Mathematics and Algorithms for Computer Algebra

July 9, 2003

4: Polynomial algebra

1 Definitions

Let $\{x_1, x_2, \ldots, x_n\}$ be a (finite) set of n symbols and let R be a commutative ring with unity.

Definition 1 *The* polynomials *over R in the* variables $x_1, x_2, \ldots, x_n$ *are the formal sums*

$$p = \sum_{i_1,\ldots,i_n \in \mathbb{N}} p_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n},$$

where the coefficients $p_{i_1,\ldots,i_n}$ *are elements of R and only a finite number of them are nonzero. The set of all such polynomials is denoted by $R[x_1, x_2, \ldots, x_n]$.*

A polynomial of the form $p_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$ with precisely one nonzero coefficient is called a *monomial*. The monomorphism $R \to R[x_1, x_2, \ldots, x_n]$, $c \mapsto c x_1^0 \cdots x_n^0$, identifies the ground ring R with a subset of $R[x_1, x_2, \ldots, x_n]$, and therefore a monomial of the form $c x_1^0 \cdots x_n^0$ is denoted simply c and called a *constant* (or occasionally a constant polynomial).

If $m, n \in \mathbb{N}$, $1 \leq m < n$, there is a natural isomorphism

$$R[x_1, \ldots, x_m][x_{m+1}, \ldots, x_n] \sim R[x_1, x_2, \ldots, x_n].$$

The notation can be simplified by defining

$$\sum_{\mathbf{i}} p_{\mathbf{i}} x^{\mathbf{i}} = \sum_{i_1,\ldots,i_n \in \mathbb{N}} p_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n}$$

in terms of the *multi-index* $\mathbf{i} = (i_1, \ldots, i_n)$, where

$$\mathbf{i} + \mathbf{j} = (i_1 + j_1, \ldots, i_n + j_n) \quad \text{and} \quad x^{\mathbf{i}} = x_1^{i_1} \cdots x_n^{i_n}.$$

Then if $p, q \in R[x_1, x_2, \ldots, x_n]$, $c \in R$, polynomial addition and multiplication are defined by

$$p + q = \sum_{\mathbf{i}} (p_{\mathbf{i}} + q_{\mathbf{i}}) x^{\mathbf{i}}, \quad pq = \sum_{\mathbf{k}} \left(\sum_{\mathbf{i},\mathbf{j},\, \mathbf{i}+\mathbf{j}=\mathbf{k}} (p_{\mathbf{i}} q_{\mathbf{j}}) x^{\mathbf{k}} \right),$$

where special cases of multiplication are

$$cp = \sum_{\mathbf{i}} (c p_{\mathbf{i}}) x^{\mathbf{i}}, \quad -p = \sum_{\mathbf{i}} (-p_{\mathbf{i}}) x^{\mathbf{i}}.$$

With these operations $R[x_1, x_2, \ldots, x_n]$ is a commutative ring with a unity element which is the constant polynomial $p = 1$.

If $p = \sum_{\mathbf{i}} p_{\mathbf{i}} x^{\mathbf{i}}$ is nonzero (i.e. at least one coefficient is nonzero) then the *total degree* of p is

$$\deg p = \max\{i_1 + \cdots + i_n \mid p_{\mathbf{i}} \neq 0\}$$

and the *partial degree* of p with respect to x_j is

$$\deg_{x_j} p = \max\{i_j \mid p_{\mathbf{i}} \neq 0\},$$

which is sometimes also denoted $\deg_j p$.

2 Arithmetic, simplification and rational expressions

In an algebraic context there is no difference between performing arithmetic operations and simplifying arithmetic expressions. Assuming that a canonical representation is being used, then an arithmetic expression composed of rational expressions is in general not in canonical form, even if the component rational expressions themselves are. The process of simplifying such an expression requires any arithmetic operations to be performed first. If necessary, the result must then be put into canonical representation, although probably the result of performing the arithmetic will already be canonical.

Arithmetic on rational functions is completely analogous to arithmetic on rational numbers, if the integer operations underlying rational arithmetic

are replaced by operations on (possibly multivariate) polynomials. Hence, arithmetic on rational expressions is straightforward, given algorithms to perform multivariate polynomial addition, subtraction, multiplication and gcd computations, and I will not discuss it in further detail.

Univariate polynomial arithmetic is analogous to integer arithmetic in which the integer base B is replaced by the polynomial variable, which I will call x. However, a polynomial has more structure than an integer, which actually makes polynomial arithmetic simpler because there are no carries between different powers of x. I will assume that the polynomials have coefficients in some ring R, which might in fact be a field, and might be either infinite or finite. Assuming the existence of algorithms to perform arithmetic in the coefficient ring, the algorithms to perform the polynomial ring operations are independent of the details of the coefficient ring R. I will discuss the complexity of polynomial arithmetic in terms of the number of arithmetic operations required in R, but this is not the true total complexity, which does depend on the details of the R, and may depend on the size of the coefficients, as we saw would be the case if $R = \mathbb{Z}$.

2.1 Polynomial addition and subtraction

In the univariate case, the algorithm to add two polynomials p, q is essentially the same as the integer algorithm, and just requires a loop to run through the terms of the polynomials, in which the number of coefficient operations is $\min(\deg p, \deg q) + 1$. (Since there are no carries there is no compelling reason to increase this bound to $\max(\deg p, \deg q) + 1$ as there was in the integer case.) Just as for integers, subtraction can be performed using $p - q = p + (-q)$. If it is reasonable to assume that negating the coefficients is a trivial operation then the number of coefficient operations is $\min(\deg p, \deg q) + 1$ as for addition, otherwise it is $\deg q + 1$. The number of coefficient operations for either addition or subtraction is certainly $O(\max(\deg p, \deg q))$.

2.2 Polynomial multiplication

Classical univariate polynomial multiplication follows the first (unsatisfactory) integer multiplication algorithm. If $m = \deg p, n = \deg q$ then there are $(m+1)(n+1)$ coefficient multiplications and $(m+1)(n+1) - (m+n-1)$ coefficient additions, so the number of coefficient operations required is $O(mn)$. There are significantly faster polynomial multiplication algorithms based on the fast Fourier transform (FFT), and analogous to the fastest integer multi-

plication algorithms. However, these fast algorithm are only asymptotically faster, which in practice means only for very high degrees.

2.3 Multivariate polynomials

Arithmetic on multivariate polynomials (multinomials) is accomplished by taking advantage of the natural isomorphism referred to above, and regarding $R[x_1, x_2, \ldots, x_r]$ as $R[x_1, \ldots, x_{r-1}][x_r]$. Then arithmetic on polynomials in r variables is carried out by recursion on r; the routines call themselves if $r \neq 0$ or call their analogues for performing arithmetic in the appropriate coefficient domain if $n = 0$ (which is the base case that terminates the recursion). Note that this recursive approach to multinomial arithmetic matches nicely with the recursive representation discussed in Notes 1.

In order to bound the complexity of multinomial arithmetic, let us assume that the *partial* degrees of $p, q \in R[x_1, x_2, \ldots, x_r]$ with respect to any variable are bounded respectively by m, n; then p, q contain respectively at most $(m+1)^r$ and $(n+1)^r$ monomials. The complexity of multivariate arithmetic depends on the number of terms in each polynomial as in the univariate case, and hence the number of coefficient operations required to add or subtract p and q is $O(\max(m, n)^r)$, and to multiply them is $O((mn)^r)$.

3 Euclidean division and pseudo-division

Addition, subtraction and multiplication of polynomials is fairly straightforward, but as in the case of the integers division is a little trickier. For one thing, it makes a significant difference when dividing what domain the coefficients lie in, because it determines whether or not they divide in general. Initially I will consider univariate polynomials whose coefficients are elements of a field, i.e. elements of $F[x]$, which we know is a Euclidean domain. Hence the Euclidean division property applies to all pairs of such polynomials, and the task is to construct an algorithm to compute the quotient and remainder polynomials. In a general Euclidean domain the quotient and remainder are not unique, but in $F[x]$ they are. The division algorithm is essentially the same as for long integers, but simpler, assuming that we have algorithms to perform the required coefficient arithmetic, in particular division.

Theorem 1 *Given two polynomials $u(x)$ and $v(x)$ over a field, with $v(x) \neq 0$, there exist* unique *polynomials $q(x)$ and $r(x)$ satisfying the* Division Prop-

erty

$$u(x) = q(x)v(x) + r(x), \quad deg(r) < deg(v).$$

Proof Existence is proved by the algorithm given below (which could be phrased more formally).

To prove that the division property is unique, suppose to the contrary that it holds for both $(q_1(x), r_1(x))$ and $(q_2(x), r_2(x))$. Then $q_1(x)v(x) + r_1(x) = q_2(x)v(x) + r_2(x)$, so $(q_1(x) - q_2(x))v(x) = r_2(x) - r_1(x)$. Now if $(q_1(x) - q_2(x))$ is nonzero, we have $\deg((q_1 - q_2)v) = \deg(q_1 - q_2) + \deg(v) \geq \deg(v) > \deg(r_2 - r_1)$, a contradiction; hence $q_1(x) - q_2(x) = 0$ and $r_2(x) - r_1(x) = 0$.[1] □

Here is the Euclidean division algorithm:

input: $u(x) = u_m x^m + \cdots + u_0, v(x) = v_n x^n + \cdots + v_0 \in F[x]$, $v_n \neq 0$
$q_0 := 0$; for $i := 0$ to m do $r_i := u_i$;
for $k := m - n$ step -1 to 0 do
begin
    $q_k := r_{n+k}/v_n$; {the only coefficient division}
    {Division loop:}
    for $j := n + k - 1$ step -1 to k do $r_j := r_j - q_k v_{j-k}$
    {$r_{n+k} := 0$, but it will be ignored}
end.
output: $q(x) = q_{m-n}x^{m-n} + \cdots + q_0, r(x) = r_{n-1}x^{n-1} + \cdots + r_0 \in F[x]$
    such that $u(x) = q(x)v(x) + r(x)$, $\deg(r) < \deg(v)$

Note that the division loop amounts to replacing $u(x)$ by $u(x) - q_k x^k v(x)$, a polynomial of degree $< n + k$. This algorithm is usually called "synthetic division", for no very obvious reason!

The main loop in the above algorithm is executed $m - n + 1$ times, and consists of one coefficient division and n coefficient multiplications and n coefficient subtractions in the inner loop. Hence the complexity in terms of coefficient operations is $O(n(m - n))$. As in the long integer case, this is essentially the time to compute the product $q(x)v(x)$.

The only actual divisions that occur anywhere in this algorithm are divisions by the leading coefficient v_n of the divisor polynomial. Hence it is not necessary for division that the polynomials have coefficients in a field and it is sufficient that v_n be a unit; it may be preferable to invert v_n once

[1] Knuth writes $r_1(x) = 0$ here, but I believe that is an error!

and then multiply by the inverse. Hence two polynomials over an arbitrary commutative ring with identity can be divided if the leading coefficient of the divisor is a unit, and in particular if the divisor is monic, whereas two polynomials over a field can be divided provided only that the divisor is non-zero.

The importance of monic polynomials is that any polynomial can be divided by a monic polynomial, so that the remainder in the division is always defined, and therefore equivalence modulo a monic polynomial is well defined regardless of the coefficient domain.

Division of multivariate polynomials can in principle be performed by regarding $R[x_1, \ldots, x_{r-1}, x_r]$ as $R[x_1, \ldots, x_{r-1}][x_r]$ and applying the univariate division algorithm recursively. However, it is necessary that the leading coefficient of the divisor be a unit (at every level of recursion) and hence an element of the ground ring R. Therefore, actual polynomial division is only necessary with respect to one (distinguished) polynomial variable, and the only multivariate operations will be the multiplications and subtractions required on the coefficients of powers of the distinguished variable, so the polynomial division algorithm itself can never be applied recursively. Hence, division is possible only if the leading coefficient with respect to at least one of the variables is a unit of R (and hence independent of the other variables), and divisions with respect to two such variables will generally be different.

3.1 Pseudo-division

Euclidean division is always possible by extending from $D[x]$ to $Q(D)[x]$, i.e. by computing in the field of fractions of the coefficient ring if it is not already a field. But this is inelegant in principle, because a problem that can be posed and has a solution within a particular algebraic system should be solvable by computing within that algebraic system without the need to extend it. Moreover, it is undesirable in practice because computations in a field of fractions are more complex than in the underlying integral domain, mainly because of the need to perform gcd computations in order to make representations canonical, and we have seen that gcd computations are inherently complex.

Frequently, one is really interested in the variable (primitive) parts of polynomials rather than their overall numerical factors (contents), in which case there is no reason to be constrained by the fact that the leading coefficient of $v(x)$ is not a unit when attempting to divide $u(x)$ by $v(x)$. In fact, it need not be a unit at all provided that it divides the leading coefficient

of the remainder *at each stage of the division*. A total of $m-n+1$ distinct intermediate remainders occur during the algorithm, including the initial remainder $u(x)$.

Let ℓ denote the leading coefficient of $v(x)$. If $u(x)$ is multiplied by ℓ^{m-n+1} then each of the $m-n+1$ divisions by ℓ must be possible. This is equivalent to multiplying the dividend by the common denominator that we expect to arise if we work in the field of quotients, and then working in the ring. In fact, the algorithm does not require any explicit multiplications at all, which proves that it is possible over any ring. The following version is essentially as given by Knuth:

input: $u(x) = u_m x^m + \cdots + u_0, v(x) = v_n x^n + \cdots + v_0 \in R[x]$, $v_n \neq 0$
$q_0 := 0$;
for $i := 0$ to $m-n-1$ do $r_i := v_n^{m-n-i} u_i$;
for $i := m-n$ to m do $r_i := u_i$;
for $k := m-n$ step -1 to 0 do
begin
    $q_k := r_{n+k} v_n^k$;
    {Multiplication loop:}
    for $j := n+k-1$ step -1 to k do $r_j := v_n r_j - r_{n+k} v_{j-k}$
    {$r_{n+k} := 0$, but it will be ignored}
end.
output: $q(x) = q_{m-n} x^{m-n} + \cdots + q_0, r(x) = r_{n-1} x^{n-1} + \cdots + r_0 \in F[x]$
    such that $v_n{}^{m-n+1} u(x) = q(x)v(x) + r(x)$, $\deg(r) < \deg(v)$

The division of $v_n{}^{m-n+1} u(x)$ (or some other suitable multiple of $u(x)$) by $v(x)$ is called *pseudo-division*, and produces a *pseudo-quotient* and *pseudo-remainder*. Of course, this will frequently produce results with larger numerical factors (contents) than necessary, but to avoid this would require gcd computations. One of the main uses of pseudo-division is as a basis for a generalized or "pseudo-Euclidean" algorithm to compute gcds of polynomials over rings, which will be considered further in the next set of notes.

The following theorem can be proved in an analogous way to that describing the division property for polynomials over a field:

Theorem 2 *Given two polynomials $u(x)$ and $v(x)$ over an integral domain, with $v(x) \neq 0$, where ℓ is the leading coefficient of $v(x)$ and $\delta = \max\{\deg u - \deg v + 1, 0\}$ there exist* unique *polynomials $q(x)$ and $r(x)$ satisfying the* Pseudo-Division Property

$$\ell^\delta u(x) = q(x)v(x) + r(x), \quad deg(r) < deg(v).$$

4 Irreducible factorization and polynomial content

The purpose of this section is to revise some properties of polynomials which I introduced earlier in a more abstract setting and which are important in the theory of general (multivariate) polynomial factorization and gcd computation.

Let us assume that the coefficient ring is an integral domain D, in which case $D[x]$ is also an integral domain. In general, a nonzero element $d \in D$ is said to be *irreducible* (or *prime*) if and only if the quotient ring $D/(d)$ is also an integral domain. Then, in particular, a polynomial $p \in D[x]$ is irreducible if and only if $D[x]/(p)$ is an integral domain.

Two elements $d_1, d_2 \in D$ are said to be *associates* or *associated* if and only if $d_1 = ud_2$ where u is a unit in D, i.e. u is invertible. Then, in particular, two polynomials $p_1, p_2 \in D[x]$ are associated if and only if $p_1 = up_2$, where u is a unit in $D[x]$ and hence must also be a unit element of the ground ring D.

4.1 Irreducible factorization when D is a field F

A polynomial $p \in F[x]$ is irreducible if and only if $\deg p > 0$ (i.e. p is not a constant) and p has no divisor $q \in F[x]$ such that $\deg p > \deg q > 0$. Therefore every non-constant polynomial $f \in F[x]$ has at least one irreducible divisor, namely any divisor of minimal nonzero degree, which might be f itself if f is irreducible.

Proposition 3 *A non-constant polynomial $p \in F[x]$ is irreducible if and only if the quotient ring $F[x]/(p)$ is a field.*

Theorem 4 *Any nonzero polynomial $f \in F[x]$ admits a decomposition of the form*

$$f = c \prod_{i=1}^{k} p_i^{\alpha_i}, \quad c \in F, \ \alpha_i > 0,$$

where the polynomials p_i are irreducible and unassociated.

Moreover, if

$$f = c' \prod_{j=1}^{k'} q_j^{\beta_j}, \quad c' \in F, \quad \beta_j > 0,$$

then $k' = k$ and there exists a permutation σ of the index set $I = \{1, \ldots, k\}$ such that $\beta_{\sigma(i)} = \alpha_i$ and $q_{\sigma(i)}$ is associated to p_i for each $i \in I$.

4.2 Polynomial content

An integral domain is said to be a *unique factorization domain* (UFD) (or a *factorial* ring) if each of its nonzero elements admits a unique decomposition into a product of irreducible elements, up to the order of the factors and up to associates. Hence, for example, by Theorem 4, $F[x]$ is a UFD.

Let D be a UFD, $Q = Q(D)$ be its field of fractions, and p be a given irreducible (i.e. prime) element of D. Then any nonzero element $a \in Q$ can be written *uniquely* as

$$a = p^r b, \quad b \in Q, \ r \in \mathbb{Z},$$

where b is the quotient of two elements of D, neither of which is divisible by p. This follows from the unique prime factorization of the numerator n and denominator d of a, which must be relatively prime. Then for any given prime p precisely one of the following is true: p appears

1. as p^r for some $r \in \mathbb{Z}^+$ in the factorization of n, or
2. as $p^{r'}$ for some $r' \in \mathbb{Z}^+$ in the factorization of d giving $r = -r'$, or
3. in neither giving $r = 0$.

For example, $\mathbb{Z}$ is a UFD and $\mathbb{Q}$ is its field of fractions. Then for successive small primes 2, 3, 5, 7:

$$a = \frac{9}{11} = 2^0\frac{9}{11} = 3^2\frac{1}{11} = 5^0\frac{9}{11} = 7^0\frac{9}{11};$$

$$a = \frac{11}{15} = 2^0\frac{11}{15} = 3^{-1}\frac{11}{5} = 5^{-1}\frac{11}{3} = 7^0\frac{11}{15}.$$

The integer r is called the *order of a at p* and denoted $\mathrm{ord}_p(a)$, and is somewhat analogous to $\log_p(a)$. For example, it satisfies

$$\mathrm{ord}_p(aa') = \mathrm{ord}_p(a) + \mathrm{ord}_p(a').$$

[By convention, the definition is extended so that $\mathrm{ord}_p(0) = +\infty$, and therefore the above product rule is satisfied if a or a' is zero.]

If $f = f_n x^n + \cdots + f_0$ is a nonzero polynomial with coefficients in the field of fractions Q then its order is defined by

$$\mathrm{ord}_p(f) = \min\{\mathrm{ord}_p(f_i) \mid f_i \neq 0\}.$$

[In fact, just $\mathrm{ord}_p(f) = \min\{\mathrm{ord}_p(f_i)\}$ suffices with the definition $\mathrm{ord}_p(0) = +\infty$.]

Now let $\wp$ denote a *system of representatives of the irreducible elements of* D, meaning that every element of $\wp$ is irreducible, every irreducible element of D has an associate in $\wp$ and no two elements of $\wp$ are associated. For example, if $D = \mathbb{Z}$ then $\wp$ could be taken as the set of all prime integers, which by convention are positive.

Definition 2 *The* content *of* $f \in Q[x]$ *(relative to* $\wp$*) is defined by*

$$\mathrm{cont}_\wp(f) = \prod_{p \in \wp} p^{\mathrm{ord}_p(f)}.$$

(The subscript $\wp$ *is omitted if it is unambiguous.)*

To see how this works, consider

$$f = \frac{3}{2}x^2 + \frac{15}{4} \in \mathbb{Q}[x], \quad \mathbb{Q} = Q(\mathbb{Z})$$

and take $\wp$ to be the positive prime integers. Then

$$\mathrm{ord}_2\left(\frac{3}{2}\right) = -1, \ \ \mathrm{ord}_2\left(\frac{15}{4}\right) = -2 \ \Rightarrow \ \mathrm{ord}_2(f) = \min(-1,-2) = -2;$$

$$\mathrm{ord}_3\left(\frac{3}{2}\right) = 1, \ \ \mathrm{ord}_3\left(\frac{15}{4}\right) = 1 \ \Rightarrow \ \mathrm{ord}_3(f) = \min(1,1) = 1;$$

$$\mathrm{ord}_5\left(\frac{3}{2}\right) = 0, \ \ \mathrm{ord}_5\left(\frac{15}{4}\right) = 1 \ \Rightarrow \ \mathrm{ord}_5(f) = \min(0,1) = 0;$$

and $\mathrm{ord}_p(f) = 0$ for $p > 5$. Hence

$$\mathrm{cont}_\wp(f) = 2^{-2} \times 3^1 \times 5^0 \times 7^0 \times \cdots = \frac{3}{4}.$$

Now note that by regarding $f \in \mathbb{Q}[x]$ as $f \in Q(\mathbb{Z}[x])$ and then factoring out the gcd of the coefficients of the numerator polynomial we have

$$f = \frac{3}{2}x^2 + \frac{15}{4} = \frac{3.2x^2 + 15}{4} = \frac{3}{4}(2x^2 + 5),$$

which has precisely the form $\mathrm{cont}(f)$ times a polynomial over $\mathbb{Z}$.

The following result is central to the computation of gcds of multinomials:

Lemma 5 (Gauss' Lemma) *Let D be a UFD and Q its field of fractions. If f, g are nonzero polynomials with coefficients in Q then*

$$\mathrm{cont}(fg) = \mathrm{cont}(f)\,\mathrm{cont}(g).$$

Proof From the definition of content, the lemma follows if for every irreducible element $p \in D$ we can prove that

$$\mathrm{ord}_p(fg) = \mathrm{ord}_p(f) + \mathrm{ord}_p(g).$$

Let us put $r = \mathrm{ord}_p(f)$, $s = \mathrm{ord}_p(g)$. Then from the definition of the order of a polynomial, the coefficients of f have a common factor of p^r which can be divided out to leave a polynomial with coefficients in Q that have no factors of p in their denominators (because the order of a polynomial is the minimum order of its coefficients). This polynomial can therefore be written (non-uniquely) in the form uf_1 where $u \in Q$, $\mathrm{ord}_p(u) = 0$ and f_1 has coefficients in D, $\mathrm{ord}_p(f_1) = 0$.

Applying a similar argument to g, there exist nonzero polynomials f_1, g_1 with coefficients in D and nonzero $u, v \in Q$ such that

$$f = up^r f_1, \quad g = vp^s g_1,$$

where

$$\mathrm{ord}_p(f_1) = \mathrm{ord}_p(g_1) = 0, \quad \mathrm{ord}_p(u) = \mathrm{ord}_p(v) = 0.$$

Then $\mathrm{ord}_p(fg) = \mathrm{ord}_p(uvp^r p^s f_1 g_1) = r + s + \mathrm{ord}_p(f_1 g_1)$ and it remains to prove that $\mathrm{ord}_p(f_1 g_1) = 0$.

Map D to its homomorphic image $D/(p)$, which is also an integral domain because (p) is a prime ideal.[2] The images f_1', g_1' of f_1, g_1 under the epimorphism defined to be the natural map $D \to D/(p)$ are nonzero because $D/(p)$ has characteristic p and at least one coefficient of each of f_1 and g_1 has no factor of p. Moreover, because $D/(p)$ is an integral domain, the product $f_1' g_1'$ must also be nonzero, and therefore at least one coefficient of $f_1 g_1$ has no factor of p, and hence $\mathrm{ord}_p(f_1 g_1) = 0$. □

We will make serious use of Gauss' Lemma to compute gcds in the next set of notes. One can also use it to prove, for example:

Theorem 6 *If D is a unique factorization domain (UFD) and F is a field then $D[x_1, \ldots, x_n]$ and $F[x_1, \ldots, x_n]$ are both UFDs for any $n \in \mathbb{Z}^+$.*

[2] Actually, this is not true in every integral domain, but it is certainly true in a PID such as the integers, which in practice is our main interest, when in fact $D/(p)$ is a field. I do not know whether it is also sufficient for D to be a UFD.

5 Polynomial functions and roots of polynomials

The main purpose of this section is to raise the important logical distinction between a polynomial, in which the variables are purely symbolic objects, and a polynomial function, in which the variables take values in some set.

Let $f \in A[x_1, \ldots, x_n]$,

$$f = \sum f_{i_1,\ldots,i_n} x_1^{i_1} \cdots x_n^{i_n},$$

and let B be an extension ring of A.

Definition 3 *The* polynomial function $f^* : B^n \to B$ *is defined by*

$$f^*(b_1, \ldots, b_n) = \sum f_{i_1,\ldots,i_n} b_1^{i_1} \cdots b_n^{i_n}.$$

In the polynomial function, a value b_i has been *substituted* for each variable x_i. However, to simplify the notation the * is often omitted.

The following properties are obvious, and relate operations in the ring $A[x_1, \ldots, x_n]$ (on the left) to those in the ring $B \geq A$ (on the right). If $f, g \in A[x_1, \ldots, x_n]$ then

$$(f+g)^* = f^* + g^*, \quad (fg)^* = f^* g^*, \quad (af)^* = af^* \ \text{ if } \ a \in A \leq B.$$

As an example of the difference between f and f^*, note that the map $\varphi : f \mapsto f^*$ may not be one-to-one. For example, if $A = B$ is the ring of integers mod 2, i.e. $\mathbb{Z}_2 = \{0, 1\}$, and $f = x^2 - x$ then f^* takes the value 0 on every element of its domain, and hence it *is* the zero function $0 : x \mapsto 0$, as is the image under φ of any multiple of f. However, if A is an infinite integral domain then φ is one-to-one, in which case one can safely identify a polynomial with its associated polynomial function.

5.1 Roots of polynomials

Let $f(x) \in R[x]$ and $\alpha \in R$, R a commutative ring. If $f(\alpha) = 0$ then we call α a *root* or *zero* of $f(x)$. Roots are intimately related to *linear* factors, as follows.

Theorem 7 (Factor Theorem) *Let $f \in R[x]$ and $\alpha \in R$; then*

$$\alpha \text{ is a root of } f(x) \iff (x - \alpha) \mid f(x).$$

Proof ($\Rightarrow$) Let $f(\alpha) = 0$. Since the leading coefficient of $x - \alpha$ is a unit we can divide $f(x)$ by $x - \alpha$ to obtain

$$f(x) = (x - \alpha)q(x) + r(x)$$

where $\deg r(x) < \deg(x - \alpha) = 1$. Hence $r(x) \in R$. Moreover, since $r(x) = r$ is independent of x, we can compute it by substituting any value for x in the division relation, in particular $x = \alpha$, giving

$$r = r(\alpha) = f(\alpha) - (\alpha - \alpha)q(\alpha) = 0 - 0 = 0.$$

Hence $(x - \alpha) \mid f(x)$.

($\Leftarrow$) $f(x) = (x - \alpha)q(x) \;\Rightarrow\; f(\alpha) = 0q(\alpha) = 0.$ □

Corollary 8 (Remainder Theorem) *Let $f \in R[x]$ and $\alpha \in R$; then the remainder on dividing $f(x)$ by $(x - \alpha)$ is $f(\alpha)$.*

Proof By the Factor Theorem, $(x - \alpha) \mid f(x) - f(\alpha)$, so that $f(x) = (x - \alpha)q(x) + f(\alpha)$, and hence $f(\alpha)$ is the unique remainder. □

5.2 Order and multiplicity of roots

If $f(x) \in R[x]$, $\alpha \in R$ is called a *root of order m* of f is f is divisible by $(x - \alpha)^m$ but not by $(x - \alpha)^{m+1}$. Since a linear factor must be irreducible, this terminology is consistent with its usage in the previous section, because (in $Q(R[x])$ and assuming that R is a UFD)

$$\operatorname{ord}_{(x-\alpha)} f(x) = m.$$

If $m \geq 1$ it is called the *multiplicity* of the root. If $m = 1$ then α is called a *simple* root; otherwise it is called a *multiple* root.

Proposition 9 *If α is a root of $f(x)$ of order m then*

$$f(x) = (x - \alpha)^m q(x), \quad q(\alpha) \neq 0.$$

Proof By the definition of order, $f(x) = (x - \alpha)^m q(x)$. If $q(\alpha) = 0$ then, by the Factor Theorem, $(x - \alpha)$ divides $q(x)$, and hence $(x - \alpha)^{m+1}$ divides $f(x)$, contrary to the definition of order. Hence $q(\alpha) \neq 0$. □

Theorem 10 *Let D be an integral domain and let $f(x) \in D[x]$ be a nonzero polynomial. Then the sum of the multiplicities of the roots of $f(x)$ that lie in D is at most equal to the degree of $f(x)$.*

Proof is by induction on $n = \deg f(x)$.

Basis ($n = 0$). A nonzero polynomial of degree 0 is a (nonzero) constant in D, and hence trivially has no roots.

Induction. Assume the statement of the theorem for all polynomials of degree $< n$, and let $f(x) \in D[x]$ have degree n. If $f(x)$ has a root α of multiplicity $m \geq 1$ then

$$f(x) = (x - \alpha)^m q(x), \quad q(\alpha) \neq 0, \quad \deg q(x) = n - m.$$

If $f(x)$ has another root $\beta \neq \alpha$ then

$$0 = f(\beta) = (\beta - \alpha)^m q(\beta).$$

Since $\beta \neq \alpha$ we must have $q(\beta) = 0$ *because D has no zerodivisors.* Hence the only roots of $f(x)$ are α of multiplicity m and those of $q(x)$. By the induction hypothesis, $q(x)$ has roots of total multiplicity at most $n - m$. Hence $f(x)$ has roots of total multiplicity at most $(n - m) + m = n$. □

Corollary 11 *A nonzero polynomial in $D[x]$ of degree n has at most n distinct roots.*

This theorem shows the importance of a polynomial coefficient domain being integral. For example, over $\mathbb{Z}_{16}$, which is not an integral domain, the quadratic polynomial $x^2 + 12$ has the 4 roots 2, 6, 10, 14.

5.3 Formal derivatives

Let R be a commutative ring, and let $f(x) \in R[x]$ be

$$f(x) = \sum_{i=0}^{n} f_i x^i.$$

Definition 4 *The (formal)* derivative *of $f(x)$ is*

$$f'(x) = \sum_{i=1}^{n} i f_i x^{i-1} = \sum_{i=0}^{n-1} (i+1) f_{i+1} x^i.$$

Note that this definition agrees with the derivative arising from infinitessimal calculus, but does not involve any arguments relying on continuity, which are not appropriate in arbitrary commutative rings. The standard properties of derivatives of sums, products, numerical multiples and composed polynomials can easily be established from the above definition.

The formal derivative relates to multiple roots as follows.

Theorem 12 *Let $\alpha \in R$ be a root of $f(x) \in R[x]$; then α is a multiple root of $f(x)$ if and only if $f'(\alpha) = 0$.*

Proof By Proposition 9

$$f(x) = (x - \alpha)^m q(x), \quad m \geq 1, \ q(\alpha) \neq 0,$$

and formally differentiating gives

$$f'(x) = (x - \alpha)^m q'(x) + m(x - \alpha)^{m-1} q(x).$$

If $m > 1$ then $f'(\alpha) = 0$ trivially. If $m = 1$ then $f'(\alpha) = q(\alpha) \neq 0$. □

More generally, put $f^0 = f$, $f^{(k)} = (f^{(k-1)})'$ for any $k \in \mathbb{Z}^+$.

Proposition 13 *Let $\alpha \in R$ be a root of multiplicity m of $f(x) \in R[x]$; then $f^{(k)}(\alpha) = 0$ for $0 \leq k < m$.*

Proof From the proof of the previous proposition

$$f'(x) = (x - \alpha)^{m-1}((x - \alpha)q'(x) + mq(x)).$$

Hence $(x - \alpha)^m \mid f^0(x) \quad \Rightarrow \quad (x - \alpha)^{m-1} \mid f^1(x)$, and by induction $(x - \alpha)^{m-k} \mid f^k(x)$ for $0 \leq k < m$. □

6 The resultant and discriminant

It is frequently necessary to find a condition that two univariate polynomial functions $f(x)$ and $g(x)$ have a non-trivial common factor or common zero, without explicitly finding it. This condition is provided by the vanishing of the *resultant* of $f(x)$ and $g(x)$, denoted $\mathrm{Res}(f, g)$, or $\mathrm{Res}_x(f, g)$ if the polynomial variable would otherwise be ambiguous. If the condition is met then the common factor or common zero could be found, e.g. by a gcd

computation, and hence a resultant computation is related to, but not the same as, a gcd computation.

It is also frequently necessary to find a condition that a univariate polynomial function $f(x)$ has a multiple root. This condition is provided by the vanishing of the *discriminant* of $f(x)$, denoted $\mathrm{Disc}(f)$ or $\mathrm{Disc}_x(f)$. But a multiple root of $f(x)$ is also a root of its derivative $f'(x)$, in which case $\mathrm{Res}_x(f, f') = 0$. There is therefore an intimate relationship between $\mathrm{Disc}_x(f)$ and $\mathrm{Res}_x(f, f')$.

If $f(x) = 0$ and $g(x) = 0$ for the same value of x then one can eliminate x between the two equations to give one equation that is independent of x and expresses a condition on the functions f and g that the common root x exists, e.g. it provides an equation that must be satisfied by the coefficients of the polynomials f and g. However, whilst this is an interpretation of a resultant, it provides neither a general systematic formulation nor a good computational algorithm, which is the main purpose of this section.

6.1 The resultant and the Sylvester matrix

Let f, g be polynomials in x over a coefficient ring R:

$$f(x) = \sum_{i=0}^{m} f_i x^i, \quad g(x) = \sum_{i=0}^{n} g_i x^i,$$

where $f_i, g_i \in R$. Suppose that f, g have a non-trivial common factor h, so that $f = Fh$, $g = Gh$. Then the equation

$$uf + vg = (uF + vG)h = 0$$

(subject to $\deg u < \deg g$, $\deg v < \deg f$ to avoid trivial common factors of u and v) has the solution $u = G, v = -F$, and conversely the existence of such a solution implies that f, g have a common factor. Written out fully, the equation $uf + vg = 0$ becomes

$$\begin{aligned} (u_{n-1}x^{n-1} + \cdots + u_1 x + u_0)(f_m x^m + \cdots + f_1 x + f_0) &+ \\ (v_{m-1}x^{m-1} + \cdots + v_1 x + v_0)(g_n x^n + \cdots + g_1 x + g_0) &= 0, \end{aligned}$$

and hence

$$\begin{aligned}
u_{n-1}(f_m x^{m+n-1} + f_{m-1}x^{m+n-2} + \cdots + f_0 x^{n-1}) + \\
u_{n-2}(f_m x^{m+n-2} + \cdots + f_1 x^{n-1} + f_0 x^{n-2}) + \\
\ddots \\
u_0(f_m x^m + \cdots + \cdots + f_1 x + f_0) + \\
v_{m-1}(g_n x^{m+n-1} + g_{n-1}x^{m+n-2} + \cdots + g_0 x^{m-1}) + \\
v_{m-2}(g_n x^{m+n-2} + \cdots + g_1 x^{m-1} + g_0 x^{m-2}) + \\
\ddots \\
v_0(g_n x^n + \cdots + \cdots + g_1 x + g_0) = 0.
\end{aligned}$$

Equating coefficients of $x^i, m+n-1 \geq i \geq 0$, gives the following system of $m+n$ linear homogeneous equations expressed in matrix form, where column j of the coefficient matrix corresponds to x^{m+n-j} and ^{T} denotes the transpose of a matrix:

$$\begin{pmatrix} u_{n-1} \\ u_{n-2} \\ \vdots \\ u_1 \\ u_0 \\ v_{m-1} \\ v_{m-2} \\ \vdots \\ v_1 \\ v_0 \end{pmatrix}^T \begin{pmatrix} f_m & f_{m-1} & \cdots & f_1 & f_0 & 0 & \cdots & 0 & 0 \\ 0 & f_m & f_{m-1} & \cdots & f_1 & f_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & f_m & f_{m-1} & \cdots & f_1 & f_0 & 0 \\ 0 & 0 & \cdots & 0 & f_m & f_{m-1} & \cdots & f_1 & f_0 \\ g_n & g_{n-1} & \cdots & g_1 & g_0 & 0 & \cdots & 0 & 0 \\ 0 & g_n & g_{n-1} & \cdots & g_1 & g_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & g_n & g_{n-1} & \cdots & g_1 & g_0 & 0 \\ 0 & 0 & \cdots & 0 & g_n & g_{n-1} & \cdots & g_1 & g_0 \end{pmatrix} = \mathbf{0}^T.$$

The $(m+n) \times (m+n)$ coefficient matrix is the *Sylvester matrix* $S(f,g)$ of f and g, which consists of n rows of the coefficients f_i of f in "eschelon formation", followed by m rows of the coefficients g_i of g in "eschelon formation". More precisely, the elements of S are given by the following algorithmic specification:

$s_{ij} =$ if $1 \leq i \leq n$ then
  if $m+i-j \bmod m+n \leq m$ then f_{m+i-j} else 0
else $\{n+1 \leq i \leq n+m\}$
  if $i-j \bmod m+n \leq n$ then g_{i-j} else 0.

This system of linear homogeneous equations has a non-trivial solution for the coefficients of the polynomials u and v if and only if the determinant of the Sylvester matrix is zero. This leads to the following

Definition 5 *The* resultant *of f and g is the determinant of the Sylvester matrix of f and g, i.e.*

$$\mathrm{Res}(f,g) = \det S(f,g).$$

6.2 Properties of resultants

These properties are useful in relating resultants to discriminants, and in developing efficient algorithms to compute resultants and discriminants.

Proposition 14 *If $f, g \in R[x]$ then $\mathrm{Res}_x(f,g) \in R$.*

Proof A determinant of a matrix over a ring R is defined using only ring operations to be a sum of products of elements, and so is itself an element of the ring R. □

Proposition 15 *If c is a constant then $\mathrm{Res}(f,c) = c^m$ where $m = \deg f$, and in particular $\mathrm{Res}(f,0) = 0$.*

Proof By definition, if $g(x) = g_0 = c$ then the Sylvester matrix of f, g consists of zero rows composed of the coefficients of f followed by m rows with $g_0 = c$ on the leading diagonal. □

Proposition 16 $\mathrm{Res}(f,g) = (-1)^{mn}\,\mathrm{Res}(g,f)$ *where $m = \deg f$, $n = \deg g$.*

Proof This follows from exchanging the f and g rows in the Sylvester matrix. It requires $m+n-1$ exchanges of adjacent rows to move the top row of an $(m+n)\times(m+n)$ matrix to the bottom, without changing the order of the remaining rows, and it therefore requires $n\times(m+n-1)$ exchanges to move all n of the f rows below all of the g rows. Each exchange introduces a factor of (-1), so the overall factor is $(-1)^{nm+n(n-1)} = (-1)^{nm}(-1)^{n(n-1)}$. But $n(n-1)$ is even for all n. □

Proposition 17 *If $R[x]$ is a Euclidean domain, $1 \le m \le n$ where $m = \deg f$, $n = \deg g$, and h is the remainder in the Euclidean division of g by f then*[3]

$$\mathrm{Res}(f,g) = f_m^{n-p}\,\mathrm{Res}(f,h) \text{ where } p = \deg h \le m-1.$$

[3]Mignotte states this proposition with p replaced by m, but I believe that both his proposition and his proof of it are wrong! Davenport *et al.* give a very brief but correct discussion.

Remark If the coefficient ring R is a field then $R[x]$ is a Euclidean domain. The above result also holds in the special case that the Euclidean division is possible in a non-Euclidean domain, as discussed earlier; otherwise a very similar result holds where the division is replaced by a pseudo-division.

Proof The Euclidean division property gives $g = fq + h$, $\deg h < \deg f$. If the coefficients of g in $S(f, g)$ are replaced by the coefficients of $h = g - fq$, without otherwise changing the structure of the matrix, then the determinant is not changed because the operation corresponds to subtracting linear combinations of the rows of f coefficients, as I will show in more detail below. Then the elements that were the leading $n - p$ coefficients of b in each row become zero.

The result is a matrix of the form

$$S' = \begin{pmatrix} T_{n-p} & * \\ 0 & S(f,h) \end{pmatrix}$$

where T_{n-p} is an $(n-p)\times(n-p)$ upper triangular submatrix with every diagonal element equal to f_m and $S(f, h)$ is an $(m+p)\times(m+p)$ submatrix which is the Sylvester matrix of f and h. Then clearly $\det S' = f_m^{n-p} \det S(f, h)$, which proves the proposition. (The submatrix denoted by $*$ is irrelevant because it does not contribute to the determinant.) □

As a more detailed example of the transformation $S \to S'$, suppose $m = n = 3$ giving

$$S = \begin{pmatrix} f_3 & f_2 & f_1 & f_0 & 0 & 0 \\ 0 & f_3 & f_2 & f_1 & f_0 & 0 \\ 0 & 0 & f_3 & f_2 & f_1 & f_0 \\ g_3 & g_2 & g_1 & g_0 & 0 & 0 \\ 0 & g_3 & g_2 & g_1 & g_0 & 0 \\ 0 & 0 & g_3 & g_2 & g_1 & g_0 \end{pmatrix}.$$

If $p = 2$ then

$$S \to S' = \left(\begin{array}{c|ccccc} f_3 & f_2 & f_1 & f_0 & 0 & 0 \\ \hline 0 & f_3 & f_2 & f_1 & f_0 & 0 \\ 0 & 0 & f_3 & f_2 & f_1 & f_0 \\ 0 & h_2 & h_1 & h_0 & 0 & 0 \\ 0 & 0 & h_2 & h_1 & h_0 & 0 \\ 0 & 0 & 0 & h_2 & h_1 & h_0 \end{array}\right),$$

whereas if $p = 1$ then

$$S \to S' = \left(\begin{array}{cc|cccc} f_3 & f_2 & f_1 & f_0 & 0 & 0 \\ 0 & f_3 & f_2 & f_1 & f_0 & 0 \\ \hline 0 & 0 & f_3 & f_2 & f_1 & f_0 \\ 0 & 0 & h_1 & h_0 & 0 & 0 \\ 0 & 0 & 0 & h_1 & h_0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_0 \end{array}\right).$$

Now let us see why replacing the coefficients of g in $S(f, g)$ by the coefficients of $h = g - fq$ corresponds to subtracting linear combinations of the rows of f coefficients. The quotient polynomial q has degree at most $n - m$, and so the polynomial product qf has the form

$$(q_{n-m}x^{n-m} + \cdots + q_1x + q_0)(f_mx^m + \cdots + f_1x + f_0) =$$

$$\begin{aligned} q_{n-m}(f_mx^n + f_{m-1}x^{n-1} + \cdots + f_0x^{n-m}) + \\ q_{n-m-1}(f_mx^{n-1} + \cdots + f_1x^{n-m} + f_0x^{n-m-1}) + \\ \ddots \\ q_0(f_mx^m + \quad \cdots \quad + \quad \cdots \quad + f_1x + f_0). \end{aligned}$$

Letting column number represent the power of x as in the derivation of the Sylvester matrix above, the row matrix representing this product polynomial is a linear combination of the bottom $n - m + 1 \leq n$ adjacent rows of f coefficients in S. Hence $S'(f, g)$ can be derived from $S(f, g)$ by subtracting this linear combination from the last row of g coefficients, and then repeating the operation with *all* rows involved shifted up by 1, because there are precisely m distinct blocks of $n - m + 1$ adjacent rows within the total of n rows of f coefficients, one block for each of the m rows of g coefficients.

Proposition 18 $\mathrm{Res}(f, g) = 0$ *if and only if* f *and* g *have a non-trivial common factor.*

Proof Construct the polynomial pseudo-remainder sequence generated by f and g, which will terminate with a degree-zero polynomial (i.e. a constant) c that is zero if and only if f and g have a non-trivial common factor. By Propositions 16 and 17, $\mathrm{Res}(f, g)$ is proportional to $\mathrm{Res}(p, c)$ where p is the penultimate element of the remainder sequence. Then by Proposition 15, $\mathrm{Res}(p, c) = c^{\deg p}$ is zero if and only if $c = 0$. □

Proposition 19 *If the coefficient ring R is an integral domain, $\{\alpha_i\}_{i=1}^m$ are the roots of f and $\{\beta_j\}_{j=1}^n$ are the roots of g, then (in the appropriate extension field of R)*

$$\mathrm{Res}(f,g) = f_m^n \prod_{i=1}^m g(\alpha_i) = (-1)^{mn} g_n^m \prod_{j=1}^n f(\beta_j) = f_m^n g_n^m \prod_{i=1}^m \prod_{j=1}^n (\alpha_i - \beta_j).$$

Remark The requirement that the coefficient ring R be an integral domain (i.e. has no zerodivisors) is so that it can be extended first to its quotient field, and then to a field containing the root fields of f and g.

Proof [Davenport *et al.* attribute this proof to Dominique Duval.] Denote

$$R_\alpha(f,g) = f_m^n \prod_{i=1}^m g(\alpha_i), \quad R_\beta(f,g) = (-1)^{mn} g_n^m \prod_{j=1}^n f(\beta_j),$$

$$R_{\alpha\beta}(f,g) = f_m^n g_n^m \prod_{i=1}^m \prod_{j=1}^n (\alpha_i - \beta_j).$$

Clearly, $R_\beta(f,g) = R_\alpha(f,g)$ by exchanging the rôles of f and g and applying Proposition 16. In its root field,

$$g(x) = g_n \prod_{j=1}^n (x - \beta_j) \;\Rightarrow\; g(\alpha_i) = g_n \prod_{j=1}^n (\alpha_i - \beta_j).$$

Hence

$$R_\alpha(f,g) = f_m^n \prod_{i=1}^m g(\alpha_i) = f_m^n \prod_{i=1}^m \left(g_n \prod_{j=1}^n (\alpha_i - \beta_j) \right) = R_{\alpha\beta}(f,g).$$

Therefore $R_\beta(f,g) = R_\alpha(f,g) = R_{\alpha\beta}(f,g)$. Their equality to $\mathrm{Res}(f,g)$ is proved by induction on $\min(m,n)$ as follows.

If $n = 0$ then $g(x) = c$ is constant and $R_\beta(f,c) = c^m$, and also $\mathrm{Res}(f,c) = c^m$ by Proposition 15. This is the base case for the induction.

If f and g are exchanged, then (obviously) $R_{\alpha\beta}(f,g) = (-1)^{mn} R_{\alpha\beta}(g,f)$ and also $\mathrm{Res}(f,g) = (-1)^{mn} \mathrm{Res}(g,f)$ by Proposition 16. Therefore we need now consider only the case that $1 \le m \le n$.

If f and g have a non-trivial common factor then they have at least one root in common, so that $\alpha_i = \beta_j$ for some i, j and therefore $R_{\alpha\beta}(f,g) = 0$,

and also $\mathrm{Res}(f,g) = 0$ by Proposition 18. Otherwise, $g = fq + h$ where $h \neq 0$ and $p = \deg h \leq m-1$, because we are working in an extension field F of R which ensures that $F[x]$ is a Euclidean domain. Then $g(\alpha_i) = f(\alpha_i)q(\alpha_i) + h(\alpha_i) = h(\alpha_i)$ since $f(\alpha_i) = 0$, and hence

$$R_\alpha(f,g) = f_m^n \prod_{i=1}^{m} g(\alpha_i) = f_m^{n-p} f_m^p \prod_{i=1}^{m} h(\alpha_i) = f_m^{n-p} R_\alpha(f,h),$$

and also $\mathrm{Res}(f,g) = f_m^{n-p}\,\mathrm{Res}(f,h)$ by Proposition 17.

Then from $R_\beta(f,g) = R_\alpha(f,g) = R_{\alpha\beta}(f,g) = \mathrm{Res}(f,g)$ as induction hypothesis for $\deg g \leq m-1$ we have proved that it is true for $\deg g = m$. □

6.3 The discriminant

When considering the multiplicity of roots of polynomials, as mentioned at the beginning of this section, the following function is useful.

Definition 6 *The* discriminant *of a polynomial f of degree m having roots $\{\alpha_i\}_{i=1}^m$ (in its root field) is*

$$\mathrm{Disc}(f) = f_m^{2m-2} \prod_{i=1}^{m} \prod_{j=1, j\neq i}^{m} (\alpha_i - \alpha_j),$$

where f_m is the leading coefficient of f.

The discriminant clearly vanishes if and only if two (or more) roots are equal, because it is explicitly defined as the product of the differences between all pair of roots.

Proposition 20 $\mathrm{Res}(f,f') = f_m\,\mathrm{Disc}(f)$ *and* $\mathrm{Disc}(f) \in R$, *the coefficient ring of f.*

The latter property is one reason for the choice of the constant factor in the definition of the discriminant. Note also that a discriminant is always a perfect square.

Proof In terms of its m roots, the polynomial f can be expressed as

$$f(x) = f_m \prod_{j=1}^{m} (x - \alpha_j).$$

The product rule for a derivative gives the derivative of f to be

$$f'(x) = f_m \sum_{k=1}^{m} \prod_{j=1, j\neq k}^{m} (x - \alpha_j) \;\Rightarrow\; f'(\alpha_i) = f_m \sum_{k=1}^{m} \prod_{j=1, j\neq k}^{m} (\alpha_i - \alpha_j).$$

But each product within this sum has a factor with $j = i$ that therefore vanishes, unless $k = i$ so that the factor that vanishes is absent. Therefore

$$f'(\alpha_i) = f_m \prod_{j=1, j\neq i}^{m} (\alpha_i - \alpha_j).$$

Using in Proposition 19 the fact that $\deg f' = m - 1$ gives

$$\begin{aligned} \mathrm{Res}(f, f') &= f_m^{m-1} \prod_{i=1}^{m} f'(\alpha_i) = f_m^{m-1} \prod_{i=1}^{m} \left(f_m \prod_{j=1, j\neq i}^{m} (\alpha_i - \alpha_j) \right) \\ &= f_m^{2m-1} \prod_{i=1}^{m} \prod_{j=1, j\neq i}^{m} (\alpha_i - \alpha_j) = f_m \,\mathrm{Disc}(f). \end{aligned}$$

Now consider the definition $\mathrm{Res}(f, f') = \det S(f, f')$. If the leading term of $f(x)$ is $f_m x^m$ then the leading term of $f'(x)$ is $m f_m x^{m-1}$, and so the first column of the Sylvester matrix $S(f, f')$ is $(f_m, 0, \ldots, 0, m f_m, 0, \ldots, 0)^T$. Then a cofactor expansion of $\det S(f, f')$ about the first column leads to $\det S(f, f') = f_m A + m f_m B$ where $A, B \in R$ are the appropriate cofactors, and hence $\mathrm{Res}(f, f') = f_m C$ where $C = A + mB \in R$. Therefore $\mathrm{Disc}(f) = \mathrm{Res}(f, f')/f_m = C \in R$. □

6.4 Computation of resultants and discriminants

It is trivial to compute a discriminant in terms of a resultant using Proposition 20, so I will explicitly consider only resultant computation. The resultant is defined in terms of a determinant, but general determinant evaluation has a rather high complexity, and so a method that takes account of the special structure of the Sylvester matrix is preferable. One way to evaluate a determinant is to reduce the matrix to triangular form by Gaussian elimination, and the analogue of this for a Sylvester matrix is provided by Proposition 17.

If $R[x]$ is a Euclidean domain, then Propositions 15, 16 and 17 immediately lead to the following recursive algorithm (essentially as given by Davenport *et al.*), which is very similar to Euclid's algorithm for computing

a gcd. I assume that the return instruction both terminates execution of the procedure and assigns a value to it, thereby avoiding a deeply nested if-then-else construct.

input: $f, g \in R[x]$, a Euclidean domain
procedure Resultant(f, g);
$m := \deg f$; $n := \deg g$;
if $m > n$ then return $(-1)^{mn}$Resultant(g, f);
$\ell := \mathsf{lc}(f)$; {leading coefficient}
if $m = 0$ then return ℓ^n;
$h := \mathsf{remainder}(g, f)$; {$g$ mod f}
if $h = 0$ then return 0;
$p := \deg h$;
return ℓ^{n-p} Resultant(f, h).
output: Res(f, g)

If $R[x]$ is not a Euclidean domain, but R is an integral domain, then one could work in the quotient field of R within the algorithm, because Proposition 14 ensures that the result is in R, or one could use an analogue of the above algorithm based on pseudo-division. The second approach will generally be more efficient, because it avoids the gcd computations required when computing in a quotient field. More sophisticated techniques are also possible – see the remarks and references in Davenport *et al.*

6.5 Complexity of resultants

Here I consider not the complexity of resultant computation but the complexity of the resultant itself, because regardless of how they are calculated, resultants can be quite complex. For example, if the polynomial coefficients f_i, g_j are integers bounded in magnitude respectively by A, B then the magnitude of the resultant is bounded by $(m+1)^{n/2}(n+1)^{m/2}A^nB^m$, and can in practice come quite close to this bound. Similarly, if f_i, g_j are themselves polynomials of degrees bounded respectively by α, β then the resultant is a polynomial of degree bounded by $n\alpha + m\beta$. This potential explosion in the complexity of resultants is something to watch out for when using them in practice, and reducing the complexity of the input polynomials f and g in any way possible is worth considering!

The above bounds are stated without proof in Davenport *et al.* One way to see where they comes from is to consider the Sylvester matrix $S(f, g)$. Suppose every coefficient of f is $\pm A$ and every coefficient of g is $\pm B$. Then

each of the n rows of f coefficients of $S(f,g)$ contains a common factor of A, and each of the m rows of g coefficients contains a common factor of B. Therefore $\det S(f,g) = A^n B^m \det S'$ where S' is the matrix $S(f,g)$ with every coefficient replaced by ± 1. Now if A is a polynomial of degree α and B is a polynomial of degree β then $\det S(f,g)$ is clearly a polynomial bounded by degree $n\alpha + m\beta$, as asserted above.

If A, B are positive integers bounding the magnitudes of the coefficients of f, g respectively, then we still need to bound the magnitude of $\det S'$. In order to deal with the unknown signs, compute the matrix $S' \times S'^T$, whose determinant is $(\det S')^2$. The elements on the leading diagonal of this matrix are all > 0 and have the largest magnitude within each row, because all the non-zero elements in the corresponding row of S' and column of S'^T match up – elsewhere some non-zero elements are multiplied by zero elements thereby giving a smaller sum.

Moreover, the n diagonal elements corresponding to f-rows each have the value $m+1$, because there are $m+1$ non-zero elements in each f-row, and similarly the m diagonal elements corresponding to g-rows have the value $n+1$. A determinant is defined to be the sum of all possible distinct products of elements, one from each row and column, with a sign determined by the choice of element. The product of elements on the leading diagonal always contributes with a positive sign, and in this case is the largest term in the sum. Other terms contribute smaller magnitudes with varying signs, and therefore it is plausible that the value of $(\det S')^2$ is bounded by $(m+1)^n(n+1)^m$, from which follows the overall bound on the magnitude of the resultant quoted above.

7 Exercises

The assessed questions in this set of exercises are the first three.

1. (** **Assessed** **)
   Showing all details of the computation, perform the Euclidean division of $x^2 + x + 1$ by $3x - 2$ over $\mathbb{Q}$, and then the pseudo-division over $\mathbb{Z}$.

2. (** **Assessed** **)
   From their basic definitions, compute all the non-zero orders of $2/3$ and $4/15$ (with respect to the conventional prime numbers), and hence

compute the content of the polynomial

$$\frac{2}{3}x^3 - \frac{4}{15}x.$$

3. (** **Assessed** **)
If $f(x) = ax^2 + bx + c$, express $\mathrm{Res}(f, f')$ as a determinant and hence compute its value. Compute the value also by applying the recursive algorithm (by hand). Show that the condition $\mathrm{Res}(f, f') = 0$ is the same as that obtained by explicitly eliminating x from $f(x) = f'(x) = 0$. Finally, express a, b, c in terms of the roots α, β of $f(x) = 0$ and hence prove that $\mathrm{Res}(f, f') = 0$ if and only if $\alpha = \beta$.

4. In $\mathbb{Z}_8[x]$, prove that $u(x) = 5x^4 + 2x^3 + 4x^2 + 7x + 2$ can be divided by $v(x) = 3x^2 + 5$, and compute the quotient $q(x)$ and remainder $r(x)$ by performing *by hand* and displaying the details of the Euclidean division. [You can *check* it by computer if you wish!]

5. Show that there is only one way to divide $x^2 + y^2$ by $x - 2y$ over $\mathbb{Z}$, but that over $\mathbb{Q}$ there are two way, and perform the two divisions.