cap J = \langle x^4y, x^2y^2, x^4y^3, xy^3 \rangle$. Notice that we can throw out the element x^4y^3 from the list of generators, because it is divisible by another monomial in the list, namely xy^3 . Thus $\{x^4y, x^2y^2, xy^3\}$ is a **minimal** generating set for $I \cap J$.

3.4 Dickson's Lemma

The next result is a major step towards the proof of the Hilbert Basis Theorem.

Theorem 3.6 *Let F be a field, let S be a set of monomials in $P = F[x_1, \dots, x_n]$, and let $I = \langle S \rangle$. Then I is generated by a finite subset of the monomials in S . In particular, every monomial ideal in P is finitely generated.*

Proof We first prove that I is generated by some finite set of monomials, which need not belong to S . For this, we use induction on n , the number of variables. The case $n = 1$ is trivial, since every monomial ideal is of the form $\langle x^k \rangle$ for some $k \geq 0$. Thus we assume that the result is true for $n - 1$ variables, and let I be a monomial ideal in $P = F[x_1, \dots, x_n]$. We shall write the variable x_n as y , so that monomials in P can be written as $\mathbf{x}^\alpha y^m$, where $\mathbf{x} = (x_1, \dots, x_{n-1})$ and $\alpha = (\alpha_1, \dots, \alpha_{n-1})$.

Given a monomial ideal I in P , let J be the ideal in $F[x_1, \dots, x_{n-1}]$ generated by the monomials \mathbf{x}^α such that $\mathbf{x}^\alpha y^k \in I$ for some $k \geq 0$. Then J is a monomial ideal in $F[x_1, \dots, x_{n-1}]$. By the induction hypothesis, we can choose a finite set of monomials $\{\mathbf{x}^{\alpha(1)}, \dots, \mathbf{x}^{\alpha(s)}\}$ which generates J .

For $1 \leq i \leq s$, by definition of J there is an integer m_i such that $\mathbf{x}^{\alpha(i)} y^{m_i} \in I$. Let $m = \max m_i$. For $0 \leq k \leq m - 1$, let J_k be the ideal in $F[x_1, \dots, x_{n-1}]$ generated by the monomials \mathbf{x}^α such that $\mathbf{x}^\alpha y^k \in I$. Then J_k is a monomial ideal in $F[x_1, \dots, x_{n-1}]$, so again the induction hypothesis allows us to choose a finite set of these monomials $\{\mathbf{x}^{\alpha_k(1)}, \dots, \mathbf{x}^{\alpha_k(s_k)}\}$ which generates J_k .

We claim that I is generated by the monomials $\mathbf{x}^{\alpha(i)} y^m$, for $1 \leq i \leq s$, together with the monomials $\mathbf{x}^{\alpha_k(i)} y^k$, for $1 \leq i \leq s_k$ and $0 \leq k \leq m - 1$. This is a finite set M of monomials.

Clearly all these monomials are in I . Let $\mathbf{x}^\alpha y^k$ be a monomial in I . If $k \geq m$, then since $\mathbf{x}^\alpha \in J$ and J is the monomial ideal generated by $\{\mathbf{x}^{\alpha(1)}, \dots, \mathbf{x}^{\alpha(s)}\}$, \mathbf{x}^α is divisible by $\mathbf{x}^{\alpha(i)}$ for some i with $1 \leq i \leq s$ by Theorem 3.2. Hence $\mathbf{x}^\alpha y^k$ is divisible by $\mathbf{x}^{\alpha(i)} y^m$. On the other hand, if $0 \leq k \leq m - 1$ then since $\mathbf{x}^\alpha \in J_k$ a similar argument proves that $\mathbf{x}^\alpha y^k$ is divisible by $\mathbf{x}^{\alpha_k(i)} y^k$ for some i with $1 \leq i \leq s_k$. Thus every monomial in I is divisible by some monomial in the set M , and it follows that $I = \langle M \rangle$.

To complete the proof, we must prove that a finite generating set for I can be selected from any given set of monomials S which generate the ideal I . Let m_1, \dots, m_r be a finite set of monomials which generates I : such a set exists by the first part of the proof. By Theorem 3.2, each m_i is divisible by some monomial $m'_i \in S$, for $1 \leq i \leq r$. Then m'_1, \dots, m'_r lie in S and generate I . \square

4 Orderings on Monomials

We would like to have a division algorithm for polynomials in n variables. That is, we would like to have an algorithm that has, as inputs, two polynomials f and g and, as outputs, two polynomials q (the quotient of f by g) and r (the remainder when f is divided by g). The polynomials q and r should satisfy

- (1) $f = qg + r$, and
- (2) r is ‘smaller’ in some sense than g .

For this to work, we must choose an **ordering** on monomials. To see why, consider the example $f = x^2y^2 + xy^2$, $g = x^2y$. Here it is natural to choose $q = y$ and $r = xy^2$. So we would like to regard xy^2 as ‘smaller’ than x^2y . However, if we interchange the variables and divide $x^2y^2 + x^2y$ by xy^2 , by symmetry we should regard x^2y as the remainder and hence ‘smaller’ than xy^2 . To escape from this contradiction, we must first agree whether to have $x^2y > xy^2$ or $x^2y < xy^2$.

Definition 4.1 *Let F be a field and let M be the set of all monomials \mathbf{x}^α in $P = F[x_1, \dots, x_n]$. A **monomial ordering** on P is a total ordering $<$ on M which is compatible with multiplication, i.e. the following four axioms hold.*

- **(trichotomy rule)** *Given $\mathbf{x}^\alpha, \mathbf{x}^\beta \in M$, exactly one of the three statements $\mathbf{x}^\alpha < \mathbf{x}^\beta$, $\mathbf{x}^\beta < \mathbf{x}^\alpha$ and $\mathbf{x}^\alpha = \mathbf{x}^\beta$ is true.*
- **(transitive rule)** *If $\mathbf{x}^\alpha, \mathbf{x}^\beta, \mathbf{x}^\gamma \in M$ satisfy $\mathbf{x}^\alpha < \mathbf{x}^\beta$ and $\mathbf{x}^\beta < \mathbf{x}^\gamma$, then $\mathbf{x}^\alpha < \mathbf{x}^\gamma$.*
- **(initialisation rule)** *If $\mathbf{x}^\alpha \in M$ and $\mathbf{x}^\alpha \neq 1$, then $1 < \mathbf{x}^\alpha$.*
- **(multiplication rule)** *If $\mathbf{x}^\alpha, \mathbf{x}^\beta, \mathbf{x}^\gamma \in M$ and $\mathbf{x}^\alpha < \mathbf{x}^\beta$, then $\mathbf{x}^\alpha \mathbf{x}^\gamma < \mathbf{x}^\beta \mathbf{x}^\gamma$.*

Along with the $<$ symbol, we shall use the symbols $\leq, >, \geq$ with the obvious meanings.

We give three examples. In these examples, we shall assume (as part of the ordering) that the variables are ordered so that $x_1 > x_2 > \dots > x_n$. Let $\mathbf{x}^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$, $\mathbf{x}^\beta = x_1^{\beta_1} \dots x_n^{\beta_n}$.

(i) The lexicographic ordering, Lex

Define $\mathbf{x}^\alpha < \mathbf{x}^\beta$ in Lex if and only if $\alpha_1 = \beta_1, \dots, \alpha_{k-1} = \beta_{k-1}$, $\alpha_k < \beta_k$ for some $k = 1, \dots, n$. In other words, reading from **left** (x_1) to **right** (x_n), the first time the exponents are different the exponent in \mathbf{x}^α is **smaller**.

(ii) The degree lexicographic ordering, DegLex

Define $\mathbf{x}^\alpha < \mathbf{x}^\beta$ in DegLex if and only if $|\alpha| < |\beta|$ or if $|\alpha| = |\beta|$ and $\mathbf{x}^\alpha < \mathbf{x}^\beta$ in Lex. In other words, we use the **degree** (the sum of the exponents) as the first test, and use **Lex** to order monomials of the same degree.

(iii) **The degree reverse lexicographic ordering, DegRevLex**

Define $\mathbf{x}^\alpha < \mathbf{x}^\beta$ in DegRevLex if and only if $|\alpha| < |\beta|$ or if $|\alpha| = |\beta|$ and for some k , $\alpha_n = \beta_n, \dots, \alpha_{k+1} = \beta_{k+1}$, $\alpha_k > \beta_k$. In other words, we use the **degree** as the first test, and order monomials of the same degree as follows: $\mathbf{x}^\alpha < \mathbf{x}^\beta$ if and only if when reading from **right** (x_n) to **left** (x_1), the first time the exponents are different the exponent in \mathbf{x}^α is **larger**.

These orderings, especially DegRevLex, take a bit of getting used to. In particular, it is important to realise that the order of the variables must be specified. For example, if the variables are x and y and we want to use Lex in $F[x, y]$, we must specify whether $x < y$ or $y < x$. With $x < y$ the Lex order on monomials looks like

$$1 < x < x^2 < x^3 < \dots < y < xy < x^2y < \dots < y^2 < \dots \quad (9)$$

and the DegLex order looks like

$$1 < x < y < x^2 < xy < y^2 < x^3 < x^2y < xy^2 < y^3 < x^4 < \dots \quad (10)$$

If $n = 2$ in our ‘standard notation’ for n variables, we have $x_2 < x_1$, and so the DegLex order looks like

$$1 < x_2 < x_1 < x_2^2 < x_1x_2 < x_1^2 < x_2^3 < x_1x_2^2 < x_1^2x_2 < x_1^3 < x_2^4 < \dots \quad (11)$$

The difference between DegLex and DegRevLex can be seen as follows. Let $x_1 > x_2 > x_3$. Then $x_1^2x_2x_3 > x_1x_2^3$ in DegLex, but $x_1^2x_2x_3 < x_1x_2^3$ in DegRevLex.

An important feature of all monomial orderings is that **there is no infinite strictly decreasing sequence of monomials**. This is obvious for DegLex and DegRevLex, since there are only finitely many monomials smaller than a given monomial in these orderings. Checking the property for Lex will help you to understand this ordering.

Having fixed a choice of monomial ordering, we can write a given polynomial $f \in F[x_1, \dots, x_n]$ in a **standard form** with respect to the ordering. This means that we write the terms in f in **decreasing** order.

Example 4.2 Let $n = 5$, $x_1 > x_2 > x_3 > x_4 > x_5$, and let $f = x_1^3(1 + 3x_2x_5) + x_1^2(x_5^2 - x_2^2x_4)$. Then the standard forms for f with respect to the three orderings above are tabulated as follows.

Ordering	Standard Form
Lex	$3x_1^3x_2x_5 + x_1^3 - x_1^2x_2^2x_4 + x_1^2x_5^2$
DegLex	$3x_1^3x_2x_5 - x_1^2x_2^2x_4 + x_1^2x_5^2 + x_1^3$
DegRevLex	$-x_1^2x_2^2x_4 + 3x_1^3x_2x_5 + x_1^2x_5^2 + x_1^3$

In multi-index notation, the standard form of a typical nonzero polynomial $f \in P$ (with respect to a given monomial ordering) can be written as

$$f = a_{\alpha(1)}\mathbf{x}^{\alpha(1)} + \dots + a_{\alpha(r)}\mathbf{x}^{\alpha(r)} \quad (12)$$

where $\mathbf{x}^{\alpha(1)} > \dots > \mathbf{x}^{\alpha(r)}$ and the coefficients $a_{\alpha(1)}, \dots, a_{\alpha(r)}$ are nonzero elements of F .

We call $\mathbf{x}^{\alpha(1)}$ the **leading monomial** of f , $\text{LM}(f)$, $a_{\alpha(1)}$ the **leading coefficient** of f , $\text{LC}(f)$, and $a_{\alpha(1)}\mathbf{x}^{\alpha(1)}$ the **leading term** of f , $\text{LT}(f)$.

Thus in Example 4.2, if the ordering is Lex or DegLex then $\text{LM}(f) = x_1^3x_2x_5$, $\text{LC}(f) = 3$ and $\text{LT}(f) = 3x_1^3x_2x_5$, and if the ordering is DegRevLex then $\text{LM}(f) = x_1^2x_2^2x_4$, $\text{LC}(f) = -1$ and $\text{LT}(f) = -x_1^2x_2^2x_4$.

Note that if f and g are nonzero polynomials, then $\text{LT}(f \cdot g) = \text{LT}(f) \cdot \text{LT}(g)$, and similarly for LM and LC.

To complete this section, we use Dickson's Lemma to prove the important property of monomial orderings stated above.

Proposition 4.3 *Let $<$ be a monomial ordering on the monomials in $P = F[x_1, \dots, x_n]$, and let S be any set of monomials in P . Then S has a minimum element.*

Proof Let $I = \langle S \rangle$ be the ideal generated by S . By Dickson's Lemma 3.6, there is a finite subset $\{\mathbf{x}^{\alpha(1)}, \dots, \mathbf{x}^{\alpha(r)}\}$ of S which generates I . Let \mathbf{x}^α be the minimum of these monomials. We claim that \mathbf{x}^α is the minimum element of S .

To see this, let $\mathbf{x}^\beta \in S$, so that \mathbf{x}^β is divisible by some $\mathbf{x}^{\alpha(i)}$, $1 \leq i \leq r$, say $\mathbf{x}^\beta = \mathbf{x}^{\alpha(i)}\mathbf{x}^\gamma$. Since $1 \leq \mathbf{x}^\gamma$, $\mathbf{x}^{\alpha(i)} \leq \mathbf{x}^\beta$ and hence $\mathbf{x}^\alpha \leq \mathbf{x}^\beta$. \square

In particular, it follows that there is no infinite strictly decreasing sequence of monomials in a monomial ordering, since such a sequence would have no minimum element.

5 A division algorithm in n variables

In this section we will formulate a division algorithm in $P = F[x_1, \dots, x_n]$. We begin by reviewing the familiar 1-variable case. Given f and $g \neq 0$ in $F[x]$, the algorithm produces a quotient q and remainder r with $f = qg + r$ and $\deg r < \deg g$ or $r = 0$. Thus $r = f - qg \in f + I$, where $I = \langle g \rangle$ is the ideal generated by g , and $f + I = \{f + h \mid h \in I\}$ is the **coset** of I which contains f . Thus we can think of the division algorithm as a procedure for obtaining the ‘simplest’ representative for that coset containing the given polynomial f : note that $r + I = f + I$, since $f - r = qg \in I$. We would like to imitate this situation in the n variable case, by associating to a polynomial f and a nonzero ideal I of P a ‘remainder’ r such that $f - r \in I$. However, the ideal I need not be principal. If $I = \langle g_1, \dots, g_r \rangle$, then $f - r \in I$ if and only if f can be written in the form

$$f = q_1 g_1 + \dots + q_s g_s + r, \quad (13)$$

where the ‘quotients’ q_i and ‘remainder’ r are in P .

As an example, if $n = 2$ and $I = \langle x_1^2, x_1 x_2, x_2^2 \rangle$, we would want r to be the polynomial $a + bx_1 + cx_2$ obtained by deleting all terms of degree ≥ 2 from f . More generally, if I is a monomial ideal, then we can obtain a uniquely defined remainder r by deleting from f all terms involving monomials in I . Note that the q_i will not be uniquely determined by (13) when $s > 1$.

We must first fix a monomial ordering in P . To see the need for this, consider the case $f = 2x + 3y + 4z$, $g = x + y + z$ in $F[x, y, z]$. With the ordering $x > y > z$, the quotient is 2 and the remainder is $y + 2z$, but with the ordering $z > y > x$, the quotient is 4 and the remainder is $-2x - y$. In each case, all monomials which appear in the remainder are lower than the leading monomial in g , and this is a condition we wish to use to characterise the remainder r in (13).

Here is an example to show what is involved. The basic idea is the same as for division in the case $n = 1$, i.e. at each step we cancel the leading term of f with respect to the given monomial order, by multiplying some divisor g_i by a suitable monomial and subtracting a scalar multiple of this product from f .

Example 5.1 Using Lex order in $F[x, y]$ with $x > y$, we will divide $f = xy^2 + 1$ by $g_1 = xy + 1$ and $g_2 = y + 1$. The leading terms xy of g_1 and y of g_2 both divide $\text{LT}(f) = xy^2$. We use g_1 since it is listed first. This gives $f - yg_1 = (xy^2 + 1) - y(xy + 1) = -y + 1 = f_1$ say. The leading term $\text{LT}(f_1) = -y$ which is not divisible by $\text{LT}(g_1)$ but is divisible by $\text{LT}(g_2)$. This gives $f_1 - (-1)g_2 = (-y + 1) + (y + 1) = 2$. Since $\text{LT}(g_1)$ and $\text{LT}(g_2)$ do not divide 2 we stop: the remainder $r = 2$.

Unfortunately, things do not always go so smoothly as this, as the following example shows.

Example 5.2 Again using Lex order in $F[x, y]$ with $x > y$, we will divide $f = x^2y + xy^2 + y^2$ by $g_1 = xy - 1$ and $g_2 = y^2 - 1$. The first two steps go as before (remember that we use g_1 when both leading terms divide):

$$\begin{aligned} f - x(xy - 1) &= xy^2 + x + y^2 = f_1 \\ f_1 - y(xy - 1) &= x + y^2 + y = f_2 \end{aligned}$$

The leading term of f_2 is x which is not divisible by $\text{LT}(g_1) = xy$ or $\text{LT}(g_2) = y^2$. However f_2 is **not** the remainder because we can continue to ‘reduce’ it using g_2 . So the next step is to ‘move’ $\text{LT}(f_2) = x$ to the remainder: $f_2 = r_1 + f_3$ where $r_1 = x$, $f_3 = y^2 + y$. Now continue as before: $f_3 - 1(y^2 - 1) = y + 1 = f_4$. This time no monomial in f_4 is divisible by $\text{LT}(g_1)$ or $\text{LT}(g_2)$, and so both terms in f_4 go into the remainder: $r_2 = r_1 + y$, $f_4 = y + f_5$ where $f_5 = 1$; $r_3 = r_2 + 1$, $f_5 = 1 + f_6$ where $f_6 = 0$. The algorithm stops when the polynomial to be divided is zero, and the remainder r is the current remainder $r_3 = x + y + 1$. The quotients q_1 and q_2 can be recovered from the steps of the algorithm involving g_1 and g_2 respectively: in this case we have $q_1 = x + y$, $q_2 = 1$.

Theorem 5.3 (Division Algorithm in $P = F[x_1, \dots, x_n]$) Fix a monomial order in P and let g_1, \dots, g_s be a finite list of nonzero polynomials in P . Then every $f \in P$ can be written in the form $f = q_1g_1 + \dots + q_s g_s + r$ for some quotients q_1, \dots, q_s and remainder r in P such that for $1 \leq i \leq s$

- no monomial which occurs in r is divisible by $\text{LT}(g_i)$,
- $\text{LM}(q_i g_i) \leq \text{LM}(f)$.

Proof We prove the theorem by giving an explicit algorithm for construction of the quotients q_1, \dots, q_s and the remainder r .

Input: g_1, \dots, g_s, f
Output: q_1, \dots, q_s, r
 $q_1 := 0; \dots; q_s := 0; r := 0$
 $p := f$

```

WHILE  $p \neq 0$  DO
   $i := 1$ 
  div := false
  WHILE  $i \leq s$  AND div = false DO
    IF  $\text{LT}(g_i)$  divides  $\text{LT}(p)$  THEN
       $q_i := q_i + (\text{LT}(p)/\text{LT}(g_i))$ 
       $p := p - (\text{LT}(p)/\text{LT}(g_i))g_i$ 
      div := true
    ELSE

```

$$i := i + 1$$

IF div = false THEN

$$r := r + \text{LT}(p)$$

$$p := p - \text{LT}(p)$$

The rest of the proof consists of checking that this algorithm operates correctly. The variable p represents the intermediate polynomial still to be divided, and the algorithm terminates when $p = 0$. The Boolean variable ‘div’ tells us whether the leading term of some g_i divides the leading term of p .

To prove that the algorithm works, we first check that the equation $f = q_1g_1 + \dots + q_sg_s + p + r$ holds at each stage. There are two cases, depending on whether a ‘division step’ (div = true) or a remainder step (div = false) has occurred. It follows that $f = q_1g_1 + \dots + q_sg_s + r$ when the algorithm stops. Since the leading monomial of p decreases in the given monomial order at each step, and since every strictly decreasing sequence in the monomial order is finite, the algorithm does always stop. You should check that the polynomials output by the algorithm have the required properties. \square

The next example shows that this algorithm does not have all the properties we would like.

Example 5.4 We repeat Example 5.2, interchanging the divisors. That is, using Lex in $F[x, y]$ with $x > y$, we will divide $f = x^2y + xy^2 + y^2$ by $g_1 = y^2 - 1$ and $g_2 = xy - 1$.

This time the algorithm produces $q_1 = x + 1$, $q_2 = x$, $r = 2x + 1$. So the output depends on the order of the input polynomials g_1, \dots, g_s ; in particular, the remainder is not uniquely specified by the properties stated in Theorem 5.3.

The next example shows that this problem can also affect the question of whether the remainder is zero.

Example 5.5 Using Lex in $F[x, y]$ with $x > y$, we will divide $f = xy^2 - x$ by $g_1 = xy + 1$ and $g_2 = y^2 - 1$. The result is $q_1 = y$, $q_2 = 0$, $r = -x - y$. However, if we take $g_1 = y^2 - 1$ and $g_2 = xy + 1$, we obtain instead $q_1 = x$, $q_2 = 0$ and $r = 0$.

Recall that f is in the ideal $I = \langle g_1, \dots, g_s \rangle$ if and only if $f = q_1g_1 + \dots + q_sg_s$ for *some* choice of q_1, \dots, q_s . Thus Example 5.5 shows that $xy^2 - x$ is in the ideal I generated by $xy + 1$ and $y^2 - 1$. Indeed, though $xy^2 - x = x(y^2 - 1)$ is in the smaller ideal generated by $y^2 - 1$ alone, the ‘bad’ choice of basis $\{xy + 1, y^2 - 1\}$ disguises the fact that $xy^2 - x \in I$.

To remedy this, we seek ‘good’ generating sets for an ideal I in P . For such a set, we want the remainder r to be independent of the order in which the generators are input to the division algorithm, and in particular we want to have $r = 0$ if and only if $f \in I$. We will see that Gröbner bases have precisely these nice properties.

Chapter 2: Computing with Polynomials

6 Gröbner bases

6.1 Leading term ideals

We have seen that **leading terms** play a central part in the division process, and that a choice of **monomial ordering** provides every nonzero polynomial f with a uniquely defined leading term $\text{LT}(f)$.

Definition 6.1 *Let I be a nonzero ideal in $F[x_1, \dots, x_n]$ and let $\text{LT}(I)$ be the set of all leading terms of elements of I . Then the ideal $\langle \text{LT}(I) \rangle$ generated by $\text{LT}(I)$ is called the **leading term ideal** of I .*

Notice that $\langle \text{LT}(I) \rangle$ is a **monomial ideal**, since $\text{LT}(f) = a\mathbf{x}^\alpha$ where $a \neq 0$ and $\mathbf{x}^\alpha = \text{LM}(f)$ is a monomial. Thus we could equally well have defined $\langle \text{LT}(I) \rangle$ to be the ideal generated by the leading monomials of elements of I .

Example 6.2 Using any monomial ordering in $F[x, y]$, let $I = \langle x - 1, y + 2 \rangle$. Then $\text{LT}(x - 1) = x$ and $\text{LT}(y + 2) = y$, so $\text{LT}(I)$ certainly contains $\langle x, y \rangle$, the ideal consisting of all polynomials with zero constant term. In fact we can prove $\text{LT}(I) = \langle x, y \rangle$: if not, then I must contain a polynomial whose leading term is constant, and such a polynomial is itself constant. But it is easy to see that I contains no non-zero constant polynomials.

The next example shows that it is **not** always true that $\langle \text{LT}(f_1), \text{LT}(f_2) \rangle$ is the leading term ideal of $\langle f_1, f_2 \rangle$.

Example 6.3 Let $I = \langle f_1, f_2 \rangle$, where $f_1 = x^3 - 2xy$, $f_2 = x^2y - 2y^2 + x$, and use the DegLex order in $F[x, y]$ with $x > y$. Then

$$x \cdot (x^2y - 2y^2 + x) - y \cdot (x^3 - 2xy) = x^2,$$

so that $x^2 \in I$. Hence $x^2 \in \langle \text{LT}(I) \rangle$. However $\text{LT}(f_1) = x^3$, $\text{LT}(f_2) = x^2y$ so the monomial ideal $\langle \text{LT}(f_1), \text{LT}(f_2) \rangle = \langle x^3, x^2y \rangle$ is strictly smaller than $\langle \text{LT}(I) \rangle$.

6.2 The Hilbert Basis Theorem

We now combine the idea of the leading term ideal with Dickson's Lemma to obtain a proof of the Hilbert Basis Theorem.

Proposition 6.4 *Let I be a nonzero ideal in $F[x_1, \dots, x_n]$. Then there exists a finite set of polynomials $g_1, \dots, g_s \in I$ such that $\langle \text{LT}(I) \rangle = \langle \text{LM}(g_1), \dots, \text{LM}(g_s) \rangle$.*

Proof Since $\langle \text{LT}(I) \rangle$ is a monomial ideal, this follows immediately from Dickson's Lemma (Theorem 3.6). \square

Theorem 6.5 (Hilbert Basis Theorem) *Let F be a field and let I be an ideal in $F[x_1, \dots, x_n]$. Then I is finitely generated, i.e. $I = \langle g_1, \dots, g_s \rangle$ for some $g_1, \dots, g_s \in I$.*

Proof If $I = 0$, we take $\{0\}$ as the generating set. Otherwise we take g_1, \dots, g_s as in Proposition 6.4. Since each $g_i \in I$, it is clear that $\langle g_1, \dots, g_s \rangle \subseteq I$. We claim that $I \subseteq \langle g_1, \dots, g_s \rangle$, so that in fact $I = \langle g_1, \dots, g_s \rangle$, and so g_1, \dots, g_s is the required finite generating set for I .

Thus let $f \in I$. By the division algorithm (Theorem 5.3), we can write f in the form $f = q_1g_1 + \dots + q_sg_s + r$, where no monomial appearing in the remainder r is divisible by any of the leading monomials $\text{LM}(g_1), \dots, \text{LM}(g_s)$. We shall prove by contradiction that $r = 0$.

If $r \neq 0$, then $\text{LT}(r) \in \langle \text{LT}(I) \rangle = \langle \text{LM}(g_1), \dots, \text{LM}(g_s) \rangle$. Since this is a monomial ideal, it follows from Theorem 3.2 that $\text{LT}(r)$ is divisible by $\text{LM}(g_i)$ for some i . This is a contradiction. Hence $r = 0$, and so $f = q_1g_1 + \dots + q_sg_s$. Thus $f \in \langle g_1, \dots, g_s \rangle$. \square

6.3 Noetherian rings

To complete our discussion of the Hilbert Basis Theorem, we shall relate it to the general structure theory of commutative rings.

Theorem 6.6 (Noetherian rings) *Let R be a commutative ring with 1. Then the following conditions are equivalent.*

- (i) *Every ideal I in R has a finite generating set.*
- (ii) *There is no infinite strictly increasing sequence of ideals in R ; more precisely, given a sequence $\{I_n\}_{n \geq 1}$ of ideals in R such that $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$, there exists a positive integer n_0 such that $I_n = I_{n_0}$ for all $n \geq n_0$.*

Condition (ii) is often called the **ascending chain condition (ACC)**. A ring R satisfying either, and hence both, of these conditions is called a **Noetherian ring**.

Proof (i) \Rightarrow (ii): Let $I_1 \subseteq I_2 \subseteq I_3 \subseteq \dots$ be an ascending chain. Define $I = \bigcup_{n=1}^{\infty} I_n$. It is easy to check that I is an ideal. By (i), I has a finite generating set f_1, \dots, f_s . Choose n_i such that $f_i \in I_{n_i}$ for $1 \leq i \leq s$, and define $n_0 = \max n_i$. Then f_1, \dots, f_s are all in I_{n_0} . Hence $I \subseteq I_{n_0}$. But $I_{n_0} \subseteq I$, hence $I_{n_0} = I$. Hence $I_n = I_{n_0}$ for all $n \geq n_0$.

(ii) \Rightarrow (i): We prove the contrapositive, i.e. 'not (i)' \Rightarrow 'not (ii)'. Let I be an ideal which is not finitely generated. Define an ascending chain $I_1 \subset I_2 \subset \dots \subset I_n$

as follows, by induction on n . Let r_1 be any element of I , and let $I_1 = \langle r_1 \rangle$. Since I is not finitely generated, $I \neq I_1$. Let r_2 be any element of $I \setminus I_1$, and let $I_2 = \langle r_1, r_2 \rangle$. Then $I_1 \subset I_2$. Assume as induction hypothesis that a strictly ascending chain $I_1 \subset I_2 \dots \subset I_n$ has been constructed, so that $I_s = \langle r_1, \dots, r_s \rangle$ and $r_s \notin I_{s-1}$ for $2 \leq s \leq n$. Since I is not finitely generated, $I \neq I_n$. Let r_{n+1} be any element of $I \setminus I_n$, and let $I_{n+1} = \langle r_1, \dots, r_{n+1} \rangle$. This completes the inductive construction of the ascending chain $\{I_n\}_{n \geq 1}$. It is clear from the construction that the ACC **(ii)** is false for this chain. \square

7 Gröbner bases

Definition 7.1 Let I be an ideal in P , and let $<$ be a monomial order. A finite set of polynomials g_1, \dots, g_s in I is called a **Gröbner basis** (or **standard basis**) of I if and only if

$$\langle \text{LT}(g_1), \dots, \text{LT}(g_s) \rangle = \langle \text{LT}(I) \rangle. \quad (14)$$

Our proof of Theorem 6.5 shows that

- a Gröbner basis for I is indeed a generating set for I , and
- every nonzero ideal I in $P = F[x_1, \dots, x_n]$ has a Gröbner basis.

However, the proof does **not** tell us

- how to find a Gröbner basis for a given ideal I , or
- how to check that a given set of polynomials in I is a Gröbner basis.

We shall return to these questions in the next section, but one special case is worth mentioning now.

Proposition 7.2 If $I = \langle g \rangle$ is the principal ideal generated by a nonzero $g \in P$, then $\{g\}$ is a Gröbner basis for I with respect to any monomial ordering.

Proof Clearly $\langle \text{LT}(g) \rangle \subseteq \langle \text{LT}(I) \rangle$. If $f \in I$ and $f \neq 0$, then $f = gh$ for some nonzero $h \in P$. Hence $\text{LT}(f) = \text{LT}(g) \cdot \text{LT}(h)$ and so $\text{LT}(f) \in \langle \text{LT}(g) \rangle$. Hence $\langle \text{LT}(I) \rangle \subseteq \langle \text{LT}(g) \rangle$ and so $\langle \text{LT}(I) \rangle = \langle \text{LT}(g) \rangle$. \square

7.1 Normal forms and quotient rings

A key property of Gröbner bases is that the remainder in the division algorithm is uniquely determined when we divide by a Gröbner basis.

Proposition 7.3 Fix a monomial ordering in $P = F[x_1, \dots, x_n]$, let I be an ideal in P , and let $f \in P$ be a polynomial. Then there is a unique $r \in P$ such that

- (i) $f = g + r$ for some $g \in I$, and
- (ii) no term of r is in $\langle \text{LT}(I) \rangle$.

The remainder r is sometimes called the **normal form** of f with respect to I .

Proof Let g_1, \dots, g_s be a Gröbner basis for I . By the division algorithm, Theorem 5.3, we can write f in the form $f = q_1g_1 + \dots + q_sg_s + r$ where the quotients q_1, \dots, q_s and the remainder r are in P . Thus $f = g+r$ where $g = q_1g_1 + \dots + q_sg_s \in I$, so that (i) is satisfied. Theorem 5.3 also tells us that no term of r is divisible by any of $\text{LT}(g_1), \dots, \text{LT}(g_s)$. Since $\langle \text{LT}(I) \rangle = \langle \text{LT}(g_1), \dots, \text{LT}(g_s) \rangle$, a monomial \mathbf{x}^α is in $\langle \text{LT}(I) \rangle$ if and only if it is divisible by some $\text{LT}(g_i)$. Hence if $\mathbf{x}^\alpha \in \langle \text{LT}(I) \rangle$, then \mathbf{x}^α does not divide any term of r .

Finally we must show that if r' also satisfies (i) and (ii) then $r = r'$. By (i), we have $f = g' + r'$ where $g' \in I$, so $r' - r = g - g' \in I$. If $r' - r \neq 0$, then $\text{LT}(r' - r) \in \langle \text{LT}(I) \rangle$. But no monomial in r or r' is divisible by any monomial in $\langle \text{LT}(I) \rangle$. This is a contradiction, hence $r = r'$. \square

The remainder r of f is sometimes written as \overline{f} . It is particularly convenient for calculations in the **quotient ring** P/I (Section 2.4).

There is an important familiar case of the quotient ring construction, when $F = \mathbf{R}$ and $I = \langle x^2 + 1 \rangle$. The ring $\mathbf{R}[x]/\langle x^2 + 1 \rangle$ is isomorphic to the field of complex numbers \mathbf{C} , the isomorphism being determined by the correspondence $x + I \longleftrightarrow i$. In this case, the normal form of a polynomial $f(x)$ is the remainder $r(x) = a + bx$ on division of $f(x)$ by $g(x) = x^2 + 1$. (Recall that, by Proposition 7.2, the generator of a principal ideal is a Gröbner basis for the ideal.) Thus $\langle \text{LT}(I) \rangle = \langle x^2 \rangle$. Now the usual method of calculating with complex numbers is equivalent to calculating with these remainders, subject to the rule that $x^2 = -1$, i.e. the remainder of $f(x) = x^2$ is $r(x) = -1$.

The following result shows how to generalise this type of calculation in quotient rings.

Proposition 7.4 *Fix a monomial ordering in $P = F[x_1, \dots, x_n]$, and let I be an ideal in P . Then*

- (i) *the correspondence $r \longleftrightarrow r + I$ is a bijection between the set of all remainders r of polynomials in P and the set P/I of all cosets of I in P ;*
- (ii) *the set of cosets which correspond to monomials $\mathbf{x}^\alpha \in P$ such that $\mathbf{x}^\alpha \notin \langle \text{LT}(I) \rangle$ is a basis for P/I as a vector space over F ;*
- (iii) *the quotient ring P/I is isomorphic to the ring whose elements are the remainders of polynomials in P , with the same addition operation as in P but with the product of r_1 and r_2 taken as the remainder of r_1r_2 .*

Proof By taking $r = 0$, Proposition 7.3 shows that $f \in I$ if and only if $\overline{f} = 0$; and by taking $g = 0$, Proposition 7.3 shows that $f = \overline{f}$ if and only if no term of f is in $\langle \text{LT}(I) \rangle$. It is also easy to see from Proposition 7.3 again that $\overline{f_1 + f_2} = \overline{f_1} + \overline{f_2}$ and that the remainder of $\overline{f_1} \cdot \overline{f_2}$ is $\overline{f_1f_2}$. The details are left as an exercise. \square

8 S -polynomials

In this section we shall see how to check that a given set of polynomials in an ideal I of P is a Gröbner basis. This in turn will lead to an algorithm for the construction of Gröbner bases.

Definition 8.1 Let f, g be nonzero polynomials and let

$$\mathbf{x}^\gamma = \text{LCM}(\text{LM}(f), \text{LM}(g))$$

be the least common multiple of their leading monomials. The S -polynomial $S(f, g)$ is defined by

$$S(f, g) = \frac{\mathbf{x}^\gamma}{\text{LT}(f)} \cdot f - \frac{\mathbf{x}^\gamma}{\text{LT}(g)} \cdot g. \quad (15)$$

Note that $S(f, f) = 0$ and $S(g, f) = -S(f, g)$ for all $f, g \in P$.

Example 8.2 Let $f = x^2y + y^2 + x$, $g = 2xy^2 - x^2$, where the ordering is DegLex with $x > y$. Then $\text{LT}(f) = x^2y$, $\text{LT}(g) = 2xy^2$ and $\mathbf{x}^\gamma = x^2y^2$. Thus

$$\begin{aligned} S(f, g) &= \frac{x^2y^2}{x^2y}(x^2y + y^2 + x) - \frac{x^2y^2}{2xy^2}(2xy^2 - x^2), \\ &= \frac{1}{2}x^3 + y^3 + xy. \end{aligned}$$

The important thing to notice about this example is that the leading terms of f and g have cancelled, and that although $S(f, g) = yf - \frac{1}{2}xg \in I$ where $I = \langle f, g \rangle$, the leading term $\text{LT}(S(f, g)) = \frac{1}{2}x^3 \notin \langle x^2y, xy^2 \rangle$, the monomial ideal generated by the leading terms of f and g .

Since $S(f, g) \in I$ and $\text{LT}(S(f, g)) \notin \langle \text{LT}(f), \text{LT}(g) \rangle$, it follows that f and g do not form a Gröbner basis for the ideal I . It is natural to add $h = S(f, g)$ to the generating set $\{f, g\}$ for I , and to ask whether $\{f, g, h\}$ is a Gröbner basis for I .

We shall see that this method of adding S -polynomials to a given generating set for an ideal I in P always succeeds in producing a Gröbner basis for I in a finite number of steps. Somehow, S -polynomials of pairs of polynomials in the set $\{g_1, \dots, g_r\}$ account for all cancellations of leading terms in forming elements of the ideal $\langle g_1, \dots, g_r \rangle$. The next result is a key property of S -polynomials.

Proposition 8.3 Let $f_1, \dots, f_s \in P$ be polynomials with the same leading monomial \mathbf{x}^δ , let $g = \sum_{i=1}^s c_i f_i$ be a linear combination of the f_i with coefficients $c_i \in F$, and suppose that $\text{LM}(g) < \mathbf{x}^\delta$. Then g is a linear combination of the S -polynomials $S(f_i, f_{i+1})$, $1 \leq i \leq s-1$, and $\text{LM}(S(f_i, f_j)) < \mathbf{x}^\delta$ for all pairs $i \neq j$.

Proof First we do the special case where the f_i are **monic**, i.e. $\text{LC}(f_i) = 1$ for $1 \leq i \leq s$. Since $\text{LT}(f_i) = \text{LT}(f_j) = \mathbf{x}^\delta$ in this case, $S(f_i, f_j) = f_i - f_j$. Clearly $\text{LM}(S(f_i, f_j)) < \mathbf{x}^\delta$ since the leading terms cancel.

Consider the ‘telescoping’ sum

$$\sum_{i=1}^s (c_1 + \dots + c_i)(f_i - f_{i+1}),$$

where $f_{s+1} = 0$. For $1 \leq i \leq s$ the coefficient of f_i in this sum is $(c_1 + \dots + c_i) - (c_1 + \dots + c_{i-1}) = c_i$, so the sum reduces to $\sum_{i=1}^s c_i f_i = g$. The leading coefficient of g is $\sum_{i=1}^s c_i$ and so the condition $\text{LM}(g) < \mathbf{x}^\delta$ is equivalent to $\sum_{i=1}^s c_i = 0$. If this holds, then the last term in the telescoping sum vanishes, and we have $g = \sum_{i=1}^{s-1} (c_1 + \dots + c_i)S(f_i, f_{i+1})$.

In the general case, let $\text{LC}(f_i) = a_i$ for $1 \leq i \leq s$, so that $g = \sum_{i=1}^s c_i a_i (f_i/a_i)$ and $\text{LC}(g) = \sum_{i=1}^s c_i a_i$. Since f_i/a_i is monic and $S(f_i, f_j) = S(f_i/a_i, f_j/a_j)$, the general case follows by applying the special case to the polynomials f_i/a_i . \square

The following property of S -polynomials will be needed in the next section.

Proposition 8.4 *Given $f, g \in P$ and monomials $\mathbf{x}^\alpha, \mathbf{x}^\beta$,*

$$S(\mathbf{x}^\alpha f, \mathbf{x}^\beta g) = \mathbf{x}^\gamma S(f, g)$$

for some monomial \mathbf{x}^γ .

Proof Let $\text{LT}(f) = a\mathbf{x}^\rho$, $\text{LT}(g) = b\mathbf{x}^\sigma$, $a, b \in F$, so that $\text{LCM}(\text{LM}(f), \text{LM}(g)) = \mathbf{x}^\tau$, where $\tau_i = \max(\rho_i, \sigma_i)$. Then $\text{LT}(\mathbf{x}^\alpha f) = a\mathbf{x}^{\alpha+\rho}$, $\text{LT}(\mathbf{x}^\beta g) = b\mathbf{x}^{\beta+\sigma}$, so that $\text{LCM}(\text{LT}(\mathbf{x}^\alpha f), \text{LT}(\mathbf{x}^\beta g)) = \mathbf{x}^\delta$ where $\delta_i = \max(\alpha_i + \rho_i, \beta_i + \sigma_i)$.

With this notation, we have

$$\begin{aligned} S(f, g) &= \frac{\mathbf{x}^\tau}{a\mathbf{x}^\rho} f - \frac{\mathbf{x}^\tau}{b\mathbf{x}^\sigma} g \\ S(\mathbf{x}^\alpha f, \mathbf{x}^\beta g) &= \frac{\mathbf{x}^\delta}{a\mathbf{x}^{\alpha+\rho}} (\mathbf{x}^\alpha f) - \frac{\mathbf{x}^\delta}{b\mathbf{x}^{\beta+\sigma}} (\mathbf{x}^\beta g) \\ &= \frac{\mathbf{x}^\delta}{a\mathbf{x}^\rho} f - \frac{\mathbf{x}^\delta}{b\mathbf{x}^\sigma} g \\ &= \mathbf{x}^\gamma S(f, g) \end{aligned}$$

where $\gamma_i = \delta_i - \tau_i$.

We must check that \mathbf{x}^γ really is a monomial, i.e. $\gamma_i \geq 0$ for all i . For this, note that the inequalities $\alpha_i + \rho_i \geq \rho_i$, $\beta_i + \sigma_i \geq \sigma_i$ imply that $\delta_i \geq \rho_i$ and $\delta_i \geq \sigma_i$, and hence $\delta_i \geq \tau_i$ as required.

9 Buchberger's criterion

We next prove Buchberger's S -polynomial criterion for a set of polynomials to be a Gröbner basis for the ideal that it generates. Recall from the division algorithm that if the remainder on division of f by a set of polynomials $G = \{g_1, \dots, g_s\}$ taken in some order is zero, then f can be written in the form $f = \sum_{i=1}^s q_i g_i$ where $\text{LM}(q_i g_i) \leq \text{LM}(f)$ for $1 \leq i \leq s$.

Theorem 9.1 (Buchberger) *Let $g_1, \dots, g_s \in P$ and let $I = \langle g_1, \dots, g_s \rangle$. Then $G = \{g_1, \dots, g_s\}$ is a Gröbner basis for the ideal I if and only if for all $i \neq j$ the remainder of $S(g_i, g_j)$ on division by G (taking the elements of G in some order) is zero.*

Proof If G is a Gröbner basis for I , then by Proposition 7.3 the remainder of any element of I with respect to G is 0. Since $S(g_i, g_j) \in I$, this proves the necessity of Buchberger's criterion.

To prove sufficiency, let $f \in I$, so that

$$f = \sum_{i=1}^s h_i g_i \quad (16)$$

for some polynomials $h_1, \dots, h_s \in P$. Among all such expressions, we choose one where $\max_i \text{LM}(h_i g_i)$ is minimal in the given monomial order. Let us denote this monomial by \mathbf{x}^δ . By (16), $\text{LM}(f) \leq \mathbf{x}^\delta$.

If $\text{LM}(f) = \mathbf{x}^\delta$, then $\text{LM}(f) = \text{LM}(h_i g_i) = \text{LM}(h_i) \cdot \text{LM}(g_i)$ for some i , so that $\text{LM}(g_i)$ divides $\text{LM}(f)$, and hence $\text{LM}(f) \in \langle \text{LM}(g_1), \dots, \text{LM}(g_s) \rangle$. If this is true for all $f \in I$, then we have proved that $\langle \text{LM}(I) \rangle = \langle \text{LM}(g_1), \dots, \text{LM}(g_s) \rangle$, i.e. G is a Gröbner basis for I , as required.

Thus it remains to prove that $\text{LM}(f) = \mathbf{x}^\delta$. We shall assume that $\text{LM}(f) < \mathbf{x}^\delta$ and obtain a contradiction using Proposition 8.3.

Let us renumber the polynomials g_1, \dots, g_s so that the terms in (16) with leading monomial \mathbf{x}^δ come before those with smaller leading monomials, i.e. $\text{LM}(h_i g_i) = \mathbf{x}^\delta$ for $1 \leq i \leq t$ and $\text{LM}(h_i g_i) < \mathbf{x}^\delta$ for $i > t$. For $1 \leq i \leq t$, let $\text{LT}(h_i) = c_i \mathbf{x}^{\alpha(i)}$ where $c_i \in F$, and let $f_i = \mathbf{x}^{\alpha(i)} g_i$. Then we can write (16) as

$$f = \sum_{i=1}^t c_i f_i + f_0, \quad (17)$$

where $f_0 = \sum_{i=1}^t (h_i - \text{LT}(h_i)) g_i + \sum_{i=t+1}^s h_i g_i$. Thus $\text{LM}(f_i) = \mathbf{x}^\delta$ for $1 \leq i \leq t$, while $\text{LM}(f_0) < \mathbf{x}^\delta$. It follows from (17) that $\text{LM}(g) < \mathbf{x}^\delta$, where $g = \sum_{i=1}^t c_i f_i$.

Thus the leading terms of the f_i cancel when we form the linear combination $g = \sum_{i=1}^t c_i f_i$. By Proposition 8.3 we can write g in the form

$$g = \sum_{i=1}^{t-1} b_i S(f_i, f_{i+1})$$

where $b_i \in F$. Since $f_i = \mathbf{x}^{\alpha(i)} g_i$, Proposition 8.4 shows that $S(f_i, f_{i+1}) = \mathbf{x}^{\gamma(i)} S(g_i, g_{i+1})$ for some monomial $\mathbf{x}^{\gamma(i)}$. Hence

$$g = \sum_{i=1}^{t-1} b_i \mathbf{x}^{\gamma(i)} S(g_i, g_{i+1}).$$

We are now ready to use the hypothesis that all the S -polynomials of pairs of polynomials in the set G have remainder 0. Thus by the division algorithm (Theorem 5.3) we can write $S(g_i, g_{i+1})$ in the form

$$S(g_i, g_{i+1}) = \sum_{j=1}^s s_{i,j} g_j$$

for some polynomials $s_{i,j} \in P$ such that $\text{LM}(s_{i,j} g_j) \leq \text{LM}(S(g_i, g_{i+1}))$ for all i, j . Hence

$$g = \sum_{i=1}^{t-1} b_i \mathbf{x}^{\gamma(i)} \sum_{j=1}^s s_{i,j} g_j = \sum_{j=1}^s k_j g_j,$$

where $k_j = \sum_{i=1}^{t-1} b_i \mathbf{x}^{\gamma(i)} s_{i,j}$. Now notice that

$$\begin{aligned} \text{LM}(k_j g_j) &\leq \max_{1 \leq i \leq t-1} \text{LM}(\mathbf{x}^{\gamma(i)} s_{i,j} g_j) \\ &\leq \max_{1 \leq i \leq t-1} \text{LM}(\mathbf{x}^{\gamma(i)} S(g_i, g_{i+1})) \\ &\leq \max_{1 \leq i \leq t-1} \text{LM}(S(f_i, f_{i+1})) \\ &< \mathbf{x}^\delta, \end{aligned}$$

since the terms in \mathbf{x}^δ in f_i and f_{i+1} cancel in $S(f_i, f_{i+1})$.

Now substitute $\sum_{i=1}^t c_i f_i = \sum_{j=1}^s k_j g_j$ in (17), and notice that $f_0 = \sum_{i=1}^s h'_i g_i$ where $\text{LM}(h'_i g_i) < \mathbf{x}^\delta$ for all i . After collecting terms in g_1, \dots, g_s , we get an expression $f = \sum_{i=1}^s h''_i g_i$ where $h''_i \in P$ and $\text{LM}(h''_i g_i) < \mathbf{x}^\delta$ for $1 \leq i \leq s$. But this contradicts the definition of \mathbf{x}^δ . Thus Buchberger's Criterion is proved. \square

10 Buchberger's algorithm

Buchberger's S -polynomial criterion actually gives an algorithm for constructing Gröbner bases. Given a monomial ordering on P and a finite set of polynomials $f_1, \dots, f_r \in P$, we can construct a Gröbner basis G for the ideal $I = \langle f_1, \dots, f_r \rangle$ by adding S -polynomials to the original list of generators f_1, \dots, f_r in a systematic way.

Theorem 10.1 (Buchberger's Algorithm) *Let $I = \langle f_1, \dots, f_t \rangle$ be a nonzero ideal in $P = F[x_1, \dots, x_n]$ and let $<$ be a monomial ordering in P . Then a Gröbner basis G for I can be constructed in a finite number of steps by the following algorithm.*

Input: (f_1, \dots, f_t) , a list of polynomials

Output: $G = (g_1, \dots, g_s)$, a Gröbner basis G for $I = \langle f_1, \dots, f_t \rangle$

$G := (f_1, \dots, f_t)$

REPEAT

$G' := G$

FOR each pair $\{p, q\}$, $p \neq q$, in G' DO

$r :=$ remainder of $S(p, q)$ on division by G'

IF $r \neq 0$ THEN $G := G \cup \{r\}$

UNTIL $G = G'$

Proof First note that if $p, q \in I$ then $S(p, q) \in I$. Initially, the list G is the given list of generators (f_1, \dots, f_t) for I . As the algorithm proceeds, we add to G all the remainders r of S -polynomials of pairs of elements in the current list G' on division by polynomials which are already in G' . These remainders must also be in I , and so, at every stage of the algorithm, G is a list of generators for the same ideal I .

The algorithm terminates when all these remainders are 0. It follows immediately from Theorem 9.1 that at this point the list G is a Gröbner basis for I .

It remains to prove that the algorithm does terminate for all possible inputs. Consider the monomial ideal $\langle \text{LT}(G) \rangle$ generated by the leading monomials of all the elements of G . The new remainder r contains no monomials which are divisible by the leading terms of elements of G , so in particular $\text{LT}(r) \notin \langle \text{LT}(G) \rangle$. Thus the sequence of ideals $\langle \text{LT}(G) \rangle$ which arise as the algorithm proceeds is an ascending chain. By Theorem 6.6, such a chain must stop after a finite number of steps. Hence the algorithm terminates. \square

Example 10.2 In $F[x, y]$ with Lex and $x < y$, let $f_1 = xy - x$, $f_2 = -y + x^2$, $I = \langle f_1, f_2 \rangle$.

We start with $g_1 = f_1$, $g_2 = f_2$, $G = (g_1, g_2)$. Then $S(g_1, g_2) = (xy - x) + x(-y + x^2) = -x + x^3$. Since $\text{LM}(-x + x^3) = x^3$ is not divisible by $\text{LM}(g_1) = xy$ or $\text{LM}(g_2) = -y$, $r = S(g_1, g_2) = x^3 - x$ in this case.

We add r to the basis: $g_3 = x^3 - x$, and REPEAT. This time $S(g_1, g_2) = -x + x^3$ (of course! — this algorithm is not very efficient). But now the remainder $r = 0$, since we have included g_3 as a divisor. However, we still have to deal with $S(g_1, g_3)$ and $S(g_2, g_3)$.

We get $S(g_1, g_3) = x^2(xy - x) - y(x^3 - x) = -x^3 + xy$. The leading monomial is $xy = \text{LM}(g_1)$ so this is NOT the remainder. In fact it is easy to see that $S(g_1, g_3) = g_1 - g_3$, so $r = 0$.

We get $S(g_2, g_3) = x^3(-y + x^2) + y(x^3 - x) = x^5 - xy$. Again division is possible and we find $x^5 - xy = (-xy + x) + (x^2 + 1)(x^3 - x) = -g_1 + (x^2 + 1)g_3$. So once again $r = 0$.

Since no new polynomials have been added in this pass through the REPEAT loop, the algorithm stops. The resulting Gröbner basis is $\{xy - x, -y + x^2, x^3 - x\}$.

Example 10.3 As in Example 6.3, let $I = \langle f_1, f_2 \rangle$, where $f_1 = x^3 - 2xy$, $f_2 = x^2y - 2y^2 + x$, and use the DegLex order in $F[x, y]$ with $x > y$. Then

$$S(f_2, f_1) = x \cdot (x^2y - 2y^2 + x) - y \cdot (x^3 - 2xy) = x^2,$$

and we see that the remainder is still x^2 , so, as expected from Example 6.3, we must include x^2 in the generating set, giving $g_1 = x^3 - 2xy$, $g_2 = x^2y - 2y^2 + x$, $g_3 = x^2$.

Now we REPEAT starting with $G = (g_1, g_2, g_3)$.

This time $S(g_1, g_3) = (x^3 - 2xy) - x(x^2) = -2xy$, and $S(g_2, g_3) = (x^2y - 2y^2 + x) - y(x^2) = -2y^2 + x$. The leading terms are not divisible by any of $\text{LM}(g_1), \text{LM}(g_2), \text{LM}(g_3)$, so we take $g_4 = -2xy$, $g_5 = -2y^2 + x$. One more run through the REPEAT loop shows that we finally have a Gröbner basis

$$\{x^3 - 2xy, x^2y - 2y^2 + x, x^2, -2xy, -2y^2 + x\}.$$

The algorithm in its most basic form is not very efficient. In practice it is better to add the new remainders to G **one at a time**. Then we only have to check the S -polynomials involving the current set G and the new remainder to be added to it. In Example 10.3, this gives the following sequence of calculations.

$S(g_1, g_2)$, getting g_3 :

$S(g_1, g_3)$, getting g_4 ; $S(g_2, g_3)$, getting g_5 :

$S(g_1, g_4)$, $S(g_2, g_4)$, $S(g_3, g_4)$, getting no new polynomials:

$S(g_1, g_5)$, $S(g_2, g_5)$, $S(g_3, g_5)$, $S(g_4, g_5)$, getting no new polynomials:

DONE!

11 Minimal and reduced Gröbner bases

A nonzero polynomial ideal I has infinitely many Gröbner bases, since we can add any element of I to an existing Gröbner basis to get a new one, and we can also multiply any of the polynomials in the Gröbner basis by a nonzero scalar. Thus we are interested in making our Gröbner bases as small (and tidy) as possible. In particular, the Gröbner basis G output by Buchberger's algorithm may contain 'redundant' generators g_i whose leading term $\text{LT}(g_i)$ is contained in the ideal generated by the leading terms of the other elements of G . Thus in Example 10.3 we can drop out g_1 and g_2 from G , since their leading terms x^3 and x^2y are divisible by the leading term x^2 of g_3 . This gives a smaller Gröbner basis $\{x^2, -2xy, -2y^2 + x\}$.

Definition 11.1 A Gröbner basis $G = \{g_1, \dots, g_s\}$ is **minimal** if and only if

- (i) each g_i is **monic**, i.e. $\text{LC}(g_i) = 1$ for all i , and
- (ii) $\text{LT}(g_i)$ does not divide $\text{LT}(g_j)$ if $i \neq j$.

If (ii) is replaced by the stronger condition

- (iii) $\text{LT}(g_i)$ does not divide any term of g_j if $i \neq j$,

we say that G is a **reduced** Gröbner basis.

Thus in Example 10.3 the Gröbner basis $\{x^2, -2xy, -2y^2 + x\}$ leads us easily to the minimal Gröbner basis $\{x^2, xy, y^2 - \frac{1}{2}x\}$, which is in fact reduced. In the same way, the Gröbner basis $\{xy - x, -y + x^2, x^3 - x\}$ of Example 10.2 leads to the reduced Gröbner basis $\{y - x^2, x^3 - x\}$.

We now have the following amazing result.

Theorem 11.2 Every nonzero polynomial ideal I in $F[x_1, \dots, x_n]$ has a unique reduced Gröbner basis with respect to a given monomial ordering.

Proof First we prove **existence** of a reduced basis. It is clear how to obtain a minimal Gröbner basis: starting from any Gröbner basis, we first omit any g_i such that $\text{LM}(g_i)$ is divisible by $\text{LM}(g_j)$ for some $j \neq i$, and then divide each remaining g_i by $\text{LC}(g_i)$ to make it monic.

Thus let $G = \{g_1, \dots, g_s\}$ be a minimal Gröbner basis for I . The idea is to work through this basis, replacing each element by its remainder with respect to the other elements of the (current) basis. We start by replacing g_1 by h_1 , its remainder with respect to $\{g_2, \dots, g_s\}$, then replace g_2 by h_2 , its remainder with respect to $\{h_1, g_3, \dots, g_s\}$, and so on until we finally replace g_s by h_s , its remainder with respect to $\{h_1, h_2, \dots, h_{s-1}\}$. We claim that this results in a reduced basis $\{h_1, \dots, h_s\}$.

To see this, notice that this reduction process does not change the leading terms, i.e. $LT(h_i) = LT(g_i)$. Hence we have a minimal Gröbner basis at each stage of the reduction. Finally note that, in the division algorithm, any term which appears in the remainder is not divisible by the leading monomials of any of the divisors.

Next we prove **uniqueness**. Thus let $G = \{g_1, \dots, g_s\}$ and $H = \{h_1, \dots, h_t\}$ be two reduced Gröbner bases for I . We first show that $s = t$ and that the leading terms of the two bases are the same, i.e. we can renumber the h 's so that $LM(g_i) = LM(h_i)$ for all i . Consider g_1 . Since $g_1 \in I$ and H is a Gröbner basis for I , there is an i such that $LM(h_i)$ divides $LM(g_1)$. But since $h_i \in I$ and G is a Gröbner basis for I , there is a j such that $LM(g_j)$ divides $LM(h_i)$. Hence $LM(g_j)$ divides $LM(g_1)$. Since G is minimal, it follows that $j = 1$ and hence $LM(h_i) = LM(g_1)$. Renumber H so that h_i becomes h_1 . Now consider g_2 . In the same way there is an i such that $LM(h_i)$ divides $LM(g_2)$. This cannot be h_1 , since $LM(h_1) = LM(g_1)$ and $LM(g_1)$ does not divide $LM(g_2)$, by minimality of G . Now $LM(g_j)$ divides $LM(h_i)$ for some j , and again minimality of G implies that $j = 2$. Hence $LM(g_2) = LM(h_i)$, and we renumber H_i as h_2 . It is clear that this process can be continued until all the elements of G and H are paired off.

Finally we must prove that $g_i = h_i$ for $1 \leq i \leq s$. Let $f = g_i - h_i$, and assume $f \neq 0$ so that $LM(f)$ is a monomial. Since we have proved that $LT(g_i) = LT(h_i)$, we have $LM(f) < LM(g_i)$. Since $f \in I$ and G is a Gröbner basis for I , $LM(g_j)$ divides $LT(f)$ for some j . Clearly $j \neq i$ since $LM(f) < LM(g_i)$. But $LM(g_j) = LM(h_j)$ does not divide any term of g_i or h_i , since the Gröbner bases G and H are both reduced. Hence $LM(g_j)$ does not divide any term of f , and in particular it does not divide $LT(f)$. This is a contradiction. Hence $f = 0$. \square

This result gives a way to test whether two sets of polynomials are generators for the same ideal. We need only compute a reduced Gröbner basis starting from each set in turn. The ideals will be equal if and only if the reduced Gröbner bases are the same set.

Many computer algebra systems (e.g. **Maple**, **Mathematica**) can be used to compute Gröbner bases. These always return the reduced Gröbner basis with respect to a given monomial ordering (except that, to avoid fractions, the basis may not be monic). Thus it is easy to compare results obtained using different systems. These systems use various implementations of (refinements of) Buchberger's algorithm. Because of the wide applications of Gröbner bases in algebra and geometry, these packages are useful to scientists and engineers as well as to mathematicians.

It is worth mentioning that when the given generating set for I consists of **linear** polynomials, the reduced Gröbner basis corresponds to the 'reduced row-echelon form' of the coefficient matrix of the corresponding system of linear equations.

12 Applications of Gröbner bases

12.1 The ideal membership problem

Problem: Given an ideal $I = \langle f_1, \dots, f_s \rangle$ in P , how can we decide whether a given polynomial f lies in I ?

Here is how.

1. Choose a monomial ordering $<$ in P .
2. Compute a Gröbner basis G for I with respect to $<$.
3. Find the remainder r of f on division by G .

Then $f \in I$ if and only if $r = 0$.

Example 12.1 Determine whether (i) $x^2y^4 - x^6$ and (ii) $x^4y^2 + y^2$ are in the ideal I of $F[x, y]$ generated by $xy - x$ and $-y + x^2$.

In Example 10.2 we have already calculated a Gröbner basis of I with respect to the Lex order with $x < y$. We also noted above that $G = \{g_1, g_2\}$, where $g_1 = y - x^2$, $g_2 = x^3 - x$ is a reduced Gröbner basis in this case, so we may as well use it. (But the method works with any Gröbner basis.)

Using the division algorithm, we find that

$$x^2y^4 - x^6 = (x^2y^3 + x^4y^2 + x^6y + x^8)(y - x^2) + (x^7 + x^5)(x^3 - x).$$

The remainder $r = 0$. We conclude that $x^2y^4 - x^6 \in I$.

A similar calculation gives

$$x^4y^2 + y^2 = (x^4y + y + x^6 + x^2)(y - x^2) + (x^5 + x^3 + 2x)(x^3 - x) + 2x^2,$$

so that the remainder $r = 2x^2$ is nonzero. We conclude that $x^4y^2 + y^2 \notin I$.

Notice that in both cases we have arranged the terms of the quotients q_1 and q_2 in decreasing Lex order.

There is another way to think about the ideal membership problem. If we have a set of equations $f_1 = 0, \dots, f_s = 0$ corresponding to the generators of the ideal I , we can ask whether the equation $f = 0$ is a consequence of these. If $f \in I$, then $f = q_1f_1 + \dots + q_sf_s = 0$ when all $f_i = 0$. In Example 12.1, the equations $xy = x$ and $x^2 = y$ imply $x^2y^4 - x^6 = (xy)^2 \cdot y^2 - x^6 = x^2 \cdot (x^2)^2 - x^6 = 0$. Gröbner bases provide a way of doing calculations of this kind systematically, so that they can be done by a computer.

12.2 The elimination problem

Problem: Given a set of polynomial equations $f_1 = 0, \dots, f_t = 0$ in $F[x_1, \dots, x_r, y_1, \dots, y_s]$, how can we eliminate the variables x_1, \dots, x_r from these equations so as to obtain equations involving only y_1, \dots, y_s ?

To get an idea of what is involved here, we consider a simple example.

Example 12.2 Consider the equations $x = t^4, y = t^3, z = t^2$. We can think of these as parametric equations of a curve C in \mathbf{R}^3 . Elimination of the parameter t between pairs of equations gives equations $x = z^2, y^2 = z^3, x^3 = y^4$. Geometrically, each of these 3 equations represents a surface which contains the curve C . Notice also that we can eliminate z from the equations $x = z^2, y^2 = z^3$ to get the equation $x^3 = y^4$.

From the point of view of ideal theory, the elimination problem is equivalent to looking for polynomials in the ideal $\langle f_1, \dots, f_t \rangle$ involving only the variables y_1, \dots, y_s . In Example 12.2, it is easy to check that $x - z^2, y^2 - z^3$ and $x^3 - y^4$ are in the ideal $I = \langle x - t^4, y - t^3, z - t^2 \rangle$, e.g. $x - z^2 = (x - t^4) - (z + t^2)(z - t^2)$, $y^2 - z^3 = (y + t^3)(y - t^3) - (z^2 + zt^2 + t^4)(z - t^2)$.

If we compute the reduced Gröbner basis of I with respect to Lex order with $t > x > y > z$, we obtain

$$\{t^2 - z, ty - z^2, tz - y, x - z^2, y^2 - z^3\}.$$

Notice that the polynomials here which do not involve t are $x - z^2$ and $y^2 - z^3$. Since t was chosen as the largest variable in the ordering, t is eliminated first in computing the Gröbner basis.

Theorem 12.3 Let I be an ideal in $F[x_1, \dots, x_r, y_1, \dots, y_s]$, let G be a Gröbner basis of I with respect to Lex order with $x_i > y_j$ for all i, j , and let $G' = G \cap F[y_1, \dots, y_s]$. Then G' is a Gröbner basis of $I' = I \cap F[y_1, \dots, y_s]$.

Proof Clearly $G' \subseteq I'$, so we need only prove that if $f \in I'$ then $\text{LT}(f)$ is divisible by $\text{LT}(g)$ for some $g \in G'$. Since $f \in I$ and G is a Gröbner basis for I , $\text{LT}(f)$ is divisible by $\text{LT}(g)$ for some $g \in G$. Since $f \in F[y_1, \dots, y_s]$, it follows that $\text{LT}(g) \in F[y_1, \dots, y_s]$. Now since we are using Lex order with $x_i > y_j$ for all i, j , any monomial involving x 's is $>$ any monomial involving only y 's. Hence g cannot have any term involving the x 's, i.e. $g \in G'$. \square

12.3 The ideal intersection problem

Problem: Given ideals $I = \langle g_1, \dots, g_k \rangle$ and $J = \langle h_1, \dots, h_l \rangle$ in $F[x_1, \dots, x_n]$, find a generating set for the ideal $I \cap J$.

We saw in Section 3.3 how to solve this problem in the case where I and J are **monomial** ideals. The elimination theorem (Theorem 12.3) provides a trick to solve this problem in general, by introducing a dummy variable which we then eliminate.

Theorem 12.4 *Given ideals I and J in $P = F[x_1, \dots, x_n]$, let t be a new variable. Then $I \cap J = K \cap P$, where K is the ideal $\langle tI, (1-t)J \rangle$ in $P[t]$.*

Proof Given $f \in I \cap J$, by writing f in the form $tf + (1-t)f$ we see that $f \in tI + (1-t)J$. Hence $f \in K \cap P$. Conversely, if $f \in K \cap P$, then we can write f in the form

$$f = \sum_{i=1}^k tg_i p_i + \sum_{j=1}^l (1-t)h_j q_j,$$

where $I = \langle g_1, \dots, g_k \rangle$ and $J = \langle h_1, \dots, h_l \rangle$, and the p_i 's and q_j 's are polynomials in $P[t]$.

This is a polynomial identity and so it remains true when we specialise the variables to elements of F . In particular, setting $t = 1$ we obtain

$$f = \sum_{i=1}^k g_i p_i(x_1, \dots, x_n, 1) \in I,$$

and $t = 0$ gives

$$f = \sum_{j=1}^l h_j q_j(x_1, \dots, x_n, 0) \in J,$$

so that $f \in I \cap J$. □

This method in fact gives a Gröbner basis for $I \cap J$, starting from any generating sets for I and J .

Example 12.5 Find a Gröbner basis for $I \cap J$ in $F[x, y]$, where $I = \langle x^2 + y^2, xy \rangle$ and $J = \langle x^2 - y^2 \rangle$.

The ideal $K = \langle tI, (1-t)J \rangle \subset F[x, y, t]$ is given by

$$K = \langle t(x^2 + y^2), txy, (1-t)(x^2 - y^2) \rangle.$$

Using Lex order with $t > x > y$, a Gröbner basis for K is

$$\{2tx^2 - x^2 + y^2, txy, 2ty^2 + x^2 - y^2, x^3 - xy^2, x^2y - y^3\}.$$

By the Elimination Theorem 12.3, it follows that $\{x^3 - xy^2, x^2y - y^3\}$ is a Gröbner basis for $K \cap F[x, y]$. By Theorem 12.4, $K \cap F[x, y] = I \cap J$.

Chapter 3: Symmetric Polynomials

13 Elementary symmetric functions

Definition 13.1 A polynomial $f \in P = F[x_1, \dots, x_n]$ is **symmetric** if and only if f is invariant under any permutation of x_1, \dots, x_n , i.e. if $\sigma : \{1, 2, \dots, n\} \rightarrow \{1, 2, \dots, n\}$ is any bijection, then $f(x_{\sigma(1)}, \dots, x_{\sigma(n)}) = f(x_1, \dots, x_n)$.

The following examples of symmetric polynomials will be important.

- the r th elementary symmetric function

$$e_r = \sum_{i_1 < i_2 < \dots < i_r} x_{i_1} x_{i_2} \cdots x_{i_r}$$

is the sum of all products of the n variables x_1, \dots, x_n taken r at a time. In particular, $e_1 = x_1 + \dots + x_n$ is the sum of all the variables, and $e_n = x_1 x_2 \cdots x_n$ is their product.

- the r th complete symmetric function

$$h_r = \sum_{\alpha_1 + \alpha_2 + \dots + \alpha_n = r} x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$$

is the sum of all monomials of total degree r in the n variables x_1, \dots, x_n .

- the r th power sum

$$p_r = x_1^r + \dots + x_n^r$$

is the sum of the r th powers of the variables.

Other symmetric polynomials are easily constructed by noting that the sum, difference and product of two symmetric polynomials are again symmetric. Thus the set of all symmetric polynomials in the variables x_1, \dots, x_n is a subring S of the polynomial ring P .

Symmetric polynomials arise naturally when studying the roots of a polynomial. For example if the cubic polynomial $f(x) = x^3 + ax^2 + bx + c$ has roots α, β, γ , then

$$x^3 + ax^2 + bx + c = (x - \alpha)(x - \beta)(x - \gamma)$$

and equating coefficients we get

$$\alpha + \beta + \gamma = -a, \quad \alpha\beta + \alpha\gamma + \beta\gamma = b, \quad \alpha\beta\gamma = -c.$$

Thus the coefficients of $f(x)$ can be regarded as polynomials in the ‘variables’ α, β, γ . If we permute the roots in any way, e.g. exchange α and β , we do not change the polynomial $f(x)$ and so the polynomials $\alpha + \beta + \gamma$, $\alpha\beta + \alpha\gamma + \beta\gamma$, $\alpha\beta\gamma$ are also unchanged. In general, if the monic polynomial

$$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$$

factors as the product

$$f(x) = (x - x_1)(x - x_2) \cdots (x - x_n),$$

then the relations $a_{n-r} = (-1)^r e_r$ hold for $1 \leq r \leq n$, and express each coefficient a_{n-r} of the polynomial f as (plus or minus) a corresponding elementary symmetric function e_r of its roots x_1, \dots, x_n .

Theorem 13.2 (Fundamental Theorem of Symmetric Polynomials) *Every symmetric polynomial in $F[x_1, \dots, x_n]$ can be written uniquely as a polynomial in the elementary symmetric functions e_1, \dots, e_n . Thus $S = F[e_1, \dots, e_n]$.*

Proof We use DegLex order with $x_1 > x_2 > \dots > x_n$. Let $f \in P$ be a nonzero symmetric polynomial, and let $\text{LM}(f) = \mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \cdots x_n^{\alpha_n}$. Note that $\alpha_1 \geq \alpha_2 \geq \dots \geq \alpha_n$; if $\alpha_i < \alpha_{i+1}$ for any i , then we can obtain a higher monomial by switching the variables x_i and x_{i+1} . Since f is symmetric, this monomial will also be a term of f .

Now let $g = e_1^{\beta_1} e_2^{\beta_2} \cdots e_n^{\beta_n}$ be a ‘monomial’ in the ‘variables’ e_1, e_2, \dots, e_n . Since $\text{LM}(e_r) = x_1 x_2 \cdots x_r$,

$$\begin{aligned} \text{LM}(g) &= x_1^{\beta_1} (x_1 x_2)^{\beta_2} \cdots (x_1 \cdots x_n)^{\beta_n} \\ &= x_1^{\beta_1 + \beta_2 + \dots + \beta_n} x_2^{\beta_2 + \dots + \beta_n} \cdots x_n^{\beta_n}. \end{aligned}$$

Thus $\text{LM}(g) = \text{LM}(f)$ if and only if $\alpha_r = \beta_r + \beta_{r+1} + \dots + \beta_n$ for $1 \leq r \leq n$, i.e. if and only if $\beta_r = \alpha_r - \alpha_{r+1}$ for $r < n$ and $\beta_n = \alpha_n$.

Now let $f_1 = f - ag$, where g is determined as above and $a = \text{LC}(f)$. Then f_1 is again a symmetric polynomial, and if $f_1 \neq 0$ then $\text{LM}(f_1) < \mathbf{x}^\alpha$, since the terms in \mathbf{x}^α cancel. If $f_1 = 0$ we are done, otherwise we repeat the procedure with f_1 . Since there are only a finite number of monomials $< \mathbf{x}^\alpha$ in DegLex order, the process must eventually stop, and we obtain an expression for f as a polynomial in the elementary symmetric functions.

The uniqueness of this expansion follows immediately from what we have already proved: the ‘monomials’ $\mathbf{e}^\beta = e_1^{\beta_1} e_2^{\beta_2} \cdots e_n^{\beta_n}$ in the ‘variables’ e_1, \dots, e_n all have distinct leading terms when expressed as polynomials in x_1, \dots, x_n , and hence they are linearly independent over the field of coefficients F . \square

14 Families of Symmetric Functions

14.1 Monomial symmetric functions

The proof of Theorem 13.2 gives an algorithm for computing a given symmetric polynomial f as a polynomial in the elementary symmetric functions. For practical calculations, it is convenient to group together all monomials in f with the same **set** of exponents, as follows.

Definition 14.1 Let $\alpha = (\alpha_1, \dots, \alpha_n)$ where $\alpha_1 \geq \dots \geq \alpha_n \geq 0$. Then the **monomial symmetric function** m_α is the sum

$$m_\alpha = \sum_{\beta} \mathbf{x}^\beta$$

of all monomials whose list of exponents $\beta = (\beta_1, \dots, \beta_n)$ is a permutation of $\alpha = (\alpha_1, \dots, \alpha_n)$.

Example 14.2 If the variables are x, y, z , then

$$\begin{aligned} m_{(3,2,0)} &= x^3y^2 + x^2y^3 + x^3z^2 + x^2z^3 + y^3z^2 + y^2z^3, \\ m_{(3,1,1)} &= x^3yz + xy^3z + xyz^3, \\ m_{(2,2,1)} &= x^2y^2z + x^2yz^2 + xy^2z^2. \end{aligned}$$

It is easy to see that m_α is symmetric, and that every symmetric polynomial can be uniquely expressed as $f = \sum_{\alpha} c_{\alpha} m_{\alpha}$, $c_{\alpha} \in F$. Thus the monomial symmetric functions form a basis for the ring of symmetric functions S , regarded as a vector space over F . It follows from Theorem 13.2 that the monomials $\mathbf{e}^{\beta} = e_1^{\beta_1} \cdots e_n^{\beta_n}$ in the elementary symmetric functions form another vector space basis for S .

Example 14.3 Let $f = m_{(3,2,0)} = x^3y^2 + x^2y^3 + x^3z^2 + x^2z^3 + y^3z^2 + y^2z^3$. to express f in terms of e_1, \dots, e_n , we first note that $\alpha = (3, 2, 0) \Rightarrow \beta = (1, 2, 0)$, and so $\mathbf{e}^{\beta} = e_1e_2^2$. Expanding \mathbf{e}^{β} in terms of monomial symmetric functions, we have

$$e_1e_2^2 = (x + y + z)(xy + xz + yz)^2 = m_{(3,2,0)} + 2m_{(3,1,1)} + 5m_{(2,2,1)}.$$

Then $\text{LM}(f - e_1e_2^2) = \text{LM}(m_{(3,1,1)}) = x^3yz$, so the new $\alpha = (3, 1, 1)$, the new $\beta = (2, 0, 1)$, and the new $\mathbf{e}^{\beta} = e_1^2e_3$. Then

$$e_1^2e_3 = (x + y + z)^2 \cdot xyz = m_{(3,1,1)} + 2m_{(2,2,1)}.$$

Noting that $m_{(2,2,1)} = x^2y^2z + x^2yz^2 + xy^2z^2 = e_2e_3$, we finally obtain

$$f = e_1e_2^2 - 2(e_1^2e_3 - 2e_2e_3) - 5e_2e_3 = e_1e_2^2 - 2e_1^2e_3 - e_2e_3.$$

14.2 Complete symmetric functions

We shall prove that the complete symmetric functions h_1, \dots, h_n provide another set of polynomial generators for the ring of symmetric functions S . The following identities provide a recursive procedure for expressing each h_k as a polynomial in e_1, \dots, e_n . For this we extend the definition of e_r by defining $e_0 = 1$ and $e_r = 0$ for $r > n$, where n is the number of variables.

Proposition 14.4 *For $1 \leq r \leq n$ and $i \geq 0$, let h_i be the i th complete symmetric function in the variables x_1, \dots, x_r and let e_i be the i th elementary symmetric function in the variables x_1, \dots, x_n . Then*

$$\sum_{i+j=k} (-1)^i e_i h_j = 0, \text{ for } k > n - r.$$

In particular, when $r = n$ this identity holds for all $k > 0$.

Proof We introduce a new variable t , and consider the **generating functions**

$$E(t) = \sum_{i=0}^n (-1)^i e_i t^i, \quad H(t) = \sum_{j=0}^{\infty} h_j t^j.$$

Note that $E(t)$ is a polynomial in $P[t]$, but $H(t)$ is a **formal power series** in t (with coefficients in P).

As we noticed at the beginning of Section 13, the elementary symmetric functions e_1, \dots, e_n (with alternating signs) are the coefficients of the polynomial with roots x_1, \dots, x_n . Hence

$$E(t) = (1 - x_1 t)(1 - x_2 t) \cdots (1 - x_n t).$$

On the other hand

$$\begin{aligned} H(t) &= 1 + h_1 t + h_2 t^2 + \dots \\ &= \frac{1}{1 - x_1 t} \frac{1}{1 - x_2 t} \cdots \frac{1}{1 - x_r t} \cdots \\ &= \frac{1}{(1 - x_1 t)} \cdot \frac{1}{(1 - x_2 t)} \cdots \frac{1}{(1 - x_r t)}. \end{aligned}$$

Hence $E(t)H(t) = (1 - x_{r+1} t) \cdots (1 - x_n t)$, a polynomial of degree $n - r$. But if we expand the product $E(t)H(t)$ in powers of t , the coefficient of t^k is $\sum_{i+j=k} (-1)^i e_i h_j$, so this sum is zero for $k > n - r$. \square

Theorem 14.5 *Every symmetric polynomial in $F[x_1, \dots, x_n]$ can be expressed uniquely as a polynomial in the complete symmetric functions h_1, \dots, h_n . Thus $S = F[h_1, \dots, h_n]$.*

Proof We can solve the system of equations in Proposition 14.4 with $r = n$ to express e_i as a polynomial in h_1, \dots, h_i for $1 \leq i \leq n$. Hence we can express any polynomial in e_1, \dots, e_n as a polynomial in h_1, \dots, h_n . Hence by Theorem 13.2 every symmetric polynomial can be written as a polynomial in h_1, \dots, h_n .

To prove uniqueness, we use the fact that the monomials \mathbf{e}^β in e_1, \dots, e_n of total degree d in x_1, \dots, x_n form an F -vector space basis for the homogeneous symmetric polynomials S^d of total degree d in x_1, \dots, x_n . Since we have seen that the monomials $\mathbf{h}^\beta = h_1^{\beta_1} \cdots h_n^{\beta_n}$ in S^d form a spanning set, and since there are the same number of \mathbf{h}^β 's as \mathbf{e}^β 's in S^d , the \mathbf{h}^β 's are also a F -vector space basis. \square

14.3 Power sums and Newton's identities

The r th **power sum** $p_r = x_1^r + \dots + x_n^r$ was introduced in Section 13. For some purposes, it turns out to be useful to express symmetric functions as polynomials in the p_r 's rather than the e_r 's or the h_r 's. However, since denominators arise in the coefficients, we must work over a field F of characteristic zero. First we obtain relations between the p_r 's and the e_r 's.

Proposition 14.6 (Newton's identities) For $r \geq 1$,

$$r e_r = \sum_{k=1}^r (-1)^{k-1} p_k e_{r-k}.$$

Proof Let $P(t)$ be the formal power series

$$\begin{aligned} P(t) &= p_1 + p_2 t + \dots + p_r t^{r-1} + \dots \\ &= \sum_{i=1}^n (x_i + x_i^2 t + \dots + x_i^r t^{r-1} + \dots) \\ &= \sum_{i=1}^n x_i (1 + x_i t + \dots + (x_i t)^{r-1} + \dots) \\ &= \sum_{i=1}^n \frac{x_i}{1 - x_i t}. \end{aligned}$$

As in Section 13, let $E(t) = 1 - e_1 t + e_2 t^2 - \dots + (-1)^n e_n t^n = (1 - x_1 t)(1 - x_2 t) \cdots (1 - x_n t)$. Then $\log E(t) = \sum_{i=1}^n \log(1 - x_i t)$. Differentiation with respect to t gives

$$\frac{E'(t)}{E(t)} = \sum_{i=1}^n \frac{-x_i}{1 - x_i t} = -P(t).$$

Hence $P(t)E(t) = -E'(t) = e_1 - 2e_2 t + \dots + (-1)^{n-1} n e_n t^{n-1}$. Equating the coefficients of t^{r-1} in this equation gives Newton's identities: note that these are valid for all $r \geq 1$ if we let $e_r = 0$ for $r > n$. \square

Theorem 14.7 *If $\text{char}(F) = 0$ then every symmetric polynomial in $F[x_1, \dots, x_n]$ can be written uniquely as a polynomial in the power sums p_1, \dots, p_n , i.e. $S = F[p_1, \dots, p_n]$.*

Proof We prove by induction on r that e_r is a polynomial in p_1, \dots, p_r . This holds for $r = 1$ since $e_1 = p_1$. The inductive step is given by Newton's identity. Note that we can divide by r in F , since $\text{char}(F) = 0$. It follows from Theorem 13.2 that every symmetric polynomial is a polynomial in p_1, \dots, p_r . Uniqueness is proved as for the complete symmetric functions (see Theorem 14.5). \square

14.4 Determinantal Formulae

The above relations between the elementary, complete and power sum symmetric functions can be solved so as to express one of these functions as an explicit polynomial in the members of one of the other families. Thus Proposition 14.4 is equivalent to either of the determinantal formulae

$$e_n = \begin{vmatrix} h_1 & h_2 & h_3 & \cdots & h_n \\ 1 & h_1 & h_2 & \cdots & h_{n-1} \\ 0 & 1 & h_1 & \cdots & h_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & h_1 \end{vmatrix}, \quad h_n = \begin{vmatrix} e_1 & e_2 & e_3 & \cdots & e_n \\ 1 & e_1 & e_2 & \cdots & e_{n-1} \\ 0 & 1 & e_1 & \cdots & e_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e_1 \end{vmatrix}.$$

Similarly, Newton's identities 14.6 can be written

$$p_n = \begin{vmatrix} e_1 & 2e_2 & 3e_3 & \cdots & ne_n \\ 1 & e_1 & e_2 & \cdots & e_{n-1} \\ 0 & 1 & e_1 & \cdots & e_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & e_1 \end{vmatrix}, \quad n!e_n = \begin{vmatrix} p_1 & p_2 & \cdots & p_{n-1} & p_n \\ 1 & p_1 & \cdots & p_{n-2} & p_{n-1} \\ 0 & 2 & \cdots & p_{n-3} & p_{n-2} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & n-1 & p_1 \end{vmatrix}.$$

The equivalence with Propositions 14.4 and 14.6 can be seen by using suitable row or column expansions of the determinants and induction. There are similar formulae relating the h 's and p 's (Examples 7, Question 6.) Although these explicit formulae are elegant, in practice it is easier to use the recursive formulae of Propositions 14.4 and 14.6 directly in calculations.

15 Invariants and Coinvariants

In this section we consider the **ideal** $I = \langle e_1, \dots, e_n \rangle$ generated by the elementary symmetric polynomials in $P = F[x_1, \dots, x_n]$, and the quotient ring P/I . In view of Theorem 13.2, we may equivalently define I as the ideal generated by all symmetric polynomials with zero constant term. There are many polynomials in I which are not symmetric: for example, the polynomial identity

$$x_i^n - e_1 x_i^{n-1} + e_2 x_i^{n-2} + \dots + (-1)^n e_n = 0 \quad (18)$$

shows that the n th power x_i^n of every variable x_i is in I . This identity follows from the relation between the coefficients of a polynomial and its roots which we noticed in Section 13.

In classical invariant theory, the ring S is called the **ring of invariants** of the **symmetric group** S_n consisting of all permutations of the variables x_1, \dots, x_n . The ring P/I is called the **ring of coinvariants**. By Theorem 14.5, the complete symmetric functions h_1, \dots, h_n also generate I .

We have seen that a Gröbner basis for I is helpful in understanding the structure of P/I . Unfortunately, none of the bases e_1, \dots, e_n , h_1, \dots, h_n or p_1, \dots, p_n is a Gröbner basis for I for any monomial ordering. This follows from our observation above that $x_i^n \in I$ for all variables x_i . A Gröbner basis for I must therefore contain a polynomial g_i whose leading term divides x_i^n and is thus a power of x_i . If we order the variables so that $x_1 > x_2 > \dots > x_n$, then $\text{LM}(e_r) = x_1 x_2 \cdots x_r$ and $\text{LM}(h_r) = \text{LM}(p_r) = x_1^r$, so these choices do not work.

Theorem 15.1 *For the monomial order DegLex with $x_n > \dots > x_2 > x_1$,*

$$H = \{h_1(x_1, \dots, x_n), h_2(x_1, \dots, x_{n-1}), \dots, h_{n-1}(x_1, x_2), h_n(x_1)\}$$

is a (reduced) Gröbner basis of I .

For example, when $n = 3$ and $x > y > z$, this basis is $\{x + y + z, y^2 + yz + z^2, z^3\}$.

Proof Note that the leading term of the polynomial $h_{n-r+1}(x_1, \dots, x_r)$ in H is x_r^{n-r+1} . Thus each leading monomial in H is a power of a different variable, so that the leading monomials are all coprime. Thus if we can show that H is a basis of I , it will follow from our work on Buchberger's algorithm that all S -polynomials of pairs of elements of H reduce to zero, and hence H is a Gröbner basis for S . It is also easy to see that H is reduced.

Note that $h_n(x_1) = x_1^n$ is in I by equation (18), and $h_1(x_1, \dots, x_n) = x_1 + \dots + x_n$ is clearly in I since it is symmetric. To prove that the other elements of H are in I , we use the identity of Proposition 14.4:

$$\sum_{i+j=k} (-1)^i e_i(x_1, \dots, x_n) h_j(x_1, \dots, x_r) = 0 \text{ for } k > n - r. \quad (19)$$

By putting $k = n - r + 1$ in (19), we see that $h_{n-r+1}(x_1, \dots, x_r)$ is an element of the ideal $\langle e_1, e_2, \dots, e_{n-r+1} \rangle$. Hence $\langle H \rangle \subseteq I$. Conversely, by solving the equations (19) recursively, we can express e_{n-r+1} as an element of the ideal $\langle h_1(x_1, \dots, x_n), h_2(x_1, \dots, x_{n-1}), \dots, h_{n-r+1}(x_1, \dots, x_r) \rangle$. Since e_1, e_2, \dots, e_n generate I , it follows that $I \subseteq \langle H \rangle$. Hence $\langle H \rangle = I$. \square

Theorem 15.1 allows us to find a normal form for any polynomial in P in the ring of coinvariants P/I , and to perform calculations in this ring. Thus we obtain a basis for P/I as a vector space over F , as follows.

Theorem 15.2 (Artin basis for the ring of coinvariants) *The cosets of the monomials $\mathbf{x}^\alpha = x_1^{\alpha_1} \dots x_n^{\alpha_n}$ such that $\alpha_r \leq n - r$ form a F -vector space basis for the ring of coinvariants P/I of the symmetric group in $P = F[x_1, \dots, x_n]$.*

Proof Since the leading terms of the elements of the Gröbner basis H of Theorem 15.1 are $x_n, x_{n-1}^2, \dots, x_1^n$, a monomial $\mathbf{x}^\alpha \in P$ is divisible by one of these leading terms if and only if $\alpha_r \geq n - r + 1$ for some r . Thus the result follows from Theorem 7.4(ii). \square

This result shows that the ring of coinvariants has dimension $n!$ as a vector space over F . This is equal to the order of the symmetric group S_n . We can calculate efficiently in P/I by using the elements of our Gröbner basis as relations. For example in the case $n = 3$ we have relations $x + y + z = 0$, $x^2 + xy + y^2 = 0$, $x^3 = 0$ (where x is the coset $\bar{x} = x + I \in P/I$, and similarly for y and z).

The ring P/I has further interesting structure which we describe briefly. By considering the coinvariants of a fixed degree d , we obtain a vector space on which the symmetric group S_n acts by permutation of the variables x_1, \dots, x_n . These give important examples of **matrix representations** of the group S_n : with each permutation in S_n we associate the matrix which describes its action on the Artin basis elements. In particular, the action of S_n on the whole ring P/I gives the **regular representation** of S_n . These ideas are developed in **group representation theory**.

16 Alternating Polynomials

In this section we consider an important family of polynomials which are not symmetric but are nearly so, in the sense that they change sign when two of the variables are exchanged. The key example of such a polynomial is the **Vandermonde determinant**

$$\Delta_n = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ x_1^2 & x_2^2 & \cdots & x_n^2 \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix}.$$

Such polynomials are said to be **alternating**. **Throughout Section 16, we assume that the characteristic of the field F is not 2, so that $f \neq -f$ for $f \in P$.**

Every permutation of x_1, \dots, x_n can be obtained by a finite sequence of exchanges of two variables. Hence every permutation $\sigma \in S_n$ transforms an alternating polynomial f either into f or into $-f$. Notice that the product of two alternating polynomials is a symmetric polynomial, since if f and g are transformed into $-f$ and $-g$ by some exchange, then fg is transformed into $(-f)(-g) = fg$. Thus fg is fixed by an exchange of two variables, and so it is fixed by any finite sequence of exchanges. Likewise the product gh of a symmetric polynomial g and an alternating polynomial h is alternating. Thus we can obtain a plentiful supply of alternating polynomials by multiplying Δ_n by symmetric polynomials. We next prove that **all** alternating polynomials are obtained in this way.

We can sometimes find factors of polynomials using the Remainder Theorem. This theorem is usually found in textbooks only for polynomials in one variable over a field, but we need something stronger. We begin by proving a stronger version of the division theorem, Theorem 1.1.

Theorem 16.1 *Let R be a commutative ring with 1, and let $f, g \in R[x]$, where the leading coefficient of g is invertible in R . Then there are unique polynomials q and r in $R[x]$ such that **(i)** $f = qg + r$, and **(ii)** either $\deg r < \deg g$ or $r = 0$.*

Proof We use induction on the degree of f . Let $f = \sum_{r=0}^n a_r x^r$, where $a_n \neq 0$, so that $\deg f = n \geq 0$. (If $f = 0$, we simply take $q = r = 0$.) Let $g = \sum_{r=0}^m b_r x^r$, where $b_m \neq 0$, so that $\deg g = m \geq 0$. If $m > n$ we simply take $q = 0$, $r = f$, so we may assume that $m \leq n$.

For the base of induction, let $n = 0$. Then $m = 0$ and f and g are constant polynomials, i.e. nonzero elements of R . We have $f = a_0$, $g = b_0$ and $f = qg$ where $q = b_0^{-1}a_0 \in R$ since b_0 is invertible in R .

Now let $n > 0$, and assume as induction hypothesis that the result holds for any polynomial f_1 of degree $< n$. In particular we may take $f_1 = b_m f - a_n x^{n-m} g$,

since clearly $\deg f_1 \leq n$ and the coefficient of x^n in f_1 is $b_m a_n - a_n b_m = 0$, so that in fact $\deg f_1 < n$. Hence the induction hypothesis applies to f_1 , to give $f_1 = q_1 g + r_1$ where $q_1, r_1 \in R[x]$ and $\deg r_1 < m$ or $r_1 = 0$. Hence $b_m f = a_n x^{n-m} g + f_1 = (a_n x^{n-m} + q_1)g + r_1$. Hence $f = qg + r$, where $q = b_m^{-1}(a_n x^{n-m} + q_1)$ and $r = b_m^{-1}r_1$, and $\deg r = \deg r_1 < m$ or $r = 0$, as required.

Uniqueness is proved by the same argument as when R is a field (exercise!).

Theorem 16.2 (Remainder Theorem) *Let R be a commutative ring with 1 and let $a \in R$. Then a polynomial $f(x) \in R[x]$ is divisible by $x - a$ if and only if $f(a) = 0$, i.e. if and only if a is a **zero** (or **root**) of $f(x)$.*

Proof By Theorem 16.1 we can divide an arbitrary polynomial $f(x) \in R[x]$ by the monic polynomial $g(x) = x - a$.

We now use a standard argument: we have $f(x) = q(x)(x - a) + r$ where the remainder r is constant, i.e. $r \in R$. Substitute $x = a$ to get $r = f(a)$. Thus $f(x)$ is divisible by $x - a$ iff $f(a) = 0$. \square

As a first application, we obtain the well-known factorisation of the Vandermonde determinant Δ_n . If $x_i = x_j$, then $\Delta_n = 0$ since the determinant has two equal columns. By Theorem 16.2 it follows that Δ_n is divisible by $x_i - x_j$ for all i, j with $1 \leq i < j \leq n$. Hence

$$\Delta_n = \prod_{1 \leq i < j \leq n} (x_j - x_i).$$

Proposition 16.3 *Every alternating polynomial h is divisible in $F[x_1, \dots, x_n]$ by the Vandermonde determinant Δ_n .*

Proof Since $\Delta_n = \prod_{i < j} (x_j - x_i)$, h is divisible by Δ_n if and only if $x_i - x_j$ divides h for all $i \neq j$. Since h is transformed into $-h$ when x_i and x_j are exchanged, $h = -h$ when $x_i = x_j$. But $h = -h$ implies that $h = 0$, since $\text{char}(F) \neq 2$. By Theorem 16.2, $x_i - x_j$ is a factor of h . \square

16.1 The alternating group A_n

We recall some standard facts from group theory. The **symmetric group** S_n consists of all permutations of x_1, \dots, x_n . Thus S_n has order $n!$. A permutation $\sigma \in S_n$ is **even** if it can be obtained by an even number of exchanges of two variables, and **odd** if an odd number of exchanges is required. The set of even permutations is denoted by A_n . The matrix of the linear transformation $x_i \mapsto \sigma(x_i)$ has determinant 1 if σ is even, -1 if σ is odd. Since the determinant is a homomorphism from S_n on to the multiplicative group $\{1, -1\}$ if $n > 1$, and A_n is the kernel of this homomorphism, A_n is a subgroup of S_n of order $\frac{1}{2}n!$. It is called the **alternating group** on the n variables x_1, \dots, x_n .

It is easy to show that the effect of a permutation $\sigma \in S_n$ on an alternating polynomial f is given by $\sigma(f) = f$ if σ is even, $\sigma(f) = -f$ if σ is odd. Thus alternating polynomials, as well as symmetric polynomials, are **invariant** under the action of the alternating group A_n .

Proposition 16.4 *Let $f \in F[x_1, \dots, x_n]$ be invariant under A_n , i.e. $\sigma(f) = f$ for all $\sigma \in A_n$. Then f can be uniquely expressed as the sum of a symmetric polynomial and an alternating polynomial.*

Proof Clearly a polynomial f of the given form is A_n -invariant: we prove the converse. Let f be invariant under A_n , and let σ be an odd permutation. Then $\sigma^2 \in A_n$, and hence $\sigma^2(f) = f$. Let $g = \frac{1}{2}(f + \sigma(f))$, $h = \frac{1}{2}(f - \sigma(f))$. (Note that we can divide by 2, since $\text{char}(F) \neq 2$.) Then using $\sigma^2(f) = f$ we have $\sigma(g) = \frac{1}{2}(\sigma(f) + f) = g$, $\sigma(h) = \frac{1}{2}(\sigma(f) - f) = -h$.

Since f is A_n -invariant, so also is $\sigma(f)$, and hence so are g and h . Now any odd permutation can be written as $\sigma \circ \rho$, where $\rho \in A_n$. Thus $\sigma \circ \rho(g) = \sigma(g) = g$ and $\sigma \circ \rho(h) = \sigma(h) = -h$. Hence g is symmetric and h is alternating, and $f = g + h$ is in the required form.

To prove uniqueness, suppose that $f = g_1 + h_1$ where g_1 is symmetric and h_1 is alternating. Then $\sigma(f) = \sigma(g_1) + \sigma(h_1) = g_1 - h_1$. Hence $g_1 = \frac{1}{2}(f + \sigma(f)) = g$ and $h_1 = \frac{1}{2}(f - \sigma(f)) = h$, and so the decomposition of f is unique. \square

The set A of all A_n -invariant polynomials is closed under addition and multiplication, i.e. it is a subring of $F[x_1, \dots, x_n]$. Clearly A contains the Vandermonde determinant Δ_n , and also all symmetric polynomials. Proposition 16.3 now allows us to give a complete description of the ring A .

Theorem 16.5 *If $\text{char}(F) \neq 2$, then*

- (i) *every A_n -invariant polynomial $f \in F[x_1, \dots, x_n]$ can be uniquely expressed as $f = g + \Delta_n h$, where g and h are symmetric polynomials, and*
- (ii) *the ring A of all A_n -invariant polynomials is given by*

$$A \cong F[e_1, \dots, e_n, \Delta_n] / \langle \Delta_n^2 - \delta(e_1, \dots, e_n) \rangle$$

where $\delta(e_1, \dots, e_n)$ is the polynomial obtained by writing Δ_n^2 in terms of e_1, \dots, e_n .

Proof Part (i) follows from Propositions 16.4 and 16.3, noting that the quotient of the ‘alternating part’ of f by Δ_n is a symmetric polynomial h . Part (ii) follows from (i) if we observe that since $\langle \Delta_n^2 - \delta(e_1, \dots, e_n) \rangle$ is a principal ideal, every element of the quotient ring has a unique representative of the form $g + \Delta_n h$, where g and h are polynomials in e_1, \dots, e_n . (Recall that the monic generator of a principal ideal is the reduced Gröbner basis, for any monomial ordering.) \square

16.2 The discriminant of a polynomial

We can apply these ideas to obtain a necessary and sufficient condition for a polynomial in one variable to have equal roots. Thus let $f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0$, with $a_i \in F$ and $\text{char}(F) \neq 2$. Suppose that $f(x) = (x - x_1)(x - x_2) \cdots (x - x_n)$ where the roots x_1, x_2, \dots, x_n lie in a suitable extension field of F . (This can always be done.) Then f has **equal roots** if and only if $x_i = x_j$ for some i, j with $i \neq j$. Since $\Delta_n = \prod_{i < j} (x_j - x_i)$, f has equal roots if and only if $\Delta_n = 0$, or equivalently $\Delta_n^2 = \delta(e_1, \dots, e_n) = 0$. Since we may identify the elementary symmetric functions e_1, \dots, e_n in the roots x_1, \dots, x_n with the coefficients of f by $e_r = (-1)^r a_{n-r}$ for $1 \leq r \leq n$, this gives the required condition on f .

In the case $n = 2$, for $f(x) = ax^2 + bx + c$ we have $\Delta^2 = (x_1 - x_2)^2 = e_1^2 - 4e_2 = (-b/a)^2 - 4(c/a)$, giving the familiar condition $b^2 - 4ac = 0$ for equal roots. The polynomial $\delta(e_1, \dots, e_n)$ (when expressed in terms of the coefficients of f) is called the **discriminant** of a polynomial f of degree n . Thus the discriminant of $f(x) = ax^2 + bx + c$ is $b^2 - 4ac$.

The discriminant of the cubic polynomial $x^3 + ax^2 + bx + c$ is a much more formidable beast:

$$\Delta^2 = a^2b^2 - 4a^3c - 4b^3 + 18abc - 27c^2.$$

This takes a more familiar form if we assume $a = 0$ (we can always reduce to this case by a linear change of variable). Thus the equation $x^3 + bx + c = 0$ has at least two equal roots if and only if $4b^3 + 27c^2 = 0$, or equivalently $(c/2)^2 = -(b/3)^3$. As an example, $b = -3, c = 2$ gives $x^3 - 3x + 2 = (x - 1)^2(x + 2)$.

16.3 Conclusion

In this course we have only scratched the surface of the vast subject of symmetric functions, which plays an important role in many areas of mathematics today, including algebraic combinatorics and group representation theory. There are also connections to applied mathematics and theoretical physics here. A good textbook for students wishing to know more about this topic is Chapter 1 of I. G. Macdonald's book *Symmetric Functions and Hall Polynomials*. We have also hinted at developments in a different direction, that of the invariant theory of finite groups. This topic motivated much of the development of algebra in the 19th century, and it has undergone a significant revival in recent years. The book by Larry Smith, *Polynomial Invariants of Finite Groups*, is recommended.