Lecture VII
Sturm Theory

We owe to Descartes the problem of counting the number of real roots of a polynomial, and to Waring (1762) and Lagrange (1773) the problem of separating these roots. Lagrange gave the first complete algorithm for separating roots, which Burnside and Panton [2] declared “practically useless”, a testimony to some implicit efficiency criteria. The decisive technique was found by Sturm in 1829. It superseded the research of his contemporaries, Budan (1807) and Fourier (1831) who independently improved on Descartes and Lagrange. In one sense, Sturm's work culminated a line of research that began with Descartes' rule of sign. According to Burnside and Panton, the combination of Horner and Sturm gives the best root separation algorithm of their day. Hurwitz, Hermite and Routh all made major contributions to the subject. Sylvester was especially interested in Sturm's work, as part of his interest in elimination theory and theory of equations [16]. In [14], he alludes to a general theory encompassing Sturm theory. This is apparently the tome of an article [15]. Uspensky [17] rated highly a method of root separation based on a theorem of Vincent (1836). Of course, all these evaluations of computational methods are based on some implicit model of the human-hand-computer. With the advent of complexity theory we have more objective methods of evaluating algorithms.

One profound generalization of Sturm's theorem is obtained by Tarski, in his famous result showing the decidability of elementary algebra and geometry (see [7]). Hermite had interest in generalizing Sturm's theory to higher dimensions, and considered some special cases; the general case has recently been achieved in the theses of Pedersen [11] and Milne [9].

§1. Sturm Sequences from PRS

We introduce Sturm's remarkable computational tool for counting the real zeros of a real function. We also show a systematic construction of such sequences from a PRS (§III.2). Our next definition is slightly more general than the usual.

Let $A(X), B(X) \in \mathbb{R}[X]$ be non-zero polynomials. By a *(generalized) Sturm sequence* for $A(X)$, $B(X)$ we mean a PRS

$$\overline{A} = (A_0, A_1, \ldots, A_h), \quad h \geq 1,$$

for $A(X)$, $B(X)$ such that for all $i = 1, \ldots, h$, we have

$$\beta_i \, A_{i+1} = \alpha_i \, A_{i-1} + Q_i \, A_i \tag{1}$$

$(\alpha_i, \beta_i \in \mathbb{R}, Q_i \in \mathbb{R}[X])$ such that $A_{h+1} = 0$ and $\alpha_i \, \beta_i < 0$.

We call $\overline{A}$ a *Sturm sequence for* A if it is a Sturm sequence for A, A' where A' denotes the derivative of A.

Note that we do not assume $\deg A \geq \deg B$ in this definition. However, if $\deg A < \deg B$ then it is clear that $A = A_0$ and A_2 are equal up to a negative constant factor. In any case, the degrees of all subsequent polynomials are strictly decreasing, $\deg A_1 > \deg A_2 > \cdots > \deg A_h \geq 0$. Note that the relation (1) exists by the definition of PRS.

Connection between a PRS and a Sturm sequence. Essentially, a Sturm sequence differs from a PRS only by virtue of the special sign requirements on the coefficients of similarity α_i, β_i.

Although this connection is well-known, the actual form of this connection has not been clearly elucidated. Our goal here is to do this, and in a way that the transformation of a PRS algorithm into a Sturm sequence algorithm can be routine.

Assume that we are given a PRS $\overline{A} = (A_0, \ldots, A_h)$. We need not know the values α_i, β_i or Q_i in equation (1), but we do require knowledge of the product

$$s_i := -\mathtt{sign}(\alpha_i \beta_i) \tag{2}$$

of signs, for $i = 1, \ldots, h-1$. Here $\mathtt{sign}(x)$ is a real function defined as expected,

$$\mathtt{sign}(x) := \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ +1 & \text{if } x > 0 \end{cases}. \tag{3}$$

In the known PRS algorithms, these signs can be obtained as a byproduct of computing the PRS. We will now construct a sequence

$$(\sigma_0, \sigma_1, \ldots, \sigma_h),$$

of signs where $\sigma_0 = \sigma_1 = +1$ and $\sigma_i \in \{-1, 0, +1\}$ such that

$$(\sigma_0\, A_0,\, \sigma_1\, A_1, \ldots, \sigma_h\, A_h) \tag{4}$$

is a Sturm sequence. From (1) we see that

$$(\beta_i \sigma_{i+1})(\sigma_{i+1} A_{i+1}) = (\alpha_i \sigma_{i-1})(\sigma_{i-1} A_{i-1}) + Q_i A_i.$$

Hence (4) is a Sturm sequence provided that $\mathtt{sign}(\alpha_i \sigma_{i+1} \beta_i \sigma_{i-1}) = -1$ or, using equation (2),

$$\mathtt{sign}(s_i \sigma_{i+1} \sigma_{i-1}) = 1.$$

Multiplying together j $(2 \le 2j \le h)$ of these equations,

$$(\sigma_0 s_1 \sigma_2)(\sigma_2 s_3 \sigma_4)(\sigma_4 \sigma_5 \sigma_6) \cdots (\sigma_{2j-2} s_{2j-1} \sigma_{2j}) = 1.$$

Telescoping, we obtain the desired formula for σ_{2j}:

$$\sigma_{2j} = \prod_{i=1}^{j} s_{2i-1}. \tag{5}$$

Similarly, we have the formula for σ_{2j+1} $(2 \le 2j+1 \le h)$:

$$\sigma_{2j+1} = \prod_{i=1}^{j} s_{2i}. \tag{6}$$

Thus the sequence $(\sigma_1, \ldots, \sigma_h)$ of signs splits into two alternating subsequences whose computation depends on two disjoint subsets of $\{s_1, \ldots, s_{h-1}\}$. Also (5) and (6) can be rapidly computed in parallel, using the so-called parallel prefix algorithm.

Descartes' Rule of Sign. As noted in the introduction, the theory of Sturm sequences basically supersedes Descartes' Rule of Sign (or its generalizations) as a tool for root counting. The rule says:

> *The sign variation in the sequence $(a_n, a_{n-1}, \ldots, a_1, a_0)$ of coefficients of the polynomial $P(X) = \sum_{i=0}^{n} a_i X^i$ is more than the number of positive real roots of $P(X)$ by some non-negative even number.*

The proof of this and its generalization is left to an exercise.

Exercises

Exercise 1.1: Suppose a student computes a sequence $(A_0, A_1, \ldots, A_h)$ where $A_{i+1} = A_{i-1} \,\mathbf{mod}\, A_i$ for $i = 1, \ldots, h-1$ and $A_h | A_{h-1}$. This was supposed to be a Sturm sequence (a common mistake!). What is sign sequence $(\sigma_0, \ldots, \sigma_h)$ so that $(\sigma_0 A_0, \ldots, \sigma_h A_h)$ is a Sturm sequence for (A_0, A_1)? □

Exercise 1.2: Modify the subresultant algorithm (§III.5) of Collins to produce a Sturm Sequence. NOTE: in §III.5, we assume that the input polynomials P, Q satisfy $\deg P > \deg Q$. A small modification must now be made to handle the possibility that $\deg P \le \deg Q$. □

Exercise 1.3: Prove Descartes' Rule of Sign. HINT: let $Q(X)$ be a real polynomial and α a positive real number. The number of sign variations in the coefficient sequence of $(X-\alpha)Q(X)$ is more than that of the coefficient sequence of $Q(X)$ by a positive odd number. □

Exercise 1.4: (i) Give the analogue of Descartes' rule of sign for negative real roots.
(ii) Prove that if $P(X)$ has only real roots, then the number of sign variations in $P(X)$ and $P(-X)$ is exactly n.
(iii) Let $(a_n, \ldots, a_1, a_0)$ be the sequence of coefficients of $P(X)$. If $a_n a_0 \neq 0$ and $P(X)$ has only real roots, then the sequence has the property that $a_i = 0$ implies $a_{i-1}a_{i+1} < 0$. □

Exercise 1.5: Newton's rule for counting the number of imaginary roots (see quotation preceding this lecture) is modified in case a polynomial has a block of two or more consecutive terms that are missing. Newton specifies the following rule for such terms:

> *If two or more terms are simultaneously lacking, beneath the first of the deficient terms, the sign − must be placed, beneath the second, +, etc., except that beneath the last of the terms simultaneously lacking, you must always place the sign + when the terms next on either sides of the deficient ones have contrary signs.*

He gives the following examples:

$$\begin{array}{ccccccccccc}
 & & \frac{2}{5} & & \frac{1}{2} & & \frac{1}{2} & & \frac{2}{5} & & \\
X^5 & + & aX^4 & + & 0 & + & 0 & + & 0 & + & a^5 \text{ (4 imaginary roots)} \\
+ & & + & & - & & + & & - & & + \\
 & & \frac{2}{5} & & \frac{1}{2} & & \frac{1}{2} & & \frac{2}{5} & & \\
X^5 & + & aX^4 & + & 0 & + & 0 & + & 0 & - & a^5 \text{ (2 imaginary roots)} \\
+ & & + & & - & & + & & + & & +
\end{array}$$

(i) Restate Newton's rule in modern terminology.
(ii) Count the number of imaginary roots of the polynomials $X^7 - 2X^6 + 3X^5 - 2X^4 + X^3 - 3 = 0$, and $X^4 + 14X^2 - 8X + 49$. □

§2. A Generalized Sturm Theorem

Let $\overline{\alpha} = (\alpha_0, \ldots, \alpha_h)$ be a sequence of real numbers. We say there is a *sign variation* in $\overline{\alpha}$ *at position* i $(i = 1, \ldots, h)$ if for some $j = 0, \ldots, i-1$ we have

(i) $\alpha_j\,\alpha_i < 0$

(ii) $\alpha_{j+1} = \alpha_{j+2} = \cdots = \alpha_{i-1} = 0$.

The *sign variation* of $\overline{\alpha}$ is the number of positions in $\overline{\alpha}$ where there is a sign variation.

For instance, the sequence $(0, -1, 0, 3, 8, -7, 9, 0, 0, 8)$ has sign variations at positions $3, 5$ and 6. Hence its sign variation is 3.

For any sequence $\overline{A} = (A_0, \ldots, A_h)$ of polynomials and $\alpha \in \mathbb{R}$, let $\overline{A}(\alpha)$ denote the sequence $(A_0(\alpha), \ldots, A_h(\alpha))$. Then the sign variation of $\overline{A}(\alpha)$ is denoted

$$\mathtt{Var}_{\overline{A}}(\alpha),$$

where we may omit the subscript when $\overline{A}$ is understood. If $\overline{A}$ is the Sturm sequence for A, B, we may write $\mathtt{Var}_{A,B}(\alpha)$ instead of $\mathtt{Var}_{\overline{A}}(\alpha)$. If $\alpha < \beta$, we define the *sign variation difference* over the interval $[\alpha, \beta]$ to be

$$\mathtt{Var}_{\overline{A}}[\alpha, \beta] := \mathtt{Var}_{\overline{A}}(\alpha) - \mathtt{Var}_{\overline{A}}(\beta). \tag{7}$$

There are different forms of “Sturm theory”. Each form of Sturm theory amounts to giving an interpretation to the sign variation difference (7), for a suitable notion of the “Sturm sequence” $\overline{A}$. In this section, we prove a general (apparently new) theorem to encompass several known Sturm theories.

In terms of counting sign variations, Exercise 7.2.1 indicates that all Sturm sequences for A, B are equivalent. Hence, we may loosely refer to *the* Sturm sequence of A, B.

Let $r \geq 0$ be a non-negative integer. Recall that α is a root of *multiplicity* r (equivalently, α is an *r-fold root*) of an r-fold differentiable function $f(X)$ if

$$f^{(0)}(\alpha) = f^{(1)}(\alpha) = \cdots = f^{(r-1)}(\alpha) = 0, \qquad f^{(r)}(\alpha) \neq 0.$$

So we refer (awkwardly) to a non-root of f as a 0-fold root. However, if we simply say ‘α is a root of f’ then it is understood that the multiplicity r is positive. If h is sufficiently small and α is an r-fold root, then Taylor’s theorem with remainder gives us

$$f(\alpha + h) = \frac{h^r}{r!} \cdot f^{(r)}(\alpha + \theta h)$$

for some θ, $0 \leq \theta \leq 1$. So for $h > 0$, $f(\alpha + h)$ has the sign of $f^{(r)}(\alpha)$; for $h < 0$, $f(\alpha + h)$ has the sign of $(-1)^r f^{(r)}(\alpha)$. Hence:

> *If r is odd, $f(X)$ changes sign in the neighborhood of α;*
> *If r is even, $f(X)$ maintains its sign in the neighborhood of α.*

Let $\overline{A} = (A_0, \ldots, A_h)$ be a sequence of non-zero polynomials and α a real number.
i) We say α is *regular for $\overline{A}$* if each $A_i(X) \in \overline{A}$ is non-vanishing at $X = \alpha$; otherwise, α is *irregular*.
ii) We say α is *degenerate for $\overline{A}$* if each $A_i(X) \in \overline{A}$ vanishes at $X = \alpha$; otherwise α is *nondegenerate*.
iii) A closed interval $[\alpha, \beta]$ where $\alpha < \beta$ is called a *fundamental interval* (at γ_0) for $\overline{A}$ if α, β are non-roots of A_0 and there exists $\gamma_0 \in [\alpha, \beta]$ such that for all $\gamma \in [\alpha, \beta]$, if $\gamma \neq \gamma_0$ then γ is regular for $\overline{A}$. Note that γ_0 can be equal to α or β.

Hence α may be neither regular nor degenerate for $\overline{A}$, *i.e.*, it is both irregular and nondegenerate for $\overline{A}$. The following characterizes nondegeneracy.

Lemma 1 *Let $\overline{A} = (A_0, \ldots, A_h)$ be a Sturm sequence.*
a) The following are equivalent:

(i) *α is degenerate for $\overline{A}$.*
(ii) *Two consecutive polynomials in $\overline{A}$ vanish at α.*
(iii) *A_h vanishes at α.*

b) If α is nondegenerate and $A_i(\alpha) = 0$ ($i = 1, \ldots, h-1$) then $A_{i-1}(\alpha)A_{i+1}(\alpha) < 0$.

Proof.
a) If α is degenerate for $\overline{A}$ then clearly any two consecutive polynomials would vanish at α. Conversely, if $A_{i-1}(\alpha) = A_i(\alpha) = 0$, then from equation (1), we see that $A_{i+1}(\alpha) = 0$ ($i + 1 \leq h$) and $A_{i-2}(\alpha) = 0$ ($i - 2 \geq 0$). Repeating this argument, we see that every A_j vanishes at α. Thus α is degenerate for $\overline{A}$. This proves the equivalence of (i) and (ii). The equivalence of (ii) and (iii) is easy once we recall that A_h divides A_{h-1}, by definition of a PRS. Hence A_h vanishes at α implies A_{h-1} vanishes at α.
b) This follows from the fact that $\alpha_i\beta_i < 0$ in equation (1). **Q.E.D.**

The importance of fundamental intervals arises as follows. Suppose we want to evaluate $\mathtt{Var}_{A,B}[\alpha, \beta]$ where α, β are non-roots of A. Clearly, there are only a finite number of irregular values in the interval $[\alpha, \beta]$. If there are no irregular values in the interval, then trivially $\mathtt{Var}_{A,B}[\alpha, \beta] = 0$. Otherwise, we can find values

$$\alpha = \alpha_0 < \alpha_1 < \cdots < \alpha_k = \beta$$

such that each $[\alpha_{i-1}, \alpha_i]$ is a fundamental interval. Clearly

$$\mathtt{Var}_{A,B}[\alpha, \beta] = \sum_{i=1}^{k} \mathtt{Var}_{A,B}[\alpha_{i-1}, \alpha_i].$$

So we have reduced our problem to sign variation difference on fundamental intervals.

Given real polynomials $A(X), B(X)$, we say $A(X)$ *dominates* $B(X)$ if for each root α of $A(X)$, we have

$$r \geq s \geq 0$$

where α is an r-fold root of $A(X)$ and an s-fold root of $B(X)$.

Note that $r \geq 1$ here since α is a root of $A(X)$. Despite the terminology, "domination" is neither transitive nor asymmetric as a binary relation on real polynomials. We use the concept of domination in the following four situations, where in each case $A(X)$ dominates $B(X)$:

- $B(X)$ is the derivative of $A(X)$.
- $A(X)$ and $B(X)$ are relatively prime.
- $A(X)$ and $B(X)$ are both square-free.
- $B(X)$ divides $A(X)$.

We have invented the concept of domination to unify these We come to our key lemma.

Lemma 2 *Let $\overline{A} = (A_0, \ldots, A_h)$ be a Sturm sequence for A, B where A dominates B. If $[\alpha, \beta]$ is a fundamental interval at γ_0 for $\overline{A}$ then*

$$\mathtt{Var}_{\overline{A}}[\alpha,\beta] = \begin{cases} 0 & \text{if } r = 0 \text{ or } r+s \text{ is even} \\ \\ \operatorname{sign}(A^{(r)}(\gamma_0)B^{(s)}(\gamma_0)) & \text{if } r \geq 1 \text{ and } r+s \text{ is odd,} \end{cases}$$

where γ_0 is an r-fold root of $A(X)$ and also an s-fold root of $B(X)$.

Proof. We break the proof into two parts, depending on whether γ_0 is degenerate for $\overline{A}$.

Part I. Suppose γ_0 is nondegenerate for $\overline{A}$. Then $A_h(\gamma_0) \neq 0$. We may define the unique sequence

$$0 = \pi(0) < \pi(1) < \cdots < \pi(k) = h, \qquad (k \geq 1)$$

such that for all $i > 0$, $A_i(\gamma_0) \neq 0$ iff $i \in \{\pi(1), \pi(2), \ldots, \pi(k)\}$. Note that $\pi(0) = 0$ has special treatment in this definition. Define for each $j = 1, \ldots, k$, the subsequence $\overline{B}_j$ of $\overline{A}$:

$$\overline{B}_j := (A_{\pi(j-1)}, A_{\pi(j-1)+1}, \ldots, A_{\pi(j)}).$$

Since two consecutive polynomials of $\overline{A}$ cannot vanish at a nondegenerate γ_0, it follows that since $\pi(j) - \pi(j-1)$ equals 1 or 2 (*i.e.*, each $\overline{B}_j$ has 2 or 3 members). Indeed, $\overline{B}_j$ has 3 members iff its middle member vanishes at γ_0. Then the sign variation difference can be expressed as

$$\mathtt{Var}_{A,B}[\alpha,\beta] = \sum_{i=1}^{k} \mathtt{Var}_{\overline{B}_i}[\alpha,\beta]. \tag{8}$$

Let us evaluate $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta]$ in two cases:

CASE 1: $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta]$ has three members. The signs of the first and third member do not vary in the entire interval $[\alpha,\beta]$. In fact, the signs of the first and third member must be opposite. On the other hand, the signs of the middle member at α and at β are different (one of them can be the zero sign). But regardless, it is now easy to conclude $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta] = 1 - 1 = 0$.

CASE 2: $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta]$ has two members. There are two possibilities, depending on whether the first member of the sequence $\overline{B}_i$ vanishes at γ_0 or not. In fact, the first member vanishes iff $i = 1$ (so $\overline{B}_1 = (A, B)$ and $A(\gamma_0) = 0$). If $A(\gamma_0) \neq 0$, then the signs of both members in $\overline{B}_i$ do not vary in the entire interval $[\alpha,\beta]$. This proves $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta] = 0$, as required by the lemma when $A(\gamma_0) \neq 0$.

Before we consider the remaining possibility where $A(\gamma_0) = 0$, we may simplify equation (8), using the fact that all the cases we have considered until now yield $\mathtt{Var}_{\overline{B}_i}[\alpha,\beta] = 0$:

$$\mathtt{Var}_{A,B}[\alpha,\beta] = \begin{cases} \mathtt{Var}_{\overline{B}_1}[\alpha,\beta] & \text{if } A(\gamma_0) = 0, \\ \\ 0 & \text{else.} \end{cases} \tag{9}$$

Note that if $A(\gamma_0) \neq 0$ then $r = 0$. Thus equation (9) verifies our lemma for the case $r = 0$.

Hence assume $A(\gamma_0) = 0$, *i.e.*, $r \geq 1$. We have $s = 0$ because γ_0 is assumed to be nondegenerate for $\overline{A}$. Also $\alpha < \gamma_0 < \beta$ since $A(X)$ does not vanish at α or β (definition of fundamental interval). There are two subcases.

SUBCASE: r is even. Then $A(X)$ and $B(X)$ both maintain their signs in the neighborhood of γ_0 (except temporarily vanishing at γ_0). Then we see that

$$\mathtt{Var}_{\overline{B}_1}(\alpha) = \mathtt{Var}_{\overline{B}_1}(\beta),$$

proving the lemma in this subcase.

SUBCASE: r is odd. Then $A(X)$ changes sign at γ_0 while $B(X)$ maintains its sign in $[\alpha, \beta]$. Hence $\mathtt{Var}_{\overline{B}_1}[\alpha,\beta] = \pm 1$. In fact, the following holds:

$$\mathtt{Var}_{\overline{B}_1}[\alpha,\beta] = \mathtt{sign}(A^{(r)}(\gamma_0)B^{(s)}(\gamma_0)), \tag{10}$$

proving the lemma when $s = 0$ and $r \ge 1$ is odd. [Let us verify equation (10) in case $B(X) > 0$ throughout the interval. There are two possibilities: if $A^{(r)}(\gamma_0) < 0$ then we get $\mathtt{Var}_{\overline{B}_1}(\alpha) = 0$ and $\mathtt{Var}_{\overline{B}_1}(\beta) = 1$ so that $\mathtt{Var}_{\overline{B}_1}[\alpha,\beta] = \mathtt{sign}(A^{(r)}(\gamma_0))$. If $A^{(r)}(\gamma_0) > 0$ then $\mathtt{Var}_{\overline{B}_1}(\alpha) = 1$ and $\mathtt{Var}_{\overline{B}_1}(\beta) = 0$, and again $\mathtt{Var}_{\overline{B}_1}[\alpha,\beta] = \mathtt{sign}(A^{(r)}(\gamma_0))$.]

Part II. Now assume γ_0 is degenerate. This means $\alpha < \gamma_0 < \beta$. Let

$$\overline{C} = (A_0/A_h, A_1/A_h, \ldots, A_h/A_h)$$

be the *depressed sequence* derived from $\overline{A}$. This is a Sturm sequence for $C_0 = A_0/A_h$, $C_1 = A_1/A_h$. Moreover, γ_0 is no longer degenerate for $\overline{C}$, and we have

$$\mathtt{Var}_{\overline{A}}(\gamma) = \mathtt{Var}_{\overline{C}}(\gamma),$$

for all $\gamma \in [\alpha,\beta]$, $\gamma \ne \gamma_0$. Since $[\alpha,\beta]$ remains a fundamental interval at γ_0 for $\overline{C}$, the result of part I in this proof can now be applied to $\overline{C}$, showing

$$\mathtt{Var}_{\overline{C}}[\alpha,\beta] = \begin{cases} 0 & \text{if } r^* = 0 \text{ or } r^* + s^* \text{ is even,} \\ \\ \mathtt{sign}(C_0^{(r^*)}(\gamma_0)C_1^{(s^*)}(\gamma_0)) & \text{if } r^* \ge 1 \text{ and } r^* + s^* \text{ is odd.} \end{cases} \tag{11}$$

Here r^*, s^* are the multiplicities of γ_0 as roots of C_0, C_1 (respectively). Clearly, if γ_0 is an m-fold root of $A_h(X)$, then $r = r^* + m, s = s^* + m$. Hence $r^* + s^* =$ even iff $r + s =$ even. This shows

$$\mathtt{Var}_{\overline{A}}[\alpha,\beta] = \mathtt{Var}_{\overline{C}}[\alpha,\beta] = 0$$

when $r + s =$ even, as desired. If $r^* + s^* =$ odd and $r^* \ge 1$, we must show

$$\mathtt{sign}(C_0^{(r^*)}(\gamma_0)C_1^{(s^*)}(\gamma_0)) = \mathtt{sign}(A^{(r)}(\gamma_0)B^{(s)}(\gamma_0)). \tag{12}$$

For clarity, let $A_h(X)$ be rewritten as $D(X)$ so that

$$\begin{aligned} A(X) &= C_0(X) \cdot D(X) \\ A^{(r)}(X) &= \sum_{i=0}^{r} \binom{r}{i} C_0^{(i)}(X) D^{(r-i)}(X) \\ A^{(r)}(\gamma_0) &= \binom{r}{r^*} C_0^{(r^*)}(\gamma_0) D^{(m)}(\gamma_0) \end{aligned}$$

since $C_0^{(i)}(\gamma_0) = 0$ for $i < r^*$, and $D^{(r-i)}(\gamma_0) = 0$ for $i > r^*$. Similarly,

$$B^{(s)}(\gamma_0) = \binom{s}{s^*} C_1^{(s^*)}(\gamma_0) D^{(m)}(\gamma_0).$$

This proves (12).

Finally suppose $r^* = 0$. But the assumption that A dominates B implies $s^* = 0$. [This is the only place where domination is used.] Hence $s^* + r^*$ is even and $\mathtt{Var}_{\overline{C}}[\alpha,\beta] = 0$. Hence $s + r$ is also even and $\mathtt{Var}_{\overline{A}}[\alpha,\beta] = 0$. This completes the proof. **Q.E.D.**

This lemma immediately yields the following:

Theorem 3 (Generalized Sturm) *Let A dominate B and let $\alpha < \beta$ so that $A(\alpha)A(\beta) \neq 0$. Then*

$$\mathtt{Var}_{A,B}[\alpha,\beta] = \sum_{\gamma,r,s} \mathtt{sign}(A^{(r)}(\gamma)B^{(s)}(\gamma)) \tag{13}$$

where γ ranges over all roots of A in $[\alpha,\beta]$ of multiplicity $r \geq 1$, and B has multiplicity s at γ, and $r+s =$ odd.

The statement of this theorem can be generalized in two ways without modifying the proof:
(a) We only need to assume that A dominates B within the interval $[\alpha,\beta]$, *i.e.*, at the roots of A in the interval, the multiplicity of A is at least that of the multiplicity of B.
(b) The concept of domination can be extended to mean that at each root γ of A (restricted to $[\alpha,\beta]$ as in (a) if we wish), if A, B have multiplicities r, s (respectively) at γ, then $\max\{0, s-r\}$ is even.

Exercises

Exercise 2.1: Suppose $\overline{A}$ and $\overline{B}$ are both Sturm sequences for $A, B \in \mathbb{R}[X]$. Then they have the same length and corresponding elements of $\overline{A}$ and $\overline{B}$ are related by positive factors: $A_i = \alpha_i B_i$ where α_i is a positive real number. □

Exercise 2.2: The text preceding Lemma 7.2 specified four situations were $A(X)$ dominates $B(X)$. Verify domination in each case. □

Exercise 2.3: (Budan-Fourier) Let $A_0(X)$ be a polynomial, $\alpha < \beta$ and $A_0(\alpha)A_0(\beta) \neq 0$. Let $\overline{A} = (A_0, A_1, \ldots, A_h)$ be the sequence of non-zero derivatives of A_0, *viz.*, A_i is the ith derivative of A_0. Then the number of real zeros of $A_0(X)$ in $[\alpha,\beta]$ is less than the $\mathtt{Var}_{\overline{A}}[\alpha,\beta]$ by an even number. HINT: Relate the location of zeros of $A(X)$ and its derivative $A'(X)$. Use induction on $\deg A_0$. □

Exercise 2.4: a) Deduce Descartes' Rule of Sign (§1) from the Budan-Fourier Rule (see previous exercise).
b) (Barbeau) Show that Descartes' Rule gives a sharper estimate for the number of negative zeros than Budan-Fourier for the polynomial $X^4 + X^2 + 4X - 3$. □

§3. Corollaries and Applications

We obtain four useful corollaries to the generalized Sturm theorem. The first is the classic theorem of Sturm.

Corollary 4 (Sturm) *Let $A(X) \in \mathbb{R}[X]$ and suppose $\alpha < \beta$ are both non-roots of A. Then the number of distinct real roots of $A(X)$ in the interval $[\alpha,\beta]$ is given by $\mathtt{Var}_{A,A'}[\alpha,\beta]$.*

Proof. With $B(X) = A'(X)$, we see that $A(X)$ dominates $B(X)$ so that the generalized Sturm theorem gives:

$$\mathtt{Var}_{A,B}[\alpha,\beta] = \sum_{\gamma,r,s} \mathtt{sign}(A^{(r)}(\gamma)B^{(s)}(\gamma)),$$

where γ is an r-fold root of A in (α, β), γ is an s-fold root of B and $r \geq 1$ with $r+s$ being odd. But at every root of A, these conditions are satisfied since $r = s+1$. Hence the summation applies to every root γ of A. Furthermore, we see that $A^{(r)}(\gamma) = B^{(s)}(\gamma)$ so that $\mathtt{sign}\,(A^{(r)}(\gamma)B^{(s)}(\gamma)) = 1$. So the summation yields the number of roots of A in $[\alpha, \beta]$. **Q.E.D.**

Note that it is computationally convenient that our version of Sturm's theorem does not assume $A(X)$ is square-free (which is often imposed).

Corollary 5 (Schwartz-Scharir) *Let $A(X), B(X) \in \mathbb{R}[X]$ be square-free polynomials. If $\alpha < \beta$ are both non-roots of A then*

$$\mathtt{Var}_{A,B}[\alpha, \beta] = \sum_{\gamma} \mathtt{sign}(A'(\gamma)B(\gamma))$$

where γ ranges over all roots of $A(X)$ in $[\alpha, \beta]$.

Proof. We may apply the generalized Sturm theorem to evaluate $\mathtt{Var}_{A,B}[\alpha, \beta]$ in this corollary. In the sum of (13), consider the term indexed by the triple (γ, r, s) with $r \geq 1$ and $r+s$ is odd. By square-freeness of A and B, we have $r \leq 1$ and $s \leq 1$. Thus $r = 1$, $s = 0$ and equation (13) reduces to

$$\mathtt{Var}_{A,B}[\alpha, \beta] = \sum_{\gamma} \mathtt{sign}(A'(\gamma)B(\gamma)),$$

where the summation is over roots γ of A in $[\alpha, \beta]$ which are not roots of B. But if γ is both a root of A and of B then $\mathtt{sign}(A'(\gamma)B(\gamma)) = 0$ and we may add these terms to the summation without any effect. This is the summation sought by the corollary. **Q.E.D.**

The next corollary will be useful in §7:

Corollary 6 (Sylvester, revisited by Ben-Or, Kozen, Reif) *Let $\overline{A}$ be a Sturm sequence for $A, A'B$ where $A(X)$ is square-free and $A(X), B(X)$ are relatively prime. Then for all $\alpha < \beta$ which are non-roots of A,*

$$\mathtt{Var}_{\overline{A}}[\alpha, \beta] = \sum_{\gamma} \mathtt{sign}(B(\gamma))$$

where γ ranges over the roots of $A(X)$ in $[\alpha, \beta]$.

Proof. Again note that A dominates $A'B$ and we can proceed as in the proof of the previous corollary. But now, we get

$$\begin{aligned}\mathtt{Var}_{\overline{A}}[\alpha, \beta] &= \sum_{\gamma} \mathtt{sign}(A'(\gamma) \cdot A'(\gamma)B(\gamma)) \\ &= \sum_{\gamma} \mathtt{sign}(B(\gamma)),\end{aligned}$$

as desired. **Q.E.D.**

In this corollary, the degree of $A_0 = A$ is generally less than the degree of $A_1 = A'B$ so that the remainder sequence typically looks like this: $\overline{A} = (A, A'B, -A, \ldots)$.

Our final corollary concerns the concept of the Cauchy index of a rational function. Let $f(X)$ be a real continuous function defined in an open interval (α, β) where $-\infty \le \alpha < \beta \le +\infty$. We allow $f(X)$ to have isolated poles in the interval (α, β). Recall that $\gamma \in (\alpha, \beta)$ is a *pole* of $f(X)$ if $1/f(X) \to 0$ as $X \to \gamma$. The *Cauchy index* of f at a pole γ is defined[1] to be

$$\frac{\mathtt{sign}(f(\gamma^-)) - \mathtt{sign}(f(\gamma^+))}{2}.$$

For instance, the index is -1 if $f(X)$ changes from $-\infty$ to $+\infty$ as X increases through γ, and the index is 0 if the sign of $f(X)$ does not change in passing through γ. The *Cauchy index* of f over an interval (α, β) is then

$$I_\alpha^\beta f(X) := \sum_\gamma \frac{\mathtt{sign}(f(\gamma^-)) - \mathtt{sign}(f(\gamma^+))}{2}$$

where the sum is taken over all poles $\gamma \in (\alpha, \beta)$. Typically, $f(X)$ is a rational function $A(X)/B(X)$ where $A(X), B(X)$ are relatively prime polynomials.

Corollary 7 (Cauchy Index) *Let $A(X), B(X) \in \mathbb{R}[X]$ be relatively prime and $f(X) = A(X)/B(X)$. Then*

$$I_\alpha^\beta f(X) = -\mathtt{Var}_{A,B}[\alpha, \beta].$$

Proof. Let (γ, r, s) index a summation term in (13). We have $s = 0$ since A, B are relatively prime. This means that r is odd, and

$$\begin{aligned} \mathtt{sign}(A^{(r)}(\gamma)) &= \frac{\mathtt{sign}(A(\gamma^+)) - \mathtt{sign}(A(\gamma^-))}{2}, \\ \mathtt{sign}(A^{(r)}(\gamma)B^{(0)}(\gamma)) &= \frac{\mathtt{sign}(A(\gamma^+)B(\gamma^+)) - \mathtt{sign}(A(\gamma^-)B(\gamma^-))}{2} \\ &= \frac{\mathtt{sign}(f(\gamma^+)) - \mathtt{sign}(f(\gamma^-))}{2}. \end{aligned}$$

Summing the last equation over each (γ, r, s), the left-hand side equals $\mathtt{Var}_{A,B}[\alpha, \beta]$, by the generalized Sturm theorem. But the right-hand side equals $I_\alpha^\beta f$. **Q.E.D.**

This result is used in §5. For now, we give two applications of the corollary of Schwartz-Scharir (cf. [13]).

A. The sign of a real algebraic number. The first problem is to determine the sign of a number β in a real number field $\mathbb{Q}(\alpha)$. We assume that β is represented by a rational polynomial $B(X) \in \mathbb{Q}[X]$: $\beta = B(\alpha)$. Assume α is represented by the isolating interval representation (§VI.9)

$$\alpha \cong (A, [a, b])$$

where $A \in \mathbb{Z}[X]$ is a square-free polynomial. First let us assume $B(X)$ is square-free. To determine the sign of β, first observe that

$$\mathtt{sign}(A'(\alpha)) = \mathtt{sign}(A(b) - A(a)). \tag{14}$$

[1] Here, $\mathtt{sign}(f(\gamma^-))$ denotes the sign of $f(X)$ for when $\gamma - X$ is positive but arbitrarily small. When $f(X)$ is a rational function, this sign is well-defined. Similarly $\mathtt{sign}(f(\gamma^+))$ is the sign of $f(X)$ when $X - \gamma$ is positive but arbitrarily small.

Using the corollary of Schwartz-Sharir,

$$\texttt{Var}_{A,B}[a,b] = \texttt{sign}(A'(\alpha)\cdot B(\alpha)).$$

Hence,

$$\begin{aligned}\texttt{sign}(B(\alpha)) &= \texttt{sign}((\texttt{Var}_{A,B}[a,b])\cdot A'(\alpha))\\ &= \texttt{sign}((\texttt{Var}_{A,B}[a,b])\cdot (A(b)-A(a))).\end{aligned}$$

If $B(X)$ is not square-free, we can first decompose it into a product of square-free polynomials. That is, B has a *square-free decomposition* $B_1 \cdot B_2 \cdot \ldots \cdot B_k$ where B_1 is the square-free part of B and $B_2 \cdot \ldots \cdot B_k$ is recursively the square-free decomposition of B/B_1. Then $\texttt{sign}(B(\alpha)) = \prod_{i=1}^k \texttt{sign}(B_i(\alpha))$.

Exercise 3.1: Alternatively, use the Sylvester corollary to obtain the sign of $B(\alpha)$. □

B. Comparing two real algebraic numbers. Given two real algebraic numbers

$$\alpha \cong (A, I), \qquad \beta \cong (B, J)$$

represented as indicated by isolating intervals, we wish to compare them. Of course, one method is to determine the sign of $\alpha - \beta$, by a suitable reduction to the problem in Section 7.3.1. But we give a more direct reduction. If $I \cap J = \emptyset$ then the comparison is trivially done. Otherwise, if either $\alpha \notin I \cap J$ or $\beta \notin I \cap J$ then again we can easily determine which of α or β is bigger. Hence assume α and β are both in a common isolating interval $I \cap J = [a,b]$.

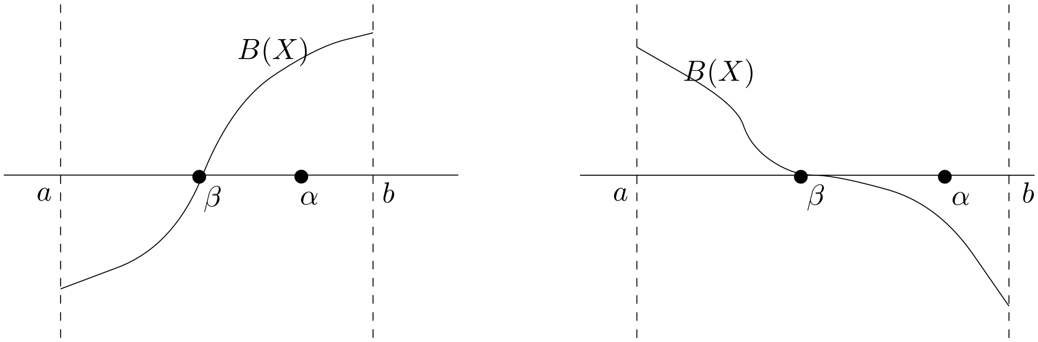

Figure 1: Two cases for $\alpha > \beta$ in isolating interval $[a,b]$.

It is not hard to verify (see Figure 1) that

$$\alpha \ge \beta \quad \Leftrightarrow \quad B(\alpha)\cdot B'(\beta) \ge 0,$$

with equality on the left-hand side if and only if equality is attained on the right-hand side (note that $B'(\beta) \neq 0$ by square-freeness of B). Since we already know how to obtain the signs of $B(\alpha)$ and of $B'(\beta)$ (Section 7.3.1) we are done:

$$B(\alpha)\cdot B'(\beta) \ge 0 \quad \Leftrightarrow \quad (\texttt{Var}_{A,B}[a,b])\cdot(A(b)-A(a))\cdot(B(b)-B(a)) \ge 0.$$

Complexity of one incremental-bit of an algebraic number. Let α be an algebraic number, given as the ith real root of a square-free polynomial $A(X) \in \mathbb{Z}[X]$. Consider the following question: what is the complexity of finding out one *incremental-bit* of α? More precisely, suppose we already know that α lies within an interval I. How much work does it take to halve the interval? There are three stages. *Sturm stage:* Initially, I can be taken to be $[-M, M]$ where $M = 1 + \|A\|_\infty$ is Cauchy's bound. We can halve I by counting the number of real roots of A in the interval $[-M, 0]$ and $[0, M]$. This takes two "Sturm queries" as given by corollary 4. Subsequently, assuming we already know the number of real roots inside I, each incremental-bit of α costs only one Sturm query. This continues until I is an isolating interval. *Bisection stage:* Now we may assume that we know the sign of $A(X)$ at the end-points of I. Henceforth, each incremental-bit costs only one polynomial evaluation, *viz.*, evaluating the sign of $A(X)$ at the mid-point of I. We continue this until the size Δ of I is within the range of guaranteed Newton convergence. *Newton stage:* According to §VI.11, it suffices to have $\Delta \le m^{-3m-9}M^{-6m}$ where $m = \deg A$ and $M = 2 + \|A\|_\infty$. Let X_0 be the midpoint of I when Δ first reaches this bound. If Newton iteration transforms X_i to X_{i+1}, then the point X_i is within distance 2^{-2^i} of α (§VI.10). The corresponding interval I_i may be taken to have size $2^{1-2^i}\Delta$, centered at X_i. That is, we obtain about 2^i incremental-bits for i Newton steps. Each Newton step is essentially two polynomial evaluations. In an amortized sense, the cost is about 2^{-i+1} polynomial evaluations per incremental-bit for the ith Newton iteration.

Exercises

Exercise 3.2: Isolate the roots of:
(a) $(X^2+7)^2 - 8X = X^4 + 14X^2 - 8X + 49$.
(b) $X^{16} - 8X^{14} + 8X^{12} + 64X^{10} - 98X^8 - 184X^6 + 200X^4 + 224X^2 - 113$.

These are the minimal polynomials of $\sqrt{2}+\sqrt{5}$ and $\sqrt{1+\sqrt{5-3\sqrt{1+\sqrt{2}}}}$, respectively. □

Exercise 3.3: Isolate the roots of the following polynomials:

$$\begin{aligned} P_2(X) &= \frac{3}{2}X^2 - \frac{1}{2}, \\ P_3(X) &= \frac{5}{2}X^3 - \frac{3}{2}X, \\ P_4(X) &= \frac{35}{8}X^4 - \frac{15}{4}X^2 + \frac{3}{8}. \end{aligned}$$

These are the Legendre polynomials, which have all real and distinct roots lying in the interval $[-1, 1]$. □

Exercise 3.4: Give an algorithm for the square-free decomposition of a polynomial $B(X) \in \mathbb{Z}[X]$: $B(X) = B_1B_2\cdots B_k$ as described in the text. Analyze the complexity of your algorithm. □

.

Exercise 3.5: What does $\mathtt{Var}_{A,A''}[\alpha, \beta]$ count, assuming $\alpha < \beta$ and $A(\alpha)A(\beta) \ne 0$? □

.

Exercise 3.6: (a) Let $Q(Y) \in \mathbb{Q}(\alpha)[Y]$, where α is a real root of $P(X) \in \mathbb{Q}[X]$. Assume that we have an isolating interval representation for α (relative to $P(X)$) and the coefficients of $Q(Y)$ are represented by rational polynomials in α. Show how to carry out a Sturm sequence computation to isolate the real roots of $Q(Y)$. Analyze the complexity of your algorithm.
(b) This gives us a method of representing elements of the double extension $\mathbb{Q}(\alpha)(\beta)$. Extend the method to multiple (real) extensions: $\mathbb{Q}(\alpha_1)\cdots(\alpha_k)$. Explain how arithmetic in such representations might be carried out. □

Exercise 3.7: (Schwartz-Sharir) Given an integer polynomial $P(X)$ (not necessarily square-free) and an isolating interval I of $P(X)$ for one of its real roots α, determine the multiplicity of $P(X)$ at α. □

Exercise 3.8: In order for all the roots of $P(X)$ to be real, it is necessary that the leading coefficients of a Sturm sequence of $P(X)$ be all positive. □

Exercise 3.9: Give a version of the generalized Sturm's theorem where we replace the condition that α, β are non-roots of A by the condition that these are nondegenerate. □

Exercise 3.10: Let $\alpha_1, \ldots, \alpha_k$ be real algebraic numbers with isolating interval representations. Preprocess this set of numbers so that, for any subsequently given integers $n_1, \ldots, n_k \in \mathbb{Z}$, you can efficiently test if $\sum_{i=1}^k n_i\alpha_i$ is zero. □

Exercise 3.11: (Sederberg and Chang)
(a) Let $P(X), B(X)$ and $C(X)$ be non-zero real polynomials and define

$$A(X) := B(X)P'(X) + C(X)P(X).$$

Then between any two adjacent real roots of $P(X)$ there is at least one real root of $A(X)$ or $B(X)$. (This statement can be interpreted in the natural way in case the two adjacent roots coincide.) In general, any pair $A(X), B(X)$ of polynomials with this property is called an *isolator pair* for $P(X)$.
(b) Let $P(X) = X^3 + aX^2 + bX + c$. Construct two linear polynomials $A(X)$ and $B(X)$ which form an isolator pair for $P(X)$. What are the roots $A(X)$ and $B(X)$? HINT: choose $B(X) = \frac{1}{3}(X + \frac{a}{3})$ and $C(X) = -1$.
(c) Relate the concept of isolator pairs to the polynomial remainder sequence of $P(X)$. □

Exercise 3.12*: Is there a simple method to decide if an integer polynomial has only real roots? □

§4. Integer and Complex Roots

We discuss the special cases of integer and rational roots, and the more general case of complex roots.

Integer and Rational Roots. Let $A(X) = \sum_{i=0}^{n} a_i X^i$ be an integer polynomial of degree n. We observe that if u is an integer root of $A(X)$ then

$$a_0 = -\sum_{i=1}^{n} a_i u^i = -u \left(\sum_{i=1}^{n} a_i u^{i-1} \right)$$

and hence u divides a_0. Hence, checking if $A(X)$ has any integer roots it can be reduced to factorization of integers: we factor a_0 and for each integer factor u, we check if $A(u) = 0$. Similarly, if u/v is a rational root of $A(X)$ with $\mathtt{GCD}(u, v) = 1$ it is easily checked that u divides a_0 and v divides a_n. [Thus, if u/v is a rational root of a monic integer polynomial then $v = 1$, *i.e.*, the set of algebraic integers that are rational is precisely $\mathbb{Z}$.] We can thus reduce the search for rational roots to the factorization of a_0 and a_n.

Hilbert's 10th problem asks for an algorithm to decide if an input integer polynomial has any integer roots. Matiyasevich (1970), building on the work of Davis, Putnam and Robinson [5], proved that no such algorithm exists, by showing that this is (many-one) equivalent to the Halting Problem. For an exposition of this result, see the book of Davis [4, Appendix 2] or [8]. It is an open problem whether there is an algorithm to decide if an input integer polynomial has any rational roots. This can be shown to be equivalent to restricting the inputs to Hilbert's 10th problem to homogeneous polynomials.

Complex Roots. We reduce the extraction of complex roots to the real case. The real and complex component of a complex algebraic number may be separately represented using isolating intervals. Suppose $P(X) \in \mathbb{C}[X]$ and $\overline{P}(X)$ is obtained by complex conjugation of each coefficient of $P(X)$. Then for $\alpha \in \mathbb{C}$, $\overline{P(\alpha)} = \overline{P}(\overline{\alpha})$. So $P(\alpha) = 0$ iff $\overline{P}(\overline{\alpha}) = 0$. It follows that if $P(X) = \prod_{i=1}^{n}(X - \alpha_i)$ then

$$P(X) \cdot \overline{P}(X) = (\prod_{i=1}^{n} X - \alpha_i)(\prod_{i=1}^{n} X - \overline{\alpha_i}).$$

Hence $P(X) \cdot \overline{P}(X)$ is a real polynomial, as $(X - \alpha_i)(X - \overline{\alpha_i}) \in \mathbb{R}[X]$. This shows that even when we are interested in complex roots, we may only work with real polynomials. But it may be more efficient to allow polynomials with complex coefficients (cf. next section). In practice, we assume that $P(X)$ has Gaussian integers $\mathbb{Z}[\mathbf{i}]$ as coefficients.

If $F(X) \in \mathbb{C}[X]$ and $\alpha + \mathbf{i}\beta \in \mathbb{C}$ $(\alpha, \beta \in \mathbb{R})$ is a root of $F(X)$ then we may write

$$F(\alpha + \mathbf{i}\beta) = P(\alpha, \beta) + \mathbf{i}Q(\alpha, \beta)$$

where $P(X, Y), Q(X, Y)$ are bivariate real polynomials determined by F. This reduces the problem of finding α, β to solving the simultaneous system

$$\begin{array}{rcl} P(\alpha, \beta) & = & 0, \\ Q(\alpha, \beta) & = & 0. \end{array}$$

We solve for α using resultants:

$$R(X) := \mathtt{res}_Y(P(X, Y), Q(X, Y)).$$

For each real root α of $R(X)$, we can plug α into $P(\alpha, Y)$ to solve for $Y = \beta$. (We have not explicitly described how to handle polynomials with algebraic coefficients but in principle we know how to perform arithmetic operations for algebraic numbers.) Alternatively, we can find β among the real roots of $\mathtt{res}_X(P, Q)$ and check for each pair α, β that may serve as a root $\alpha + \mathbf{i}\beta$ of $F(X)$. This will be taken up again in the next section.

It is instructive to examine the above polynomials P, Q in greater detail. To this end, let us write $F(X)$ as

$$F(X) = A(X) + \mathbf{i}B(X), \quad A(X), B(X) \in \mathbb{R}[X].$$

Then by Taylor's expansion,

$$A(\alpha + \mathbf{i}\beta) = A(\alpha) + \frac{A'(\alpha)}{1!} \cdot (\mathbf{i}\beta) + \frac{A''(\alpha)}{2!} \cdot (\mathbf{i}\beta)^2 + \cdots + \frac{A^{(n)}(\alpha)}{n!}(\mathbf{i}\beta)^n$$

where $n = \max\{\deg A, \deg B\}$. Similarly,

$$B(\alpha + \mathbf{i}\beta) = B(\alpha) + \frac{B'(\alpha)}{1!}(\mathbf{i}\beta) + \cdots + \frac{B^{(n)}}{n!}(\mathbf{i}\beta)^n.$$

Hence the real and imaginary parts of $F(\alpha + \mathbf{i}\beta)$ are, respectively,

$$\begin{aligned} P(\alpha, \beta) &= A(\alpha) + \frac{B'(\alpha)}{1!}(-\beta) + \frac{A^{(2)}(\alpha)}{2!}(-\beta^2) + \cdots, \\ Q(\alpha, \beta) &= B(\alpha) + \frac{A'(\alpha)}{1!}(\beta) + \frac{B^{(2)}(\alpha)}{2!}(-\beta^2) + \cdots. \end{aligned}$$

So $P(\alpha, \beta)$ and $Q(\alpha, \beta)$ are polynomials of degree $\le n$ in β with coefficients that are polynomials in α of degree $\le n$. Hence $R(\alpha)$ is a polynomial of degree n^2 in α. Moreover, the bit-size of $R(X)$ remains polynomially bounded in the bit-size of $A(X)$, $B(X)$. Hence, any polynomial-time solution to real root isolation would lead to a polynomial-time solution to complex root isolation.

Remarks: See Householder [6] for more details on this approach.

Exercises

Exercise 4.1: Work out the algorithmic details of the two methods for finding complex roots as outlined above. Determine their complexity. □

Exercise 4.2: Express $P(\alpha, \beta)$ and $Q(\alpha, \beta)$ directly in terms of $F^{(i)}(\alpha)$ and β^i by a different Taylor expansion, $F(\alpha + \mathbf{i}\beta) = F(\alpha) + F'(\alpha)(\mathbf{i}\beta) + \cdots$. □

Exercise 4.3: A **Diophantine polynomial** is a polynomial $D(X_1, \ldots, X_n)$ with (rational) integer coefficients and whether the X_i's are integer variables. Hilbert's 10th Problem asks whether a given Diophantine polynomial $D(X_1, \ldots, X_n)$ is solvable. Show that the decidability of Hilbert's 10th Problem is equivalent to the decidability of each of the following problems:
(i) The problem of deciding if a system of Diophantine equations is solvable.
(ii) The problem of deciding if a Diophantine equation of total degree 4 is solvable. **Remark:** It is an unknown problem whether '4' here can be replaced by '3'. HINT: First convert the single Diophantine polynomial to an equivalent system of polynomials of total degree at most 2.
(iii) The problem of deciding if a Diophantine equation of degree 4 has solution in non-negative integers. HINT: In one direction, use the fact that every non-negative integer is the sum of four squares of integers. □

Exercise 4.4: A **Diophantine set of dimension** n is one of the form

$$\{(a_1, \ldots, a_n) \in \mathbb{Z}_n : (\exists b_1, \ldots, b_m \in \mathbb{Z}) D(a_1, \ldots, a_n, b_1, \ldots, b_m) = 0\}$$

where $D(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ is a Diophantine polynomial. A Diophantine set $S \subseteq \mathbb{Z}^n$ can be viewed as **Diophantine relation** $R(X_1, \ldots, X_n)$ where $R(a_1, \ldots, a_n)$ holds iff $(a_1, \ldots, a_n) \in S$.
(i) Show that the following relations are Diophantine: $X_1 \neq X_2$, $X_1 = (X_2 \,\mathbf{mod}\, X_3)$, $X_1 = \mathtt{GCD}(X_2, X_3)$
(ii) A set $S \subseteq \mathbb{Z}$ is Diophantine iff

$$S = \{D(a_1, \ldots, a_m) : (\exists a_1, \ldots, a_n \in \mathbb{Z}\}$$

for some Diophantine polynomial $D(Y_1, \ldots, Y_m)$.
(iii) Show that Diophantine sets are closed under union and intersection.
(iv) (M.Davis) Diophantine sets are not closed under complement. The complementation is with respect to $\mathbb{Z}^n$ if the dimension is n.
(v) (Y.Matijasevich) The exponentiation relation $X = Y^Z$, where X, Y, Z are restricted to natural numbers, is Diophantine. This is a critical step in the solution of Hilbert's 10th Problem. □

§5. The Routh-Hurwitz Theorem

We now present an alternative method for isolating complex zeros using Sturm's theory. First we consider a special subproblem: to count the number of complex roots in the upper complex plane. This problem has independent interest in the theory of stability of dynamical systems, and was first solved by Routh in 1877, using Sturm sequences. Independently, Hurwitz in 1895 gave a solution based on the theory of residues and quadratic forms. Pinkert [12] exploited this theory to give an algorithm for isolating complex roots. Here, we present a variant of Pinkert's solution.

In this section we consider complex polynomials as well as real polynomials.

We begin with an elementary result, a variant of the so-called *principle!of argument.* Let $F(Z) \in \mathbb{C}[Z]$ and L be an oriented line in the complex plane. Consider the *increase* in the argument of $F(Z)$ as Z moves along the entire length of L, denoted

$$\Delta_L \arg F(Z).$$

Note that if $F = G \cdot H$ then

$$\Delta_L \arg F = (\Delta_L \arg G) + (\Delta_L \arg H). \tag{15}$$

Lemma 8 *Suppose no root of $F(Z)$ lies on L, $p \geq 0$ of the complex roots of $F(Z)$ lie to the left-hand side of L, and $q \geq 0$ of the roots lie to the right-hand side, multiplicity counted. Then $\Delta_L \arg F(Z) = \pi(p - q)$.*

Proof. Without loss of generality, let $F(Z) = \prod_{i=1}^{p+q}(Z - \alpha_i)$, $\alpha_i \in \mathbb{C}$. Then $\arg F(Z) = \sum_{i=1}^{p+q} \arg(Z - \alpha_i)$. Suppose α_i lies to the left of L. Then as Z moves along the entire length of L, $\arg(Z - \alpha_i)$ increases by π *i.e.*, $\Delta_L \arg(Z - \alpha_i) = \pi$. Similarly, if α_i lies to the right of L, $\Delta_L \arg(Z - \alpha_i) = -\pi$. The lemma follows by summing over each root. **Q.E.D.**

Since $p + q = \deg F(Z)$, we conclude:

Corollary 9

$$\begin{aligned} p &= \frac{1}{2}\left[\deg F + \frac{1}{\pi}\Delta_L \arg F(Z)\right], \\ q &= \frac{1}{2}\left[\deg F - \frac{1}{\pi}\Delta_L \arg F(Z)\right]. \end{aligned}$$

Number of roots in the upper half-plane. Our immediate goal is to count the number of roots above the real axis. Hence we now let L be the real axis. By the foregoing, the problem amounts to deriving a suitable expression for $\Delta_L \arg F(Z)$. Since Z is going to vary over the reals, we prefer to use 'X' to denote a real variable. Let

$$F(X) = F_0(X) + \mathbf{i}F_1(X)$$

where $F_0(X), F_1(X) \in \mathbb{R}[X]$. Observe that α is a real root of $F(X)$ iff α is a real root of $G = \mathtt{GCD}(F_0, F_1)$. Before proceeding, we make three simplifications:

- We may assume $F_0(X)F_1(X) \neq 0$. If $F_1 = 0$ then the complex roots of $F(X)$ come in conjugate pairs and their number can be determined from the number of real roots. Similarly if $F_0 = 0$ then the same argument holds if we replace F by $\mathbf{i}F$.
- We may assume F_0, F_1 are relatively prime, since we can factor out any common factor $G = \mathtt{GCD}(F_0, F_1)$ from F, and and apply equation (15) to F/G and G separately.
- We may assume $\deg F_0 \geq \deg F_1$. Otherwise, we may replace F by $\mathbf{i}F$ which has the same set of roots. This amounts to replacing (F_0, F_1) by $(-F_1, F_0)$ throughout the following.

We define

$$\rho(X) := \frac{F_0(X)}{F_1(X)}.$$

Thus $\rho(X)$ is well-defined for all X (we never encounter $0/0$). Clearly $\arg F(X) = \cot^{-1}\rho(X)$. Let

$$\alpha_1 < \alpha_2 < \cdots < \alpha_k$$

be the real roots of $F_0(X)$. They divide the real axis L into $k+1$ segments,

$$L = L_0 \cup L_1 \cup \cdots \cup L_k, \qquad (L_i = [\alpha_i, \alpha_{i+1}])$$

where $\alpha_0 = -\infty$ and $\alpha_{k+1} = +\infty$. Thus,

$$\Delta_L \arg F(X) = \sum_{i=0}^{k} \Delta_{\alpha_i}^{\alpha_{i+1}} \cot^{-1}\rho(X).$$

Here the notation

$$\Delta_\alpha^\beta f(Z)$$

denotes the increase in the argument of $f(Z)$ as Z moves along the line segment from α to β. Since $F(X)$ has no real roots, $\rho(X)$ is defined for all X (we do not get $0/0$) and $\rho(X) = 0$ iff $X \in \{\alpha_i : i = 1, \ldots, k\}$. We will be examining the signs of $\rho(\alpha_i^-)$ and $\rho(\alpha_i^+)$, and the following graph of the cotangent function is helpful:

Note that $\cot^{-1}\rho(\alpha_i) = \cot^{-1}0 = \pm\pi/2$ (taking values in the range $[-\pi, +\pi]$), and

$$\Delta_{\alpha_i}^{\alpha_{i+1}} \cot^{-1}\rho(X) = \lim_{\epsilon \to 0} \Delta_{\alpha_i+\epsilon}^{\alpha_{i+1}-\epsilon} \cot^{-1}\rho(X).$$

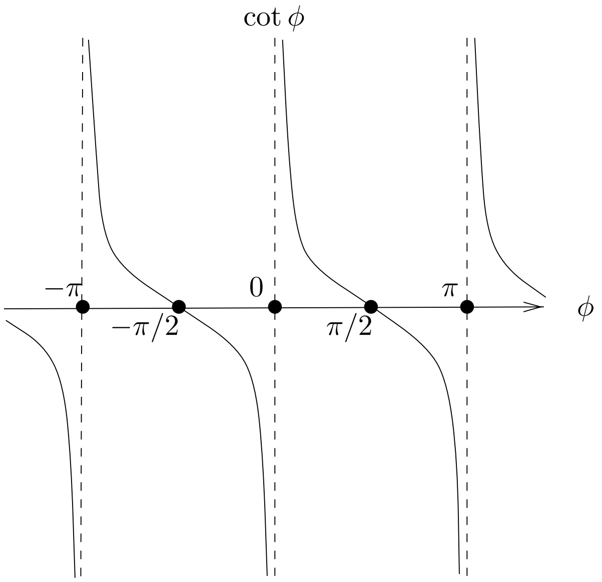

Figure 2: The cotangent function.

But $\rho(X)$ does not vanish in the interval $[\alpha_i+\epsilon, \alpha_{i+1}-\epsilon]$. Hence for $i = 1,\ldots,k-1$,

$$\begin{aligned}\Delta_{\alpha_i}^{\alpha_{i+1}} \cot^{-1}\rho(X) &= \begin{cases} 0 & \text{if } \ \rho(\alpha_i^+)\rho(\alpha_{i+1}^-) > 0 \\ \pi & \text{if } \ \rho(\alpha_i^+) < 0, \quad \rho(\alpha_{i+1}^-) > 0 \\ -\pi & \text{if } \ \rho(\alpha_i^+) > 0, \quad \rho(\alpha_{i+1}^-) < 0 \end{cases} \\ &= \pi\left[\frac{\mathtt{sign}(\rho(\alpha_{i+1}^-)) - \mathtt{sign}(\rho(\alpha_i^+))}{2}\right]. \end{aligned} \tag{16}$$

This is seen by an examination of the graph of $\cot\phi$. For $i = 0, k$, we first note that if $\deg F_0 > \deg F_1$ then $\rho(-\infty) = \pm\infty$ and $\rho(+\infty) = \pm\infty$. It follows that

$$\begin{aligned}\Delta_{-\infty}^{\alpha_1} \cot^{-1}\rho(X) &= \frac{\pi}{2}\mathtt{sign}(\rho(\alpha_1^-)), \\ \Delta_{\alpha_k}^{+\infty} \cot^{-1}\rho(X) &= -\frac{\pi}{2}\mathtt{sign}(\rho(\alpha_k^+)),\end{aligned}$$

and so

$$\Delta_{-\infty}^{\alpha_1} \cot^{-1}\rho(X) + \Delta_{\alpha_k}^{+\infty} \cot^{-1}\rho(X) = \frac{\pi}{2}\mathtt{sign}(\rho(\alpha_1^-)) - \frac{\pi}{2}\mathtt{sign}(\rho(\alpha_k^+)). \tag{17}$$

If $\deg F_0 = \deg F_1$ then $\rho(-\infty) = \rho(+\infty) = (\mathtt{lead}(F_0))/(\mathtt{lead}(F_1))$ and again (17) holds. Combining equations (16) and (17), we deduce:

Lemma 10

$$\Delta_L \arg F(X) = \pi \sum_{i=1}^{k} \frac{\mathtt{sign}(\rho(\alpha_i^-)) - \mathtt{sign}(\rho(\alpha_i^+))}{2}.$$

But α_i is a pole of $\rho^{-1} = F_1/F_0$. Hence the expression $\frac{\mathtt{sign}(\rho(\alpha_i^-)) - \mathtt{sign}(\rho(\alpha_i^+))}{2}$ is the Cauchy index of ρ^{-1} at α_i. By Corollary 7 (§3), this means $-\mathtt{Var}_{F_1,F_0}[-\infty,+\infty]$ gives the Cauchy index of ρ^{-1} over the real line L. Thus $\Delta_L \arg F(X) = -\mathtt{Var}_{F_1,F_0}[-\infty,+\infty]$. Combined with corollary 9, we obtain:

Theorem 11 (Routh-Hurwitz) *Let $F(X) = F_0(X) + \mathbf{i}F_1(X)$ be monic with $\deg F_0 \ge \deg F_1 \ge 0$ and F_0, F_1 relatively prime. The number of roots of $F(X)$ lying above the real axis L is given by*

$$\frac{1}{2}\left(\deg F - \mathtt{Var}_{F_1,F_0}[-\infty, +\infty]\right).$$

To exploit this result for a complex root isolation method, we proceed as follows.

1. Counting Roots to one side of the imaginary axis. Suppose we want to count the number p of roots of $F(Z)$ to the right of the imaginary axis, assuming $F(Z)$ does not have any purely imaginary roots. Note that α is a root of $F(Z)$ to the right of the imaginary axis iff $\mathbf{i}\alpha$ is a root of $F(Z/\mathbf{i}) = F(-\mathbf{i}Z)$ lying above the real axis. It is easy (previous section) to construct the polynomial $G(Z) := F(-\mathbf{i}Z)$ from $F(Z)$.

2. Roots in two opposite quadrants. We can count the number of roots in the first and third quadrant as follows: from $F(Z)$ construct a polynomial $F^*(Z)$ whose roots are precisely the squares of roots of $F(Z)$. This means that α is a root of $F(Z)$ in the first (I) or third (III) quadrant iff α^2 is a root of $F^*(Z)$ in the upper half-plane (which we know how to count). Similarly, the roots of $F(Z)$ in (II) and (IV) quadrants are sent into the lower half-plane. It remains to construct $F^*(Z)$. This is easily done as follows: Let $F(Z) = F_o(Z) + F_e(Z)$ where $F_o(Z)$ consists of those monomials of odd degree and $F_e(Z)$ consisting of those monomials of even degree. This means $F_o(Z)$ is an odd function (*i.e.*, $F_o(-Z) = -F_o(Z)$), and $F_e(Z)$ is an even function (*i.e.*, $F_e(-Z) = F_e(Z)$). Consider

$$\begin{aligned} G(Z) &= F_e(Z)^2 - F_o(Z)^2 \\ &= (F_e(Z) + F_o(Z))(F_e(Z) - F_o(Z)) \\ &= F(Z)(F_e(-Z) + F_o(-Z)) \\ &= F(Z)F(-Z). \end{aligned}$$

If $F(Z) = c\prod_{i=1}^n (Z - \beta_i)$ where β_i are the roots of $F(Z)$ then

$$F(Z)F(-Z) = c^2 \prod_{i=1}^n (Z - \beta_i)(-Z - \beta_i) = (-1)^n c^2 \prod_{i=1}^n (Z^2 - \beta_i^2).$$

Hence, we may define our desired polynomial $F^*(Y)$ by the relation $F^*(Z^2) = G(Z)$. In fact, $F^*(Y)$ is trivially obtained from the coefficients of $G(Z)$.

3. Roots inside a quadrant. We can count the number $\#(I)$ of roots in the first quadrant, since

$$\#(I) = \frac{1}{2}\left[(\#(I) + \#(II)) + (\#(I) + \#(IV)) - (\#(II) + \#(IV))\right]$$

where $\#(I) + \#(II)$ and $\#(I) + \#(IV)$ are half-plane counting queries, and $\#(II) + \#(IV)$ is a counting query for an opposite pair of quadrants. But we have shown how to answer such queries.

4. Roots in a translated quadrant. If the origin is translated to a point $\alpha \in \mathbb{C}$, we can count the number of roots of $F(Z)$ in any of the four quadrants whose origin is at α, by counting the number of roots of $F(Z + \alpha)$ in the corresponding quadrant.

5. Putting these together. In the last section, we have shown how to isolate a sequence $x_1 < x_2 < \cdots < x_k$ of real numbers that contain among them all the real parts of complex roots of $F(Z)$. Similarly, we can isolate a sequence $y_1 < y_2 < \cdots < y_\ell$ of real numbers that contains among them all the imaginary parts of complex roots of $F(Z)$. So finding all roots of $F(Z)$ is reduced to testing if each $x_i + \mathbf{i}y_j$ is a root. We may assume from the root isolation that we know (rational) numbers a_i, b_j such that

$$x_1 < a_1 < x_2 < a_2 < \cdots < a_{k-1} < x_k < a_k, \quad y_1 < b_1 < y_2 < b_2 < \cdots < b_{\ell-1} < y_\ell < b_\ell.$$

Then for $j = 1, \ldots, \ell$ and for $i = 1, \ldots, k$, we determine the number $n(i,j)$ of roots of $F(Z)$ in the quadrant (III) based at $a_i + \mathbf{i}b_j$. Note that $n(1,1) = 1$ or 0 depending on whether $x_1 + \mathbf{i}y_1$ is a root or not. It is easy to work out a simple scheme to similarly determine whether each $x_i + \mathbf{i}y_j$ is a root or not.

Exercises

Exercise 5.1: Determine the complexity of this procedure. Exploit the fact that the testings of the various $x_i + \mathbf{i}y_j$'s are related. □

Exercise 5.2: Isolate the roots of $F(Z) = (Z^2-1)(Z^2+0.16)$ using this procedure. [This polynomial has two real and two non-real roots. Newton iteration will fail in certain open neighborhoods (attractor regions).] □

Exercise 5.3: Derive an algorithm to determine if a complex polynomial has all its roots inside any given circle of the complex plane.
HINT: the transformation $w \mapsto z = r\frac{1+w}{1-w}$ (for any real $r > 0$) maps the half-plane $\mathtt{Re}(w) < 0$ into the open disc $|z| < r$. □

Exercise 5.4: If $F(X)$ is a real polynomial whose roots have no positive real parts then the coefficients of $F(X)$ have no sign variation.
HINT: write $F(X) = \prod_{i=1}^n (X - \alpha_i)$ and divide the n roots into the k real roots and 2ℓ complex roots ($n = k + 2\ell$).

□

Exercise 5.5: Let $F_n(X), F_{n-1}(X), \ldots, F_0(X)$ be a sequence of real polynomials where each $F_i(X)$ has degree i and positive leading coefficient. Moreover, $F_i(x) = 0$ implies $F_{i-1}(x)F_{i+1}(x) < 0$ (for $i = 1, 2, \ldots, n-1$, and $x \in \mathbb{R}$). Then each $F_i(X)$ ($i = 1, \ldots, n$) has i simple real roots and between any two consecutive roots is a root of F_{i-1}. □

Exercise 5.6: (Hermite, Biehler) If all the roots of $F(X) = A(X) + \mathbf{i}B(X)$ ($A(X), B(X) \in \mathbb{R}[X]$) lie on one side of the real axis of the complex plane, then $A(X)$ and $B(X)$ have only simple real roots, and conversely. □

§6. Sign Encoding of Algebraic Numbers: Thom's Lemma

We present an alternative representation of real algebraic numbers as suggested by Coste and Roy [3]. If $\overline{A} = [A_1(X), A_2(X), \ldots, A_m(X)]$ is a sequence[2] of real polynomials, then a *sign condition* of $\overline{A}$ is any sequence of signs,

$$[s_1, s_2, \ldots, s_m], \qquad s_i \in \{-1, 0, +1\}.$$

We say $[s_1, s_2, \ldots, s_m]$ is the *sign condition* of $\overline{A}$ at $\alpha \in \mathbb{R}$ if $s_i = \mathtt{sign}(A_i(\alpha))$ for $i = 1, \ldots, m$. This will be denoted

$$\mathtt{sign}_\alpha(\overline{A}) = [s_1, \ldots, s_m].$$

A sign condition of $\overline{A}$ is *consistent* if there exists such an α. Define the sequence

$$\mathrm{Der}[A] := [A(X), A'(X), A^{(2)}(X), \ldots, A^{(n)}(X)], \qquad \deg A = n,$$

of derivatives of $A(X) \in \mathbb{R}[X]$. The representation of algebraic numbers is based on the following "little lemma" of Thom. Let us call a subset of $\mathbb{R}$ *simple* if it is empty, a singleton or an open interval.

Lemma 12 (Thom) *Let $A(X) \in \mathbb{R}[X]$ have degree $n \ge 0$ and let $s = [s_0, s_1, \ldots, s_n] \in \{-1, 0, +1\}^{n+1}$ be any sign condition. Then the set*

$$S := \{x \in \mathbb{R} : \mathtt{sign}(A^{(i)}(x)) = s_i, \text{for all } \ i = 0, \ldots, n\}$$

is simple.

Proof. We may use induction on n. If $n = 0$ then $A(X)$ is a non-zero constant and S is either empty or equal to $\mathbb{R}$. So let $n \ge 1$ and let $s' = [s_1, \ldots, s_n]$. Then the set

$$S' := \{x \in \mathbb{R} : \mathtt{sign}(A^{(i)}(x)) = s_i, i = 1, \ldots, n\}$$

is simple, by the inductive hypothesis for $A'(X)$. Note that $S = S' \cap S_0$ where $S_0 := \{x \in \mathbb{R} : \mathtt{sign}(A(x)) = s_0\}$. Now the set S_0 is a disjoint union of simple sets. In fact, viewing $A(X)$ as a continuous real function, S_0 is equal to $A^{-1}(0)$, $A^{-1}(\mathbb{R}_{>0})$ or $A^{-1}(\mathbb{R}_{<0})$, depending on whether $s_0 = 0, +1$ or -1. In any case, we see that if $S' \cap S_0$ is a connected set, then it is simple. So assume it is disconnected. Then S' contains two distinct roots of $A(X)$. By Rolle's theorem (§VI.1), $A'(X)$ must have a root in S'. This implies S' is contained in the set $\{x \in \mathbb{R} : \mathtt{sign}(A'(x)) = 0\}$, which is a finite set. Since S' is connected, it follows that S' is empty or a singleton. This contradicts the assumption that $S' \cap S_0$ is disconnected. **Q.E.D.**

Lemma 13 *Let α, β be distinct real roots of $A(X)$, $\deg A(X) = n \ge 2$. Let $s = [s_0, \ldots, s_n]$ and $s' = [s'_0, \ldots, s'_n]$ be the sign conditions of Der$[A]$ at α and at β (respectively).*
(i) s and s' are distinct.
(ii) Let i be the largest index such that $s_i \ne s'_i$. Then $0 < i < n$ and $s_{i+1} = s'_{i+1} \ne 0$. Furthermore, $\alpha < \beta$ iff one of the following conditions holds:

$$\begin{array}{ll} (a) & s_{i+1} = +1 \text{ and } s_i < s'_i; \\ (b) & s_{i+1} = -1 \text{ and } s_i > s'_i. \end{array}$$

Proof. Let I be the open interval bounded by α, β.
(i) If $s = s'$ then by Thom's lemma, every $\gamma \in I$ also achieves the sign condition s. In particular, this means $A(\gamma) = 0$. Since there are infinitely many such γ, $A(X)$ must be identically zero,

[2] In this section, we use square brackets '[...]' as a stylistic variant of the usual parentheses '(...)' for writing certain sequences.

contradiction.
(ii) It is clear that $0 < i < n$ since $s_0 = s'_0 = 0$ and $s_n = s'_n$. Thom's lemma applied to the polynomial $A^{(i+1)}(X)$ implies that $A^{(i+1)}(\gamma)$ has constant sign throughout the interval I. If $s_{i+1} = s'_{i+1} = 0$ then we obtain the contradiction that $A^{(i+1)}(X)$ is identically zero in I. So suppose $s_{i+1} = s'_{i+1} = +1$ (the other case being symmetrical). Again by Thom's lemma, we conclude that $A^{(i+1)}(\gamma) > 0$ for all $\gamma \in I$, *i.e.*, $A^{(i)}(X)$ is strictly increasing in I. Thus $\alpha < \beta$ iff

$$A^{(i)}(\alpha) < A^{(i)}(\beta). \tag{18}$$

Since the signs of $A^{(i)}(\alpha)$ and $A^{(i)}(\beta)$ are distinct, the inequality (18) amounts to $s_i < s'_i$. **Q.E.D.**

This result suggests that we code a real algebraic number α by specifying a polynomial $A(X)$ at which α vanishes, and by specifying its sign condition at Der$[A']$, written

$$\alpha \cong (A(X), \mathtt{sign}(\mathrm{Der}[A'])).$$

This is the same notation ($\cong$) used when α is represented by an isolating interval (§VI.9), but it should not lead to any confusion. We call $(A(X), \mathtt{sign}(\mathrm{Der}[A']))$ a *sign encoding* of α. For example, $\sqrt{2} \cong (X^2 - 2, [+1, +1])$ and $-\sqrt{2} \cong (X^2 - 2, [-1, +1])$.

This encoding has some advantages over the isolating interval representation in that, once A is fixed, the representation is unique (and we can make A unique by choosing the distinguished minimal polynomial of α). It's discrete nature is also desirable. On the other hand, the isolating intervals representation gives an explicit numerical approximation, which is useful. Coste and Roy [3] also generalized the sign encoding to the multivariate situation.

Exercises

Exercise 6.1: Let $s = [s_0, \ldots, s_n]$ be a sequence of *generalized sign condition* that is, s_i belongs to the set $\{< 0, \leq 0, 0, \geq 0, > 0\}$ of generalized signs (rather than $s_i \in \{-1, 0, +1\}$). If $A(X)$ has degree $n \geq 0$, show that the set $\{x \in \mathbb{R} : s = \mathtt{sign}_x(\mathrm{Der}[A])\}$ is connected (possibly empty). □

Exercise 6.2: Give an algorithm to compare two arbitrary real algebraic numbers in this representation. □

§7. Problem of Relative Sign Conditions

Uses of the sign encoding of real algebraic numbers depend on a key algorithm from Ben-Or, Kozen and Reif [1]. This algorithm has come to be known as the "BKR algorithm". We first describe the problem solved by this algorithm.

Let $\overline{B} = [B_1, B_2, \ldots, B_m]$ be a sequence of real polynomials, and A another real polynomial. A sign condition $s = [s_1, \ldots, s_m]$ of $\overline{B}$ is *consistent relative to* A (or, A-*consistent*) if $[0, s_1, \ldots, s_m]$ is consistent for the sequence $[A, B_1, \ldots, B_m]$. In other words, s is A-consistent if $s = \mathtt{sign}_\alpha[\overline{B}]$ for some root α of A. The *weight of* s *relative to* A is the number of roots of A at which $\overline{B}$ achieves the sign condition s. Thus s is relatively consistent iff $[0, s_1, \ldots, s_m]$ has positive weight. If A is understood, we may simply call s a *relatively consistent sign condition* of $\overline{B}$.

The *problem of relative sign consistency*, on input $A, \overline{B}$, asks for the set of all A-consistent sign conditions of $\overline{B}$; a stronger version of this problem is to further ask for the weight of each A-consistent sign condition.

There are numerous other applications of this problem, but we can see immediately its applications to the sign encoding representation:

- To determine the sign encoding of all roots of $A(X)$, it suffices to call the BKR algorithm on $A, \overline{B}$ where $\overline{B} = \mathrm{Der}[A']$.
- To determine the sign of a polynomial $P(X)$ at the roots of A, we call BKR on $A, \overline{B}$ where $\overline{B} = [P, A', A^{(2)}, \ldots, A^{(m-1)}]$.

The original BKR algorithm is described only for the case where $A, B_1, \ldots, B_m$ are relatively prime, as the general case can be reduced to this special case. Still, it is convenient to give a direct algorithm. Mishra and Pedersen [10] observed that corollary 6 used in the original BKR algorithm in fact holds without any conditions on the polynomials A, B:

Lemma 14 *Let $A, B \in \mathbb{R}[X]$ such that $A(\alpha)A(\beta) \neq 0$, $\alpha < \beta$. Then*

$$\mathtt{Var}_{A,A'B}[\alpha, \beta] = \sum_{\gamma} \mathtt{sign}(B(\gamma))$$

where γ ranges over the distinct real roots of A.

Proof. Again, it suffices to prove this for a fundamental interval $[\alpha, \beta]$ at some $\gamma_0 \in [\alpha, \beta]$. Let γ_0 be an r-fold root of A and an s-fold root of $A'B$. If $r \geq s$, then this has been proved in corollary 6. So assume $s > r$. The sign variation difference over $[\alpha, \beta]$ in the Sturm sequence $[A_0, A_1, \ldots, A_h]$ for $A, A'B$ is evidently equal to that in the depressed sequence $[A_0/A_h, A_1/A_h, \ldots, 1]$. But the sign variation difference in the depressed sequence is 0 since γ_0 is a non-root of A_0/A_h (here we use the fact that γ_0 is an r-fold root of A_h). Since $B(\gamma_0) = 0$ (as $s > r$), we have verified

$$\mathtt{Var}_{A,A'B}[\alpha, \beta] = 0 = \mathtt{sign}(B(\gamma_0)).$$

Q.E.D.

In the following, we fix A and $\overline{B} = [B_1, \ldots, B_m]$. If ε is a sign condition of $\overline{B}$, write

$$W^{\varepsilon} := \{\alpha : A(\alpha) = 0, \mathtt{sign}_{\alpha}[\overline{B}] = \varepsilon\} \tag{19}$$

for the set of real roots α of A at which $\overline{B}$ achieves the condition ε. So the weight of ε is given by

$$w^{\varepsilon} := |W^{\varepsilon}|.$$

For instance, when $m = 1$, the roots of A are partitioned into W^0, W^+, W^-. When $m = 3$, w^{+-0} is the number of roots of A at which B_1 is positive, B_2 is negative and B_3 vanishes.

So the BKR algorithm amounts to determining these weights. First consider some initial cases of the BKR algorithm (for small m).

CASE $m = 0$: In this case, the A-consistent sign condition is $[\,]$ (the empty sequence) and its weight is (by definition) just the number of real roots of A. By the original Sturm theorem (§3), this is given by

$$v_A(1) := \mathtt{Var}_{A,A'}[-\infty, +\infty].$$

In general, we shall abbreviate $\mathtt{Var}_{A,A'B}[-\infty,+\infty]$ by $v_A(B)$, or simply, $v(B)$ if A is understood. In this context, computing $v(B)$ is sometimes called "making a Sturm query on B".

CASE $m = 1$: By the preceding lemma,

$$v_A(B_1) = w^+ - w^-, \qquad v_A(B_1^2) = w^+ + w^-.$$

Case $m = 0$ shows that

$$v_A(1) = w^0 + w^+ + w^-.$$

We put these together in the matrix format,

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} w^0 \\ w^+ \\ w^- \end{bmatrix} = \begin{bmatrix} v(1) \\ v(B_1) \\ v(B_1^2) \end{bmatrix}. \tag{20}$$

Thus we can solve for w^0, w^+, w^- since we know the right hand side after making the three Sturm queries $v(1), v(B_1), v(B_1^2)$.

CASE $m = 2$: If we let M_1 be the matrix in equation (20), it is not hard to verify

$$\begin{bmatrix} M_1 & M_1 & M_1 \\ \mathbf{0} & M_1 & -M_1 \\ \mathbf{0} & M_1 & M_1 \end{bmatrix} \cdot \begin{bmatrix} w^{00} \\ w^{0+} \\ w^{0-} \\ w^{+0} \\ w^{++} \\ w^{+-} \\ w^{-0} \\ w^{-+} \\ w^{--} \end{bmatrix} = \begin{bmatrix} v(1) \\ v(B_1) \\ v(B_1^2) \\ v(B_2) \\ v(B_1B_2) \\ v(B_1^2B_2) \\ v(B_2^2) \\ v(B_1B_2^2) \\ v(B_1^2B_2^2) \end{bmatrix}. \tag{21}$$

Again, we can solve for the weights after making some Sturm queries. The case $m = 2$ will illustrate the general development of the BKR algorithm below. If the square matrix in (21) is denoted M_2 then M_2 can be viewed as the "Kronecker product" of M_1 with itself.

Exercises

Exercise 7.1: Let α have the sign encoding $E = (A(X), [s_1, \ldots, s_m])$.
(i) What is the sign encoding of $-\alpha$ in terms of E?
(ii) Give a method to compute the sign encoding E' of $1/\alpha$. Assume that the polynomial in E' is $X^m A(1/X)$. HINT: consider $\mathrm{Der}[A](1/X)$ instead of $\mathrm{Der}[X^m A(1/X)]$. □

§8. The BKR algorithm

We now develop the BKR algorithm.

Let $M \in R^{m\times n}$ and $M' \in R^{m'\times n'}$ where R is any ring. The *Kronecker product* $M \otimes M'$ of M and M' is the $mm' \times nn'$ matrix partitioned into $m \times n$ blocks, with the (i,j)th block equal to

$$(M)_{ij} \cdot M'.$$

In other words, $M \otimes M'$ is defined by

$$\begin{array}{l}(M \otimes M')_{(i-1)m'+i',(j-1)m'+j'} = M_{ij} M_{i'j'}, \\ \qquad i \in \{1, \ldots, m\}, j \in \{1, \ldots, n\}, i' \in \{1, \ldots, m'\}, j' \in \{1, \ldots, n'\}.\end{array}$$

For instance, the matrix M_2 in (21) can be expressed as $M_1 \otimes M_1$. Again, if u, u' are m-vectors and m'-vectors, respectively, then $u \otimes u'$ is a (mm')-vector. We leave it as an exercise to show that the Kronecker product is associative.

Lemma 15 *Let $M \in R^{m \times m}$ and $M' \in R^{m' \times m'}$ and u, u' be m-vectors and m'-vectors, respectively.*
(i) $(M \otimes M')(u \otimes u') = (Mu) \otimes (M'u')$.
(ii) If M, M' are invertible, so is $M \otimes M'$, with inverse $M^{-1} \otimes M'^{-1}$.

Proof. (i) This is a straightforward exercise.
(ii) Consider the action of the matrix product $(M^{-1} \otimes M'^{-1}) \cdot (M \otimes M')$ on $u \otimes u'$:

$$\begin{array}{rcl}(M^{-1} \otimes M'^{-1}) \cdot (M \otimes M') \cdot u \otimes u' & = & (M^{-1} \otimes M'^{-1}) \cdot (M \cdot u \otimes M' \cdot u') \\ & = & (M^{-1} \cdot M \cdot u) \otimes (M'^{-1} \cdot M' \cdot u'). \\ & = & u \otimes u'.\end{array}$$

As u, u' are arbitrary, this proves that $(M^{-1} \otimes M'^{-1}) \cdot (M \otimes M')$ is the identity matrix. **Q.E.D.**

The real algebra of vectors. We describe the BKR algorithm by "shadowing" its action in the ring $R = \mathbb{R}^k$ of k-vectors over $\mathbb{R}$. This notion of shadowing will be clarified below; but it basically makes the correctness of the algorithm transparent.

Note that $R = \mathbb{R}^k$ is a ring under component-wise addition and multiplication. The real numbers $\mathbb{R}$ are embedded in R under the correspondence $\alpha \in \mathbb{R} \mapsto (\alpha, \alpha, \ldots, \alpha) \in R$. Thus R is a real algebra[3].

To describe the BKR algorithm on inputs $A(X)$ and $\overline{B} = [B_1, \ldots, B_m]$, we first choose the k in the definition of R to be the number of distinct real roots of the polynomial $A(X)$; let these roots be

$$\overline{\alpha} = (\alpha_1, \ldots, \alpha_k). \tag{22}$$

We shall use R in two distinct ways:

- A vector in R with entries from $-1, 0, +1$ will be called a *root sign vector*. Such vectors[4] represent the signs of a polynomial $Q(X)$ at the k real roots of $A(X)$ in the natural way:

  $$\mathtt{sign}_{A(X)}(Q(X)))$$

  denotes the sign vector $[s_1, \ldots, s_k]$ where $s_i = \mathtt{sign}(Q(\alpha_i))$. If $s_i = \mathtt{sign}_A(Q_i)$ $(i = 0, 1)$ then notice that $s_0 \cdot s_1 = \mathtt{sign}_A(Q_0 Q_1)$.

  In the BKR algorithm, Q will be a power product of $B_1, \ldots, B_m$.

[3] In general, a ring R containing a subfield K is called a K-algebra.

[4] Although root sign vectors are formally sign conditions, notice that root sign vectors arise quite differently, and hence the new terminology. By the same token, Boolean vectors are formally a special type of sign condition, but they are interpreted very differently.

- A 0/1 vector in R will be called a *Boolean vector*. Such a vector u represents a subset U of the roots of $A(X)$ in the natural way: the i-th component of u is 1 iff $\alpha_i \in U$. If the Boolean vectors $u_0, u_1 \in R$ represent the subsets U_0, U_1 (respectively) then observe that $U_0 \cap U_1$ is represented by the vector product $u_0 \cdot u_1$.

  In the BKR algorithm, the subsets U are determined by sign conditions of $\overline{B}$: such subsets have the form W^ε (see equation (19)) where $\varepsilon = [s_1, \ldots, s_\ell]$ is a sign condition of $\overline{C} = [C_1, \ldots, C_\ell]$ and $\overline{C}$ is a subsequence of $\overline{B}$. Note that ε is not to be confused with the root sign vectors in R. In fact, we define a rather different product operation on such sign conditions: let $\varepsilon = [s_1, \ldots, s_\ell]$ be a sign condition of $\overline{C} = [C_1, \ldots, C_\ell]$ and $\varepsilon' = [s_{\ell+1}, \ldots, s_{\ell'}]$ be a sign condition of $\overline{C}' = [C_{\ell+1}, \ldots, C_{\ell'}]$, $\ell < \ell'$. Assuming that $\overline{C}$ and $\overline{C}'$ are disjoint, we define

  $$\varepsilon \cdot \varepsilon' :=[s_1, \ldots, s_\ell, s_{\ell+1}, \ldots, s_{\ell'}],$$

  i.e., the concatenation of ε with ε'. This definition of product is consistent with the product in R in the following sense: if $u_0, u_1 \in R$ represent $W^\varepsilon, W^{\varepsilon'}$ (respectively) then $u_0 \cdot u_1$ (multiplication in R) represents

  $$W^{\varepsilon \cdot \varepsilon}.$$

We come to a key definition: let $\overline{C} = [C_1, \ldots, C_\ell]$ be a subsequence of $\overline{B}$. Let $M \in \mathbb{R}^{\ell \times \ell}$, $\overline{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_\ell]$ where each ε_i is a sign condition for $\overline{C} = [C_1, \ldots, C_\ell]$, and $\overline{Q} = [Q_1, \ldots, Q_\ell]$ be a sequence of real polynomials. We say that

$$(M, \overline{\varepsilon}, \overline{Q})$$

is a *valid triple* for $\overline{C}$ if the following conditions hold:

- M is invertible.
- Every A-consistent sign condition for $\overline{C}$ occurs in $\overline{\varepsilon}$ (so $\overline{\varepsilon}$ may contain relatively inconsistent sign conditions).
- The equation

  $$M \cdot u = s \qquad (23)$$

  holds in R where $u = (u_1, \ldots, u_\ell)^T$ with each u_i a Boolean vector representing W^{ε_i}, and $s = (s_1, \ldots, s_\ell)^T$ with s_i equal to the root sign vector $\mathtt{sign}_A(Q_i) \in R$. Equation (23) is called the *underlying equation* of the triple.

We can view the goal of the BKR algorithm to be the computation of valid triples for $\overline{B}$ (note that A is implicit in our definition of valid triples).

Example: $(M_1, ([0], [+], [-]), [1, B_1, B_1^2])$ is a valid triple for $\overline{B} = [B_1]$. The underlying equation is

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ 0 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} u^0 \\ u^+ \\ u^- \end{bmatrix} = \begin{bmatrix} \mathtt{sign}_A(1) \\ \mathtt{sign}_A(B_1) \\ \mathtt{sign}_A(B_1^2) \end{bmatrix}. \qquad (24)$$

where we write u^0, u^+, u^- for the Boolean vectors representing the sets W^0, W^+, W^-. Compare this equation to equation (20).

We define the "Kronecker product" of two triples $(M, \overline{\varepsilon}, \overline{Q})$ and $(M', \overline{\varepsilon}', \overline{Q}')$ as

$$(M \otimes M', \overline{\varepsilon} \otimes \overline{\varepsilon}', \overline{Q} \otimes \overline{Q}')$$

where the underlying "multiplication" in $\overline{\varepsilon} \otimes \overline{\varepsilon}'$ and $\overline{Q} \otimes \overline{Q}'$ are (respectively) concatenation of sign conditions and multiplication of polynomials. For example,

$$(0,+,-) \otimes (+-,-0) = (0+-, 0-0, ++-, +-0, -+-, --0)$$

and

$$[Q_1, Q_2] \otimes [Q_3, Q_4] = [Q_1Q_3, Q_1Q_4, Q_2Q_3, Q_3Q_4].$$

Lemma 16 *Suppose* $(M, \overline{\varepsilon}, \overline{Q})$ *is valid for* $[B_1, \ldots, B_\ell]$ *and* $(M', \overline{\varepsilon}', \overline{Q}')$ *is valid for* $[B_{\ell+1}, \ldots, B_{\ell+\ell'}]$. *Then*

$$(M \otimes M', \overline{\varepsilon} \otimes \overline{\varepsilon}', \overline{Q} \otimes \overline{Q}') \tag{25}$$

is valid for $[B_1, \ldots, B_\ell, B_{\ell+1}, \ldots, B_{\ell+\ell'}]$.

Proof. (i) First we note that $M \otimes M'$ is invertible.
(ii) Next note that every A-consistent sign condition for $[B_1, \ldots, B_{\ell+\ell'}]$ is listed in $\overline{\varepsilon} \otimes \overline{\varepsilon}'$.
(iii) Let the underlying equations of $(M, \overline{\varepsilon}, \overline{Q})$ and $(M', \overline{\varepsilon}', \overline{Q}')$ be $M \cdot u = s$ and $M' \cdot u' = s'$, respectively. By lemma 15(i),

$$(M \otimes M')(u \otimes u') = s \otimes s'. \tag{26}$$

Then it remains to see that equation (26) is the underlying equation for equation (25). This follows since for each i, $(u \otimes u')_i$ represents the set $W^{(\overline{\varepsilon} \otimes \overline{\varepsilon}')_i}$, and $(s \otimes s')_i = \mathtt{sign}_A((\overline{Q} \otimes \overline{Q}')_i)$. **Q.E.D.**

Pruning. It follows from this lemma that

$$(M_2, ([0],[+],[-]) \otimes ([0],[+],[-]), [1, B_1, B_1^2] \otimes [1, B_2, B_2^2])$$

is a valid triple for $[B_1, B_2]$. We can repeat this formation of Kronecker product m times to obtain a valid triple $(M, \overline{\varepsilon}, \overline{Q})$ for $[B_1, \ldots, B_m]$. But the size of the matrix M would be $3^m \times 3^m$, which is too large for practical computation. This motivates the idea of "pruning". Observe that the number of A-consistent sign conditions cannot be more than k. This means that in the underlying equation $Mu = s$, all but k of the Boolean vectors $(u)_i$ must be the zero vector $\mathbf{0}$ (representing the empty set). The following steps reduces the matrix M to size at most $k \times k$:

```
Pruning Procedure for the equation $Mu = s$:
    1. Detect and eliminate the zero vectors in $u$.
         Call the resulting vector $u'$.
         So the length of $u'$ is $\ell$ where $\ell \leq k$.
    2. Omit the columns in $M$ corresponding to eliminated entries of $u$.
         We get a new matrix $M''$ satisfying $M''u' = s$.
    3. Since $M$ is invertible, find $\ell$ rows in $M''$ that form
         an invertible $\ell \times \ell$ matrix $M'$.
    4. If $s'$ are the entries corresponding to these rows,
         we finally obtain the "pruned equation" $M'u' = s'$.
```

After we have pruned the underlying equation of the valid triple $(M, \overline{\varepsilon}, \overline{Q})$, we can likewise "prune" the valid triple to a new triple $(M', \overline{\varepsilon}', \overline{Q}')$ whose underlying equation is $M'u' = s'$. It is not hard to verify that that this new triple is valid. The resulting matrix M' has size at most $k \times k$.

Shadowing. The Pruning Procedure above is not intended to be effective because we have no intention of computing over R. Instead, we apply the linear map

$$\lambda : R \to \mathbb{R}$$

defined by $\lambda(x) = \sum_{i=1}^{k} x_i$ for $x = (x_1, \ldots, x_k)$. Notice

- If x is a Boolean vector representing W^{ε} then $\lambda(x) = w^{\varepsilon}$.
- If x is a root sign condition for a polynomial Q then $\lambda(x) = v_A(Q)$, a Sturm query on Q.

If $u \in R^{\ell}$, then $\lambda(u) \in \mathbb{R}^{\ell}$ is defined by applying λ component-wise to u. The underlying equation is transformed by λ into the real matrix equation,

$$M \cdot \lambda(u) = \lambda(s).$$

This equation is only a "shadow" of the underlying equation, but we can effectively compute with this equation. More precisely, we can compute $\lambda(s)$ since it is just a sequence of Sturm queries:

$$\lambda(s) = (v_A(Q_1), \ldots, v_A(Q_{\ell}))^T$$

where $\overline{Q} = (Q_1, \ldots, Q_{\ell})$. From this, we can next compute $\lambda(u)$ as $M^{-1} \cdot \lambda(s)$. The A-inconsistent sign conditions in $\overline{\varepsilon}$ correspond precisely to the 0 entries in $\lambda(u)$. Thus step 1 in the Pruning Procedure can be effectively carried out. The remaining steps of the Pruning Procedure can now be carried out since we have direct access to the matrix M (we do not need u or s). Finally we can compute the pruned valid triple.

All the ingredients for the BKR algorithm are now present:

```
BKR Algorithm
    Input: $A(X)$ and $\overline{B} = [B_1, \ldots, B_m]$.
    Output: a valid triple $(M, \overline{\varepsilon}, \overline{Q})$ for $\overline{B}$.
1.  If $m = 1$, we output $(M_1, ([0], [+], [-]), (1, B_1, B_1^2))$ as described above.
2.  If $m \geq 2$, recursively compute $(M', \overline{\varepsilon}', \overline{Q}')$ valid for $[B_1, \ldots, B_{\ell}]$ ($\ell = \lfloor m/2 \rfloor$),
        and also $(M'', \overline{\varepsilon}'', \overline{Q}'')$ valid for $[B_{\ell+1}, \ldots, B_m]$.
3.  Compute the Kronecker product of $(M', \overline{\varepsilon}', \overline{Q}')$ and $(M'', \overline{\varepsilon}'', \overline{Q}'')$.
4.  Compute and output the pruned Kronecker product.
```

The correctness of this algorithm follows from the preceding development. The algorithm can actually be implemented efficiently using circuits.

Exercises

Exercise 8.1: Show that Kronecker products is associative: $M_1 := (M \otimes M') \otimes M''$ is equal to $M_2 = M \otimes (M' \otimes M'')$. HINT: let M be $m \times n$, M' be $m' \times n'$ and M'' be $m'' \times n''$. Write $I = (i, i', i'')$ and $J = (j, j', j'')$ where (i, j) range over the indices of M, (i', j') range over the indices of M', etc. Then interpret (I, J) to run over the indices of the triple products M_1 and M_2 (how?). Express $(M_1)_{I,J}$ as a function of $(M)_{i,j}, (M')_{i',j'}$, etc. □

Exercise 8.2: Analyze the complexity of the BKR algorithm. □

End Exercises

References

[1] M. Ben-Or, D. Kozen, and J. Reif. The complexity of elementary algebra and geometry. *J. of Computer and System Sciences*, 32:251–264, 1986.

[2] W. S. Burnside and A. W. Panton. *The Theory of Equations*, volume 1. Dover Publications, New York, 1912.

[3] M. Coste and M. F. Roy. Thom's lemma, the coding of real algebraic numbers and the computation of the topology of semi-algebraic sets. *J. of Symbolic Computation*, 5:121–130, 1988.

[4] M. Davis. *Computability and Unsolvability.* Dover Publications, Inc., New York, 1982.

[5] M. Davis, H. Putnam, and J. Robinson. The decision problem for exponential Diophantine equations. *Annals of Mathematics, 2nd Series*, 74(3):425–436, 1962.

[6] A. S. Householder. *Principles of Numerical Analysis.* McGraw-Hill, New York, 1953.

[7] N. Jacobson. *Basic Algebra 1.* W. H. Freeman, San Francisco, 1974.

[8] Y. V. Matiyasevich. *Hilbert's Tenth Problem.* The MIT Press, Cambridge, Massachusetts, 1994.

[9] P. S. Milne. On the solutions of a set of polynomial equations. In B. R. Donald, D. Kapur, and J. L. Mundy, editors, *Symbolic and Numerical Computation for Artificial Intelligence*, pages 89–102. Academic Press, London, 1992.

[10] B. Mishra and P. Pedersen. Arithmetic of real algebraic numbers is in *NC*. Technical Report 220, Courant Institute of Mathematical Sciences, Robotics Laboratory, New York University, Jan 1990.

[11] P. Pedersen. Counting real zeroes. Technical Report 243, Courant Institute of Mathematical Sciences, Robotics Laboratory, New York University, 1990. PhD Thesis, Courant Institute, New York University.

[12] J. R. Pinkert. An exact method for finding the roots of a complex polynomial. *ACM Trans. on Math. Software*, 2:351–363, 1976.

[13] S. M. Rump. On the sign of a real algebraic number. *Proceedings of 1976 ACM Symp. on Symbolic and Algebraic Computation (SYMSAC 76)*, pages 238–241, 1976. Yorktown Heights, New York.

[14] J. J. Sylvester. On a remarkable modification of Sturm's theorem. *Philosophical Magazine*, pages 446–456, 1853.

[15] J. J. Sylvester. On a theory of the syzegetic relations of two rational integral functions, comprising an application to the theory of Sturm's functions, and that of the greatest algebraical common measure. *Philosophical Trans.*, 143:407–584, 1853.

[16] J. J. Sylvester. *The Collected Mathematical Papers of James Joseph Sylvester*, volume 1. Cambridge University Press, Cambridge, 1904.

[17] J. V. Uspensky. *Theory of Equations.* McGraw-Hill, New York, 1948.

Contents