Analytic Theory of Polynomials

Q. I. Rahman
Université de Montréal

and

G. Schmeisser
Universität Erlangen – Nürnberg

CLARENDON PRESS · OXFORD
2002

by a division scheme. Repeating this process with the pair f, g replaced by g, r, etc., we arrive at the *Euclidean algorithm.* It may formally be described by

$$\begin{aligned} f_0 &:= f(z)\,, \qquad f_1(z) := g(z)\,, \\ f_j(z) &=: q_j(z) f_{j+1}(z) - f_{j+2}(z) \qquad (j = 0, 1, \ldots)\,. \end{aligned} \tag{1.1.3}$$

Since the degrees of the polynomials f_j are strictly decreasing, we finish with an equation $f_{k-1}(z) = q_{k-1}(z) f_k(z)$, where $f_k(z) \not\equiv 0$. We may also allow the possibility that $\deg g > \deg f$. Then (1.1.2) holds with $q(z)$ being identically zero and $r(z) = -f(z)$, that is, $f_2 = -f$. Hence, except for a sign, the first step of the Euclidean algorithm interchanges f and g. The relations (1.1.3) show that any common divisor of f_0 and f_1 is a divisor of all the successors $f_2, \ldots, f_k$. Conversely, f_k is a divisor of all its predecessors $f_{k-1}, \ldots, f_0$. Hence f_k is a *greatest common divisor* of f and g. Moreover, we have factorizations

$$f(z) = f_k(z)\varphi(z) \quad \text{and} \quad g(z) = f_k(z)\psi(z)\,,$$

where φ and ψ are appropriate polynomials.

A point $\zeta \in \mathbb{C}$ is a *zero* of *multiplicity* m, or of *order* m, or an m*-fold* zero of f if

$$f(\zeta) = f'(\zeta) = \cdots = f^{(m-1)}(\zeta) = 0 \quad \text{and} \quad f^{(m)}(\zeta) \neq 0\,.$$

A zero of multiplicity 1 is said to be *simple.* If $f(\zeta) \neq 0$, then we may consider ζ as a zero of multiplicity 0. In the case where $m \geq 1$, the division transformation with $g(z) := (z - \zeta)^m$ yields that $f(z) = q(z)(z - \zeta)^m$. In fact, if the polynomial r in (1.1.2) were not identically zero, then either r itself or one of its derivatives of order less than m would be a non-vanishing constant. However, $r^{(\mu)}(\zeta) = 0$ for $\mu = 0, \ldots, m-1$; this is a contradiction.

In this way we also conclude that a polynomial of degree $n \geq 1$ cannot have more than n zeros, counted according to their multiplicities. However, it is a non-trivial question whether an arbitrary polynomial of positive degree has a zero at all. Its answer is crucial for the whole book, and is the main result of this section.

Theorem 1.1.1 (Fundamental theorem of algebra) *Every polynomial of positive degree has a complex zero.*

Proof Let f be a polynomial of positive degree. Then $|f(z)|$ tends to infinity as $|z| \to \infty$. Hence there exists an $R > 0$ such that $|f(z)| > |f(0)|$ for $|z| > R$. As a continuous function, $|f(\cdot)|$ attains a minimum on the compact disc $\overline{\mathcal{D}}(0; R)$. If this occurs at ζ, then $|f(\zeta)| \leq |f(0)|$, and so $|f(\zeta)| = \min_{z \in \mathbb{C}} |f(z)|$.

Now, assuming that f has no zero in $\mathbb{C}$, we may write

$$f(\zeta + z) = a + bz^k + z^{k+1} g(z)$$

with some integer $k \geq 1$, coefficients $a, b \in \mathbb{C} \setminus \{0\}$, and a polynomial $g(z)$. Corresponding to ω, chosen as a k-th root of $-a/b$, there exists a $t \in (0, 1)$ so that $t \left|\omega^{k+1} g(t\omega)\right| < |a|$. For this pair t, ω, we obtain

$$\begin{aligned} |f(\zeta + t\omega)| = \left|a + b(t\omega)^k + (t\omega)^{k+1} g(t\omega)\right| &= \left|a(1 - t^k) + (t\omega^{k+1} g(t\omega))t^k\right| \\ &< |a|\,(1 - t^k) + |a|\, t^k = |a| = |f(\zeta)|\,. \end{aligned}$$

This permits us to determine the signature. We first introduce some notation. Let $a_0, \dots, a_n$ be a sequence of non-vanishing real numbers. Then we count a variation of sign at the index k, where $1 \le k \le n$, if $a_{k-1}a_k < 0$, and denote by $V(a_0, \dots, a_n)$ the total number of variations of sign in this sequence. For example, $V(-3, -2, 1, 5, -2) = 2$.

It is easily seen that $V(1, D_1, \dots, D_m)$ is the number of squares in (1.7.10) which have a negative coefficient. This leads us to the following result.

Theorem 1.7.14 (Jacobi) *Let A be a symmetric matrix of rank m. Suppose that the principal minors $D_1, \dots, D_m$ of A are all different from zero. Then the quadratic form $Q(\boldsymbol{x}) := \boldsymbol{x}^\top A\boldsymbol{x}$ has signature $\sigma(Q) = m - 2V(1, D_1, \dots, D_m)$.* □

In the case where some of the principal minors $D_1, \dots, D_m$ vanish, attempts were made to define the counting function V in such a way that Theorem 1.7.14 remains valid. They were successful provided that not more than two consecutive principal minors vanish. If three or more consecutive principal minors are allowed to be zero, then appropriately constructed examples show that the validity of Theorem 1.7.14 cannot be preserved. However, it was discovered by Frobenius that for certain classes of matrices, which cover some important types, it is possible to establish an analogue of Theorem 1.7.14 without any restriction on the minors. We shall present one of his results, but only to the extent needed for our applications, see (Gantmacher 1986, § 10.10) for details. Let us start with some preparations.

A real *Hankel* matrix of order n is of the form

$$S := \begin{pmatrix} s_1 & s_2 & \cdots & s_n \\ s_2 & s_3 & \cdots & s_{n+1} \\ \vdots & \vdots & & \vdots \\ s_n & s_{n+1} & \cdots & s_{2n-1} \end{pmatrix},$$

where $s_1, s_2, \dots, s_{2n-1}$ is a sequence of real numbers. We now extend the counting function V to vanishing intermediate elements.

Definition 1.7.15 *Let $a_0, \dots, a_n$ be a sequence of real numbers such that $a_0 a_n \ne 0$. For $\nu = 0, \dots, n$, define*

$$a_\nu^* := \begin{cases} a_\nu & \text{if } a_\nu \ne 0\,, \\ (-1)^{j(j-1)/2}\, a_{\nu-j} & \text{if } a_\nu = \cdots = a_{\nu-j+1} = 0,\ a_{\nu-j} \ne 0,\ j \in \mathbb{N}. \end{cases}$$

Then $V^*(a_0, \dots, a_n) := V(a_0^*, \dots, a_n^*)$, which is the number of variations of sign in the sequence of non-vanishing numbers $a_0^*, \dots, a_n^*$.

Note that this extension of V to vanishing elements is different from that given in Chapter 10 (see Definition 10.1.1).

Theorem 1.7.16 (Frobenius) *Let S be a real Hankel matrix of rank m. Denote by $D_1, \dots, D_m$ the principal minors of S, and suppose that $D_m \ne 0$. Then the quadratic form Q defined by S has signature $\sigma(Q) = m - 2V^*(1, D_1, \dots, D_m)$.* □

10

Number of zeros in an interval

10.1 THE BUDAN–FOURIER THEOREM AND DESCARTES' RULE

Let f be a polynomial of degree n with real coefficients. Expanding f about $c \in \mathbb{R}$, we obtain

$$f(x) = \sum_{\nu=0}^{n} \frac{f^{(\nu)}(c)}{\nu!}(x-c)^{\nu}. \tag{10.1.1}$$

Hence, if at $x = c$ all the polynomials

$$f(x),\ f'(x),\ f''(x), \ldots\ldots,\ f^{(n-1)}(x),\ f^{(n)}(x) \tag{10.1.2}$$

are of the same sign, then the right-hand side of (10.1.1) cannot vanish for $x > c$. In other words, for the polynomial f to have a zero in (c, ∞) it is necessary that, at $x = c$, there is at least one variation of sign in the sequence (10.1.2). Refinements of this simple observation have become a powerful tool for estimating the number of zeros in a given interval. Let us first be more precise about variations of sign.

In a sequence of non-vanishing real numbers $a_0, a_1, \ldots, a_n$, we count a variation of sign at the index k, where $1 \le k \le n$, if $a_{k-1}a_k < 0$.

Definition 10.1.1 *Let $a_0, a_1, \ldots, a_n$ be a sequence of real numbers. Then we define $V(a_0, a_1, \ldots, a_n)$ as the total number of variations of sign in the reduced sequence obtained by ignoring all vanishing elements. We agree that V assumes the value zero if only one or none of the elements is different from zero.*

According to this definition, we have $V(0,2,0,0,3,1) = V(2,3,1) = 0$ and $V(3,0,-2,0,0,0,-4,5,1) = V(3,-2,-4,5,1) = 2$.

Next, we list some useful properties of the counting function V.

Proposition 10.1.2 *For real numbers $a_0, a_1, \ldots, a_n$ the following holds.*

(i) *Let $\lambda_0, \ldots, \lambda_n$ be non-vanishing numbers which are all of the same sign. Then*

$$V(a_0, \ldots, a_n) = V(\lambda_0 a_0, \ldots, \lambda_n a_n).$$

(ii) *Suppose that $a_k \neq 0$ for some k, where $0 < k < n$. Then*

$$V(a_0, \ldots, a_n) = V(a_0, \ldots, a_k) + V(a_k, \ldots, a_n).$$

(iii) *Cancellation of an element does not increase the value of V. If $a_0 a_n \neq 0$ and an element a_k, where $0 < k < n$, is cancelled, then the number of variations of sign remains the same or decreases by 2.*

(iv) *Insertion of a real element does not decrease the value of V. If $a_0 a_n \neq 0$ and a real element is inserted somewhere between a_0 and a_n, then the number of variations of sign remains the same or increases by 2.*

(v) *Let $\lambda \geq 0$, and let $0 \leq k < n$. Then*

$$V(a_0, \ldots, a_k, a_{k+1}, \ldots, a_n) = V(a_0, \ldots, a_k, a_k + \lambda a_{k+1}, a_{k+1}, \ldots, a_n).$$

(vi) *We have*

$$V(a_0, a_1, \ldots, a_k, \ldots, a_n) + V(a_0, -a_1, \ldots, (-1)^k a_k, \ldots, (-1)^n a_n) \leq n \tag{10.1.3}$$

with equality if none of the elements $a_0, \ldots, a_n$ vanishes.

Proof A verification of these statements may be left to the reader with (vi) as a possible exception.

For this, consider the sequences

$$a_0, a_1, \ldots, a_k, \ldots, a_n \quad \text{and} \quad a_0, -a_1, \ldots, (-1)^k a_k, \ldots, (-1)^n a_n.$$

If none of the elements vanishes, then

$$\operatorname{sgn}(a_{k-1} a_k) = -\operatorname{sgn}\left((-1)^{k-1} a_{k-1} (-1)^k a_k\right) \neq 0 \qquad (k = 1, \ldots, n).$$

Hence, at each k exactly one of the two sequences has a variation of sign, and so (10.1.3) holds with equality. This also shows that, if in (10.1.3) we replace each vanishing element a_k by a real number $a \neq 0$, then the left-hand side becomes equal to n. But in this process the left-hand side cannot decrease, as a consequence of (iv). This completes the proof of (vi). □

Throughout this chapter, we shall use the following notation.

Definition 10.1.3 *Let f be a polynomial of degree n, and let I be any interval, finite or infinite, open, semi-open, or closed. Then $N_f I$ denotes the number of zeros of f in I, counted according to their multiplicities. Furthermore,*

$$V_f(x) := V\Big(f(x), f'(x), \ldots, f^{(n)}(x)\Big).$$

We now state the main result of this section.

Theorem 10.1.4 (Budan–Fourier) *Let f be a polynomial of degree n with real coefficients. Then, for any interval $(a, b]$,*

$$N_f(a, b] = V_f(a) - V_f(b) - 2\kappa$$

for some $\kappa \in \mathbb{N}_0$.

The theorem shows that $V_f(a) - V_f(b)$ is an upper bound for the number of zeros in $(a, b]$ and can exceed this number only by an even integer. If the upper bound is attained, that is, $\kappa = 0$, then we say that the Budan–Fourier count is *exact* for the polynomial f on the interval $(a, b]$. The proof will reveal the following criterion for exactness.

Supplement to Theorem 10.1.4. *The Budan–Fourier count is exact for f on $(a, b]$ if and only if the following holds. Whenever $f^{(m)}(\xi) = 0$ for $\xi \in (a, b]$ and ξ is not a zero of f of multiplicity greater than m, then $f^{(m-1)}(\xi) f^{(m+1)}(\xi) < 0$.*

Proof Consider the function $D : x \mapsto V_f(a) - V_f(x)$ on $(a, b]$. It is piecewise constant, and $D(x)$ can change its value only when x crosses one of the zeros of f or of its derivatives. In view of Proposition 10.1.2(ii), only the following two cases can contribute to an alteration in $D(x)$.

Case 1 Let $\xi \in [a, b]$ be a zero of f of multiplicity k. Then

$$f(\xi) = f'(\xi) = \cdots = f^{(k-1)}(\xi) = 0\,, \qquad f^{(k)}(\xi) \neq 0\,.$$

Therefore, the Taylor formula for $f^{(j)}$ expanded at ξ gives

$$f^{(j)}(\xi + h) = \frac{h^{k-j}}{(k-j)!}\left(f^{(k)}(\xi) + \frac{h}{k-j+1} f^{(k+1)}(\xi + \theta_j h)\right)$$
$$(\text{for some } \theta_j \in (0,1)\,,\ j = 0, \ldots, k)\,.$$

Hence, for sufficiently small $h > 0$ and $\sigma := \operatorname{sgn} f^{(k)}(\xi)$, the situation in a neighbourhood of ξ is as follows.

Table 10.1

x	$\operatorname{sgn} f(x)$	$\operatorname{sgn} f'(x)$	$\ldots$	$\operatorname{sgn} f^{(k-1)}(x)$	$\operatorname{sgn} f^{(k)}(x)$
$\xi - h$	$(-1)^k\sigma$	$(-1)^{k-1}\sigma$	$\ldots$	$-\sigma$	σ
ξ	0	0	$\ldots$	0	σ
$\xi + h$	σ	σ	$\ldots$	σ	σ

This shows that $V(f(x), f'(x), \ldots, f^{(k)}(x))$ decreases by k when x moves from $\xi - h$ to $\xi + h$.

Case 2 Let $\xi \in [a, b]$ be a zero of $f^{(m)}$ of multiplicity k, but suppose that it is not a zero of $f^{(m-1)}$. Then

$$f^{(m-1)}(\xi) \neq 0\,, \quad f^{(m)}(\xi) = \cdots = f^{(m+k-1)}(\xi) = 0\,, \quad f^{(m+k)}(\xi) \neq 0$$
$$(1 \leq m < m + k \leq n)\,.$$

Introducing $\sigma_1 := \operatorname{sgn} f^{(m-1)}(\xi)$ and $\sigma_2 := \operatorname{sgn} f^{(m+k)}(\xi)$, we find, with the help of Taylor's formula, that the situation in a neighbourhood of ξ is as described in Table 10.2.

Table 10.2

x	$\operatorname{sgn} f^{(m-1)}(x)$	$\operatorname{sgn} f^{(m)}(x)$	$\ldots$	$\operatorname{sgn} f^{(m+k-1)}(x)$	$\operatorname{sgn} f^{(m+k)}(x)$
$\xi - h$	σ_1	$(-1)^k\sigma_2$	$\ldots$	$-\sigma_2$	σ_2
ξ	σ_1	0	$\ldots$	0	σ_2
$\xi + h$	σ_1	σ_2	$\ldots$	σ_2	σ_2

This shows that, if x moves from $\xi - h$ to $\xi + h$, then

$$V\left(f^{(m-1)}(x), f^{(m)}(x), \ldots, f^{(m+k)}(x)\right)$$

decreases by k if k is even and by $k + \sigma_1\sigma_2$ if k is odd. In any case it decreases by a non-negative even integer.

Altogether, we see that Case 1 contributes to an increase in $D(x)$ by k if we pass through a zero ξ of f of order k, while Case 2 may contribute to another increase by an even number. Tables 10.1 and 10.2 also show that the function V_f is continuous on the right, that is, $\lim_{h\to 0+} V_f(\xi+h) = V_f(\xi)$, and so a possible zero at b has been counted by $D(b)$. This completes the proof of the theorem.

For the supplement, it is enough to observe that Case 2 does not contribute to an increase in $D(x)$ if and only if $k=1$ and $\sigma_1 = -\sigma_2$. □

Remark 10.1.5 We have already mentioned that the function V_f is continuous on the right. To describe its behaviour on the left, we denote by $V_f^+(\xi)$ the maximum number of variations of sign in the sequence $f(\xi), \dots, f^{(n)}(\xi)$, where each vanishing element is replaced either by $+1$ or by -1, whichever makes the count largest. Studying Tables 10.1 and 10.2, we find that

$$\lim_{h\to 0+} V_f(\xi-h) = V_f^+(\xi) \geq V_f(\xi) = \lim_{h\to 0+} V_f(\xi+h).$$

Remark 10.1.6 Variants of Theorem 10.1.4 for the number $N_f I$ of zeros in an interval I which is not of the form $(a,b]$, can be easily established. Using the notation introduced in Remark 10.1.5, we have

$$\begin{aligned} N_f(a,b) &= V_f(a) - V_f^+(b) - 2\kappa_1 && (\text{for some } \kappa_1 \in \mathbb{N}_0),\\ N_f[a,b] &= V_f^+(a) - V_f(b) - 2\kappa_2 && (\text{for some } \kappa_2 \in \mathbb{N}_0),\\ N_f[a,b) &= V_f^+(a) - V_f^+(b) - 2\kappa_3 && (\text{for some } \kappa_3 \in \mathbb{N}_0). \end{aligned}$$

A simple characterization of exactness is as follows.

Theorem 10.1.7 (Loewy–Curtiss) *For a polynomial f with real coefficients, the Budan–Fourier count is exact on any interval $(a,b] \subset \mathbb{R}$ if and only if all the zeros of f are real.*

Proof Let $f(x) = \sum_{\nu=0}^{n} a_\nu x^\nu$ be of degree n. Since, for all real x of sufficiently large modulus,

$$\operatorname{sgn} f^{(\nu)}(x) = \operatorname{sgn}\left(x^{n-\nu} a_n\right) \qquad (\nu = 0,1,\dots,n),$$

there exists an $\alpha < 0$ and a $\beta > 0$ such that

$$V_f(x) = n \quad \text{for } x \leq \alpha \qquad \text{and} \qquad V_f(x) = 0 \quad \text{for } x \geq \beta.$$

Now suppose that the Budan–Fourier count is exact on every interval. Then we find on $(\alpha,\beta]$ that

$$N_f(\alpha,\beta] = V_f(\alpha) - V_f(\beta) = n.$$

Hence f has all its zeros in $(\alpha,\beta]$. In particular, it has only real zeros.

Conversely, suppose that f has only real zeros, which lie in some interval $(\alpha_1,\beta_1]$. If $(a,b]$ is an arbitrary interval and

$$\alpha_2 := \min\{a-1,\alpha,\alpha_1\} \qquad \text{and} \qquad \beta_2 := \max\{b+1,\beta,\beta_1\},$$

then we have

$$n = N_f(\alpha_2,\beta_2] = N_f(\alpha_2,a] + N_f(a,b] + N_f(b,\beta_2]$$

together with

$$n = V_f(\alpha_2) - V_f(\beta_2) = \big(V_f(\alpha_2)-V_f(a)\big) + \big(V_f(a)-V_f(b)\big) + \big(V_f(b)-V_f(\beta_2)\big)\,.$$

Due to Theorem 10.1.4, this is only possible if $N_f(a,b] = V_f(a) - V_f(b)$. Hence the Budan–Fourier count is exact on $(a,b]$. □

In conjunction with the supplement to Theorem 10.1.4, we obtain the following interesting characterization of polynomials having only real zeros.

Corollary 10.1.8 *A polynomial f with real coefficients has only real zeros if and only if, for every $\xi \in \mathbb{R}$ which is a zero of a derivative $f^{(m)}$, but is not a zero of f of multiplicity greater than m, we have $f^{(m-1)}(\xi)f^{(m+1)}(\xi) < 0$.* □

For a practical application of the Budan–Fourier theorem, it is important to know an efficient way of evaluating the sequence (10.1.2) or a modification according to Proposition 10.1.2(i). The following algorithm, due to Shaw and Traub (1974), is superior to the familiar Horner scheme since it needs only $2n-1$ multiplications and $\frac{1}{2}n(n+1)$ additions. Of course, it can be extended so that it produces the value of $V_f(x)$ immediately.

Theorem 10.1.9 *Let $f(x) = \sum_{\nu=0}^{n} a_\nu x^\nu$ be a polynomial of degree n with real coefficients. Then, for real $x \neq 0$, the algorithm*

$c_n := a_n,\ x_1 := x,\ T_{-1,n-1} := a_0,\ T_{-1,n-2} := a_1 \cdot x_1,$
for $\nu = 2$ **to** n $(x_\nu := x \cdot x_{\nu-1},\ T_{-1,n-\nu-1} := a_\nu \cdot x_\nu)$
for $\mu = 0$ **to** $n-1$ $(T_{\mu,\mu} := T_{-1,-1},$
  for $\nu = \mu+1$ **to** n $(T_{\mu,\nu} := T_{\mu-1,\nu-1} + T_{\mu,\nu-1}),$
  $c_\mu := (\operatorname{sgn} x)^\mu T_{\mu,n})$

produces $V_f(x) = V(c_0, c_1, \ldots, c_n)$.

Proof The algorithm shows that

$$T_{-1,\nu} = a_{n-1-\nu}x^{n-1-\nu} \quad \text{and} \quad T_{\nu,\nu} = a_n x^n \qquad (\nu = -1, 0, \ldots, n-1)\,.$$

From this, it is easy to verify by induction on ν that

$$T_{\mu,\nu} = \sum_{j=\mu}^{\nu} \binom{j}{\mu} a_{n+j-\nu} x^{n+j-\nu} \qquad (\mu = 0, \ldots, \nu)\,.$$

Hence

$$T_{\mu,n} = \sum_{j=\mu}^{n} \binom{j}{\mu} a_j x^j = x^\mu \frac{f^{(\mu)}(x)}{\mu!}\,,$$

and so $c_\mu = |x|^\mu f^{(\mu)}(x)/\mu!$ for $\mu = 0, \ldots, n-1$. Further, $c_n = a_n = f^{(n)}(x)/n!$, by definition. Now the proof is completed by using Proposition 10.1.2(i). □

As a consequence of Theorem 10.1.4, we obtain the following important rule of signs.

Corollary 10.1.10 (Descartes' rule of signs) *Let $f(x) = \sum_{\nu=0}^{n} a_\nu x^\nu$ be a polynomial of degree n with real coefficients. Then*

$$N_f(0,\infty) = V(a_0, a_1, \ldots, a_n) - 2\kappa$$

for some $\kappa \in \mathbb{N}_0$.

Proof Note that $V_f(0) = V(a_0, a_1, \ldots, a_n)$ and $V_f(b) = 0$ for sufficiently large real b. Therefore, the conclusion follows from Theorem 10.1.4 by setting $a = 0$ and letting b tend to infinity. □

Remark 10.1.11 The number of negative zeros of f can be counted by applying Descartes' rule to the polynomial $g(x) := f(-x)$. This gives

$$N_f(-\infty, 0) = V\big(a_0, -a_1, \ldots, (-1)^n a_n\big) - 2\kappa \qquad (\text{for some } \kappa \in \mathbb{N}_0).$$

A sufficient condition for exactness of Descartes' rule is easily obtained from Theorem 10.1.7. The hypothesis that the polynomial has only real zeros is always satisfied for the characteristic polynomial of a Hermitian matrix, a case which often occurs in applications.

Corollary 10.1.12 *Let $f(x) = \sum_{\nu=0}^{n} a_\nu x^\nu$ be a polynomial of degree n with only real zeros. Then $N_f(0, \infty) = V(a_0, a_1, \ldots, a_n)$.* □

While Descartes' rule was deduced from the Budan–Fourier theorem, the former can in turn be used to deduce another estimate for the number of zeros in a finite interval.

Corollary 10.1.13 (Jacobi's rule of signs) *Let f be a polynomial of degree n with real coefficients. For real a and b, where $a < b$, define*

$$g(x) := (1+x)^n f\left(\frac{a+bx}{1+x}\right) =: \sum_{\nu=0}^{n} b_\nu x^\nu .$$

Then

$$N_f(a, b) = V(b_0, b_1, \ldots, b_n) - 2\kappa$$

for some $\kappa \in \mathbb{N}_0$.

Proof Obviously, ξ is a positive zero of g if and only if $(a + b\xi)/(1 + \xi)$ is a zero of f lying in (a, b). Hence the conclusion follows by applying Descartes' rule to g. □

Remark 10.1.14 Jacobi's rule of signs may be viewed in comparison with the Budan–Fourier theorem. The calculation of the sequence $b_0, \ldots, b_n$ is somewhat more costly than that of $c_0, \ldots, c_n$ in Theorem 10.1.9. However, Jacobi's rule never gives a worse bound. It was shown by Schoenberg (1934) that

$$V(b_0, b_1, \ldots, b_n) \le V_f(a) - V_f(b).$$

Moreover, if $f(a)f(b) \ne 0$, then the difference $(V_f(a) - V_f(b)) - V(b_0, \ldots, b_n)$ is a non-negative even integer.

10.2 EXACT COUNT UNDER A SIDE CONDITION

The phenomenon, that the bounds following from Theorem 10.1.4 and Corollaries 10.1.10 and 10.1.13 may overestimate the number of zeros in an interval by an even integer, has to do with the occurrence of pairs of conjugate zeros, as Theorem 10.1.7 shows. In this connection, the following questions arise naturally.

Definition 10.5.1 *A sequence $f_0, f_1, \ldots, f_m$ of polynomials with real coefficients is a* (generalized) Sturm sequence *on an interval I if it has the following properties:*

(i) *if $\xi \in I$ is a zero of f_0, then $f_1(\xi) \neq 0$;*

(ii) *if $\xi \in I$ is a zero of f_μ and $1 \leq \mu \leq m-1$, then $f_{\mu-1}(\xi)f_{\mu+1}(\xi) < 0$;*

(iii) *$f_m(x) \neq 0$ for all $x \in I$.*

We define $\xi \in I$ to be a relevant zero *of f_0 if $f_0(x)$ changes sign at $x = \xi$. Relative to f_1, we classify a relevant zero ξ as being of*

- *the* first kind *if $f_0(\xi + h)f_1(\xi) > 0$,*
- *the* second kind *if $f_0(\xi + h)f_1(\xi) < 0$,*

for all sufficiently small positive numbers h.

Example 10.5.2 Let f and g be polynomials with real coefficients. Then the Euclidean algorithm, started with $f_0 := f$ and $f_1 := g$, produces a sequence of polynomials $f_0, f_1, \ldots, f_m$ satisfying a recurrence formula

$$f_\mu(x) = q_\mu(x)f_{\mu+1}(x) - f_{\mu+2}(x) \qquad (\mu = 0, \ldots, m-1, \ f_{m+1} \equiv 0)$$

and terminating in the greatest common divisor f_m of f and g. In addition, the polynomials f_μ are all divisible by f_m (see § 1.1). This readily shows that $f_0, f_1, \ldots, f_m$ is a Sturm sequence on any interval on which f_m has no zero. Moreover, the sequence $\varphi_0, \varphi_1, \ldots, \varphi_m$ defined by

$$\varphi_\mu(x) := \frac{f_\mu(x)}{f_m(x)} \qquad (\mu = 0, \ldots, m)$$

is a Sturm sequence on any subinterval of the real line. In particular, if we choose $g = f'$, then φ_0 has only simple zeros and

$$\varphi_0(x)\varphi_1(x) = \frac{f(x)f'(x)}{\big(f_m(x)\big)^2}.$$

A discussion of the right-hand side shows that $\varphi_0(x)\varphi_1(x)$ changes its sign from minus to plus when x passes through a real zero of f. Hence the real zeros of φ_0 are all relevant zeros of the first kind.

Other standard examples are given by sequences $P_0, P_1, \ldots, P_n$ of orthogonal polynomials. In fact, it is easily seen from the recurrence formula that the polynomials $f_\nu := P_{n-\nu}$, where $\nu = 0, \ldots, n$, form a Sturm sequence.

The main result on Sturm sequences is as follows.

Theorem 10.5.3 (Sturm) *Let $f := f_0, f_1, \ldots, f_m$ be a Sturm sequence on an interval $[a, b]$. Suppose that f does not vanish at a and b. Denote by $N_f^+[a, b]$ the number of distinct relevant zeros of the first kind in $[a, b]$ and by $N_f^-[a, b]$ the corresponding number concerning the relevant zeros of the second kind. Then*

$$N_f^+[a, b] - N_f^-[a, b] = V\big(f_0(a), \ldots, f_m(a)\big) - V\big(f_0(b), \ldots, f_m(b)\big).$$

Proof It suffices to study how $V(f_0(x), \dots, f_m(x))$ behaves in a neighbourhood of a zero $\xi \in [a, b]$ of one of the polynomials $f_0, \dots, f_{m-1}$. We consider two cases.

Case 1 Let $\xi \in (a, b)$ be a zero of f_0. With $\sigma := \operatorname{sgn} f_1(\xi)$ the situation in a small neighbourhood of ξ is described in Table 10.3 if ξ is a relevant zero of the first or the second kind. We see that $V(f_0(x), f_1(x))$ decreases by one if x passes through a relevant zero of the first kind, but it increases by one if x passes through a relevant zero of the second kind. It is also clear that passing through a non-relevant zero ξ does not give a change of sign.

Table 10.3

Zero of the first kind			Zero of the second kind		
x	$\operatorname{sgn} f_0(x)$	$\operatorname{sgn} f_1(x)$	x	$\operatorname{sgn} f_0(x)$	$\operatorname{sgn} f_1(x)$
$\xi - h$	$-\sigma$	σ	$\xi - h$	σ	σ
ξ	0	σ	ξ	0	σ
$\xi + h$	σ	σ	$\xi + h$	$-\sigma$	σ

Case 2 Let $\xi \in [a, b]$ be a zero of f_μ, where $1 \le \mu \le m - 1$. By the condition (ii) of Definition 10.5.1, we have, for sufficiently small $h > 0$,

$$\operatorname{sgn} f_{\mu-1}(\xi \pm h) = -\operatorname{sgn} f_{\mu+1}(\xi \pm h).$$

With $\varepsilon := \operatorname{sgn} f_\mu(\xi - h)$ and $\delta := \operatorname{sgn} f_\mu(\xi + h)$ the situation in a neighbourhood of ξ is described in Table 10.4, where either the upper or the lower signs hold throughout. We see that $V(f_{\mu-1}(x), f_\mu(x), f_{\mu+1}(x))$ does not change its value on $[\xi - h, \xi + h]$ no matter how ε and δ may be.

Table 10.4

x	$\operatorname{sgn} f_{\mu-1}(x)$	$\operatorname{sgn} f_\mu(x)$	$\operatorname{sgn} f_{\mu+1}(x)$
$\xi - h$	$\pm$	ε	$\mp$
ξ	$\pm$	0	$\mp$
$\xi + h$	$\pm$	δ	$\mp$

In view of Proposition 10.1.2 (ii), we conclude from the Cases 1 and 2 that, if x moves from a to b, then only the relevant zeros of f contribute to the total change of $W(x) := V(f_0(x), \dots, f_m(x))$. Each zero of the first kind causes a decrease in the value of $W(x)$ by one, while each zero of the second kind causes an increase by one.

This completes the proof. □

A simple consequence of Theorem 10.5.3 is the following result.

Corollary 10.5.4 *Let f be a polynomial with real coefficients. Denote by $f_0, f_1, \dots, f_m$ the sequence of polynomials produced by the Euclidean algorithm started with $f_0 = f$ and $f_1 = f'$. Suppose that f does not vanish at a and b, where $a < b$. Then*

$$N_f^\circ[a, b] = V\big(f_0(a), \dots, f_m(a)\big) - V\big(f_0(b), \dots, f_m(b)\big),$$

where $N_f^\circ[a,b]$ denotes the number of distinct zeros of f in $[a,b]$.

Proof Set $g := f'$ and consider the Sturm sequence $\varphi_0, \varphi_1, \ldots, \varphi_m$ of Example 10.5.2. Applying Theorem 10.5.3, we obtain

$$N_f^\circ[a,b] = N_{\varphi_0}^+[a,b] = V\big(\varphi_0(a), \ldots, \varphi_m(a)\big) - V\big(\varphi_0(b), \ldots, \varphi_m(b)\big).$$

Since a and b are not zeros of f, we conclude that $f_m(a)f_m(b) \neq 0$. Hence, by Proposition 10.1.2 (i), we have

$$V\big(\varphi_0(x), \ldots, \varphi_m(x)\big) = V\big(f_0(x), \ldots, f_m(x)\big)$$

at $x = a$ and $x = b$.

This completes the proof. □

Remark 10.5.5 The proof shows that, if $f(a)f(b) \neq 0$, then, as far as variations of signs are concerned, the sequence $f_0, \ldots, f_m$ produced by the Euclidean algorithm may be used instead of the Sturm sequence $\varphi_0, \ldots, \varphi_m$, even if f_m has zeros in (a,b).

In Corollary 10.5.4, the zeros are not counted according to their multiplicities. The following modification of the Euclidean algorithm removes this deficiency and provides further information on the zeros of a polynomial. Temporarily, we shall denote by $c(f)$ the leading coefficient of a polynomial f, that is,

$$c(f) := \lim_{x \to \infty} \frac{f(x)}{x^n} \qquad (n = \deg f).$$

Algorithm Let f be a polynomial of positive degree. Define

$$f_0(x) := f(x), \qquad f_1(x) := f'(x)$$

and proceed for $\nu = 1, 2, \ldots$ as follows. If f_ν is not a constant, then perform the division transformation of $f_{\nu-1}$ and f_ν to obtain

$$f_{\nu-1}(x) = q_{\nu-1}(x) f_\nu(x) - r_\nu(x) \qquad (\deg r_\nu < \deg f_\nu), \tag{10.5.1}$$

and define

$$\begin{aligned} f_{\nu+1}(x) &:= r_\nu(x), \quad c_\nu := \frac{c(f_{\nu+1})}{c(f_{\nu-1})} &\quad \text{if } r_\nu(x) \not\equiv 0, \\ f_{\nu+1}(x) &:= f_\nu'(x), \quad c_\nu := 0 &\quad \text{if } r_\nu(x) \equiv 0. \end{aligned}$$

Terminate the algorithm when f_ν is a constant.

Theorem 10.5.6 *Let f be a polynomial with real coefficients. Denote by $f_0, f_1, \ldots, f_\ell$ the sequence of polynomials produced by the preceding algorithm. Suppose that f does not vanish at a and b, where $a < b$. Then*

$$N_f[a,b] = V\big(f_0(a), \ldots, f_\ell(a)\big) - V\big(f_0(b), \ldots, f_\ell(b)\big).$$

Proof Let $m_1 < \cdots < m_k$ be the indices for which $r_{m_j}(x) \equiv 0$, where $j = 1, \ldots, k$, and define $m_0 := 0$ and $m_{k+1} := \ell$. Note that f_{m_1} is the greatest common divisor of f and f'; likewise, f_{m_2} is the greatest common divisor of f_{m_1} and f'_{m_1}, and so on. Thus, $f_{m_j}(\xi) = 0$ if and only if ξ is a zero of f with a the beginning of this section, that the quadratic forms built from the matrices $S_m, S_{m+1}, \ldots$ are all equivalent to that associated with g/f. Moreover, S_m is non-singular while $\det S_\mu = 0$ for $\mu > m$.

For the subsequent considerations, we find it convenient to introduce a new term. Note that here the coefficients of the polynomials f and g are again numbered in reverse order as in (10.6.12).

Definition 10.6.8 *Let*

$$f(x) = \sum_{\nu=0}^{n} a_\nu x^{n-\nu} \quad (a_0 \neq 0) \qquad \text{and} \qquad g(x) = \sum_{\nu=0}^{n} b_\nu x^{n-\nu}$$

be polynomials with real coefficients, and define $a_\nu = 0$ and $b_\nu = 0$ for $\nu > n$. Then the determinants

$$\Delta_{2\ell} := \det \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{2\ell-1} \\ b_0 & b_1 & b_2 & \cdots & b_{2\ell-1} \\ 0 & a_0 & a_1 & \cdots & a_{2\ell-2} \\ 0 & b_0 & b_1 & \cdots & b_{2\ell-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & b_\ell \end{pmatrix} \qquad (\ell = 1, 2, \ldots)$$

are called the Hurwitz determinants *of f and g.*

Lemma 10.6.9 *Let f and g be polynomials with real coefficients such that $\deg g \leq \deg f = n$. Then the Hurwitz determinants $\Delta_{2\ell}$ of f and g, and the Hankel matrices S_ℓ of Lemma* 10.6.6 *are connected by the equations*

$$\Delta_{2\ell} = a_0^{2\ell} \det S_\ell \qquad (\ell = 1, 2, \ldots). \tag{10.6.14}$$

Proof Multiplying both sides of (10.6.13) by $f(z)$ and equating coefficients of equal powers of z, we find that

$$b_\mu = \sum_{\nu=0}^{\mu} a_\nu s_{\mu-\nu} \qquad (\mu = 0, \ldots, n),$$

$$0 = \sum_{\nu=0}^{n} a_\nu s_{\mu-\nu} \qquad (\mu = n+1, n+2, \ldots).$$

The first equation even holds for all $\mu \in \mathbb{N}_0$ since $a_\nu = b_\nu = 0$ for $\nu > n$, by definition. With these relations, we readily verify the following matrix equation:

$$\begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{2\ell-1} \\ b_0 & b_1 & b_2 & \cdots & b_{2\ell-1} \\ 0 & a_0 & a_1 & \cdots & a_{2\ell-2} \\ 0 & b_0 & b_1 & \cdots & b_{2\ell-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & b_\ell \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ s_0 & s_1 & s_2 & \cdots & s_{2\ell-1} \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & s_0 & s_1 & \cdots & s_{2\ell-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & s_\ell \end{pmatrix} \begin{pmatrix} a_0 & a_1 & a_2 & \cdots & a_{2\ell-1} \\ 0 & a_0 & a_1 & \cdots & a_{2\ell-2} \\ 0 & 0 & a_0 & \cdots & a_{2\ell-3} \\ 0 & 0 & 0 & \cdots & a_{2\ell-4} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & a_0 \end{pmatrix}.$$

In conjunction with (10.6.2) for $a = -\infty$ and $b = \infty$, Theorem 10.6.10 implies the following result.

Corollary 10.6.12 *Let f be a polynomial of degree n with real coefficients. Denote by $\Delta_{2\ell}$ the Hurwitz determinants of f and f', and let m be the largest integer in $\{1,\ldots,n\}$ such that $\Delta_{2m} \neq 0$. Then f has exactly m distinct zeros and*

$$m - 2V^*(1,\, \Delta_2,\, \Delta_4, \ldots,\, \Delta_{2m})$$

of them are real, where V^* is the counting function introduced in Definition 1.7.15. □

As a consequence of Corollary 10.6.12 and Remark 1.7.17, the following result holds.

Corollary 10.6.13 *Let f be a polynomial of degree n with real coefficients, and denote by $\Delta_{2\ell}$ the Hurwitz determinants of f and f'. Then f has n distinct zeros if and only if $\Delta_{2n} \neq 0$. Moreover, f has n distinct real zeros if and only if $\Delta_2 > 0,\, \Delta_4 > 0, \ldots,\, \Delta_{2n} > 0$.* □

Of course, by Lemma 10.6.9, we may as well replace the Hurwitz determinants $\Delta_{2\nu}$ by $\det S_\nu$, in these results. It may be worth mentioning that, for $g = f'$, each element s_ν of the Hankel matrix S_n is equal to $p_{\nu-1}$, which is $(\nu-1)$-th power sum of the zeros of f (see § 1.2).

By combining (10.6.2) and (10.6.15) with Theorem 10.6.10, we obtain an alternative version of Corollary 10.5.4.

Corollary 10.6.14 *Let f be a polynomial of degree n with real coefficients. For $\xi \in \mathbb{R}$, set $g(x) := (x-\xi)f'(x)$, and let $\Delta_{2\ell}(\xi)$ be the Hurwitz determinants of f and g. Let m be the largest integer in $\{1,\ldots,n\}$ such that $\Delta_{2m}(\xi) \neq 0$. Using the counting function V^* introduced in Definition* 1.7.15*, define*

$$W_f(\xi) \;:=\; V^*\big(1,\, \Delta_2(\xi),\, \Delta_4(\xi), \ldots,\, \Delta_{2m}(\xi)\big)\,,$$

and denote by $N_f^\circ(a,b]$ the number of distinct zeros of f in the interval $(a,b]$. If a and b are not zeros of f, then $N_f^\circ(a,b] = W_f(b) - W_f(a)$. □

Example 10.6.15 For the polynomials $f(x) = x^3 - x$ and $g(x) = (x-\xi)f'(x)$, we find:

$$\Delta_2(\xi) \;=\; -3\xi\,, \quad \Delta_4(\xi) \;=\; 6\xi^2 + 4\,, \quad \Delta_6(\xi) \;=\; 4\xi(1-\xi^2)\,.$$

Since we have to consider only the signs, we infer that

$$W_f(\xi) \;=\; V^*\big(1,\, -\xi,\, 1,\, \xi(1-\xi^2)\big) \;=\; \begin{cases} 0 & \text{if } \xi < -1\,, \\ 1 & \text{if } \xi \in (-1,0)\,, \\ 2 & \text{if } \xi \in (0,1)\,, \\ 3 & \text{if } \xi > 1\,. \end{cases}$$

Hence, if $a < b$ and $a,\, b \notin \{-1,\, 0\,, 1\}$, then $W_f(b) - W_f(a)$ is the exact number of zeros between a and b.