Lecture 11

ALGEBRAIC BACKGROUND

This chapter introduces several computational tools for algebraic computation: Euclidean algorithm and its variants, Sturm Sequences, resultants and subresultants. We will review the standard approaches to algebraic number computation.

§1. Rings, Domains and Fields

We are interested in geometric computation. Although geometry can be highly intuitive, its precise language ultimately depends on algebra. That is the great achievement of Descartes, in his “algebraization” of geometry.

More of engineering and scientific computation involve computations in \( \mathbb{R} \). Now \( \mathbb{R} \) has analytic as well as algebraic properties. Here we focus on its algebraic properties. For a thorough understanding of these algebraic properties, we must look at substructures of \( \mathbb{R} \) as well as its extension into the complex numbers, \( \mathbb{C} \).

Here is the hierarchy of algebraic structures we consider:

\[
\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{A} \subseteq \mathbb{R} \subseteq \mathbb{C}.
\]

Let us briefly introduce these substructures. We will not develop the theory of groups, rings and fields systematically; instead we rely on your intuitions about basic properties of numbers.

\( \mathbb{N} = \{0, 1, 2, \ldots \} \) are the natural numbers. It is an example of a semiring.

\( \mathbb{Z} = \{0, \pm 1, \pm 2, \ldots \} \) are the integers. It is our main example of a ring. Most of the simple properties you can think of involving \( \mathbb{Z} \) will probably be ring properties. For instance, \( \mathbb{Z} \) has two operations + and \( \times \) which are related by the distributive law.

A ring \( R \) (with unity 1) is a set of elements with two binary operations +, \( \times \) and constants 0, 1 such that \((R, +, 0)\) is a group, and \((R, \times, 1)\) is a monoid (meaning that \( \times \) is an associative operation with identity 1). Thus a monoid is like a group, except that it lacks an inverse. If \( \times \) is commutative, then we have a commutative ring. Moreover, these two operations are related by the distributive law: \( a(b + c) = ab + ac \). A semiring is basically a ring in which + may not have an inverse.

\( \mathbb{Q} = \{m/n : m, n \in \mathbb{Z}, n \neq 0\} \) are the rational numbers. It is our main example of a field. In a field, unlike a ring, the operation \( \times \) has an partial inverse, denoted by \( \div \). It is a partial inverse because you can divide by any value, as long as the value is non-zero.

\( \mathbb{A} \) are the real algebraic numbers. Each element of \( \mathbb{A} \) is just a real root of a polynomial with integer coefficients. For instance, \( \sqrt{3} \) is algebraic because it is a real root of the polynomial \( X^2 - 3 \). In this book, the numerical quantities we handle is almost exclusively in this set. Most of the interesting algebraic properties of reals reside within \( \mathbb{A} \).
\[ \mathbb{R} \text{ are the real numbers} \quad \text{it is not as easy to describe this set. But we have} \quad \text{intuitions about it based on the real number line and its total ordering property.} \]

\[ \mathbb{C} \text{ are the complex numbers. It is easy to describe, assuming you know } \mathbb{R}: \quad \text{elements of } \mathbb{C} \text{ are just } a + ib \text{ where } i = \sqrt{-1}. \text{ Addition and multiplication on such numbers are defined as usual.} \]

**Divisibility Concepts.** If \( a, b, c \in D \), then we say \( a \) **divides** \( b \), denoted \( a | b \), in case there exists some non-zero \( c \) such that \( ac = b \). We call \( a \) a **divisor** of \( b \). Two special types of divisors are important to notice: If \( b = 0 \) then \( a \) is a **zero-divisor**; if \( b = 1 \), we call \( a \) a **unit**. For instance, in \( \mathbb{Z}, 2 | 8 \) but \( 2 \nmid 3 \). In \( \mathbb{Q}, 2 | 3 \). In \( \mathbb{Z} \) and \( \mathbb{Q} \), the only zero-divisor is 0. In \( \mathbb{Z} \) the only units are \( \pm 1 \) while in \( \mathbb{Q} \), every non-zero is a unit.

If \( a = bc \) and \( b \) is a unit, we say \( a \) and \( c \) are **associates**. Clearly, this is an equivalence class. Usually, we fix some convention whereby one member of the equivalence class is called **distinguished**. For instance, in \( \mathbb{Z} \) the equivalence classes have the form \( \{n, -n\} \) where \( n \in \mathbb{N} \), and the distinguished member is taken to be \( n \).

We call \( a \) **irreducible** if the only divisors of \( a \) are units or its associates. A **prime** element \( p \) is a non-zero, non-unit such that if \( p | ab \) then \( p | a \) or \( p | b \). For instance, in \( \mathbb{Z} \), the prime numbers are \( \pm 2, \pm 3, \pm 5, \pm 7, \ldots \). If we take distinguished elements only, then primes are exactly as we usually think of them. Note that 1 is not regarded as prime since it is a unit. REMARK: prime elements of a domain are irreducible, but irreducible elements may not be prime.

**Domains** are rings whose only zero divisor is 0 (this is the trivial zero-divisor). Since \( \mathbb{C} \) has no non-trivial zero divisors, it means all the rings in (1) are domains.

Finite rings such as \( \mathbb{Z}_n = \{0, 1, \ldots, n - 1\} \) (with operations modulo \( n \)) are domains iff \( n \) is non-prime. If \( R \) is a ring, we can consider the set of all \( n \times n \) matrices with entries from \( R \). This is a ring of matrices, with addition and multiplication defined in the usual way. For instance, if \( R \) are the \( 2 \times 2 \) matrices over \( \mathbb{Z} \), then we have another example of a non-domain.

**Fields, Quotient Fields, Algebraic Closure.** A **field** is a domain in which all non-zero elements have inverses. We can construct a unique minimal field to extend any domain \( D \) using the familiar construction that extends \( \mathbb{Z} \) to \( \mathbb{Q} \). This field is called the **quotient field** of \( D \), and denoted \( Q_D \). Element of \( Q_D \) are represented by pairs \((a, b)\) where \( a, b \in D \) and \( b \neq 0 \), under the equivalence \((ca, cb) \equiv (a, b)\) for all \( c \neq 0 \). Usually we write \( a/b \) for the equivalence class of \((a, b)\).

**Polynomial Rings.** Besides matrix rings, another canonical way to construct new rings from a given ring \( R \) is the **polynomial ring** \( R[X] \) over a variable \( X \). A typical member of \( R[X] \) is written \( A(X) = \sum_{i=0}^{\delta} a_i X^i \) \((n \geq 0)\) where \( a_i \in R \) with addition and multiplication of polynomials defined in the standard way. If \( a_n \neq 0 \), then the degree of \( A(X) \) is \( n \) and \( a_n \) is called the **leading coefficient.** By definition, if \( A = 0 \) is the zero polynomial, then \( \deg(A) = -\infty \) and \( \text{lead}(A) = 0 \). These are denoted \( \deg(A) = n \) and \( \text{lead}(A) = a_n \), respectively. If we have another variable \( Y \), we can form \( (R[X])[Y] = R[X][Y] \). Since \( XY = YX \), this ring is the same as \( R[Y][X] \), and which we simply write as \( R[X,Y] \). This can be extended to any finite number of variables, \( R[X_1, \ldots, X_n] \).

If \( D \) is a domain, then the quotient field of \( D[X] \) is denoted \( D(X) \). The elements of \( D(X) \) may be denoted by \( P(X)/Q(X) \) where \( P(X), Q(X) \in Q[X] \).
and \( Q(X) \neq 0 \). Note that \( Q_D \) and \( D[X] \) are both contained in \( D(X) \). Again, the multivariate version is denoted \( D(X_1,\ldots,X_n) \).

Suppose \( D' \) is an extension of \( D \) and \( A(X) \in D[X] \). For any \( \alpha \in D' \), we let \( A(\alpha) \in D' \) denote the value obtained by evaluating the polynomial at \( \alpha \). We call \( \alpha \) a zero of \( A(X) \) if \( \alpha = 0 \), or a root of the equation \( A(X) = 0 \). Finding zeros of polynomials, and its generalizations, may be called the “fundamental computational problems” of algebra (cf. [2, Chap. 0]).

First assume \( D \) is a field. The following captures what we learn in High School about division of polynomials:

**Theorem 1 (Remainder Theorem)**  For all \( A,B \in D[X] \) where \( B \neq 0 \), there exists unique \( Q,R \in D[X] \) such that

\[
A = QB + R
\]

such that \( \deg R < \deg A \).

We call \( Q,R \) the quotient and remainder of \( A \) divided by \( B \). The algorithm to compute \( Q \) and \( R \) is well-known in high-school algebra. If \( D \) is not a field, we have a modified result:

**Theorem 2 (Pseudo Remainder Theorem)**  \( A,B \in D[X] \) with \( \delta = \max\{0,1+\deg(A) - \deg(B)\} \) and \( \beta = \text{lead}(B) \). Then exists unique \( Q,R \in D[X] \) such that

\[
\beta^\delta A = QB + R
\]

such that \( \deg R < \deg A \).

This is proved as Lemm 3.4 in [Yap]. We now call \( Q,R \) the pseudo quotient and pseudo remainder of \( A \) divided by \( B \). In fact, \( Q,R \) can be computed by the same division algorithm above.

From this, we can deduce that if \( \alpha \in D \) is a root of \( A \) then \( A = QB \) where \( B = X - \alpha \) and \( \deg(Q) = \deg(A) - 1 \). In proof, by the pseudo remainder theorem then \( A = QB + R \) with \( \deg(R) < \deg(B) \). Since \( B(\alpha) = 0 \), we conclude that \( 0 = A(\alpha) = Q(\alpha)B(\alpha) + R(\alpha) = R(\alpha) \). Since \( \deg(R) = 0 \) or \( -\infty \), this implies \( R = 0 \). This concludes our proof. It follows by induction that, in any extension of \( D \), that \( A \) has at most \( \deg(A) \) roots.

**Unique factorization domains** (UFD). Recall the Fundamental Theorem of Arithmetic says that every \( n \in \mathbb{N} \), if \( n \geq 2 \), can be written uniquely (up to re-ordering) as a product of prime numbers. UFD’s are domains \( D \) in which the analogue of the Fundamental Theorem of Arithmetic is true: namely, every non-zero \( a \in D \) can be written as the product of the form

\[
a = up_1,\ldots,p_k
\]

where \( u \) is unit and \( p_i \) are prime numbers. Moreover, this product is unique up to re-ordering (i.e., if \( a = u'q_1,\ldots,q_k \) then \( k = \ell \) and \( u' = u \) and, after reordering the \( q_i \)'s, each \( p_i = q_i \)). Note that a field is trivially a UFD. The first theorem about UFD’s is from Gauss:

**Theorem 3**  If \( D \) is a UFD, then \( D[X] \) is a UFD.

See [Yap, Sect. 2.1] or any standard algebra book. It follows by induction that \( D[X_1,\ldots,X_n] \) is a UFD.

The concept of divisibility can be extended into the quotient field \( Q_D \). See [Yap, §3.1.3]. The irreducible elements of \( Q_D \) are \( q \) and \( q^{-1} \) where \( q \) are the
irreducible elements of $D$. A factorization of $b \in D$ is an expression of the form $b = u q_1 \cdots q_m$ where $u$ is a unit, the $e_i$ are positive integers, the $q_i$’s are distinct irreducible elements, and $q_i q_j$ is a non-unit for all $i, j$. Then it is clear that every element of $D$ has a unique factorization up to reordering and up to choice of units. If $b = u q_1 \cdots q_m$ then we say $a$ divides $b$ iff $a = u q_1 \cdots q_m$ where $0 \leq d_i \leq e_i$ ($i = 1, \ldots, m$). For instance $1/2$ divides $1/4$ in $Q$ but the only elements that divide $1/2$ are the units, and $\pm 1/2$. We can now define $\text{GCD}(a/b, a'/b') = \prod_p p^{\text{ord}(p)}$ where $p$ range over all irreducible non-units and $\text{ord}(p)$ is the integer $k$ such that $p^k|(a/b)$ and $p^k|(a'/b')$ and $|k|$ is maximized. For instance, $\text{GCD}(1/2, 4/9) = 1$. On the other hand, $\text{GCD}(18/2, 15/4) = 3/2$. We can then define $\text{GCD}$ in $Q_D$ and also in $Q_D[X]$.

**The Floating Point Domain.** Let $D$ be any UFD, and let $p$ be an irreducible element of $D$. Consider the domain $D/p := \{fp^e : f \in D, e \in \mathbb{Z}\}$ viewed as a subdomain of $Q_D$. When $D = \mathbb{Z}$ and $p = 2$, the domain $\mathbb{Z}/2$ is usually called the binary floating point numbers (or bigfloats for short).

The irreducible elements in $D/p$ are $p^{-1}$ and the irreducible elements in $D$.

**Square-Freeness.** Let $D$ be a UFD. A polynomial $A(X) \in D[X]$ is said to be square-free if $\text{deg}((\text{GCD}(A(X), A'(X)))) = 0$ where $A'(X) = \frac{dA(X)}{dx}$ is the formal differentiation of $A(X)$. In other words, $A(X)$ has no double roots in the algebraic closure of $D$. We define the square-free part of $A(X)$ to be

$$\text{SqFree}(A(X)) := \frac{A(X)}{\text{GCD}(A(X), A'(X))}.$$  

Note that this division is well defined in $D[X]$, i.e., this is an exact division with no remainders.

**Quadratic Number Fields and Failure of UFD Property.** Consider $\mathbb{Q}[\sqrt{d}]$. This is easily shown to be a field, i.e., $\mathbb{Q}[\sqrt{d}] = \mathbb{Q}(|\sqrt{d}|)$. Moreover, these fields are distinct and exhaustively enumerated if we choose $d \in \mathbb{Z}$ to be square-free. For instance, for $|d| \leq 7$, there are exactly 14 such fields, corresponding to $d = \pm 1, \pm 2, \pm 3, \pm 5, \pm 6, \pm 7$.

The concept of divisibility in $\mathbb{Q}[\sqrt{d}]$ is trivial since non-zero element is a unit. To get the interesting “arithmetical structure” of $\mathbb{Z}$, we must consider the concept of integers in $\mathbb{Q}[\sqrt{-1}]$. In the simplest case where $d = -1$, this us the Gaussian integers, $\mathbb{Z}[\sqrt{-1}]$. We usually write $i = \sqrt{-1}$. In general, the set of integers in $\mathbb{Q}[\sqrt{d}]$ is denoted $\mathbb{O}_{\sqrt{d}}$. It can be shown that $\mathbb{O}_{\sqrt{-5}} = \mathbb{Z}[\sqrt{-5}] = \{x + y\sqrt{-5} : x, y \in \mathbb{Z}\}$. We now see that this domain is not a UFD. It can be shown that the numbers $2, 3, 1 + \sqrt{-5}, 1 - \sqrt{-5}$ are irreducible. They are not associates of each other because the units in this domain are $\pm 1$. But 6 can be expressed as two distinct product of irreducible factors:

$$6 = 2 \cdot 3 = (1 + \sqrt{-5}) \cdot (1 - \sqrt{-5}).$$

We conclude that $\mathbb{O}_{\sqrt{-5}}$ is not a UFD.

**§2. Euclid’s Algorithm**

In a UFD, the concept of a greatest common divisor is well-defined. See [2, Chap. 2] for the definition of $\text{GCD}$ in a UFD. For instance, $\text{GCD}(6, 15) = 3$. 

---

© Mehlhorn/Yap Draft February 16, 2005
Euclid's algorithm for computing the integer GCD is well-known, but we briefly recall it here.

Euclid’s Algorithm: given integers \(m_0 > m_1 > 0\), we compute \(\gcd(m_0, m_1)\) as follows: form the following sequence

\[m_0, m_1, m_2, \ldots, m_k\]

where \(m_{i+1} = m_{i-1} \mod m_i\), and \(m_{k-1} \mod m_k = 0\). Then \(\gcd(m_0, m_1) = m_k\).

Extended Euclidean Algorithm: The goal is to compute two numbers \(s, t\) such that

\[\gcd(m, n) = sm + tn.\]

We call such \((s, t)\) a co-factor of \((m, n)\). Let us define \(q_i\) by the equation

\[m_{i+1} = m_{i-1} - q_i m_i\]

(so \(q_i\) is the quotient of \(m_{i-1}\) divided by \(m_i\)). Define the sequences

\[(s_0, s_1, s_2, \ldots, s_k), \quad (t_0, t_1, t_2, \ldots, t_k)\]

where \(s_0 = 1, s_1 = 0, t_0 = 1, s_1 = 0,\) and

\[s_{i+1} = s_{i-1} - q_is_i, \quad t_{i+1} = t_{i-1} - q_it_i.\]

We see that the following equation is true:

\[m_i = s_im + t_in\]

for \(i = 0, \ldots, k\). The base case \((i = 0, 1)\) is true by our choice for \(s_0, t_0, s_1, t_1\). The other cases follow by an easy induction.

How general is this method? This is captured in the concept of an Euclidean Ring [Yap, Chap 2].

APPLICATIONS. Suppose you want to compute the inverse of a number \(x \in \mathbb{Z}_p\). You can use the extended Euclidean algorithm to compute \(s\) and \(t\) such that \(sx + tp = \gcd(x, p) = 1\). So \(s\) is the inverse of \(x\) modulo \(p\).

Similarly, suppose \(\theta\) is the root of some irreducible polynomial \(P(X) \in \mathbb{Z}[X]\). Then the ring \(\mathbb{Q}[\theta]\) is actually a field (see [Yap, Chap 5.1]). Every element in \(\mathbb{Q}[\theta]\) can be written as a polynomial in \(\theta\). Given \(A(\theta) \in \mathbb{Q}[\theta]\), we want to compute the inverse of \(A(\theta) \in \mathbb{Q}[\theta]\) in exactly the same way as before, using the extended Euclidean algorithm on the polynomials \(A(X)\) and \(P(X)\).

§3. Multivariate GCD

Let us consider the problem of computing GCD in \(R = D[X_1, \ldots, X_n]\) for \(n \geq 1\) where \(D\) is a UFD. The most important case is \(D = \mathbb{Z}\). However, the best way to compute GCD in \(R\) is far from obvious.

Univariate Polynomials. Even for \(n = 1\), there are interesting issues. How do we compute \(\gcd(A, B)\) for \(A, B \in D[X]\)? The simplest way is view \(A, B\) as elements of \(Q_D[X]\) where \(Q_D\) is the quotient field of \(D\). In this case, Euclid’s algorithm is applicable, since we can define the remainder of any two polynomials in \(Q_D[X]\). Let \(C \in Q_D[X]\) be this \(\gcd\). Then \(\gcd\) in \(D[X]\) can be recovered from \(C\) (Exercise).

An alternative method proceeds in two steps:
1. Compute the primitive-decomposition of \(A\) and \(B\): let \(A = aA_1\) where \(a\) is the content of \(A\) and \(A_1 \in D[X]\) is the primitive-part of \(A\). Similarly, let
§3. Multivariate GCD

Bivariate Polynomials. The issues with multivariate polynomials are already illustrated by considering the case $n = 2$. Our goal is to reduce this to the univariate case.

First, let us generalize the concept of primitive decomposition. An element $E \in D[X,Y]$ is type 0 if $E \in D$, type X if $E \in D_X \setminus D$, type Y if $E \in D_Y \setminus D$, and otherwise type XY. Also, if $u \in D$ is a unit, we say it is type 1. We call $\tau \in \{0,1,X,Y,XY\}$ a type symbol. A quadruple

$$(E_0, E_X, E_Y, E_{XY})$$

is a primitive decomposition of an element $E \in D[X,Y]$ if

- $E = E_0 E_X E_Y E_{XY}$.
- $E_\tau$ is type $\tau$ or type 1.
- If $(E'_0, E'_X, E'_Y, E'_{XY})$ is another quadruple satisfying the above conditions then $E_{XY} | E'_{XY}$, $E_X E_{XY} | E'_X E_{XY}$ and $E_Y E_{XY} | E'_Y E_{XY}$.

Lemma 4 The primitive decomposition of $E$ is unique up to associates. I.e., if $(E_0, E_X, E_Y, E_{XY})$ and $(E'_0, E'_X, E'_Y, E'_{XY})$ are two such decompositions, then $E_i \sim E'_i$ for $i \in \{0, X, Y, XY\}$.

Proof. We may write $E = \prod_{i=1}^{m} P_i^{d_i}$ where $P_i$ are non-associated irreducible elements of $D[X,Y]$ and $d_i \geq 1$. Each $P_i$ has a type. Collecting all the $P_i$'s to type $\tau$ into $E_\tau$, and defining $E_\tau = 1$ incase there are no $P_i$'s of type $\tau$, we get a quadruple $(E_0, E_X, E_Y, E_{XY})$ that is unique up to associates. We can verify that this quadruple is a primitive decomposition. Q.E.D.

Let $A = \prod_i P_i^{d_i}$ and $B = \prod_i P_i^{e_i}$ be the unique factorization of $A, B$ where $P_i$ are irreducible elements in $D[X,Y]$ and $d_i \geq 0$ and $e_j \geq 0$. Then $\text{GCD}(A, B) = \prod_i P_i^{\min(d_i, e_i)}$. But we can express

$$\text{GCD}(A, B) = C_0 C_X C_Y C_{XY}$$

is the primitive decomposition of $\text{GCD}(A, B)$.

How can we compute the decomposition $(E_0, E_X, E_Y, E_{XY})$ in (4)?

We can view elements of $D[X,Y]$ as polynomials in $D[X][Y]$, or as polynomials in $D[Y][X]$. Suppose $\text{GCD}_X(A, B)$ is the result when $A, B \in D[Y][X]$, and $\text{GCD}_Y(A, B)$ when $A, B \in D[X][Y]$. We first note that

$$\text{GCD}(A, B) = \text{GCD}_X(A, B) = \text{GCD}_Y(A, B).$$

In proof, note that $\text{GCD}(A, B)$ divides $A$ and $B$ and for any $g$ that divides both $A$ and $B$ must also divide $\text{GCD}(A, B)$. In particular, $g = \text{GCD}_X(A, B)$ (viewed as an element of $D[X,Y]$) has this property. So $\text{GCD}_X(A, B) | \text{GCD}(A, B)$. Conversely, we
see that $\gcd(A, B) \mid \gcd_X(A, B)$ by viewing $\gcd(A, B)$ as an element of $D[Y][X]$.
This proves $\gcd(A, B) = \gcd_X(A, B)$.

Now computing in $D[X][Y]$, we can compute the primitive decomposition

$$\gcd_Y(A, B) = C_0 C_X C_{XY}$$

and similarly, computing in $D[Y][X]$ we obtain the primitive decomposition

$$\gcd_X(A, B) = D_0 D_Y D_{XY}.$$ 

It is easily verified that $D_0 = C_0 = E_0$, $C_X = E_X$ and $D_Y = E_Y$. Finally,

$$E_{XY} = \gcd_X(C_{XY}, D_{XY}).$$

Thus, with three calls to a $\gcd$ decomposition algorithm, we obtain the decomposition in (4),

§4. Resultants

Fix any UFD $D$. Assume polynomials in this section have coefficients in $D$.

It is useful to introduce the following concept of determinantal polynomials
[Yap, §3.3]: let $M$ be a $m \times n$ matrix with $m \leq n$. Then determinantal polynomial of $M$ is

$$d \text{pol}_1(M) := \det(M_m) X^{n-m} + \det(M_{m+1} X^{n-m-1} + \cdots + \det(M_n)$$

where $M_i$ is the square submatrix of $M$ comprising the first $m - 1$ columns of $M$ and the ith column of $M$. In case $m > n$, simply define $d \text{pol}_1(M) = 0$.

In case $P_1, \ldots, P_m$ are polynomials and $n = 1 + \max_i \{\deg(P_i)\}$ then $\text{mat}_n(P_1, \ldots, P_n)$ is the $m \times n$ matrix whose ith row contains the coefficients of $P_i$ listed in order of decreasing degree, treating $P_i$ as having (nominal) degree $n - 1$. Also, $d \text{pol}_n(P_1, \ldots, P_n)$ is the determinantal polynomial of this matrix.

Let $A, B \in D[X]$. If $A = \sum_{i=0}^m a_i X^i$ and $B = \sum_{i=0}^n b_j X^j$, with $a_m b_n \neq 0$, then the Sylvester Matrix of $A, B$ is defined as the following square $(m+n) \times (m+n)$ matrix

$$\text{mat}\left(\begin{array}{c}
X^{n-1} A, X^{n-2} A, \ldots, X^1 A, X^0 A, X^{m-1} B, X^{m-2} B, \ldots, X^0 B
\end{array}\right) =
\begin{bmatrix}
  a_m & a_{m-1} & \cdots & a_0 \\
  a_m & a_{m-1} & \cdots & a_0 \\
  \vdots & \vdots & \ddots & \vdots \\
  b_n & b_{n-1} & \cdots & a_m \\
  \vdots & \vdots & \ddots & \vdots \\
  b_n & b_{n-1} & \cdots & b_0
\end{bmatrix}.$$ 

Thus,

$$d \text{pol}_{m+n}(X^{n-1} P, X^{n-2} P, \ldots, X^1 P, P, X^{m-1} Q, X^{m-2} Q, \ldots, X^1 Q, Q)$$

is a constant polynomial, equal to the determinant of this matrix. It is called the resultant of $A$ and $B$, and written $\text{res}(A, B)$.

**Theorem 5** $\text{res}(A, B) = 0$ iff $\gcd(A, B)$ is not a constant.
Theorem 6 Let $A = \sum_{i=0}^{m} a_i X^i$, and $B = \sum_{j=0}^{n} b_j X^j$ be polynomials where the coefficients $a_i, b_j$ are indeterminates. We define $\deg(a_i) = m - i$ and $\deg(b_j) = n - j$, so that $A, B$ can be regarded as homogeneous polynomials. Then $\text{res}(A, B)$ is a homogeneous polynomial in $D[a, b] = D[a_0, \ldots, a_m, b_0, \ldots, b_n]$.

Proof. This is basically the proof in the next lemma. Q.E.D.

§5. Homogeneous Polynomials

Now assume $A, B \in D'[X_1]$ where $D' = D[X_2, \ldots, X_r]$, $r \geq 1$, and $D$ is a UFD. Thus $A, B \in D[X_1, \ldots, X_r]$ are multivariate polynomials. Let $\text{res}_{X_i}(A, B) \in D[X_2, \ldots, X_r]$ denote the resultant where $A, B$ are viewed as polynomials in $X_1$. To avoid double subscripts, we write $\text{res}_i(A, B)$ instead of $\text{res}_{X_i}(A, B)$.

More generally, define $\text{res}_i(A, B)$ where $i = 1, \ldots, r$ for the resultant with respect to $X_i$. Alternatively, the subscript $i$ or $X_i$ says we are “eliminating $X_i$” from the system of equations $A = B = 0$. In examples, we usually write $X$ for $X_1$, $Y$ for $X_2$ and $Z$ for $X_3$. Now, when we speak of the “degree” of $A \in D[X_1, \ldots, X_r]$, we mean its total degree in $X_1, \ldots, X_r$, still denoted $\deg(A)$. For example, $\deg(X^3Y - XY + 2) = 4$.

A multivariate polynomial can be written as

$$A = \sum_{e \in I} c_e X^e, \quad c_e \in D$$

where $I \subseteq \mathbb{N}^r$ is a finite set and $X^e = X_1^{e_1} \cdots X_r^{e_r}$ where $e = (e_1, \ldots, e_r)$. If $c_e \neq 0$, then we say $X^e$ occurs in $A$ and call $c_e X^e$ a term of $A$. The polynomial $A$ is $X_i$-regular if $X_i^{\deg(A)}$ occurs in $A$. We simply say “regular” for $X_1$-regular.

Let $|e| = e_1 + \cdots + e_r$. So $\deg(A) = \max_{e \in I} |e|$. We say $A$ is homogeneous if $|e| = |f|$ for all $e, f \in I$. The zero polynomial is, by definition, homogeneous and has degree $-\infty$.

The following are basic facts about homogeneous polynomials.

Lemma 7 Let $A \in D[X_1, \ldots, X_r]$ have degree $m$.

(i) Then $A$ is homogeneous iff for all nonzero $t \in D$,

$$A(tX_1, \ldots, tX_r) = t^m A(X_1, \ldots, X_r).$$

(ii) Let $\deg_1(A) = m' \leq m$. Write $A = \sum_{i=0}^{m} a_i X^i$ where $a_i \in D[X_2, \ldots, X_r]$ and $X = X_1$. Then $A$ is homogeneous iff each nonzero $a_i$ is homogeneous of degree $m - i$.

(iii) There is a bijection between elements in $D[X_1, \ldots, X_r]$ and the homogeneous polynomials in $D[X_0, X_1, \ldots, X_r] \setminus D[X_0]$ such that $A \in D[X_1, \ldots, X_r]$ corresponds to the homogeneous $\hat{A} \in D[X_0, \ldots, X_r] \setminus D[X_0]$ iff $\deg(A) = \deg(\hat{A})$ and $A(X_1, \ldots, X_r) = \hat{A}(1, X_1, \ldots, X_r)$.

Proof. (i) The result is clearly true if $A$ has only one term. If $A$ has more than one term, then $A$ is homogeneous iff each term has the same degree (say $m > 0$). But a term $c_e X^e$ has degree $m$ iff $t^m$ appears in the corresponding term for $A(tX_1, \ldots, tX_r)$. Hence, factoring $t^m$ from all these terms, we get $A(tX_1, \ldots, tX_r) = t^m A(X_1, \ldots, X_r)$.

(ii) If $A$ is homogeneous, then each $a_i$ is necessarily homogeneous of degree $m - i$. The converse is also clear.
(iii) The bijection \( A \in D[X_1, \ldots, X_r] \) into homogeneous \( \hat{A} \in D[X_0, \ldots, X_r] \) is obtained as follows: if \( A \in D \), then \( \hat{A} = A \). Otherwise, let \( m = \deg(A) > 0 \). For each term \( c_\tau X^\epsilon \) of \( A \), we obtain a corresponding term \( c_\tau X^\epsilon X_0^{m - \lvert \epsilon \rvert} \). This bijection between \( A \) and \( \hat{A} \) has the claimed properties. \( \text{Q.E.D.} \)

Note that we do not assume that \( A \) is regular in (ii). Thus, \( a_m \) is not necessarily a constant. To illustrate (i), if \( A(X, Y) = X^4 - 3X^3Y + X^2Y^2 - 4Y^4 \) then \( \hat{A}(X, Y) = t^4(X^4 - 3X^3Y + X^2Y^2 - 4Y^4) \). Property (i) can be taken as the definition of homogeneous polynomials. In illustration of (iii), if \( B(X, Y) = 2X^2Y + 3XY - 1 \) then \( \hat{B} = 2X^2Y + 3XYX_0 - X_0^3 \).

If \( A \in D[X_1, \ldots, X_r] \), let \( \hat{A} \in D[X_1, \ldots, X_r, W] \) denote the standard homogenization of \( A \) using a new variable \( W \), with the property that \( \deg(A) = \deg(\hat{A}) \) and \( \hat{A}(X_1, \ldots, X_r, 1) = A(X_1, \ldots, X_r) \). Also, for \( B \in D[X_1, \ldots, X_r, W] \), let \( B^\lor \) denote the operation of substituting \( W = 1 \) in \( B \). Thus, \( \hat{A}^\lor = A \).

**Lemma 8** If \( A, B \) be arbitrary polynomials (not necessarily homogeneous).

(i) \( \hat{A}B = \hat{A}\hat{B} \)

(ii) \( \hat{A} + \hat{B} = \hat{A} + \hat{B}X_0^{\deg(A) - \deg(B)} \) (where \( \deg(A) \geq \deg(B) \))

(iii) \( \text{GCD}(\hat{A}, \hat{B}) = \text{GCD}(\hat{A}, \hat{B}). \)

We are ready to prove the main result of this section.

**Theorem 9** Let \( A, B \) be regular homogeneous polynomials in \( r \geq 2 \) variables, of degrees \( m \) and \( n \) respectively. If \( \text{res}_X(A, B) \neq 0 \) then \( \text{res}_X(A, B) \) is homogeneous of degree \( mn \).

**Proof.** Write \( A = \sum_{i=0}^m a_i X^i \) and \( B = \sum_{j=0}^n b_j X^j \) where \( X = X_1 \) and \( a_i, b_j \in D[V] \) where, for simplicity, we write \( V \) for \( (X_2, \ldots, X_r) \). By regularity, either \( a_i = 0 \) or \( \deg(a_i) = m - i \) for all \( i = 0, \ldots, m \). Similarly, either \( b_j = 0 \) or \( \deg(b_j) = n - j \) for \( j = 0, \ldots, n \). Let \( R(V) = \text{res}_X(A, B) \). For any \( t \), let \( tV = (tX_2, \ldots, tX_r) \). From (5), we conclude that

\[
R(tV) = \begin{bmatrix}
a_m & t a_{m-1} & \cdots & t^m a_0 \\
a_m & t a_{m-1} & \cdots & t^m a_0 \\
\vdots & & \ddots & \vdots \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 \\
\vdots & & \ddots & \ddots & \ddots \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 \\
\end{bmatrix}.
\]

If we next multiply the \( i \)-th row of \( A \)'s by \( t^{i-1} \) and the \( j \)-th row of the \( B \)'s by \( t^{j-1} \), we obtain

\[
t^n R(tV) = \begin{bmatrix}
a_m & t a_{m-1} & \cdots & t^m a_0 & t^{m+1} a_0 \\
t a_m & t^2 a_{m-1} & \cdots & t^{m+1} a_0 \\
\vdots & & \ddots & \ddots & \vdots \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 & t^{m+n-1} a_0 \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 & \ddots \\
\vdots & & \ddots & \ddots & \ddots & \ddots \\
b_n & t b_{n-1} & \cdots & t^{n-1} b_1 & t^n b_0 & \ddots \\
\end{bmatrix}.
\]
where \( p = \binom{m}{n} + \binom{n}{2} \). But in the righthand side determinant, we can extract a factor of \( t^{i-1} \) from the \( i \) th column (for \( i = 1, \ldots, m+n \)). Hence the righthand side determinant is equal to
\[
t^{\binom{m+n}{2}} R(V).
\]
This proves that
\[
t^p R(tV) = t^{\binom{m+n}{2}} R(V).
\]
Hence \( R(V) \) is homogeneous and its degree is
\[
\binom{m+n}{2} - \binom{m}{2} - \binom{n}{2} = mn.
\]
Instead of a direct calculation, the reader may instantly see the truth of this last equation in terms of its combinatorial interpretation. Q.E.D.

**Corollary 10** Suppose \( A,B \in D[X_1, \ldots, X_r] \) are regular polynomials, not necessarily homogeneous. Then \( \text{res}_1(A,B) \), if non-zero, has degree \( \leq mn \) in \( X_2, \ldots, X_r \).

**Proof.** Our theorem says that \( R(X_0, X_2, \ldots, X_r) = \text{res}_1(\hat{A}, \hat{B}) \) is homogeneous of degree \( mn \). Since \( \text{res}_1(A,B) \) is equal to \( R(1, X_2, \ldots, X_r) \), its degree is at most \( \deg(R(X_0, X_2, \ldots, X_r)) \). Q.E.D.

Suppose the polynomial \( A = \sum_{i=0}^{m} a_i X^i \) (\( a_i \in D[X_2, \ldots, X_r] \)) is not regular, i.e., \( a_0 \notin D \). What can we do? One possibility is make divide \( A \) by \( a_0 \). The coefficients of \( A/a_0 \) can now be viewed as elements of a meromorphic series (i.e., power series with finitely many terms with negative powers). In essence, this is Newton’s approach. See [1, Lect. 9] for this development. A simpler approach is to consider the following transformation:

\[
X_i \mapsto Y_i + c_i X, \quad (i = 2, \ldots, r).
\]

The polynomial \( A'(X,Y_2, \ldots, Y_r) = A(X,Y_2+c_2 X, \ldots, Y_r+c_r X) \) will be regular for some choice of the \( c_i \)'s. In practice, this will turn a sparse polynomial into a very dense one.

The above theorem can be generalized as in the following exercise.

**Exercise 5.1:** Let \( A,B \in D[X_1, \ldots, X_r] \) be homogeneous with \( \deg(A) = m, \deg(B) = n \). If \( \deg_i(A) \) denotes the degree of \( A \) in \( X_i \), let \( \deg_1(A) = m', \deg_1(B) = n' \). Write \( \mu = m - m' \) and \( \nu = n - n' \). Thus \( A,B \) are regular iff \( \mu = \nu = 0 \). We have the following generalization of the previous theorem: if \( \text{res}_X(A,B) \neq 0 \) then \( \text{res}_X(A,B) \) is homogeneous of degree \( mn - \mu \nu \).

\[\diamondsuit\]
Sylvester matrix. Write \( S(y_0) \) for the matrix obtained from \( S \) by setting \( Y = y_0 \). So the question is whether \( S_0 = S(y_0) \). Note that \( S_0 \) depends on the degrees of \( A_0 \) and \( B_0 \). It is now easy to see:

**Lemma 11**

(i) If \( a_m(y_0) \neq 0 \) and \( b_n(y_0) \neq 0 \), then

\[ R(y_0) = \text{res}_X(A_0, B_0). \]

(ii) If \( a_m(y_0)b_n(y_0) \neq 0 \), then

\[ R(y_0) = 0 \iff \text{res}_X(A_0, B_0) = 0. \]

In this regards, it is convenient to call a curve \( A(X, Y) = 0 \) regular (or \( X \)-regular) in case \( \deg(A) = m \) and \( a_m \) is a constant. In other words,

\[ A(X, Y) = \sum_{i=0}^{m} a_i(Y)X^i \]

where \( a_m(Y) \) is a constant and \( \deg_Y(a_i) \leq m - i \). It is easy to see that by a translation, every curve can be made regular (Exercise).

We can now prove a simple version of Bezout’s theorem.

**Theorem 12 (Bezout)** Suppose \( A(X, Y), B(X, Y) \) are relatively prime and \( \deg(A) = m, \deg(B) = n \). Then the curves \( A = 0 \) and \( B = 0 \) has at most \( mn \) common points of intersection.

**Proof.** By a linear translation, we may assume that \( A \) and \( B \) are regular. Let \( R(Y) = \text{res}_X(A, B) \). Since they are relatively prime, \( R(Y) \) is non-zero. Hence \( \deg(R) \leq mn \). If \( (x_i, y_i) \ (i = 0, 1, \ldots, mn) \) are \( mn+1 \) distinct common intersections, then by a rotation of the curves, we further assume the \( y_i \)'s are distinct. It follows that \( R(y_i) = 0 \) for each \( i \). This is a contradiction since the degree of \( R(Y) \leq mn \).

\( \text{Q.E.D.} \)

**§7. Theory of Subresultants**

We can clearly apply Euclid’s algorithm to \( D[X] \) where \( D \) is a field. In case of a UFD, the search for an efficient algorithm was a major topic in early days of computer algebra. An efficient method here is based on the theory of subresultants.

The **content** of a polynomial \( A(X) \) is the GCD of all its coefficients. The **primitive part** of \( A(X) \) is obtained by dividing \( A(X) \) by its content. Two polynomials \( A(X), B(X) \) are similar if they have the same primitive part. We denote this by \( A(X) \sim B(X) \).

The **pseudo-Euclidean Polynomial Remainder Sequence** (PRS) of \( A(X), B(X) \) is

\[ (A_0, A_1, \ldots, A_k) \]

where \( A_0 = A, A_1 = B \) and \( A_{i+1} = \text{prem}(A_{i-1}, A_i) \), i.e., the pseudo-remainder of \( A_{i-1} \) divided by \( A_i \). The last term of this sequence is determined by the condition \( \text{prem}(A_{k-1}, A_k) = 0 \).

The sequence (6) is easy to define but it’s coefficients can grow exponentially large in \( k \). So what we would like is any sequence

\[ (B_0, B_1, \ldots, B_k) \]
such that $B_0 = A_0, B_1 = A_1,$ and $B_i \sim A_i$ for $i = 2, \ldots, k$. We call such a sequence a polynomial remainder sequence (PRS) for $A_0, A_1$. If $B_i = A_i/\beta_i$ where $\beta + i \in D$, for all $i \geq 0$, we say that (7) is based on $(\beta_0, \ldots, \beta_k)$.

Our goal is to present Collin’s “subresultant PRS algorithm” [?, ?]. It is considered the best algorithm in the family of algorithms based on sequences of β's.

**Subresultant Chain.** First we introduce a new sequence that contains a PRS as subsequence. Given $A, B$ as above, with $\deg(A) = m > n = \deg(B)$, define the $i$th subresultant of $A, B$ (for $i = 0, \ldots, n$) to be

$$dpol(X^{n-i}A, X^{n-i-2}A, \ldots, A, X^{m-i-1}B, X^{m-i-2}B, \ldots, B)$$

We denote this subresultant by $sres_i(A, B)$. For $i = n + 1, \ldots, m$, we define $sres_i(A, B)$ to be $A$ if $i = m$, to be $B$ if $i = m - 1$, and 0 otherwise. Observers that $sres_0(A, B)$ is just the standard resultant. The subresultant chain of $A, B$ is just the sequence

$$S = (S_m, S_{m-1}, \ldots, S_1, S_0) \quad (8)$$

where $S_i = sres_i(A, B)$.

Let $ppsc_i(A, B)$ (called the $i$th pseudo-principal subresultant coefficient of $A, B$) be the nominal leading coefficient of $sres_i(A, B)$, i.e., it is the leading coefficient of $S_i$ where we treat $sres_i(A, B)$ as a polynomial of nominal degree $i$. The true degree of $sres_i(A, B)$ may be strictly less than $i$. Hence we call $sres_i(A, B)$ regular if its nominal degree is equal to its actual degree. However, by definition, $ppsc_m(A, B) = 1$ and NOT $1_{lead}(A)$.

We are ready to present Collins’s algorithm. The algorithm uses a simple loop that resembles the usual Euclidean algorithm:

**Subresultant PRS Algorithm**

Input: Polynomials $A, B \in D[X]$ with $\deg A > \deg B > 0.$

Output: Subresultant PRS of $A, B$.

**Initialization**

$P_0 \leftarrow A; P_1 \leftarrow B;$

$a_0 \leftarrow 1_{lead}(P_0); a_1 \leftarrow 1_{lead}(P_1);$  
$\delta_0 \leftarrow \deg(P_0) - \deg(P_1);$  
$\beta_1 \leftarrow -(-1)^{\delta_0};$  
$\psi_0 \leftarrow 1; \psi_1 \leftarrow (a_0)^{\delta_0};$  
$i \leftarrow 0;$

**Loop**

While $P_1 \neq 0$ do

1. $P_{i+1} \leftarrow sres_i(P_{i-1}, P_i)$  
   \hspace{1cm} ($i = 1, \ldots, k - 1$);

2. $a_{i+1} \leftarrow 1_{lead}(P_i);$  

3. $\delta_i \leftarrow \deg(P_i) - \deg(P_{i+1});$

4. $\psi_{i+1} \leftarrow \frac{(a_{i+1})^{\delta_i}}{\psi_i^{(a_{i+1})^{\delta_i}}};$

5. $\beta_{i+1} \leftarrow (-1)^{\psi_{i+1}} \psi_{i+1}^{\delta_i} a_i;$  

$i \leftarrow i + 1;$

Return $(P_0, P_1, \ldots, P_i);$
Connection to the Block Structure Theorem. With respect to the subresultant chain (8), a subsequence

\[ B = (S_i, S_{i-1}, \ldots, S_j) \]

where \( m \leq i \geq j \geq 0 \) is called a non-zero block if \( S_i \sim S_j \neq 0 \), and for \( k = i-1, i-2, \ldots, j+1 \), \( S_k = 0 \). Let \( S_i \) and \( S_j \) be called the top and bottom of the block. Note that if \( i = j \) then the top and bottom coincide. A zero block is \( (S_i, S_{i-1}, \ldots, S_0) \) where each \( S_k = 0 \) for \( k = i-1, i-2, \ldots, 0 \) and \( S_{i+1} \neq 0 \). As a special case, the empty sequence ( ) is regarded as a zero block. The Block Structure Theorem (Theorem 3.10 [Yap, §3.7] says that the subresultant chain (8) can be divided into a concatenation of blocks,

\[ S = B_0; B_1; \cdots; B_k, B_{k+1} \]

where each \( B_i \) \((i = 1, \ldots, k)\) is a non-zero block, and \( B_{k+1} \) is a zero block. Let \( T_i \) and \( U_i \) be the top and bottom of block \( B_i \). Then we obtain two subchains:

\[ T = (T_0, T_1, \ldots, T_k), \quad U = (U_0, U_1, \ldots, U_k). \]

It can be shown that \( T \) (and thus \( U \)) is a PRS of \( A, B \). The correctness of Collin's algorithm can be inferred from this connection:

**Theorem 13** Collin's algorithm computes the sequence \((T_0, T_1, \ldots, T_k)\).

---

**Exercises**

**Exercise 7.1:** Compute the inverse of 100 modulo 263.

**Exercise 7.2:**
(i) Describe Euclid's algorithm for \( K[X] \) where \( K \) is a field.
(ii) Compute the \( \text{gcd} \) of \( A(X) = 6X^4 + 3X^3 - 2X^2 + 3X + 2 \) and \( B(X) = 6X^2 + X - 1 \) where \( A(X), B(X) \) are viewed at polynomials in \( \mathbb{Q}[X] \).
(iii) Show how to convert the \( \text{gcd} \) in a quotient field \( Q_D[X] \) to one in \( D[X] \).

**Exercise 7.3:** Let \( D \) is algebraically closed, and \( A, B \in D[X_1, \ldots, X_r] \).
(i) Suppose there is an infinite set \( E \subseteq D \) such that for all \( x_1, \ldots, x_r \in E \),
\[ A(x_1, \ldots, x_r) = 0. \]
Then \( A = 0 \).
(ii) Suppose for all \( x_1, \ldots, x_r \in D, B(x_1, \ldots, x_r) = 0 \) implies \( A(x_1, \ldots, x_r) = 0 \). Then \( A \mid B \).
(iii) Prove that if \( \deg(\text{gcd}(A, B)) = 0 \) then \( \text{res}_{X_i}(A, B) \) is non-zero.

**Exercise 7.4:**
(i) Bound the number of arithmetic operations in the subresultant PRS algorithm of \( A, B \). Obtain this bound as a function of \( m = \deg(A) \).
(ii) If the input coefficients have at most \( L \) bits, bound the bit complexity of the algorithm.

---

**§8. Algebraic Number Computation**

In this section, we review a standard approach to algebraic computing, namely to computing within a fixed number field \( \mathbb{Q}(\alpha) \) for some \( \alpha \). In theory, this approach is sufficient, since any finite algebraic computation can be embedded in some \( \mathbb{Q}(\alpha) \). In practice, this approach may not be good enough because the degree of \( \alpha \) may be exponential (or worse) and we need to know the irreducible polynomial of \( \alpha \). For more details, see [Yap, §6.3.1].
Computation in an Algebraic Number Field. If we are interested in performing the basic arithmetic operations on algebraic numbers, then the computer algebra text books inform us that we might as well be computing in a fixed number field \( \mathbb{Q}(\alpha) \). This is justified by the primitive element theorem which says that if \( \alpha_1, \alpha_2 \) are algebraic then \( \mathbb{Q}(\alpha_1, \alpha_2) = \mathbb{Q}(\alpha) \) for some \( \alpha \). In fact, we can take \( \alpha = m\alpha_1 + n\alpha_2 \) for some suitable integers \( m, n \). This generalizes to any finite algebraic extension \( \mathbb{Q}(\alpha_1, \ldots, \alpha_n) = \mathbb{Q}(\alpha) \). Moreover, any finite sequence of computation lies in such a finite extension.

Algebraic computations in \( \mathbb{Q}(\alpha) \) can be reduced to manipulating polynomials modulo the minimal polynomial \( A(X) \in \mathbb{Z}[X] \) of \( \alpha \).

1. First we claim \( \mathbb{Q}(\alpha) = \mathbb{Q}[\alpha] \). To see this, suppose \( \beta \in \mathbb{Q}(\alpha) \). By definition, \( \beta = B(\alpha) \) where \( B(X), C(X) \in \mathbb{Q}[X] \) and \( C(\alpha) \neq 0 \). But \( C(\alpha) \neq 0 \) means that \( C(X) \mod A(X) \neq 0 \). Since \( A(X) \) is irreducible, \( \gcd(C(X), A(X)) = 1 \). Hence, by the Extended Euclidean algorithm, there is \( S(X), T(X) \in \mathbb{Q}[X] \) such that

\[
S(X)C(X) + T(X)A(X) = \gcd(C(X), A(X)) = 1.
\]

(9)

Since \( A(\alpha) = 0 \), we obtain \( S(\alpha)C(\alpha) = 1 \). Therefore, \( \beta = B(\alpha) = B(\alpha)S(\alpha) \).

This proves \( \beta \in \mathbb{Q}[\alpha] \).

2. Next, \( \mathbb{Q}[\alpha] \) is a vector space over \( \mathbb{Q} \) of dimension \( d = \deg(A) \). Note that the \( d \) elements

\[
1, \alpha, \ldots, \alpha^{d-1}
\]

(10)

are linearly independent over \( \mathbb{Q} \). Otherwise, we would have constants \( c_i \)'s (not all 0) such that \( \sum_{i=0}^{d-1} c_i \alpha^i = 0 \). That means the degree of \( \alpha \) is \( \leq d - 1 \), contrary to our choice of \( d \). Moreover, these \( d \) elements span \( \mathbb{Q}[\alpha] \) because if \( \beta = B(\alpha) \) where \( B(X) \in \mathbb{Q}[X] \), and if \( B(X) \equiv B'(X) \mod A(X) \) then \( B(\alpha) = B'(\alpha) \). It follows that we might as well view \( B(X) \) as a polynomial modulo \( A(X) \). Thus \( \beta = B(\alpha) \) is a linear combination of the elements (10).

3. It follows that every element of \( \mathbb{Q}(\alpha) \) can be viewed as a polynomial \( B(X) \) modulo \( A(X) \). The addition, subtraction and multiplication of elements of \( \mathbb{Q}[\alpha] \) is just polynomial addition, subtraction and multiplication modulo \( A(X) \). The inverse of a nonzero polynomial \( C(X) \) modulo \( A(X) \) is obtained from the extended Euclidean algorithm, as the element \( S(X) \) in (9). Thus division of polynomials modulo \( A(X) \) is possible.

The preceding picture of algebraic computation is fine if you are interested in a fixed field \( \mathbb{Q}[\alpha] \) and only interested in algebraic relations in this field. In many applications, there is no á priori field of interest. The actual field is simply determined by all the algebraic numbers you encounter in the course of solving some computational problem. In this case, the degree of \( \alpha \) might be so huge as to make the entire computation infeasible. For instance, if your computation involves the algebraic numbers \( \alpha_1, \ldots, \alpha_n \), then you need \( \alpha \) such that \( \mathbb{Q}(\alpha) \) contains each of \( \alpha_1, \ldots, \alpha_n \). This means \( \deg(\alpha) \) can be as large as \( \prod_{i=1}^{n} \deg(\alpha_i) \).

Example: suppose you want to compute the shortest path amidst a set of \( n \) polygonal obstacles which are specified with integer coordinates. The shortest paths are polygonal paths whose breakpoints are corners of the obstacles. Then the length \( \ell \) of such paths are sums of square roots, \( \ell = \sum_{i=1}^{m} \sqrt{a_i} \) where \( a_i \in \mathbb{N} \). Thus \( \ell \) is an algebraic number whose degree can be as high as \( 2^m \), where \( m \) grows at least linearly with the number of obstacles. Computing the minimal polynomial of \( \ell \) is not considered feasible.

§9. Real Algebraic Computation

In this section, we review another standard approach to algebraic number computation.
The previous approach of computing within a number field \( \mathbb{Q}(\alpha) \) suffers from another restriction: we are not just interested in algebraic relations in \( \mathbb{Q}(\alpha) \), but in the metric\(^1\) or analytic properties of \( \mathbb{Q}(\alpha) \) when embedded in \( \mathbb{R} \). E.g., given \( x, y \in \mathbb{Q}(\alpha) \), we want to know if \( |x - y| < \varepsilon \). Such properties cannot be resolved by manipulating polynomials modulo \( A(X) \). Remarkably, it is possible to resolve such properties using purely algebraic methods, using the theory of Sturm sequences or its relatives.

To capture metric properties, we would view \( \sqrt{2} \), not as an element of \( \mathbb{Q}(\alpha) \), but as an element of \( \mathbb{R} \). As such, these elements can be approximated to any desired accuracy, \( \sqrt{2} = 1.4142\ldots \), by other elements in \( \mathbb{Q}(\alpha) \). However, we cannot represent such numbers by any fixed precision approximation; we need representation of \( \sqrt{2} \) that can yield more precision on demand.

§9.1. Isolating interval representation

To capture metric properties of algebraic numbers, the most commonly used representation the following. A polynomial \( A(X) \in \mathbb{Z}[X] \) is called a **defining polynomial** for \( \alpha \) if \( A(\alpha) = 0 \). A real interval \( I \) is called an **isolating interval** for \( \alpha \) modulo \( A(X) \) if (1) \( \alpha \in I \), and for any other real zero \( \beta \) of \( A(X) \), we have \( \beta \not\in I \). The pair \((A(X), I)\) is called an **isolating interval representation** for \( \alpha \). We write \( \alpha \sim (A(X), I) \) to indicate this relationship. If we need complex algebraic numbers, we can represent it as a pair of real algebraic numbers.

If \( I = [a, b] \), then we say that this representation has \(-\lg(b - a)\) **absolute bits** of precision. For instance, if \( b - a = 2^{-100} \), then we have 100 absolute bits of precision. Typically \( a, b \) are rational numbers, or binary floats.

We may make the following normalization assumption: if \( A(a)A(b) = 0 \), then \( \alpha = a = b \). In this case, the interval \( I = [a, b] = [a, a] \) is said to be **exact** and our representation has \( \infty \) bits of precision. So suppose \( A(a)A(b) \neq 0 \). It is easy to see that \( A(a)A(b) < 0 \). We assume that our representation (in the inexact case) stores one extra bit of information, so that we know whether \( A(a) > 0 > A(b) \) or \( A(a) < 0 < A(b) \).

It is easy to refine our isolating interval representation is easy: let \( c = (a + b)/2 \). We can evaluate \( A(c) \). If \( A(c) = 0 \) then we have found an exact representation. Otherwise, it is easy to determine whether \( J = [a, c] \) or \( J = [c, b] \) is an isolating interval for \( \alpha \) modulo \( A(X) \). Note that \((A(X), J)\) is a new representation for \( \alpha \) with one extra bit of precision.

**Real Root Operators.** How do we get initial isolating interval representations? We assume that if a user needs a representation of a real algebraic number \( \alpha \), then the user has access to some definition polynomial \( A(X) \), and has some means of identifying a unique real zero of \( A(X) \) of interest.

One way are to specify an integer \( i \geq 1 \) and say that we are interested in the \( i \)-th largest real zero of \( A(X) \). We may write \( \text{RootOf}(A(X), i) \) in this case. Another is to give a number (probably a rational or float) \( x \) and to say that we want to largest real root of \( A(X) \) that is larger than or equal to \( x \). We may write \( \text{RootOf}(A(X), x) \) in this case. Note that \( \text{RootOf}(A(X), i) \) \((i = i \) or \( i = x)\) may be undefined.

We call \( i \) a “root indicator”. Other kinds of indicators may be defined. For instance if \( i \leq -1 \), then we want the \((-i)\)-th smallest real root, and if \( i = 0 \), we want the smallest positive root. Instead of \( x \), we can have an interval \( I \) and we want to have the unique root of \( A(X) \) inside \( I \) (there are multiple roots, then \( \text{RootOf}(A(X), I) \) is undefined).

\(^1\)By metric, we mean the order properties of real numbers.
We now provide algorithms to construct isolating interval representations for \( \text{RootOf}(A(X),i) \). This is quite easy using Sturm sequences. But first we need some simple bounds. We use the following simple bound of Cauchy. Recall that if \( A(X) = \sum_{i=0}^{m} a_i X^i \in \mathbb{Z}[X] \) of degree \( m \geq 1 \) then the *height* of \( A(X) \) is \( \text{ht}(A) = \max\{|a_i| : i = 0, \ldots, m\} \).

**Lemma 14 (Cauchy)** Let \( A(X) \in \mathbb{Z}[X] \) and \( A(\alpha) = 0 \). If \( \alpha \neq 0 \) then

\[
\frac{1}{1 + \text{ht}(A)} < |\alpha| < 1 + \text{ht}(A).
\]

**Proof.** First we prove the upper bound on \(|\alpha|\). If \(|\alpha| < 2\), the result is immediate. Assume \(|\alpha| \geq 2\). From \( A(\alpha) = 0 \), we have \(|\alpha|^d \leq \sum_{i=0}^{d-1} |a_i| |\alpha|^i \leq \text{ht}(A)(|\alpha|^d - 1)/(|\alpha| - 1)\). Hence \(|\alpha| - 1 < \text{ht}(A)\), as desired.

For the lower bound, note that \( 1/\alpha \) is the root of the polynomial \( B(X) = X^m A(1/X) \). Hence \( |1/\alpha| < 1 + \text{ht}(B) \). Since \( \text{ht}(B) = \text{ht}(A) \), we conclude that \(|\alpha| > 1/(1 + \text{ht}(A))\). Q.E.D.

Let \( \text{Var}_A(a) \) denote the number of sign variations of the Sturm sequence of \( A(X) \) at \( X = a \), and also write \( \text{Var}_A(a,b) \) for \( \text{Var}_A(a) - \text{Var}_A(b) \). Then we can count the number of real roots of \( A(X) \) as

\[
N = \text{Var}_A[-1 - \text{ht}(A), 1 + \text{ht}(A)]).
\]

Thus \( \text{RootOf}(A(X),i) \) is defined iff \( 1 \leq i \leq N \). Further more, by a recursive subdivision of the initial interval \([−1 - \text{ht}(A), 1 + \text{ht}(A)]\), we can isolate the \( i \)th largest root. Similarly, \( \text{RootOf}(A(x),x) \) is defined iff\[
\text{Var}_A(x) > \text{Var}_A(1 + \text{ht}(A)).
\]

### §9.2. Isolating Interval Arithmetic

Suppose we want to perform the four arithmetic operations on algebraic numbers \( \alpha, \beta \). Let \( \alpha \sim (A(X), I) \) and \( \beta \sim (B(X), J) \). Say we want \( \gamma = \alpha + \beta \). So we must produce a defining polynomial \( C(X) \) for \( \gamma \) and an isolating interval \( K \) for \( \gamma \) modulo \( C(X) \).

The polynomial \( C(X) \) can be taken to be \( \text{res}_Y(A(Y), B(X - Y)) \). To see this, suppose \( \alpha_1, \ldots, \alpha_m \) are all the complex roots of \( A(Y) \) and \( \beta_1, \ldots, \beta_n \) are all the complex roots of \( B(Y) \). Then Poisson’s resultant formula [Yap, Theorem 6.15, p.158] gives

\[
\text{res}_Y(A(Y), B(Y)) = \text{res}_Y(a \prod_{i=1}^{m} (Y - \alpha_i), b \prod_{j=1}^{n} (Y - \beta_j))
\]

\[
\overset{(*)}{=} a^n \prod_{i=1}^{m} B(\alpha_i)
\]

\[
= a^n b^m \prod_{i=1}^{m} \prod_{j=1}^{n} (\alpha_i - \beta_j).
\]

Note that equation (*) here amounts to evaluating \( B(X) \) at each root of \( A(X) \). Therefore,

\[
\text{res}_Y(A(Y), B(X - Y)) = \text{res}_Y(a \prod_{i=1}^{m} (Y - \alpha_i), b \prod_{j=1}^{n} (X - \beta_j)).
\]
\[ \begin{align*}
&= (-1)^m a^m b^n \prod_{i=1}^{m} \prod_{j=1}^{n} (\alpha_i - (X - \beta_j)) \\
&= a^m b^n \prod_{i=1}^{m} \prod_{j=1}^{n} (X - (\alpha_i + \beta_j)).
\end{align*} \]

Hence this resultant is a defining polynomial for \( \gamma = \alpha + \beta \) as claimed. Next, we can let \( K = I + J \). In order for \( K \) to be an isolating interval for \( \gamma \), we can again use Sturm sequences to check that \( \gamma \) is the unique root in this interval. If now, we can subdivide the interval \( I \) and \( J \), and repeated this check for \( K \). Eventually, we will succeed.

We can generalize the above resultant construction to other arithmetic operations.

**Lemma 15** Suppose \( A(X), B(X) \) are defining polynomials for \( \alpha, \beta \) and \( \beta \neq 0 \). Let the conjugates of \( \alpha \) be \( \alpha_i \)'s and the conjugates of \( \beta \) be \( \beta_j \)'s (all non-zero). Then the defining polynomials for \( \alpha \pm \beta \), \( \alpha \beta \) and \( \alpha/\beta \) are given by the following resultants.

\[
\begin{align*}
\alpha \pm \beta & : \text{res}_Y(A(Y), B(X \mp Y)) = a^m b^n \prod_{i,j} (X - (\alpha_i \pm \beta_j)) \\
\alpha \beta & : \text{res}_Y(A(Y), Y^n B(X/Y)) = a^m b^n \prod_{i,j} (X - (\alpha_i \beta_j)) \\
\alpha/\beta & : \text{res}_Y(A(Y), X^n B(Y/X)) = a^m b^n \prod_{i,j} (X - (\alpha_i/\beta_j))
\end{align*}
\]

**Proof.**

To get the expression for \( \alpha/\beta \), we use the fact that \( \beta \) is a root of \( B(X) \) iff \( 1/\beta \) is a root of \( X^n B(1/X) \). We leave as exercise to obtain a resultant which is the defining polynomial for \( \alpha/\beta \).

**Comparing Isolating Interval Representations.** We have one more important operation: how to compare two algebraic numbers \( \alpha \sim (A, I) \) and \( \beta \sim (B, J) \). If \( I \cap J = \emptyset \) then we can decide this immediately. If \( \alpha \neq \beta \), we can refine \( I \) and \( J \) until \( I \cap J = \emptyset \). But what if \( \alpha = \beta \)? This process will not stop. The following method is described in [Yap, §7.3.2].

First suppose that we may assume that \( I = J \). To see this reduction, consider \( K = I \cap J \). If one of the intervals is empty, we can check if \( \alpha \in K \) and \( \beta \in K \). If they are not both in \( K \), then we can easily determine which is larger. If they are both in \( K \), then we replace both \( I \) and \( J \) by \( K \).

Suppose that \( I = J = [a, b] \). We can assume that \( A(X) \) and \( B(X) \) are both square-free. We can compute the “generalized Sturm sequence” for \( A \) and \( B \). Let \( \text{Var}_{A,B}(a) \) denote the number of sign variations of this sequence when evaluation at \( a \), and \( \text{Var}_{A,B}[a,b] = \text{Var}_{A,B}(a) - \text{Var}_{A,B}(b) \). Then \( \alpha \geq \beta \) if and only if

\[ \text{Var}_{A,B}[a,b] \cdot (A(b) - A(a)) \cdot (B(b) - B(a)) \geq 0. \]

See [Yap, §7.3.2]. Another way to do this comparison is to use the concept of root separation bound, as we see next.

**Other representations.** A representation of algebraic numbers based on the Thom’s lemma has been proposed in recent years. See [Yap, Chap.7]. This method has some attractiveness because of its compactness. Algorithms for
comparing any two algebraic numbers in this representation is available. How-
however, it does not appear as convenient as numerical or interval-based represen-
tations.

References
