

THE RELATIVE DISTRIBUTION OF THE REAL ROOTS OF A SYSTEM OF POLYNOMIALS*

BY

C. F. GUMMER

1. INTRODUCTION

The problem about to be discussed may be looked upon as a generalization of the classical problem solved by Sturm in 1829 with regard to the real roots of a polynomial.† In Sturm's theorem it is shown how the number of distinct roots of a single polynomial which fall within a given real interval may be determined through a process rational in the coefficients. In the present paper we shall study a system of two or more polynomials in a single variable, and our aim will be to develop a rational process by which the order of succession of the roots of the several polynomials in a given real interval may be discovered.

Let us confine ourselves, at least for the present, to the case where the end-values of the interval are not roots and where the polynomials have only simple roots in the interval. It is clear that in such a case Sturm's theorem determines the only relations of the real roots of a single polynomial to the interval that remain invariant under continuous transformation of the real number system into itself. A similar remark applies to the theory about to be developed with respect to the real root system of several polynomials; so that we are undertaking the study of a problem which, from the point of view of a one-dimensional analysis situs, may be said to be the fundamental problem of the system under consideration.

The greater part of the paper will be devoted to the case of two polynomials. If the roots of the first are denoted generally by α and those of the second by β , and if we write down the roots within the interval considered in increasing numerical order (as $\alpha\alpha\alpha\beta\beta\alpha\beta\beta\alpha\beta$), the β 's effect a certain partition of the α 's (in the present case into groups of 3, 0, 1, 0, 0, 1, 0,). The solution of the problem will consist in the determination of the numbers of α 's in the successive groups.

* Presented to the Society, September 7, 1920.

† Bulletin de Férussac, vol. 11 (1829); Mémoires des Savants étrangers, vol. 6 (1835).

Previous writers have considered a less precise form of the problem, namely that in which α - and β -groups are counted only modulo 2. Cauchy* in 1837 or earlier examined the relation of the remainders in the highest common factor process for two polynomials to the changes in sign of these polynomials, thus obtaining essentially the solution of our problem modulo 2. Sylvester† in 1853 gave to these relations the interpretation in which we are here interested, in connection with the ordinal distribution of the roots. In Sylvester's paper the arrangement of α 's and β 's is called an "intercalation scale." The arrangement obtained by removing pairs of consecutive α 's or β 's as long as any such pairs are present is called an "effective intercalation scale." Thus, in the above example, the arrangement reduces, by cancellation of the α -pairs, to $\alpha\beta\beta\alpha\beta\beta\alpha\beta$; then, by removal of β 's, to $\alpha\alpha\beta\alpha\beta$; and finally the effective intercalation scale is $\beta\alpha\beta$. This reduced arrangement is connected by Sylvester with the signs that the two functions and their Sturmiian remainders take at the ends of the interval.‡ A generalization of the theory to the case of n functions in $(n - 1)$ variables, proposed by Sylvester,§ was carried out by Kronecker|| in 1869, in his theory of the "characteristic," which determined, also to modulus 2, the relative arrangement of the intersections of one closed curve in $(n - 1)$ -space with two hypersurfaces. Neither these nor other writers appear to have considered the problem of the complete root-distribution, whether in one or more variables.¶ It is the object of the present paper to obtain a solution for the one-dimensional case, to make some generalizations, and to discuss a number of special points in connection with the theory.

The solution of the problem for two polynomials will be found in § 3, an alternative method being given also, in § 13. In both methods use is made of the sequence of Sturmiian functions derived from one of the polynomials. In the one case these serve as material for a number of additional Sturmiian processes. In the alternative method they are used in building up a number of quadratic forms on whose signatures the solution depends.

* *Calcul des indices des fonctions*, Journal de l'Ecole polytechnique, vol. 25 (1837), pp. 176-229. Part of the theory was developed in a memoir presented to the Turin Academy in 1831.

† *A theory of the syzygetic 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 Transactions of the Royal Society of London, vol. 143, part 3 (1853), pp. 407-548; *Collected Mathematical Papers*, vol. 1, pp. 429-586.

‡ Ibid., Art. 48-50; (*Collected Mathematical Papers*, vol. 1, pp. 517-521).

§ Ibid., Art. 55; (*Collected Mathematical Papers*, vol. 1, pp. 527-8).

|| *Ueber Systeme von Functionen mehrerer Variabeln*, Berliner Monatsberichte, 1869, pp. 159-183, 688-697.

Ueber die Charakteristik von Functionen-Systemen, ibid., 1878, pp. 145-152.

¶ See also Cayley, *On the geometrical presentation of Cauchy's theorems of root-limitation*, Transactions of Cambridge Philosophical Society, vol. 12, part 2 (1877), pp. 395-413; *Collected Mathematical Papers*, vol. 9, pp. 21-39.

In § 7 will be found the theory for a system of more than two polynomials.

In § 8 the simple interval previously used will be replaced by the set of intervals defined by a system of algebraic inequalities; and in § 9 will be discussed a point arising out of this notion, whose chief interest consists in the connection that it makes with the restricted theory of Cauchy and Sylvester.

Since the reader may feel some disappointment at the laborious character of the computations demanded in numerical applications of the theory, it seems proper to include (in §§5, 10, and 12) a number of remarks tending to mitigate this condition. Even so, however, the main interest of the problem must remain theoretical; nor does it seem likely that any method could be devised capable of rapid application to numerical cases.

Among a variety of special problems to which § 11 will be devoted, we may notice in particular the determination of the number of separate intervals of the real number-system defined by a set of real algebraic inequalities. Two solutions will be given for this problem (in vii and viii).

2. THE CAUCHY-SYLVESTER THEORY

It will be useful to have before us the existing theory of the reduced root-distribution in a form that can be readily applied in obtaining a complete theory.

Let a and b be two real numbers, $a < b$. Let $f_0(x)$ be a polynomial with real coefficients, which does not vanish when $x = a$ or b , and which has in the interval (a, b) no multiple roots and no roots in common with another given real polynomial $\varphi(x)$. Let $f_1(x)$ be the first derivative of $f_0(x)$. Let those real roots of $f_0(x)$ which lie in (a, b) be $\alpha_1, \alpha_2, \dots$

Let $S(f_0, \varphi)$ denote the Sturmian sequence $[f_0(x), \varphi(x), \varphi'(x), \varphi''(x), \dots]$ obtained as follows: $-\varphi'(x)$ is the polynomial found as remainder when f_0 is divided by φ , $-\varphi''$ the remainder when φ is divided by φ' , and so on, the last function in the sequence being the highest common factor of f_0 and φ . Let $V(x)$ be the number of variations of sign between consecutive elements of the sequence $S(f_0, \varphi)$ for any given x .

The Cauchy-Sylvester theorem may then be stated in the following form: *the number of α 's making the quotient $\varphi(\alpha)/f_1(\alpha)$ positive exceeds the number of α 's making it negative by $V(a) - V(b)$.*

This excess is called by Cauchy the "index" of f_0 and φ in (a, b) .

It is easy to see that, with a suitable convention as to sign, the index enumerates the α 's in the reduced or effective intercalation scale, and it is in such a form that Sylvester obtains the theorem.* It is, however, a simple matter to

* For a physical interpretation of the effective intercalations in terms of two interlacing strings, see Sylvester, *Sur l'entrelacement d'une fonction par rapport à une autre*, *Journal für Mathematik*, vol. 88 (1880), pp. 1-3; *Collected Mathematical Papers*, vol. 3, pp. 449-450.

prove the rule in its present shape by studying the continuous variation of x , and by observing that (i) the last function of the sequence vanishes for no x in (a, b) , (ii) when any other function of the sequence, except $f_0(x)$, vanishes, the two adjacent functions have opposite signs, and (iii) when $x = \alpha + \delta$, $f_0(\alpha)$ being zero, $f_0(x)/f_1(x)$ has the same sign as δ for small enough values of $|\delta|$.

An immediate corollary of this theorem will be of use to us a little later. With the same assumptions, we may replace $\varphi(x)$ by the product of $\varphi(x)$ by $f_1(x)$. Hence the number of α 's making $\varphi(\alpha)$ positive exceeds the number making it negative by $U(a) - U(b)$, where $U(x)$ is the number of variations of sign in $S\{f_0(x), \varphi(x) \cdot f_1(x)\}$.

3. COMPLETE THEORY FOR TWO POLYNOMIALS

Let a and b be real numbers of which $a < b$. Let $f_0(x)$ and $g_0(x)$ be given polynomials with real coefficients, and $f_1(x)$, $g_1(x)$ their first derivatives. Let us denote the sequence of Sturmian functions $S(g_0, g_1)$ by $g_0(x)$, $g_1(x)$, $g_2(x)$, \dots , $g_k(x)$.* Let us assume that (i) neither a nor b is a root of $f_0(x)$, (ii) $f_0(x)$ has no multiple root in (a, b) , (iii) $f_0(x)$ has no root in (a, b) in common with any one of $g_0(x)$, $g_1(x)$, \dots , $g_k(x)$. Suppose the roots of $f_0(x)$ in (a, b) to be $\alpha_1, \alpha_2, \dots$, and suppose the distinct roots of $g_0(x)$ in the interval $a < x \leq b$ to be β_1, β_2, \dots .† Our object is to show how the partition of the α 's by the β 's may be determined.

We naturally inquire how many of the α 's give to the sequence $g_0(\alpha)$, $g_1(\alpha)$, \dots , $g_k(\alpha)$ some particular number r of variations of sign. For, if we knew this variation number r for a certain α , we should be able to identify, by using Sturm's theorem, the β -interval in which α was contained. Let N_r be the number of α 's giving to g_0, g_1, \dots, g_k exactly r variations. The problem is then to determine N_r .

It will be convenient to denote the product $g_{i-1}(x) \cdot g_i(x)$ by $G_i(x)$, and the product $G_{i_1}(x) \cdot G_{i_2}(x) \cdot \dots \cdot G_{i_s}(x)$ by $G_{i_1 i_2 \dots i_s}(x)$, or briefly by $G_{(i)}(x)$. A product of g 's may in like manner be written $g_{(i)}(x)$; but, for the most part, it will be convenient to write formulas in terms of the G 's. It will be understood that $G(x)$ or $g(x)$, without suffix, represents unity.

There will be a permanence or a variation between $g_{i-1}(\alpha)$ and $g_i(\alpha)$ according as $G_i(\alpha)$ is positive or negative (by assumption (iii) it cannot be zero). Therefore N_r is the number of α 's making exactly r of the quantities $G_1(\alpha), \dots, G_k(\alpha)$ negative.

Let us denote by $L_{i_1 i_2 \dots i_s}$ or $L_{(i)}$, the number of α 's making the particular set

* Or let g_0, g_1, \dots, g_k be any general Sturmian sequence, however obtained, in the sense that its loss of variations agrees with that of the sequence defined above.

† The reader may prefer to assume throughout that g_0 has no multiple roots in the interval, and that neither a nor b is a root of g_0 .

of functions G_{i_1}, G_{i_2}, \dots negative, and the remaining G 's positive. Then it is evident that N_r is equal to the sum of those L 's that possess r suffixes; that is to say, $N_0 = L, N_1 = \sum_i L_i, \dots, N_k = L_{12} \dots k$.

In order to find the L 's, let us connect them with another set of integers which it will be possible to determine directly. Let $M_{i_1 i_2 \dots}$, or $M_{(i)}$, be the excess of the number of α 's making the product $G_{i_1 i_2} \dots$ positive over the number making it negative. In particular, M (without suffix) is the number of roots of $f_0(x)$ in the interval. It will also be understood that $m_{(i)}$ has the same meaning with reference to $g_{(i)}$.

It is clear that a particular α contributes $+1$ or -1 to $M_{i_1 i_2 \dots}$ according as it gives negative values to an even or an odd number of the functions G_{i_1}, G_{i_2}, \dots whose suffixes appear in the symbol $M_{i_1 i_2 \dots}$. If then we denote by $\{(i), (j)\}$ the number of elements common to the two sets of integers (i) and (j) , we find that the 2^k integers $M_{(j)}$ are connected with the 2^k non-negative integers $L_{(i)}$ by means of the system of 2^k equations

$$(1) \quad \sum_{(i)} (-1)^{\{(i), (j)\}} L_{(i)} = M_{(j)},$$

where the summation covers all combinations of i 's (including the null-set) and there is one such equation for each combination of j 's.

To solve (1) for the L 's we notice that, whenever (i) and (i') are not the same set,

$$\sum_{(j)} (-1)^{\{(i), (j)\}} (-1)^{\{(i'), (j)\}} = 0.$$

For, if e is some element present in one but not both of (i) and (i') , the sum on the left side is composed of pairs of equal and opposite terms differing only in the presence or absence of e in (j) .

Hence, on multiplying (1) by $(-1)^{\{(i'), (j)\}}$ and adding, we have, after writing (i) for (i') ,

$$(2) \quad L_{(i)} = 2^{-k} \sum_{(j)} (-1)^{\{(i), (j)\}} M_{(j)}.$$

The problem is now reduced to the determination of the M 's; and for this we have the method of §2 available. Denoting by $H_{(j)}(x)$ the product $G_{(j)}(x) \dots f_1(x)$, let us form by the Sturmian division process the sequence $S[f_0, H_{(j)}] = [f_0, H_{(j)}, H'_{(j)}, H''_{(j)}, \dots]$. If $V_{(j)}(x)$ represents the number of variations of sign in the sequence for any given x , the required formula is, by the corollary in §2,

$$(3) \quad M_{(j)} = V_{(j)}(a) - V_{(j)}(b).$$

An alternative process for finding $M_{(j)}$ will be given in §13.

We now have a complete chain of relations leading to the values of N_r for $r = 0, 1, \dots, k$. These relations contain the solution of the problem for two polynomials as far as the relative distribution of the α 's and the intervening β 's are concerned. They do not disclose the number of outlying β 's at either end of the interval. These numbers, however, are easily found by observing the values of g_0, g_1, \dots when $x = a$ and b . If we introduce the fresh assumption (iv) that neither a nor b is a multiple root of g_0 , and if the sequence g_0, g_1, \dots, g_k shows κ variations for $x = a$ and λ variations for $x = b$, while κ' and λ' are the highest and lowest values of r for which $N_r \neq 0$, it is clear that the numbers of marginal β 's are $\kappa - \kappa'$ and $\lambda' - \lambda$, on the understanding that the interval includes b but not a .

4. THE DISTRIBUTION FUNCTION

We are interested, not in the individual numbers $L_{(i)}$, but in the sum of those among them that involve r suffixes, since this sum is equal to N_r . It will be seen that the N 's are not capable of simple expression in terms of the M 's. For some purposes the most convenient way of exhibiting the relation between the N 's and M 's is by means of a generating function

$$E(t) = \sum_{r=0}^k N_r t^r.$$

On writing N_r in terms of the L 's and then eliminating the latter by means of (2), we find that

$$(4) \quad E(t) = 2^{-k} \sum_{(j)} \sum_{(i)} (-1)^{(i), (j)} M_{(j)} t^r,$$

where, in any term of the sum, r is the number of elements in (i) .

Now it is not difficult to see that, for any particular set (j) , s in number,

$$(5) \quad \sum_{(i)} (-1)^{(i), (j)} t^r = (1+t)^{k-s} (1-t)^s.$$

For the linear factors of the right member may be arranged in such order that the j th factor is $(1-t)$ for every j belonging to (j) . The general term of the product (if like terms are not collected) is then the positive or negative product of a number of t 's drawn from the i_1 th, i_2 th, \dots factors, the number of negative factors in this term being $\{(i), (j)\}$. The complete product contains one such term for each possible combination (i) , which shows that (5) is correct.

By means of (5), (4) may be reduced to the form

$$E(t) = 2^{-k} \sum_{(j)} M_{(j)} (1+t)^{k-s} (1-t)^s,$$

s being the number of elements in (j) .

If P_s stands for the sum of those M 's having s suffixes, we have finally the identity in t

$$(6) \quad \sum_{r=0}^k N_r t^r = E(t) = 2^{-k} \sum_{s=0}^k P_s (1+t)^{k-s} (1-t)^s.$$

The P 's being known, the expansion of the right member in descending powers of t will give, by means of its coefficients, the numbers of α 's in the successive β -intervals from left to right. We may call $E(t)$ the "distribution function."

Since the P 's are formed from the M 's in the same way as the N 's from the L 's, a reference to equations (1) and (2) will convince us that, corresponding to (6), there is another identity (in u)

$$(7) \quad \sum_{s=0}^k P_s u^s = \sum_{r=0}^k N_r (1+u)^{k-r} (1-u)^r.$$

In fact (7) may be obtained from (6) by the transformation $t = (1-u)/(1+u)$.

To illustrate the occasional advantage of the reciprocal formula (7) over (6), let us write the conditions that the root-distribution should be $\beta\beta\alpha\alpha\beta$, taking $k = 3$. It is required that $\sum_s P_s u^s \equiv 2(1+u)^2 (1-u)$, so that the conditions are $P_0 = 2, P_1 = 2, P_2 = -2, P_3 = -2$. If a pair of polynomials does not satisfy these conditions, the fact is likely to be discovered without computing all the M 's. On the other hand, the use of the t -identity (6) gives four conditions each of which involves all the eight M 's.

5. REMARKS ON THE PRECEDING THEORY

(i) The relations expressed in terms of the G 's in equations (1), (2), (3), and (6) may also be put in terms of the g 's. Equations (1), (2), and (3) will merely exchange the capitals for small letters (to be understood in a corresponding sense), except that the factor 2^{-k} will be replaced by 2^{-k-1} ; but (6) will go into a less simple form, and it is therefore convenient to keep the formulas in terms of the capital letters. In numerical work, however, the calculations will actually be made, as the following remark will show, with functions of lower degree than would appear from the general formulas.

(ii) In computing $M_{(j)}$, we have evidently the right to substitute for $G_{(j)}$ or $H_{(j)}$ any polynomial assuming the same signs for all the α 's. For example, $M_{123} = m_{03}$; hence H_{123} may be replaced by $h_{03} = g_0 g_3 f_1$.

We may also subtract from any $G_{(j)}$ or $H_{(j)}$ the product of any polynomial by f_0 , and the residue so obtained may be used in place of the original function, not only for the calculation of $M_{(j)}$, but also in the formation of higher products.

(iii) When some $M_{(i)}$ is equal to $\pm M$, every $M_{(j)}$ calculated will at the same time furnish the value of some other. For, if (i') and (i'') together make up (i) , and if (j) is any set not overlapping (i) , we know by (ii) that $M_{(i', j)} = \pm M_{(i'', j)}$, the ambiguity being in agreement with the former, where (i, j) means the set of elements belonging to either (i) or (j) .

In particular, if it is known that $M_i = \pm M$ for every i in some set (e) , it is no longer necessary to calculate independently any $M_{(j)}$ whose set of suffixes has elements in common with (e) .

These ideas will be generalized in §10.

(iv) In certain cases, as when g_0 is known to have all its roots real and distinct, g_0 and its successive derivatives will serve as the sequence (g_0, g_1, \dots, g_k) . In any case the Budan-Fourier theorem shows that the root-distribution indicated by the use of the derivatives will differ from the true distribution at most by the presence of pairs of consecutive β 's. It will be exact if no such pairs are indicated.

(v) Failure of any of the hypotheses (i), (ii), or (iii) of §3 will always lead to reducibility of f_0 , if the degree of f_0 is not less than that of g_0 .

6. NUMERICAL EXAMPLES

In the following examples, substitutions in accordance with §5 (ii) are freely made without a corresponding change of notation.

(i) Let $f_0(x) = x^3 + 3x^2 - 3x - 1$, $g_0(x) = x^2 - 4x - 7$.

Then $f_1 = x^2 + 2x - 1$; $g_1 = x - 2$, $g_2 = 1$. Therefore $G_1 = -9x^2 + 4x + 15$, $G_2 = x - 2$, $G_{12} = g_0 = x^2 - 4x - 7$. Hence

$$\begin{array}{llll} H = x^2 + 2x - 1, & H_1 = -17x^2 + 28x - 1, & H_2 = -3x^2 - 2x + 3, & H_{12} = x^2 - 12x + 1, \\ H' = x, & H'_1 = -83x + 23, & H'_2 = 8x - 3, & H'_{12} = -11x + 1 \\ H'' = 1, & H''_1 = -1, & H''_2 = -1, & H''_{12} = 1, \end{array}$$

$$\begin{array}{llll} V(-\infty) = 3, & V_1(-\infty) = 2, & V_2(-\infty) = 0, & V_{12}(-\infty) = 1, \\ V(0) = 1, & V_1(0) = 2, & V_2(0) = 2, & V_{12}(0) = 1, \\ V(\infty) = 0, & V_1(\infty) = 1, & V_2(\infty) = 3, & V_{12}(\infty) = 2. \end{array}$$

If the interval is $(-\infty, \infty)$,

$$M = 3, \quad M_1 = 1, \quad M_2 = -3, \quad M_{12} = -1.$$

Hence $E(t) = \frac{1}{4}\{3(1+t)^2 + (1+t)(1-t) - 3(1+t)(1-t) - (1-t)^2\} = t^2 + 2t$. Also $\kappa = 2$ and $\lambda = 0$; so that the arrangement is $\alpha\beta\alpha\alpha\beta$.

If the interval is $(-\infty, 0)$,

$$M = 2, \quad M_1 = 0, \quad M_2 = -2, \quad M_{12} = 0.$$

Therefore $E(t) = \frac{1}{4}\{2(1+t)^2 - 2(1-t)^2\} = t^2 + t$. Moreover, λ is now 1, showing that there are no β 's on the upper side, and the arrangement is $\alpha\beta\alpha$.

In both cases M_{12} might have been found as the negative of M_1 , as explained in §5 (iii).

$$(ii) \text{ Let } f_0 = x^2 - 3x - 1, \quad g_0 = x^3 - 3x^2 - 3x - 1; \quad a = -\infty, \quad b = \infty.$$

We find that $M = 2, M_1 = 0, M_2 = 0, M_{12} = -2, M_3 = -2$; and it follows that $M_{13} = -M_1 = 0, M_{23} = -M_2 = 0, M_{123} = -M_3 = 2$.

Hence $E(t) = \frac{1}{8}\{2(1+t)^3 - 2(1+t)^2(1-t) - 2(1+t)(1-t)^2 + 2(1-t)^3\} = 2t^2$. Since $\kappa = 2$ and $\lambda = 1$, the order of roots is $\alpha\alpha\beta$.

7. MORE THAN TWO POLYNOMIALS

It is always possible to find the relative arrangement of roots for a system of polynomials by applying the method of §§3 and 4 to the several pairs. It may, however, be interesting to exhibit the solution of the problem by means of a single formula analogous to (6).

Let one of the polynomials be $f_0(x)$, having roots $\alpha_1, \alpha_2, \dots$ in (a, b) ; and let the other polynomials (in finite number) be g'_0, g''_0, \dots . Let all the preceding notation be applied, with accents wherever necessary; and further let the product $G'_{(i')}G''_{(i'')} \dots$ be abbreviated to $G_{[i]}$.*

It will be assumed that neither a nor b is a root of f_0 , and that f_0 has in (a, b) no multiple root or root in common with any of the Sturmian functions $g'_0, g'_1, g'_2, \dots, g'_{k'}, g''_0, \dots, \dots$.

Let $N_{r', r''} \dots$ or $N_{[r]}$ be the number of α 's lying in that β -interval in which the sequence $g'_0, \dots, g'_{k'}$ has r' variations, $g''_0, \dots, g''_{k''}$ has r'' variations, and so on. By reasoning similar to that of §§3 and 4, we find that $N_{[r]}$ is the coefficient of $t^{r'}t^{r''} \dots$ in the generating function

$$(8) \quad E(t', t'', \dots) = 2^{-k'-k''-\dots} \sum_{[j]} M_{[j]} (1+t')^{k'-s'} (1-t')^{s'} (1+t'')^{k''-s''} (1-t'')^{s''} \dots,$$

where s' is the number of elements in (j') , s'' the number in (j'') , etc.

It is evident that the remarks made in §5 apply without essential modification to the present case.

* It will be noted that $[]$ is the symbol of a permutation, $()$ of a combination.

8. GENERALIZATION OF THE RANGE

Hitherto we have taken account of all the roots of f_0 that lie in a given real interval (a, b) . Let us now consider how formula (8) or (6) must be modified when only those roots are counted which, besides lying in (a, b) , satisfy a given system of algebraic inequalities.

Let us have then, in addition to the functions already assumed, a system of n auxiliary polynomials $G^1(x)$, $G^2(x)$, \dots , $G^n(x)$ without roots in common with f_0 , and let the inequalities in question be

$$\begin{aligned} G^i &< 0 \text{ for every } i \text{ in a certain set } (e); \\ G^i &> 0 \text{ for the remaining } i\text{'s}. \end{aligned}$$

Let the α 's satisfying these inequalities be denoted generally by $\bar{\alpha}$.

Let $G_{[G]}^{(i)}$ denote the product of all the G 's whose indices (upper or lower) appear in the symbol, and let us define in connection with this product the corresponding quantities of types M , H , and V . Our object being to find $\bar{N}_{[r]}$, the number of $\bar{\alpha}$'s in a particular β -interval, we shall study the generating function $\bar{E}(t', t'', \dots) = \sum_{[r]} \bar{N}_{[r]} t'^r t''^r \dots$.

Let us denote by $E^{(i)}(t', t'', \dots)$ the function

$$2^{-k'-k''-\dots} \sum_{[G]} M_{[G]}^{(i)} (1+t')^{k'-s'} (1-t')^{s'} \dots,$$

obtained from $E(t', t'', \dots)$ by affixing the set of upper indices (i) to every M , observing that $E^{(i)}$ is not in general the generating function of a distribution and that some of its coefficients may be negative. It will be called a "modified distribution function." There is clearly a set of relations, analogous with (1),

$$(9) \quad \sum_{(e)} (-1)^{[(e), (i)]} \bar{E} = E^{(i)},$$

where the summation is for all combinations (e) out of $1, 2, \dots, n$, and where the \bar{E} in any term is formed according to the inequalities associated with the (e) in question. From these we obtain, as (2) was obtained,

$$(10) \quad \bar{E} = 2^{-n} \sum_{(i)} (-1)^{[(e), (i)]} E^{(i)}.$$

The M 's appearing in $E^{(i)}$ being given by the formula

$$(11) \quad M_{[G]}^{(i)} = V_{[G]}^{(i)}(a) - V_{[G]}^{(i)}(b),$$

the function \bar{E} is determined.

In practice we can of course arrange to have all inequalities of the form $G^i > 0$, when (10) reduces to $\bar{E} = 2^{-n} \sum_{(i)} E^{(i)}$.

There are two ways in which, starting from the range $(-\infty, \infty)$, we may arrive at the interval (a, b) by the use of (10). (i) We may use the two inequalities $G^1 = x - a > 0$, $G^2 = x - b < 0$, obtaining $\bar{E} = \frac{1}{4}(E + E^1 - E^2 - E^{12})$. (ii) We may use the single inequality $G^1 = (x - a)(x - b) < 0$, obtaining $\bar{E} = \frac{1}{2}(E - E^1)$. Of these, (ii) still gives correct results when $a \geq b$, if we agree that (a, b) and (b, a) are the same interval; while (i) is correct in this case if by (a, b) we mean the interval $a < x < b$. In the method of §3 it is necessary that $a \leq b$.

9. THE MODIFIED DISTRIBUTION FUNCTION

It is interesting to examine a little the meaning of the modified function $E^{(i)}$, at least in a simple case. Without being a true distribution function, it has some relation to the distribution, as equation (9) shows.

Let us consider the case of two polynomials f_0 and g_0 with a single auxiliary function G^1 , and denote now by E^+ the generating function of the root-distribution in that part of (a, b) where G^1 is positive, and by E^- the same for G^1 negative. With E^1 as already defined (as a special case of $E^{(i)}$), we have, by (9), $E = E^+ + E^-$ and $E^1 = E^+ - E^-$; so that $E - 2E^- = E^1 = -E + 2E^+$.

It follows that the general coefficient N_r^1 in E^1 does not exceed in absolute value the corresponding N_r , and differs from it by an even number. In particular, $|N_r^1|$ will correctly enumerate the α 's in the β -interval whenever either N_r^+ or N_r^- is zero; that is, in any β -interval where the α 's are not separated by a root of G^1 .

The information given by E^1 is complete when $\sum_r |N_r^1| = M$. In other cases the total error in the enumeration of the α -groups by E^1 is $M - \sum_r |N_r^1|$.

Two special cases may be noted.

(i) If $G^1 = g_0$, all the α -groups will remain undivided, and E^1 will give exact information. In fact, for this case, if g_0 has no roots in (a, b) of even multiplicity, $E^1(t) = \pm E(-t)$.

(ii) On taking $G^1 = g_0 f_1$, we find that the function $H_{(j)}$ is replaced by $g_0 G_{(j)}$. Every N_r^1 is now 0 or ± 1 , according as N_r is even or odd. The α -groups have therefore been reduced modulo 2. To arrive at the Cauchy-Sylvester theory, we must reduce both α - and β -groups, which is accomplished by making $t = 1$. In fact, $E^1(1)$ is equal to M^1 , which is the loss of variations in the sequence $S(f_0, g_0)$; in other words, the Cauchy index.

10. THE M -INEQUALITIES

Although the $M_{(j)}$ are independent quantities, in the sense that they are connected by no identical equation, they satisfy a number of linear inequalities, which, in many cases, enable us to dispense with a part of the work in computing them. These inequalities express the fact that the $L_{(i)}$ are essentially ≥ 0 , so that they take the form

$$(12) \quad \sum_{(j)} (-1)^{\{(i), (j)\}} M_{(j)} \geq 0.$$

On account of these relations, a single equation in the M 's may be equivalent to a number of equations. For example, the equation $\sum_s P_s(k - 2s) = 0$, being the same as $\sum_i L_i = 0$, breaks up into the k equations $L_i = 0$ ($i = 1, \dots, k$), which again may be written in terms of the M 's.

For the purpose of direct numerical application of the theory, the most important consequences of (12) are contained in the following theorem.

If, for some set (e) contained in a second given set (d) ,

$$(A) \quad \sum_{(i) \text{ in } (d)} (-1)^{\{(i), (e)\}} M_{(i)} = 0,$$

it will follow that, for every (j) without elements in common with (d) ,

$$(B) \quad \sum_{(i) \text{ in } (d)} (-1)^{\{(i), (e)\}} M_{(i, j)} = 0,$$

where (i, j) is the set whose elements are those of (i) and those of (j) . The summation is for all subsets (i) of (d) . The proof is that (A) expresses the vanishing of the sum of those $L_{(e, r)}$ for which (r) has no elements in common with (d) . These $L_{(e, r)}$ must separately vanish. Hence

$$\sum_{(r)} (-1)^{\{(r), (j)\}} L_{(e, r)} = 0,$$

which reduces to (B).

It is natural to compute the M 's in some such order as $M, M_i, M_j, M_{ij}, M_p, M_{ip}, M_{jp}, M_{ijp}, \dots$. If some $M_{(d)}$ is found, with preceding M 's, to satisfy a relation of type (A), all the M 's whose suffixes include (d) are given by equations of type (B) in terms of M 's already found. The smaller the set (d) , the more numerous are the consequences. When (d) has only one element, we get the rule of §5 (iii) with respect to the condition $M_i = \pm M$, and the values of one half of the M 's are given in terms of the others. If there are two elements in (d) , (A) will have one of three possible forms; for instance, if (d) is 1, 3, and (e) is 3, (A) becomes $M + M_1 - M_3 - M_{13} = 0$, and (B) is

of the form $M_{(p)} + M_{1(p)} - M_{3(p)} - M_{13(p)} = 0$, which gives a quarter of the M 's, namely those of the form $M_{13(p)}$.

This theorem remains true when (j) is any set, and also when the individual suffixes are replaced by sets of suffixes, provided we understand by (i, j) the set of all elements occurring in one but not both of (i) and (j) . Thus, from $M + M_{12} + M_{13} + M_{23} = 0$ we deduce $M_{14} + M_{24} + M_{34} + M_{1234} = 0$.

A reason for the frequent intervention of this principle in shortening numerical work is in the limited choice of values for $M_{(d)}$, which narrows as the number of suffixes increases. Since $\sum_{(i) \text{ in } (d)} M_{(i)}$ must be divisible by 2^q , where q is the number of elements in (d) , $M_{(d)}$ is known beforehand modulo 2^q , and this alone will enable us to foretell the value of $M_{(d)}$ if $2^q > 2M$.

There is no difficulty in extending the principle of this section to the more general methods of §§7 and 8.

11. APPLICATION TO SPECIAL PROBLEMS

In the following applications (i) to (vi) the reference is to two polynomials in a single interval (finite or infinite).

(i) The question whether all the α 's are to be found in a particular set of β -intervals reduces to the question of the linear dependence of $\sum_s P_s u^s$ on $(1 + u)^{k-r_1} (1 - u)^{r_1}, (1 + u)^{k-r_2} (1 - u)^{r_2}, \dots$; but in particular cases, the conditions may take various forms through the principle discussed in §10.

Thus (under assumption (iv) of §3) the β 's will be entirely unseparated by the α 's if and only if the sum vanishes of all L 's with more than λ and less than κ suffixes, which reduces to a single equation in the M 's.

In particular, if all the roots of g_0 are simple and interior to (a, b) , so that $\kappa = k$ and $\lambda = 0$, a necessary and sufficient condition that the α 's do not separate them is that $M - L - L_{12} \dots_k = 0$; which reduces, first to $\sum_{(i) \text{ even}} [M - M_{(i)}] = 0$, and finally to the condition that all $M_{(i)}$ with even numbers of suffixes are equal. By §5 (iii), it is necessary (though not sufficient) that the $M_{(i)}$ with odd numbers of suffixes are also equal. When the roots are so arranged, the manner of partition is completely determined by $N_0 = \frac{1}{2}(M + M_1)$, $N_k = \frac{1}{2}(M - M_1)$.

(ii) A condition for k β 's, each less than any one of the α 's, is $M_i = M$ for every i .

For k β 's, each greater than any α , $M_i = -M$ for every i .*

(iii) For the α 's to occur in alternate β -intervals only, $P_s = \epsilon P_{k-s}$ for every s , where ϵ is either 1 or -1 ; which, by §5 (iii), is equivalent to the one con-

* See also the note to (vi).

dition $M = \pm m_{0k}$. This condition may also be obtained from elementary considerations.

(iv) The partition of the α 's is ordinarily symmetric about the position where the g_i 's have $k/2$ variations if and only if $P_r = 0$ for every odd r .

(v) To have an arrangement consisting of $(n+1)$ groups of r α 's each, alternating with n groups of s β 's each, where $ns = k$, it is necessary and sufficient that

$$r[(1+u)^{ns} + (1+u)^{(n-1)s}(1-u)^s + \cdots + (1-u)^{ns}] \equiv P_0 + P_1u + \cdots + P_ku^k.$$

It is therefore necessary that $P_i = 0$ for every odd i .

If s is even, the equations of condition reduce to one more than half their original number, owing to the inclusion of the equation $M = M_{12} \dots k$, and involve only one more than half the M 's. For instance, (taking $n = r = s = 2$) the roots of a sextic and a quartic occur alternately in pairs whenever $M = M_{1234} = 6$, $M_{12} + M_{13} + M_{23} = 10$, and $M_1 + M_2 + M_3 + M_4 = 0$.

When $r = s = 1$, the conditions are $P_{2i} = \binom{k+1}{2i+1}$, $P_{2i+1} = 0$; but the following method is simpler.

(vi) Let g_0 have no multiple roots in (a, b) . The modified process with the auxiliary function $G^1 = g_0f_1$ shows that a necessary and sufficient condition for k (or $k+1$) α 's alternating with k β 's is $M = k$ (or $k+1$), $M^1 = \pm M$. These are Sylvester's results, and represent the case where his restricted theory gives complete information. The tests involve very little calculation. It is easy to see that k may here be interpreted as the number of β 's*.

(vii) The theory of the generalized range in §8 enables us to find rationally the number of separate intervals of the real number system satisfying a set of algebraic inequalities

$$(A) \quad G^i > 0 \quad (i = 1, 2, \dots, n),$$

the G^i being real polynomials in x without common or multiple roots in the range $(-\infty, \infty)$.

To complete the system of §8, let f_0 denote the first derivative of the product $G^1G^2 \dots G^n$, and suppose that f_0 has no real multiple root. Let $g_0 = 1$. Since $k = 0$, the E -functions reduce to constants (the corresponding M 's). Let us introduce another auxiliary function $G^0 = f_1$, and let M^+ denote the number of real roots of f_0 making G^0 positive and satisfying (A), M^- the number making G^0 negative and satisfying (A).

* The result of (vi) may be used to prove the following theorem. Let f_i and g_i be the i th derivatives of f_0 and g_0 , and let n_i be the loss of variations between a and b in the sequence $S(f_{i-1}g_i, f_i g_{i-1})$. Let p_i be the degree of $f_{i-1}g_i$, and q the degree of the lower (or both) of f_i and g_0 . A necessary and sufficient condition that the roots of f_0 and g_0 may be all real and simple and in (a, b) , with the least root of f_0 greater than the greatest root of g_0 , is that $n_i = p_i$ ($i = 1, \dots, q$).

Equation (10) degenerates in the two respective cases to

$$M^+ = 2^{-n-1} \sum_{(i)} \{M^{(i)} + M^{0(i)}\}, \quad M^- = 2^{-n-1} \sum_{(i)} \{M^{(i)} - M^{0(i)}\},$$

(i) denoting in turn every subset of $1, 2, \dots, n$. A little consideration then shows that the number of separate finite intervals is $M^- - M^+$, that is $-2^{-n} \sum_{(i)} M^{0(i)}$. The question of infinite regions may be settled by means of the signs of the coefficients of the highest terms of the G 's.

It will be seen that, as a result of adjoining the derivative of f_0 , we may remove a square factor from $H^{0(i)}$, so that $M^{0(i)}$ is the number of variations lost between $-\infty$ and ∞ by the sequence $S(f_0, G^{(i)})$, and G^0 does not actually enter into the calculations.

As an example, let $G^1 = 1 - x^2$, $G^2 = x^2 - 2x$ ($n = 2$). Then $f_0 = -2x^3 + 3x^2 + x - 1$; $M^0 = \text{loss of variations in } S(f_0, 1) = 0 - 1 = -1$; $M^{01} = \text{loss of variations in } S(f_0, G^1) = 2 - 1 = 1$; $M^{02} = -1$; $M^{012} = -3$. Hence the number of finite intervals is $-\frac{1}{4}(-1 + 1 - 1 - 3) = 1$. There is no infinite region.

(viii) A second method for the problem of (vii) involves generally more sequences, but simpler ones, and allows f_0 to have real multiple roots. Let ${}_iM^{(j)}$ be the value of $M^{(j)}$ when G^i is used in the place of f_0 , (j) not containing i . Let I be the number of infinite regions satisfying (A), as found from the signs of the highest terms. It may be shown then that the number of separate regions, finite and infinite, defined by (A) is $2^{-n} \sum_i \sum_{(j)} {}_iM^{(j)} + \frac{I}{2}$.

Thus, if $G^1 = x^2 - 6x + 8$, $G^2 = x^2 + 6x + 8$, and $G^3 = x$ ($n = 3$), we shall find that ${}_1M = {}_1M^2 = {}_1M^3 = {}_1M^{23} = 2$, ${}_2M = {}_2M^1 = 2$, ${}_2M^3 = {}_2M^{13} = -2$, ${}_3M = {}_3M^1 = {}_3M^2 = {}_3M^{12} = 1$; and the functions are all positive to the right, but not to the left. Hence the number of regions is $\frac{1}{8}(6 \times 2 - 2 \times 2 + 4 \times 1) + \frac{1}{2} = 2$.

12. RESTRICTED DISTRIBUTIONS

It will sometimes happen that, before applying the general theory, we have a partial knowledge of the root-distribution which enables us to assume for E an expression of the type $e + \sum_{i=1}^n c_i e_i$, where the coefficients c_i are independent of t (or of t', t'', \dots) and fewer than $k + 1$ in number, and e, e_i are known functions of t (or of t', t'', \dots). In such a case it is not usually necessary to use all the M 's in determining the distribution. If, starting with the equation

$$e + \sum_i c_i e_i = 2^{-k} \sum_s P_s (1+t)^{k-s} (1-t)^s,$$

we differentiate successively, and then put t in turn equal to 1 and -1 , we get $2k + 2$ equations which, if the e_i are linearly independent, determine the n

constants c_i . From these equations we may select any n whose determinant is not zero, and we shall naturally do so in such a way as to use as few M 's as possible, noting that, when $t = 1$, the r th derivative of E involves no M with more than r suffixes, and that, when $t = -1$, it involves none with fewer than $k - r$.

Let $\Delta_{\sigma_1 \dots \sigma_{n-\nu}}^{\rho_1 \dots \rho_\nu}$ be the n -rowed determinant whose first ν rows consist of the derivatives of order $\rho_1, \rho_2, \dots, \rho_\nu$ of the e_i , with $t = 1$, and whose last $n - \nu$ rows consist of derivatives of order $\sigma_1, \sigma_2, \dots, \sigma_{n-\nu}$, with $t = -1$. Let us examine this determinant in some special cases. The problems most naturally presenting themselves are those in which e_i is of the form $t^{r_i}(1+t)^\sigma (1-t)^\rho \varphi(t)$ ($i = 1, \dots, n$), where $\varphi(t)$ is not divisible by $1+t$ or $1-t$. Let μ of the r_i be even and $n - \mu$ of them odd. It may then be shown that $\Delta_{\sigma, \sigma+1, \dots, \sigma+n-\nu-1}^{\rho, \rho+1, \dots, \rho+\nu-1}$ differs from zero when $\nu = 0$ or n , as also when the smaller of ν and $n - \nu$ is equal to the smaller of μ and $n - \mu$, but is equal to zero when the smaller of ν and $n - \nu$ is greater than the smaller of μ and $n - \mu$.

For example, whenever $\kappa - \lambda$ is less than k , we have an instance with $\rho = \sigma = 0$ and $\varphi(t) = 1$. If then g_0 has only two roots in (a, b) , the positions of the α 's with respect to them may be found by means of P_0, P_1 , and P_2 ; P_k, P_{k-1}, P_{k-2} ; P_0, P_1, P_k ; or P_0, P_k, P_{k-1} .

Another example is afforded by the case where the distributions of the α 's and the β 's with respect to some special set of numbers are already known, so that their mutual arrangement is known subject to certain transpositions of adjacent roots. In such a case $\rho = 1, \sigma = 0$, and $\varphi(t) = 1$. Thus, the arrangement being $(\alpha\beta)(\alpha\alpha\beta\beta)\beta(\alpha\alpha\beta)$, with the order in each bracketed group undetermined, we have a case where μ and $n - \mu$ are 1 and 3 (or 3 and 1), and the complete determination is made by means of $P_1, P_k, P_{k-1}, P_{k-2}$, but not by means of P_1, P_2, P_k, P_{k-1} .

A second type of restricted distribution has its generating function of the form $e + \varphi(t) \cdot \sum_{i=1}^n c_i t^{r_i}$, where now only the exponents r_i are unknown. If $\varphi(1)$ is not zero, we find, by successive differentiation and putting $t = 1$, $\sum_i c_i r_i^j$ in terms of P_0, P_1, \dots, P_j for $j = 1, 2, \dots$. The first n of the resulting equations must, if the c_i are all different from zero, be solvable rationally in the n unknown r_i , though possibly not without ambiguity. Thus, if the first two equations so obtained are $2r_1 + r_2 = 12$, $2r_1^2 + r_2^2 = 54$, either $r_1 = 3$ and $r_2 = 6$, or $r_1 = 5$ and $r_2 = 2$. The value of $2r_1^3 + r_2^3$ will decide.

Generally, it may be proved that, if the first $2n$ equations are used, there cannot be two solutions in non-negative r 's unless they are mere permutations of one another determining the same distribution.*

When the coefficients c_i are all positive, the n equations $\sum_{i=1}^n c_i r_i^j = C_j$

* This can be proved from Descartes' rule by making the transformation $j = \log y, \log r_i = p_i$.

($j = 1, 2, \dots, n$) do not admit more than one solution satisfying the conditions $r_1 \geq r_2 \geq \dots \geq r_n \geq 0$.* Thus, in the example already used, if the α 's are known to fall into a group of two to the left and a single one to the right, the fact that these are in the intervals of five and two variations (three and six being now excluded) will be discovered with the aid of P_0 , P_1 , and P_2 only.

In particular, if the c_i are all unity, the first n equations obtained will always give a unique solution of the problem. Clearly the c 's may always be taken as unity, by using enough of them.

There are two conspicuous cases to which this method applies. The first is that in which $M < k + 1$. We may then take $E(t) = \sum_{i=1}^M t^{r_i}$, and the values of P_0 , P_1 , \dots , P_M will suffice.

The other case is that of the relative distribution of the roots of a polynomial and those of its derivative over the range $(-\infty, \infty)$. If $g_0 = f_1$, E is of the form $\sum_{i=\lambda}^{\kappa} t^i - (1+t) \sum_{i=1}^n t^{r_i}$; and the present method gives $\kappa - \lambda + 1 - 2n = P_0$, $\sum_{i=\lambda}^{\kappa} i - 2 \sum_{i=1}^n r_i - n = kP_0/2 - P_1/2$, etc. The first equation gives n , and the next n equations give in succession $\sum_{i=1}^n r_i$, $\sum_{i=1}^n r_i^2$, \dots , $\sum_{i=1}^n r_i^n$ in terms of P_0 , P_1 , \dots , P_n , thus determining the positions of the $\kappa - \lambda$ turning-points of f_0 with respect to the roots.

If in this problem the range is not $(-\infty, \infty)$ but (a, b) , the formula for $E(t)$ is $\sum_{i=\lambda}^{\kappa} t^i - (1+t) \sum_{i=1}^n t^{r_i} - \frac{1}{2} \{1 - \operatorname{sgn}[f_0(b)f_1(b)]\} t^{\lambda} - \frac{1}{2} \{1 + \operatorname{sgn}[f_0(a)f_1(a)]\} t^{\kappa}$, the theory being otherwise the same.

13. METHOD BY QUADRATIC FORMS

An alternative method for the determination of $M_{[j]}^{(i)}$ depends on the theory of real quadratic forms, and corresponds to Hermite's theorem on a single polynomial, as the previous method corresponds to that of Sturm.

Let $\xi_1, \xi_2, \dots, \xi_p$ be all the roots of f_0 , real and imaginary, counted each as often as its multiplicity requires. We shall suppose still that assumptions (i), (ii), and (iii) of §3 hold good.

By §8, it will be sufficient to consider the case where the range is $(-\infty, \infty)$, and for this range we know, by Hermite's principle,† that $M_{[j]}^{(i)}$ is the excess of positive over negative coefficients in any sum of real square terms to which a real linear transformation on the T 's will reduce the quadratic form

$$\sum_{r=1}^p \left(G_{[j]}^{(i)}(\xi_r) \cdot \left\{ \sum_{s=0}^{p-1} \xi_r^s T_s \right\}^2 \right).$$

* This follows in a similar way from a theorem by Laguerre. See *Acta Mathematica*, vol. 4 (1884), pp. 111-114.

† *Sur le nombre des racines d'une équation algébrique comprises entre des limites données*, Extrait d'une lettre à M. Borchardt, *Journal für Mathematik*, vol. 52 (1856), p. 39.

The matrix of this form is $\left(\sum_{r=1}^p \xi_r^s + {}^t G_{[G]}^{(i)}(\xi_r) \right)_{\substack{s=0, \dots, p-1 \\ t=0, \dots, p-1}}.$

If any sequence of principal minors of this matrix is selected (as it may always be) such that each minor contains the preceding and no two vanish in succession, and if $W_{[G]}^{(i)}$ is the number of variations of sign in the sequence, we have the formula

$$(13) \quad M_{[G]}^{(i)} = p - 2 W_{[G]}^{(i)}.$$

In numerical work by this method, a large part of the labor will occur in evaluating the determinants. As to the determination of the elements of the matrix, the theory of partial fractions provides a convenient method. For the element in the $(s+1)$ th row and $(t+1)$ th column is the coefficient of x^{-s-t-1} in the expansion of $H_{[G]}^{(i)}(x)/f_0(x)$ in descending powers of x . Since $H_{[G]}^{(i)}$ is the product of f_1 by a number of G 's, all the necessary expansions may be obtained by multiplications from the single expansion representing $f_1(x)/f_0(x)$.

QUEEN'S UNIVERSITY,
KINGSTON, ONTARIO.