

SIAM REVIEW
Vol. 8, No. 3, July, 1966

ASSOCIATED POLYNOMIALS AND UNIFORM METHODS FOR THE SOLUTION OF LINEAR PROBLEMS*

J. F. TRAUB†

TABLE OF CONTENTS

1. Introduction	277
2. Brief survey of results	279
3. The basic orthonormality relation	279
4. The inverse of the Vandermonde matrix	281
5. Interpolation	285
6. Approximation of linear functionals	287
7. The variation of the coefficients of a polynomial with the variation of its zeros	289
8. Solution of inhomogeneous difference and differential equations with constant coefficients	290
9. Divided differences	291
10. Applications of divided difference calculations	292
11. Some interesting formulas	293
12. The inverse transformation and Wronski's aleph function	295
13. A result concerning powers of zeros	297
14. Historical survey	298
Appendix. Notation	299
References	300

1. Introduction. To every polynomial P of degree n we associate a sequence of $n + 1$ polynomials of increasing degree which we call the associated polynomials of P . The associated polynomials depend in a particularly simple way on the coefficients of P . These polynomials have appeared in many guises in the literature, usually related to some particular application and most often going unrecognized. They have been called Horner polynomials and Laguerre polynomials. (For a historical discussion see §14.) Often what occurs is not an associated polynomial itself but a number which is an associated polynomial evaluated at a zero of P .

The properties of associated polynomials have never been investigated in themselves. We shall try to demonstrate that associated polynomials provide a useful unifying concept.

Although many of the results of this paper are new, we shall also present known results in our framework.

Let $P(t)$ be a monic polynomial of degree n ,

$$(1.1) \quad P(t) = \sum_{j=0}^n a_{n-j} t^j, \quad a_0 = 1.$$

The condition that P be monic is convenient rather than essential. The a_j and t are in general complex. Let the zeros of P be labeled ρ_j . In some of our applica-

* Received by the editors June 10, 1965, and in final revised form December 20, 1965.

† Mathematics Research Center, Bell Telephone Laboratories, Incorporated, Murray Hill, New Jersey.

tions we refer to the set of ρ_j as the sample point set. We assume here that the ρ_j are distinct; the uniform extension of our results to confluent zeros will be reported elsewhere.

Let u be a complex parameter and define polynomials $A_j(u, P), j = 0, 1, \dots, n$, by

$$(1.2) \quad \frac{tP(t) - uP(u)}{t - u} = \sum_{j=0}^n A_{n-j}(u, P)t^j, \quad t \neq u.$$

The remainder theorem shows that (1.2) is a valid definition. We verify below that $A_j(u, P)$ is a polynomial in u of degree j . We call the set $A_j(u, P), j = 0, 1, \dots, n$, the associated polynomials of P . For many applications we can abbreviate $A_j(u, P)$ by $A_j(u)$.

Observe that the left side of (1.2) is the first divided difference of the polynomial $tP(t)$ with respect to t and u . Hence we may write

$$(1.3) \quad (tP)[t, u] = \sum_{j=0}^n A_{n-j}(u, P)t^j, \quad t \neq u.$$

In words, the generating function of the associated polynomials of P is the first divided difference of tP .

Multiplying (1.2) by $t - u$, comparing powers of t , and recalling that P is monic, yields the recursion

$$(1.4) \quad \begin{aligned} A_0(u) &= 1, \\ A_j(u) &= uA_{j-1}(u) + a_j, \quad j = 1, 2, \dots, n. \end{aligned}$$

If we define $A_{-1}(u) = 0$, we can write

$$(1.5) \quad A_j(u) = uA_{j-1}(u) + a_j, \quad j = 0, 1, \dots, n.$$

From (1.5) we obtain the explicit formula

$$(1.6) \quad A_j(u) = \sum_{r=0}^j a_{j-r}u^r, \quad j = 0, 1, \dots, n.$$

This verifies that $A_j(u)$ is a polynomial of degree j . In particular

$$(1.7) \quad A_n(u) = P(u).$$

It is sometimes convenient to replace the generating function for $A_j(u), j = 0, 1, \dots, n$, by a simpler generating function for $A_j(u), j = 0, 1, \dots, n - 1$.

Since

$$(1.8) \quad \begin{aligned} (tP)[t, u] &= tP[t, u] + P(u), \\ P[t, u] &= \sum_{j=0}^{n-1} A_{n-1-j}(u)t^j. \end{aligned}$$

Since $A_n(u) = P(u)$,

$$(1.9) \quad \frac{P(t)}{t - u} = \frac{A_n(u)}{t - u} + \sum_{j=0}^{n-1} A_{n-1-j}(u)t^j.$$

Hence the quotient and remainder after division by a monic linear polynomial may be calculated recursively. The nature of the recursion is such that one may work with detached coefficients. The process is often called synthetic division. (For historical comments see §14.)

From (1.9) it follows that

$$\frac{P(t)}{t - \rho_i} = \sum_{j=0}^{n-1} A_j(\rho_i) t^{n-1-j}.$$

Thus

$$(1.10) \quad A_j(\rho_i) = (-1)^j \sigma_{n,i,j},$$

where $\sigma_{n,i,j}$ denotes the j th elementary symmetric function of $\rho_1, \rho_2, \dots, \rho_n$ with ρ_i deleted. Hence any result which involves $A_j(\rho_i)$ can always be interpreted as a result involving elementary symmetric functions.

2. Brief survey of results. The orthonormality relation given by (3.5) is of basic importance. It is used in §4 to provide us with two numerical algorithms for inverting the Vandermonde matrix. A variety of row and column sums to be used for check calculations are provided.

In §5 we point out advantages of writing the representation of an interpolatory polynomial in canonical polynomial form. In §6, the canonical representation of the preceding section is used to derive formulas for the approximation of linear functionals. Equation (6.4) exhibits the approximation as a vector-matrix-vector product. The columns of the inverse to the Vandermonde matrix may be characterized as the coefficients of the approximations to $f^{(k)}(0)/k!$.

In §7 we show that the variation of the coefficients of a polynomial with the variation of its zeros is determined by associated polynomials, (7.1) being the main result. In §8 we give the general solution of an inhomogeneous difference or differential equation with constant coefficients as an explicit function of the initial conditions.

The calculation of divided differences is discussed in §9 and is applied to §10 to “nonnoisy” computation and to the calculation of Stirling numbers. A number of interesting formulas are derived in §11. A symbolic statement of the Newton relations is given by (11.7) and a sweeping generalization of an important formula of Euler is given by (11.14) or (11.16).

Consideration of the transformation between powers of t and associated polynomials leads to recognition of the relation between Wronski’s aleph function and associated polynomials. The main result is given by (12.4). In §13 we give a simple solution to the classical problem of expressing an arbitrary power of a zero of a polynomial as a linear combination of the first $n - 1$ powers of that zero. A symbolic solution is provided by (13.5). We end with a historical survey in §14 and a summary of notation in the Appendix.

3. The basic orthonormality relation. Consider the generating function

$$(3.1) \quad P[t, u] = \sum_{j=0}^{n-1} A_{n-1-j}(u) t^j.$$

Let $t \rightarrow u$. Then

$$(3.2) \quad P'(u) = \sum_{j=0}^{n-1} A_{n-1-j}(u)u^j.$$

Let ρ_i and ρ_k be distinct zeros of P . Then

$$(3.3) \quad 0 = \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i)\rho_k^j.$$

Since ρ_i is a simple zero of P , we conclude from (3.2) and (3.3),

$$(3.4) \quad \sum_{j=0}^{n-1} \frac{A_{n-1-j}(\rho_i)\rho_k^j}{P'(\rho_i)} = \delta_{ik}, \quad i, k = 1, 2, \dots, n,$$

where δ_{ik} is the Kronecker delta function which vanishes except when the indices are equal.

Let α and β denote the matrices whose elements are

$$\begin{aligned} \alpha_{ij} &= \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)}, & j &= 0, 1, \dots, n-1, \quad i = 1, 2, \dots, n; \\ \beta_{jk} &= \rho_k^j, & j &= 0, 1, \dots, n-1, \quad k = 1, 2, \dots, n. \end{aligned}$$

Thus β is just the Vandermonde matrix. Let I denote the identity matrix. Then (3.4) may be written

$$\alpha\beta = I.$$

This implies that

$$\beta\alpha = I$$

or

$$(3.5) \quad \sum_{i=1}^n \frac{A_{n-1-j}(\rho_i)\rho_i^k}{P'(\rho_i)} = \delta_{jk}.$$

The orthonormality relation given by (3.5) is a basic result which permits us to solve a variety of linear problems. It is of importance for two reasons. First it permits the theoretical investigation of "inverse problems". Secondly it gives the inverse to the Vandermonde matrix in a form well suited for numerical calculation. Two algorithms for the inversion are discussed in detail in the next section.

Note that the ranges of i, j, k are not the same. This is because we have chosen to follow the usual convention of denoting the n zeros of a polynomial by $\rho_1, \rho_2, \dots, \rho_n$. If we had labeled the zeros $\rho_0, \rho_1, \dots, \rho_{n-1}$, the difference in ranges could have been avoided.

Two important consequences of (3.5) are

$$(3.6) \quad \sum_{i=1}^n \frac{\rho_i^\nu}{P'(\rho_i)} = \delta_{n-1,\nu}, \quad \nu = 0, 1, \dots, n-1,$$

$$(3.7) \quad \sum_{i=1}^n \frac{A_\nu(\rho_i)}{P'(\rho_i)} = \delta_{n-1,\nu}, \quad \nu = 0, 1, \dots, n-1.$$

Relation (3.6) is a well-known formula of Euler. Both (3.6) and (3.7) are special cases of (11.16).

4. The inverse of the Vandermonde matrix. From (3.5) we see that the inverse of the Vandermonde matrix is given by the matrix whose elements are

$$(4.1) \quad \alpha_{ij} = \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)}.$$

Equation (4.1) not only gives an explicit formula for the inverse, but in addition this formula is well suited for hand or machine calculation. We shall describe two algorithms for forming α , given the sample points ρ_i .

As we shall show below, check calculations may easily be performed on the rows and columns of (4.1). If the ρ_i satisfy symmetry conditions, then the rows of (4.1) satisfy certain symmetry conditions. A characterization of the columns of α in terms of approximation of derivatives is given in §6. The case of most importance for applications is that of a_i and ρ_i real. For the purpose of making operation counts, we restrict ourselves to this case for the remainder of this section.

We now describe Algorithm I for calculating α . The first step in this algorithm is to calculate the coefficients of P from the ρ_i . Let $\sigma_{m,j}$ denote the j th elementary symmetric function of the numbers $\rho_1, \rho_2, \dots, \rho_m$. Then $a_j = (-1)^j \sigma_{n,j}$. The $\sigma_{m,j}$ may be calculated recursively by

$$(4.2) \quad \sigma_{m,j} = \sigma_{m-1,j} + \rho_m \sigma_{m-1,j-1}, \quad m = 2, 3, \dots, n, \quad j = 1, 2, \dots, m;$$

with

$$\sigma_{m,0} = 1, \quad j \geq 0; \quad \sigma_{1,1} = \rho_1; \quad \sigma_{m,j} = 0, \quad j > m.$$

A numerical example of the recursion is given below. The calculation of all the a_i requires $\frac{1}{2}n(n - 1)$ multiplications and the same number of additions.

For each ρ_i , $A_{n-1-j}(\rho_i)$ may be calculated by the recursion

$$(4.3) \quad \begin{aligned} A_j(\rho_i) &= \rho_i A_{j-1}(\rho_i) + a_j, & j = 1, 2, \dots, n - 1; \\ A_0(\rho_i) &= 1. \end{aligned}$$

Hence $A_j(\rho_i)$, $j = 0, 1, \dots, n - 1$, may be calculated with just $n - 2$ multiplications and $n - 1$ additions. Observe that

$$(4.4) \quad \begin{aligned} A_j'(\rho_i) &= \rho_i A_{j-1}'(\rho_i) + A_{j-1}(\rho_i), & j = 2, 3, \dots, n; \\ A_1'(\rho_i) &= 1, & A_n'(\rho_i) &= P'(\rho_i). \end{aligned}$$

Hence $P'(\rho_i)$ may be calculated with just $n - 2$ multiplications and $n - 1$ additions.

Let us assume that a division time is comparable to a multiplication time. Then α may be formed from the ρ_i with $(n/2)(7n - 9)$ multiplications and $(5/2)n(n - 1)$ additions for $n > 1$. (See also the discussion below on the decomposition of α into the product of two matrices.)

We now describe Algorithm II for forming α . This algorithm is based on the

fact that

$$(4.3) \quad \pi_i(t) = \frac{P(t)}{t - \rho_i} = \prod_{\substack{j=1 \\ j \neq i}}^n (t - \rho_j) = \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i) t^j.$$

Hence the $A_{n-1-j}(\rho_i)$, $j = 0, 1, \dots, n-1$, are given by the coefficients of $\pi_i(t)$ and these are formed by the recursion (4.2) based on the numbers $\rho_1, \rho_2, \dots, \rho_{i-1}, \rho_{i+1}, \dots, \rho_n$. The calculation of the $P'(\rho_i)$ coincides with the calculation of the $P'(\rho_i)$ in Algorithm I.

The calculation of α from the ρ_i by Algorithm II requires $n(n^2 + n - 2)/2$ multiplications and $n^2(n - 1)/2$ additions.

Hence for $n \leq 4$, Algorithm II requires fewer operations while for $n > 4$, Algorithm I requires fewer operations. Observe that in Algorithm I, α is formed by a process of "analysis" whereas in Algorithm II, α is formed by a process of "synthesis". That is, in Algorithm I we form the coefficients of P and then calculate the $A_j(\rho_i)$ by division. In Algorithm II we form the coefficients of $\pi_i(t)$ and no division is involved. It is clear that Algorithm II is numerically more stable than Algorithm I.

We can cut down the number of multiplications required by Algorithm II by reusing some of the common factors obtained in the calculation of one $\pi_i(t)$ in the calculation of another $\pi_i(t)$.

We now want to show that certain advantages are obtained by decomposing α into the product of two matrices. Let \mathbf{n} be the diagonal matrix with elements

$$(4.6) \quad \eta_{ii} = P'(\rho_i).$$

Then \mathbf{n}^{-1} has elements $1/P'(\rho_i)$ on its diagonal. Let γ be the matrix with elements

$$(4.7) \quad \gamma_{ij} = A_{n-1-j}(\rho_i).$$

Then

$$(4.8) \quad \alpha = \mathbf{n}^{-1}\gamma.$$

This decomposition of α has a number of advantages. As we shall see, α is often used to form a vector-matrix-vector product, $\mathbf{f}\alpha\mathbf{M}$. (See (6.4).) Rather than using n^2 multiplications to form α from \mathbf{n}^{-1} and γ , we can form $\gamma\mathbf{M}$ and then form $\mathbf{n}^{-1}(\gamma\mathbf{M})$ in just n multiplications. A second advantage of this approach is that if the ρ_i are integers, all elements of γ and \mathbf{n} are integers, whereas the elements of α would not in general be integers. Because of the reduction in the number of multiplications and because of the integral property, the accuracy of the calculation is improved by the use of the decomposition. If the matrices \mathbf{n} and γ are stored in a machine or in a table, one vector suffices to hold the information in \mathbf{n} . Observe on the other hand that α can be produced once and for all, and to use the decomposition $\alpha = \mathbf{n}^{-1}\gamma$ requires n extra divisions for each \mathbf{M} to which α is applied.

As a numerical example, we choose the ρ_i as $-2, -1, 0, 1, 2$ and calculate α by Algorithm I. The calculation of the a_j , $A_j(\rho_i)$, and $P'(\rho_i)$ is performed with

detached coefficients. The rows of γ are obtained by writing the elements enclosed in parentheses in reverse order. The $P'(\rho_i)$ are enclosed in circles. Note that $A_n(\rho_i)$ is calculated as a check; it should be zero. See Table 1.

Hence

$$\mathbf{n} = (24, -6, 4, -6, 24),$$

$$(4.7) \quad \gamma = \begin{bmatrix} 0 & 2 & -1 & -2 & 1 \\ 0 & 4 & -4 & -1 & 1 \\ 4 & 0 & -5 & 0 & 1 \\ 0 & -4 & -4 & 1 & 1 \\ 0 & -2 & -1 & 2 & 1 \end{bmatrix},$$

$$(4.8) \quad \alpha = \begin{bmatrix} 0 & 2 & -1 & -2 & 1 \\ 0 & -16 & 16 & 4 & -4 \\ 24 & 0 & -30 & 0 & 6 \\ 0 & 16 & 16 & -4 & -4 \\ 0 & -2 & -1 & 2 & 1 \end{bmatrix}.$$

TABLE 1

-2	1	-2					
-1	1	-3	2				
0	1	-3	2	0			
1	1	-2	-1	2	0		
2	1	0	-5	0	4	4	0
1	0	-5	0	4	0	0	-2
	-2	4	2	-4	0		
(1	-2	-1	2	0)	0		
	-2	8	-14	24			
1	-4	7	-12	(24)			
1	0	-5	0	4	0		-1
	-1	1	4	-4	0		
(1	-1	-4	4	0)	0		
	-1	2	2	-6			
1	-2	-2	6	(-6)			
1	0	-5	0	4	0		0
	0	0	0	0	0		
(1	0	-5	0	4)	0		
	0	0	0	0			
1	0	-5	0	(4)			
1	0	-5	0	4	0		1
	1	1	-4	-4			
(1	1	-4	-4	0)	0		
	1	2	-2	-6			
1	2	-2	-6	(-6)			
1	0	-5	0	4	0		2
	2	4	-2	-4			
(1	2	-1	-2	0)	0		
	2	8	14	24			
1	4	7	12	(24)			

Observe that the rows of γ and α enjoy certain symmetry properties and that certain of the elements are zero. We show that this is characteristic for symmetric arrangements of the zeros of P .

First we draw some conclusions for the case that $P(0) = 0$. In particular, let $\rho_k = 0$. Then it may be shown that

$$(4.9) \quad \gamma_{i0} = \delta_{ik} a_{n-1}, \quad i = 1, 2, \dots, n,$$

$$(4.10) \quad \alpha_{i0} = \delta_{ik}, \quad i = 1, 2, \dots, n.$$

Furthermore

$$(4.11) \quad \gamma_{kj} = a_{n-1-j}, \quad j = 0, 1, \dots, n - 1,$$

$$(4.12) \quad \alpha_{kj} = \frac{a_{n-1-j}}{a_{n-1}}, \quad j = 0, 1, \dots, n - 1.$$

Observe that $a_{n-1} \neq 0$ since double zeros are not permitted.

Assume now that n is odd and that the ρ_i are symmetrically placed about the origin. This is an important case in the applications. Arrange the ρ_i in increasing order. Then it may be shown that

$$(4.13) \quad \gamma_{ij} = (-1)^j \gamma_{n+1-i,j}, \quad j = 0, 1, \dots, n - 1, \quad i = 1, 2, \dots, \frac{1}{2}(n - 1);$$

$$\gamma_{kj} = a_{n-1-j}, \quad k = \frac{1}{2}(n + 1);$$

$$\alpha_{ij} = (-1)^j \alpha_{n+1-i,j}, \quad j = 0, 1, \dots, n - 1, \quad i = 1, 2, \dots, \frac{1}{2}(n - 1);$$

$$(4.14) \quad \alpha_{kj} = \frac{a_{n-1-j}}{a_{n-1}}, \quad k = \frac{1}{2}(n + 1).$$

Observe that P has only odd powers of t . The application of (4.9), (4.10), (4.13), and (4.14) shows that the occurrence of zeros in the first column and middle row of (4.7) and (4.8) is not accidental.

Now let n be even and let the zeros of P be symmetrically arranged about the origin. Then

$$(4.15) \quad \gamma_{ij} = (-1)^{j+1} \gamma_{n+1-i,j}, \quad j = 0, 1, \dots, n - 1, \quad i = 1, 2, \dots, \frac{1}{2}n,$$

$$(4.16) \quad \alpha_{ij} = (-1)^j \alpha_{n+1-i,j}, \quad j = 0, 1, \dots, n - 1, \quad i = 1, 2, \dots, \frac{1}{2}n.$$

A number of check formulas may be derived. First, note that $A_n(\rho_i) = 0$. Hence in performing the recursion to form the elements $A_j(\rho_i)$, $j = 0, 1, \dots, n - 1$, it may be worthwhile to perform one more recursion so as to obtain $A_n(\rho_i)$ and compare with zero.

In addition we obtain row and column check sums for γ and α . From (11.18),

$$(4.17) \quad \sum_{i=1}^n A_{n-1-j}(\rho_i) = \sum_{i=1}^n \gamma_{ij} = (j + 1)a_{n-1-j}, \quad j = 0, 1, \dots, n - 1,$$

which provides us with a column sum check for γ . From (3.7),

$$(4.18) \quad \sum_{i=1}^n \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)} = \sum_{i=1}^n \alpha_{ij} = \delta_{j0}, \quad j = 0, 1, \dots, n - 1,$$

which provides us with a column sum check for α .

To obtain row sum checks we proceed as follows. Let $\rho_i \neq 1, i = 1, 2, \dots, n$. From

$$\frac{P(t)}{t - \rho_i} = \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i)t^j,$$

it follows that

$$(4.19) \quad \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i) = \sum_{j=0}^{n-1} \gamma_{ij} = \frac{C}{1 - \rho_i}, \quad i = 1, 2, \dots, n,$$

where

$$C = \sum_{k=0}^n a_k .$$

On the other hand, suppose one of the ρ_i is unity. Let $\rho_k = 1$. Then

$$(4.20) \quad \sum_{j=0}^{n-1} \gamma_{ij} = P'(1)\delta_{ik}, \quad i = 1, 2, \dots, n .$$

The corresponding row sum check for α is

$$(4.21) \quad \sum_{j=0}^{n-1} \alpha_{ij} = \frac{C}{P'(\rho_i)} \frac{1}{1 - \rho_i}$$

if none of the ρ_i is unity and

$$(4.22) \quad \sum_{j=0}^{n-1} \alpha_{ij} = \delta_{ik}$$

if $\rho_k = 1$.

From (3.6),

$$(4.23) \quad \sum_{i=1}^n \frac{1}{P'(\rho_i)} = \sum_{i=1}^n \eta_{ii}^{-1} = 0 .$$

If n is even and the zeros are symmetric with respect to the origin, then

$$(4.24) \quad \sum_{i=1}^n P'(\rho_i) = \sum_{i=1}^n \eta_{ii} = 0 .$$

5. Interpolation. Let $\rho_1, \rho_2, \dots, \rho_n$ be n arbitrary distinct numbers and let f_1, f_2, \dots, f_n be n arbitrary numbers which are not necessarily distinct. Let $P(t) = \prod_{i=1}^n (t - \rho_i)$. Then there exists a unique polynomial of degree $n - 1$, $I_{n-1}(t, f, P)$, called the *interpolatory polynomial*, such that

$$(5.1) \quad I_{n-1}(\rho_i, f, P) = f_i, \quad i = 1, 2, \dots, n .$$

The symbol f on the left side of (5.1) symbolizes the n given numbers f_1, f_2, \dots, f_n , and does not imply the existence of a function f . Any of the coefficients of I_{n-1} may be zero and indeed I_{n-1} could be identically zero.

There are two well-known representations of the interpolatory polynomial, those which bear the names of Lagrange and Newton. One other representation has certain advantages over the two conventional representations. This is a

representation in the canonical form for a polynomial, as a linear combination of powers of t . Such a representation is advantageous if one wants to perform operations on the interpolatory polynomial, which is often the case in practice. Writing the interpolatory polynomial in canonical form is hardly a surprising step. But the advantages of doing so seem to have been largely overlooked.

Let

$$I_{n-1}(t, f, P) = \sum_{j=0}^{n-1} d_{n-1-j} t^j.$$

Thus

$$f_i = \sum_{j=0}^{n-1} d_{n-1-j} \rho_i^j, \quad i = 1, 2, \dots, n.$$

By the orthonormality relation (3.5),

$$(5.2) \quad d_{n-1-j} = \sum_{i=1}^{n-1} f_i \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)}.$$

Let α be the matrix with elements

$$\alpha_{ij} = \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)}, \quad j = 0, 1, \dots, n-1, \quad i = 1, 2, \dots, n.$$

Let \mathbf{f} be the vector with elements $f_i, i = 1, 2, \dots, n$, and let \mathbf{t} be the vector with elements $t_j = t^j, j = 0, 1, \dots, n-1$. Then

$$I_{n-1}(t, f, P) = \sum_{i=1}^n \sum_{j=0}^{n-1} f_i \alpha_{ij} t_j,$$

or

$$(5.3) \quad I_{n-1}(t, f, P) = \mathbf{f} \alpha \mathbf{t}.$$

Observe that α depends only on the sample points ρ_i , while \mathbf{f} depends only on the f_i . The matrix α may be easily calculated as discussed in §4. For certain "standard sets" of sample points this calculation may be done once and for all and tabulated. For any set of f_i , the coefficients of $I_{n-1}(t)$ may then be computed by a vector-matrix multiplication. As we shall see in the next section, the form (5.3) is well suited for the approximation of linear functionals.

Rather than using the set $1, t, \dots, t^{n-1}$ as the basic set, we can use $A_0(t), A_1(t), \dots, A_{n-1}(t)$. Let

$$(5.4) \quad I_{n-1}(t, f, P) = \sum_{j=0}^{n-1} e_{n-1-j} A_j(t).$$

Then from the orthonormality relation,

$$(5.5) \quad e_{n-1-j} = \sum_{i=1}^n \frac{f_i \rho_i^{n-1-j}}{P'(\rho_i)}.$$

6. Approximation of linear functionals. Let L be a linear functional from a function space with elements f to the real or complex numbers. Let $f(\rho_i) = f_i$. One general method for approximating Lf is by applying L to an interpolatory polynomial for f . Many well-known approximate integration and differentiation formulas are of this type. Let M_j denote the moments of L ,

$$(6.1) \quad Lt^j = M_j, \quad j = 0, 1, \dots, n-1.$$

We introduce the notation

$$(6.2) \quad Q(L, f, P) = LI_{n-1}(t, f, P).$$

If we approximate Lf by $LI_{n-1}(t, f, P)$, we write

$$(6.3) \quad Lf \sim Q(L, f, P).$$

From

$$I_{n-1}(t, f, P) = \mathbf{f}\boldsymbol{\alpha}t,$$

we have immediately that

$$(6.4) \quad Q(L, f, P) = \mathbf{f}\boldsymbol{\alpha}\mathbf{M},$$

where \mathbf{M} denotes the vector with elements M_0, M_1, \dots, M_{n-1} . Observe that \mathbf{M} depends only on L , $\boldsymbol{\alpha}$ depends only on the sample points, and \mathbf{f} depends only on f . These formulas are conventionally given as a linear combination of f_i . That is, the matrix-vector product $\boldsymbol{\alpha}\mathbf{M}$ is formed and Q is then expressed as a scalar product. The generalization of (6.4) to the case of confluent sample points will be discussed in another paper.

The columns of the $\boldsymbol{\alpha}$ matrix may be characterized by the approximation of one particular set of functionals. Let

$$(6.5) \quad L_k f = \frac{f^{(k)}(0)}{k!}, \quad k = 0, 1, \dots, n-1.$$

Then

$$L_k t^j = \delta_{jk}, \quad j, k = 0, 1, \dots, n-1.$$

For k fixed $L_k t^j = \mathbf{M}_k$ is a column vector with one nonzero element and $\boldsymbol{\alpha}\mathbf{M}_k$ is just the k th column of $\boldsymbol{\alpha}$. Hence

$$(6.6) \quad \begin{aligned} \mathbf{f}\boldsymbol{\alpha}\mathbf{M} &= \sum_{i=1}^n f(\rho_i)\boldsymbol{\alpha}_{ik}, \\ \frac{f^{(k)}(0)}{k!} &\sim \sum_{i=1}^n f(\rho_i)\boldsymbol{\alpha}_{ik}. \end{aligned}$$

Thus we can characterize the columns of $\boldsymbol{\alpha}$ as the coefficients in the approximation of $f^{(k)}(0)/k!$, where the approximation is obtained by applying L_k to the interpolatory polynomial.

As a simple numerical illustration of the results of this chapter we take the sample zeros as $-2, -1, 0, 1, 2$. We calculated in §4 that

$$\alpha = \frac{1}{24} \begin{bmatrix} 0 & 2 & -1 & -2 & 1 \\ 0 & -16 & 16 & 4 & -4 \\ 24 & 0 & -30 & 0 & 6 \\ 0 & 16 & 16 & -4 & -4 \\ 0 & -2 & -1 & 2 & 1 \end{bmatrix}.$$

Let

$$Lf = f'(0).$$

Then,

$$\mathbf{M} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \alpha\mathbf{M} = \frac{1}{12} \begin{bmatrix} 1 \\ -8 \\ 0 \\ 8 \\ -1 \end{bmatrix},$$

(6.7) $f'(0) \sim \frac{1}{12}[f(-2) - 8f(-1) + 8f(1) - f(2)].$

This is a special case of the preceding discussion on the columns of α .

Let

$$Lf = \int_{-2}^2 f(t) dt.$$

Then

$$\mathbf{M} = \frac{4}{15} \begin{bmatrix} 15 \\ 0 \\ 20 \\ 0 \\ 48 \end{bmatrix}, \quad \alpha\mathbf{M} = \frac{2}{45} \begin{bmatrix} 7 \\ 32 \\ 12 \\ 32 \\ 7 \end{bmatrix},$$

(6.8) $\int_{-2}^2 f(t) dt \sim \frac{2}{45}[7f(-2) + 32f(-1) + 12f(0) + 32f(1) + 7f(2)].$

Let

$$Lf = \int_{-3}^3 f(t) dt.$$

Then

$$\mathbf{M} = \frac{6}{5} \begin{bmatrix} 5 \\ 0 \\ 15 \\ 0 \\ 81 \end{bmatrix}, \quad \alpha\mathbf{M} = \frac{3}{10} \begin{bmatrix} 11 \\ -14 \\ 26 \\ -14 \\ 11 \end{bmatrix},$$

(6.9) $\int_{-3}^3 f(t) dt \sim \frac{3}{10}[11f(-2) - 14f(-1) + 26f(0) - 14f(1) + 11f(2)].$

Equation (6.8) is an example of a closed integration formula while (6.9) is an example of an open formula.

7. The variation of the coefficients of a polynomial with the variation of its zeros. Let

$$P(t) = \sum_{j=0}^n a_{n-j}t^j, \quad a_0 = 1,$$

$$P(t) = \prod_{i=1}^n (t - \rho_i) = P(t, \rho_1, \dots, \rho_n).$$

Denote by Δa_k the change in a_k due to a change $\Delta \rho_i$ in ρ_i . We have

$$P(t, \rho_1, \dots, \rho_i + \Delta \rho_i, \dots, \rho_n) = (t - \rho_1) \cdots (t - \rho_i - \Delta \rho_i) \cdots (t - \rho_n)$$

$$= P(t, \rho_1, \dots, \rho_i, \dots, \rho_n) - \Delta \rho_i \frac{P(t)}{t - \rho_i}.$$

Hence

$$\Delta P = - \sum_{r=0}^{n-1} \Delta \rho_i A_r(\rho_i) t^{n-1-r},$$

where

$$\Delta P = P(t, \rho_1, \dots, \rho_i + \Delta \rho_i, \dots, \rho_n) - P(t, \rho_1, \dots, \rho_i, \dots, \rho_n).$$

On the other hand

$$\Delta P = \sum_{r=0}^{n-1} \Delta a_{r+1} t^{n-1-r}.$$

Hence

$$(7.1) \quad \Delta a_{r+1} = -\Delta \rho_i A_r(\rho_i), \quad i = 1, 2, \dots, n, \quad r = 0, 1, \dots, n - 1.$$

Thus $A_r(\rho_i)$ may be characterized as the change in a_{r+1} due to a change of -1 in ρ_i . Note that (7.1) holds in the large, that is, for arbitrary changes $\Delta \rho_i$. Taking the limit as $\Delta \rho_i \rightarrow 0$,

$$(7.2) \quad \frac{\partial a_{r+1}}{\partial \rho_i} = -A_r(\rho_i).$$

The differential form *only* is given by Brioschi [2] without proof.

Since the matrix with elements $\partial \rho_i / \partial a_{r+1}$ is the inverse of the matrix with elements $\partial a_{r+1} / \partial \rho_i$, we can conclude from (3.5) that

$$\frac{\partial \rho_i}{\partial a_{r+1}} = -\frac{\rho_i^{n-1-r}}{P'(\rho_i)},$$

or

$$(7.3) \quad \frac{\partial \rho_i}{\partial a_r} = -\frac{\rho_i^{n-r}}{P'(\rho_i)}.$$

Equation (7.3) has been obtained by Brioschi [2] and applied by Olver [21, p. 412].

8. Solution of inhomogeneous difference and differential equations with constant coefficients. Consider the linear difference equation (or recursion formula)

$$(8.1) \quad \sum_{j=0}^n a_{n-j} f(\lambda + j) = g(\lambda), \quad \lambda = 0, 1, \dots,$$

with the initial conditions

$$f(\mu) = y_\mu, \quad \mu = 0, 1, \dots, n - 1.$$

Using the translation operator we may also write

$$(8.2) \quad \begin{aligned} P(E)f(\lambda) &= g(\lambda), \\ E^\mu f(0) &= y_\mu, \end{aligned} \quad \mu = 0, 1, \dots, n - 1.$$

It may be shown that the solution as an explicit function of the initial conditions is given by

$$(8.3) \quad f(\lambda) = \sum_{j=0}^{n-1} y_{n-1-j} \sum_{i=1}^n \frac{\rho_i^\lambda A_j(\rho_i)}{P'(\rho_i)} + \sum_{j=0}^{\lambda-1} g(\lambda - 1 - j) \sum_{i=1}^n \frac{\rho_i^j}{P'(\rho_i)}.$$

This formula was given by Traub [27], the proof based on a division algebra. This formula may also be derived by any of the standard transform techniques. A direct derivation based on polynomial interpolation is given by Traub [28].

The fact that the solution satisfies the initial conditions follows immediately from (3.5). That the solution satisfies the difference equation may also be verified.

Certain solutions of the homogeneous equation are particularly simple. Let $g(\lambda) = 0$. Let $y_{n-1-j} = \delta_{jk}$ and let the corresponding solution be labeled $f_k(\lambda)$. That is, the starting sequence has a one in the $(n - 1 - k)$ th position and zero elsewhere. Then

$$f_k(\lambda) = \sum_{i=1}^n \frac{\rho_i^\lambda A_k(\rho_i)}{P'(\rho_i)},$$

or using (1.6),

$$(8.4) \quad f_k(\lambda) = A_k(E)\Omega(\lambda),$$

where

$$\Omega(\lambda) = \sum_{i=1}^n \frac{\rho_i^\lambda}{P'(\rho_i)}.$$

We turn to differential equations. Let

$$(8.5) \quad \begin{aligned} P(D)f(x) &= g(x), \\ D^j f(0) &= y_0^{(j)}, \end{aligned}$$

where D is the differentiation operator. Then

$$(8.6) \quad f(x) = \sum_{j=0}^{n-1} y_0^{(n-1-j)} \sum_{i=1}^n \frac{e^{\rho_i x} A_j(\rho_i)}{P'(\rho_i)} + \int_0^x g(x - v) \sum_{i=1}^n \frac{e^{\rho_i v}}{P'(\rho_i)} dv.$$

As in the difference equation case treated above, one may verify that (8.6) satisfies the equation and initial conditions of (8.5).

The fitting of initial conditions usually involves determinants. We have avoided the determinants by inverting the Vandermonde matrix in (3.5).

9. Divided differences. It is well known that repeated synthetic division of a polynomial P by a fixed number u_0 results in the calculation of $P^{(j)}(u_0)/j!$, $j = 0, 1, \dots, n$. This fact lies at the heart of the Ruffini-Horner method for solving algebraic equations [3], [9]. This method depends on calculating from the coefficients of a given polynomial the coefficients of a second polynomial all of whose zeros are less than those of the given polynomial by an amount u_0 . Let $P(t + u_0) = Z(t)$. Then

$$Z(t) = \sum_{j=0}^n \frac{P^{(j)}(u_0)}{j!} t^j.$$

Let ρ_i be a zero of P . Then

$$0 = P(\rho_i) = Z(\rho_i - u_0).$$

Hence the polynomial with reduced zeros has coefficients $P^{(j)}(u_0)/j!$, which coefficients may be computed by iterated synthetic division by u_0 .

We generalize this result by investigating synthetic division by a sequence u_0, u_1, \dots, u_{n-1} . The results may be used to perform “nonnoisy” calculations on polynomials as shown in §10.

Divided differences are usually recursively defined as follows. Let u_0, u_1, \dots, u_{n-1} be a sequence of numbers. Let

$$P[t] = P(t),$$

$$P[t, u_0, u_1, \dots, u_i] = \frac{P[t, u_0, u_1, \dots, u_{i-1}] - P[u_0, u_1, \dots, u_i]}{t - u_i},$$

$$i = 0, 1, \dots, n - 1,$$

with

$$P[t, u_{-1}] \equiv P[t].$$

If P is a polynomial of degree n , then the i th divided difference, $P[t, u_0, u_1, \dots, u_{i-1}]$, is a polynomial of degree $n - i$ and is symmetric in all its arguments.

We introduce the simplified notation

$$P_0(t) = P(t),$$

$$P_i(t) = P[t, u_0, u_1, \dots, u_{i-1}], \quad i = 1, 2, \dots, n.$$

Then

$$(9.1) \quad P_i(t) = \frac{P_{i-1}(t) - P_{i-1}(u_{i-1})}{t - u_{i-1}}, \quad i = 1, 2, \dots, n,$$

or

$$(9.2) \quad P_i(t) = P_{i-1}[t, u_{i-1}].$$

Let

$$(9.3) \quad m = n - i.$$

From (9.1) and (1.8),

$$(9.4) \quad P_{i+1}(t) = P_i[t, u_i] = \sum_{j=0}^{m-1} A_j(u_i, P_i)t^{m-1-j}.$$

That is, the coefficients of the $(i + 1)$ st difference of $P(t)$ with respect to t, u_0, u_1, \dots, u_i are given by the associated polynomials of the i th difference of P . It follows that if, for convenience, we define the symbols

$$A_{-1}(u_i, P_i) = 0, \quad A_j(u_{-1}, P_{-1}) = a_j,$$

then

$$(9.5) \quad A_j(u_i, P_i) = u_i A_{j-1}(u_i, P_i) + A_j(u_{i-1}, P_{i-1}),$$

$$i = 0, 1, \dots, n, \quad j = 0, 1, \dots, m, \quad \text{where } m = n - i.$$

From this it follows that

$$(9.6) \quad A_j(u_i, P_i) = \sum_{k=0}^j A_k(u_{i-1}, P_{i-1})u_i^{j-k}, \quad j = 0, 1, \dots, m.$$

Furthermore

$$(9.7) \quad A_m(u_i, P_i) = P[u_0, u_1, \dots, u_i].$$

This generalizes (1.7),

$$A_n(u_0, P) = P(u_0).$$

Equation (9.7) gives the generalization of the comments at the beginning of this section concerning the calculation of derivatives by detached coefficients since if $u_0 = u_1 = \dots = u_i$, the i th divided difference is just $P^{(i)}(u_0)/i!$.

10. Applications of divided difference calculations. The importance of (9.5) and (9.7) lies in the fact that these results permit the calculation of divided differences of polynomials in a “nonnoisy mode.” By this we mean the following. If the points u_0, u_1, \dots, u_{i+1} are close together, then the calculation of $P[u_0, u_1, \dots, u_{i+1}]$ by

$$(10.1) \quad P[u_0, u_1, \dots, u_{i+1}] = \frac{P[u_0, u_1, \dots, u_i] - P[u_1, u_2, \dots, u_{i+1}]}{u_0 - u_{i+1}}$$

leads to the loss of accuracy due to the subtraction of quantities which are close together. This is the case when divided differences are being used to approximate the derivative at a point. This is easily corrected by performing the calculation using (9.5) and (9.7).

We must add one word of caution. The use of (9.5) and (9.7) permits us to avoid the subtraction of (10.1). Of course, it is possible that (9.5) also involves a

subtraction of close quantities. A simple example is given when $u_0 \sim -a_1$. However, this would be an unusual rather than a typical circumstance.

We apply this to the solution of polynomial equations by iterative methods. One common method is to use the iteration formula

$$x_{i+1} = x_i - \frac{P(x_i)}{P'(x_i)},$$

with $P(x_i)$ and $P'(x_i)$ calculated by detached coefficients. This variation of the Newton-Raphson method is often called the Birge-Vieta method. For monic polynomials, the Birge-Vieta method requires $2n - 1$ multiplications per iteration.

Another possibility is to use the formula

$$(10.2) \quad x_{i+1} = x_i - \frac{P(x_i)}{P[x_i, x_{i-1}]}$$

and to use at the next iteration the two points at which P has opposite signs. This method is commonly called *regula falsi* [26]. Let $P(\alpha) = 0$. As $x_i \rightarrow \alpha$, $P[x_i, x_{i-1}]$ becomes harder to calculate accurately by the usual formula

$$(10.3) \quad P[x_i, x_{i-1}] = \frac{P(x_i) - P(x_{i-1})}{x_i - x_{i-1}}.$$

By calculating $P(x_i)$ by synthetic division and $P[x_i, x_{i-1}]$ by a second synthetic division, the difficulty is avoided. The number of multiplications is $2n - 1$ as in the Birge-Vieta method. Regula falsi has the virtue of bracketing the zero at each step.

As a second application we consider the calculation of Stirling numbers of the second kind, $s(n, i)$, defined by

$$x^n = \sum_{i=0}^n s(n, i)[x]_i,$$

where

$$[x]_i = \prod_{j=0}^{i-1} (x - j).$$

It follows that

$$s(n, i) = (x^n)[0, 1, \dots, i], \quad i = 0, 1, \dots, n,$$

that is, the n th row of the Stirling number matrix may be obtained using repeated synthetic division by $0, 1, \dots, n$. This is a more efficient way of obtaining a number of rows of this matrix than by using either the explicit formula for Stirling numbers of the second kind or their recurrence relation [11, p. 169].

11. Some interesting formulas. A number of interesting formulas are direct consequences of the interpolation formulas of §5. They stem from the fact that if there exists a polynomial f , of degree less than or equal to $n - 1$, such that $f(\rho_i) = f_i$, then $I_{n-1}(t, f, P) \equiv f(t)$. A number of the formulas to be derived are

well known while others are new. Define

$$(11.1) \quad s(j) = \sum_{i=1}^n \rho_i^j,$$

$$(11.2) \quad \Omega(j) = \sum_{i=1}^n \frac{\rho_i^j}{P'(\rho_i)},$$

$$(11.3) \quad \omega(j) = \sum_{i=1}^n \frac{\rho_i^{n-1+j}}{P'(\rho_i)},$$

and recall that

$$(11.4) \quad A_j(t) = \sum_{r=0}^j a_{j-r} t^r.$$

Taking $f_i = P'(\rho_i)$ in (5.5) and using (5.4) yields

$$(11.5) \quad P'(t) = \sum_{j=0}^{n-1} s(j) A_{n-1-j}(t).$$

Comparing coefficients of powers of t in (11.5) yields

$$(11.6) \quad (n - r)a_r = \sum_{j=0}^r a_{r-j}s(j), \quad r = 0, 1, \dots, n - 1,$$

which is equivalent to the Newton relations. Equation (11.6) may be written in the symbolic form

$$(11.7) \quad A_r(E)s(0) = (n - r)a_r.$$

Taking $f_i = \rho_i^k, k = 0, 1, \dots, n - 1$, in (5.5) and using (5.4) yields

$$(11.8) \quad t^k = \sum_{j=0}^{n-1} \Omega(j + k) A_{n-1-j}(t).$$

In particular, we have

$$(11.9) \quad t^{n-1} = \sum_{j=0}^{n-1} \omega(j) A_{n-1-j}(t)$$

as the counterpart of (11.5). Comparing coefficients of powers of t in (11.9) yields

$$(11.10) \quad \sum_{r=0}^j a_{j-r}\omega(r) = \delta_{j,0}, \quad j = 0, 1, \dots, n - 1.$$

This may be written in symbolic form as

$$(11.11) \quad A_j(E)\omega(0) = \delta_{j,0}.$$

Let $R_m(t)$ be an arbitrary polynomial of degree $m \leq n - 1$,

$$R_m(t) = \sum_{j=0}^m g_{m-j} t^j.$$

This may be written

$$(11.12) \quad R_m(t) = \sum_{j=0}^{n-1} g_{m-j} t^j$$

if we define

$$(11.13) \quad g_{-i} = 0, \quad i = 1, 2, \dots, n - 1 - m.$$

It follows from (5.2) with $f(t) = R_m(t)$ that

$$(11.14) \quad \sum_{i=1}^n \frac{A_{n-1-j}(\rho_i) R_m(\rho_i)}{P'(\rho_i)} = g_{m-j}, \quad j, m = 0, 1, \dots, n - 1.$$

Let $R_m(t)$ be monic. Then

$$(11.15) \quad \sum_{i=1}^n \frac{A_{n-1-j}(\rho_i) R_m(\rho_i)}{P'(\rho_i)} = \delta_{mj},$$

$$j = 0, 1, \dots, n - 1, \quad m = 0, 1, \dots, j.$$

In particular, if $R_m(t)$ is monic,

$$(11.16) \quad \sum_{i=1}^n \frac{R_m(\rho_i)}{P'(\rho_i)} = \delta_{m,n-1}, \quad m = 0, 1, \dots, n - 1.$$

This includes (3.6) and (3.7) as special cases.

Taking $R_m(t) = A_m(t)$ in (11.14) yields

$$\sum_{i=1}^n \frac{A_{n-1-j}(\rho_i) A_m(\rho_i)}{P'(\rho_i)} = a_{m-j}, \quad j, m = 0, 1, \dots, n - 1,$$

with the convention of (11.13). Hence

$$(11.17) \quad \sum_{i=1}^n \frac{A_m^2(\rho_i)}{P'(\rho_i)} = a_{2m-n+1}, \quad m = 0, 1, \dots, n - 1.$$

Taking $R_m(t) = P'(t)$ yields

$$\sum_{i=1}^n A_{n-1-j}(\rho_i) = (j + 1) a_{n-1-j}, \quad j = 0, 1, \dots, n - 1.$$

Taking $R_m(t) = \sum_{i=0}^{n-1} t^i$ yields

$$(11.19) \quad \sum_{i=1}^n \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)} \frac{\rho_i^n - 1}{\rho_i - 1} = 1, \quad \rho_i \neq 1, \quad j = 0, 1, \dots, n - 1.$$

12. The inverse transformation and Wronski's aleph function. The explicit formula for $A_j(t, P)$,

$$(12.1) \quad A_j(t, P) = \sum_{k=0}^j a_{j-k} t^k, \quad a_0 = 1, \quad j = 0, 1, \dots, n,$$

We have used the fact that $\omega(0) = 1$ which is a consequence of Euler's formula (3.6). From (11.10) and (12.3),

$$(12.6) \quad \sum_{j=0}^k a_{k-j} \omega(j) = \delta_{0,k}, \quad k = 0, 1, \dots, n.$$

This equation may be written in operational form as

$$(12.7) \quad A_k(E)\omega(0) = \delta_{0,k}, \quad k = 0, 1, \dots, n.$$

Equations (11.10) and (12.3) may be used to recursively calculate the $\omega(\nu)$, $\nu = 0, 1, \dots$, in terms of the a_j , $j = 0, 1, \dots, n$. The first few $\omega(\nu)$ are given by

$$\begin{aligned} \omega(0) &= 1, \\ \omega(1) &= -a_1, \\ \omega(2) &= a_1^2 - a_2, \\ \omega(3) &= -a_1^3 + 2a_1a_2 - a_3. \end{aligned}$$

From (11.10) and (12.3) it follows that

$$(12.8) \quad \frac{1}{t^n P(1/t)} = \sum_{\nu=0}^{\infty} \omega(\nu) t^\nu.$$

Equation (12.8) is the basis for methods used to calculate the dominant zero of polynomials. We defer to another paper a general discussion of such methods. From (12.8) we have

$$\omega(\nu) = \frac{1}{\nu!} D^\nu \left[\frac{1}{t^n P(1/t)} \right]_{t=0}.$$

Observe that $P(t)$ may be considered the generating function for a_n, a_{n-1}, \dots, a_0 . On the other hand $[t^n P(1/t)]^{-1}$ is the generating function for $\omega(\nu)$, $\nu = 0, 1, \dots$.

13. A result concerning powers of zeros. It is clear that if ρ_i is any zero of P , then an arbitrary power of ρ_i may be expressed as a linear combination of $1, \rho_i, \dots, \rho_i^{n-1}$, with coefficients which are polynomials in the coefficients of P . The problem of calculating the coefficients of this linear combination is a classical one.

We may obtain the solution simply as follows. Let

$$(13.1) \quad \rho_i^\nu = \sum_{j=0}^{n-1} \beta_{\nu j} \rho_i^{n-1-j}, \quad \nu = 0, 1, \dots.$$

Our orthonormality relation immediately yields

$$(13.2) \quad \beta_{\nu j} = \sum_{i=1}^n \frac{\rho_i^\nu A_j(\rho_i)}{P'(\rho_i)}$$

or

$$(13.3) \quad \beta_{\nu j} = \sum_{r=0}^j a_{j-r} \sum_{i=1}^n \frac{\rho_i^{\nu+r}}{P'(\rho_i)} = \sum_{r=0}^j a_{j-r} \Omega(\nu + r).$$

Since $\Omega(\nu) = \omega(\nu - n + 1)$ is a polynomial in the a_j , this is the desired result. Observe that (13.3) may be written in the symbolic form

$$(13.4) \quad \beta_{\nu j} = A_j(E)\Omega(\nu).$$

From (13.1),

$$\rho_i^\nu = \sum_{j=0}^n \rho_i^{n-1-j} A_j(E)\Omega(\nu),$$

and hence

$$(13.5) \quad \rho_i^\nu = P[E, \rho_i]\Omega(\nu).$$

It is clear that an arbitrary power of ρ_i may also be expressed as a linear combination of $A_0(\rho_i)$, $A_1(\rho_i)$, \dots , $A_{n-1}(\rho_i)$. Let

$$\rho_i^\nu = \sum_{j=0}^{n-1} \alpha_{\nu j} A_{n-1-j}(\rho_i).$$

Then

$$(13.6) \quad \begin{aligned} \alpha_{\nu j} &= \Omega(\nu + j), \\ \rho_i^\nu &= \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i)\Omega(\nu + j). \end{aligned}$$

In symbolic form, we have

$$\rho_i^\nu = \sum_{j=0}^{n-1} A_{n-1-j}(\rho_i)E^j\Omega(\nu) = P[E, \rho_i]\Omega(\nu),$$

which is the result obtained in (13.5).

14. Historical survey. The recursion formula

$$(14.1) \quad A_j(u) = uA_{j-1}(u) + a_j, \quad j = 0, 1, \dots, n, \quad A_{-1}(u) = 0,$$

and $P(u) = A_n(u)$ show that a polynomial may be evaluated by "nesting." The recursion formula and

$$(14.2) \quad \frac{P(t)}{t-u} = \frac{A_n(u)}{t-u} + \sum_{j=0}^{n-1} A_{n-1-j}(u)t^j$$

lead to synthetic division.

The nested evaluation of polynomials, synthetic division, and a certain method for approximating the roots of polynomial equations are each commonly called Horner's method. Indeed what we have labeled associated polynomials are sometimes referred to as Horner polynomials. These appellations seem incorrect, as in all these matters Horner was anticipated by others.

As pointed out by Ostrowski [22], Newton (1711) observed that the polynomial $y^4 - 4y^3 + 5y^2 - 12y + 17$ could be evaluated by nesting. Horner (1819) [9] noted that the associated polynomials, introduced by their recursion, were used by Lagrange in his *Theory of Functions*. As to the method (1819) for solving

polynomial equations, which depends on evaluating the derivatives of a polynomial by repeated synthetic division (see §9), Cajori [3] points out that the same results were obtained by Ruffini in 1804. Cajori recommends naming this the Ruffini-Horner method. (De Morgan [5, pp. 66–67] observes that both men had been anticipated by Chinese mathematicians.) To sum up this section of our history, although Horner's work was at first not appreciated by the Royal Society [5, p. 188], he is today credited with results of which he was not the originator.

Lagrange (1775) [14] introduced associated polynomials by giving them in the explicit form

$$A_j(t) = \sum_{k=0}^j a_{j-k} t^k.$$

He uses them to obtain a solution of the *homogeneous* linear difference equation with constant coefficients in the case that the zeros of the indicial polynomial are distinct. The result is presented without proof or verification, and his generalization to the case that the indicial polynomial has confluent zeros is incorrect. (Lagrange finally corrected this error in 1792.)

The associated polynomials should perhaps be called Lagrange polynomials except for the confusion with the polynomials in Lagrange's interpolation formula which bear his name.

Scattered results on associated polynomials appear in a number of 19th century works. These include Schröder [24], Laguerre [15], Netto [19, p. 97], and Weber [29, pp. 34, 156–163]. (Obreschkoff [20, p. 50] calls the associated polynomials Laguerre polynomials.) Associated polynomials are implicit in Cauchy's [4] treatment of the solution of linear differential equations.

A prescription for the numerical inversion of the Vandermonde matrix is given by Kowalewski [13, pp. 2–6] who obtains the results from Cauchy's interpolatory relations. Kowalewski's prescription is the basis for Algorithm II, but Kowalewski does not concern himself with the numerical details of the prescription. Since then a number of authors have independently rediscovered variations of this method. See Stojaković [25], Macon and Spitzbart [17], Hamming [8, pp. 124–128], Legras [16, pp. 98–104], Parker [23], and Klinger [12]. Gautschi [7] has studied the norm of the inverse Vandermonde matrix. The importance of the Vandermonde inverse for the approximation of linear functionals has been pointed out by Legras [16] and Hamming [8]. See also Bragg [1]. Hamming calls the inverse matrices "universal matrices". None of these authors has studied the numerical aspects of the inversion.

Our contribution to the theory has been a unified approach which has led to many new results. We have taken the generating function as the basic definition and the orthonormality relation as the basic result.

Acknowledgment. It is a pleasure to thank my colleague A. J. Goldstein for a number of illuminating discussions.

Appendix. Notation.

1. $P(t) = \sum_{j=0}^n a_n - j t^j, \quad a_0 = 1.$
2. $P(t) = \prod_{i=1}^n (t - \rho_i).$

3. $P[t, u_0, u_1, \dots, u_{i-1}]$ is the i th divided difference of $P(t)$ with respect to $t, u_0, u_1, \dots, u_{i-1}$.
4. $P_i(t) = P[t, u_0, u_1, \dots, u_{i-1}]$, $P_0(t) = P(t)$.
5. $(tP)[t, u] = \sum_{j=0}^n A_{n-j}(u, P)t^j \equiv \sum_{j=0}^n A_{n-j}(u)t^j$.
6. $\delta_{jk} = \begin{cases} 0 & \text{if } j \neq k, \\ 1 & \text{if } j = k. \end{cases}$
7. $s(\nu) = \sum_{i=1}^n \rho_i^\nu$, $\Omega(\nu) = \sum_{i=1}^n \frac{\rho_i^\nu}{P'(\rho_i)}$, $\omega(\nu) = \sum_{i=1}^n \frac{\rho_i^{n-1+\nu}}{P'(\rho_i)}$.
8. $f(\lambda + 1) = Ef(\lambda)$, $f'(\lambda) = Df(\lambda)$.
9. $\alpha: \alpha_{ij} = \frac{A_{n-1-j}(\rho_i)}{P'(\rho_i)}$; $\beta: \beta_{ij} = \rho_j^i$. Then $\alpha\beta = I$.
10. $\mathbf{n}: \eta_{ij} = P'(\rho_i)\delta_{ij}$; $\gamma: \gamma_{ij} = A_{n-1-j}(\rho_i)$. Then $\alpha = \mathbf{n}^{-1}\gamma$.

REFERENCES

- [1] L. H. BRAGG, *A matrix approach to numerical integration*, Amer. Math. Monthly, 71 (1964), pp. 391-398.
- [2] F. BRIOSCHI, *Intorno ad alcune questioni d'algebra superiore*, Ann. di Sci. Matem. Fis., (1854), pp. 301-312.
- [3] F. CAJORI, *Horner's method of approximation anticipated by Ruffini*, Bull. Amer. Math. Soc., 17 (1911), pp. 409-414.
- [4] A. L. CAUCHY, *Sur la détermination des constantes arbitraires renfermées dans la intégrales des équations différentielles linéaires*, Oeuvres Ser. 2, vol. 7, Académie des Sciences, Paris.
- [5] A. DE MORGAN, *A Budget of Paradoxes*, vol. II, London, 1872.
- [6] Encyklop. d. Math. Wissensch., vol. I, part I, §3b, *Rationale Funktionen der Wurzeln*, pp. 450-479.
- [7] W. GAUTSCHI, *On inverses of Vandermonde and confluent Vandermonde matrices*, Numer. Math., 4(1962), pp. 117-123; 5(1963), pp. 425-430.
- [8] R. W. HAMMING, *Numerical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1962.
- [9] W. G. HORNER, *A new method of solving numerical equations of all orders by continuous approximation*, Philos. Trans. Roy. Soc. London, (1819), pp. 308-335.
- [10] ———, *On algebraic transformation*, The Mathematician, 1(1845), pp. 108-112.
- [11] C. JORDAN, *Calculus of Finite Differences*, Chelsea, New York, 1950.
- [12] A. KLINGER, *The Vandermonde matrix*, Report P3201, RAND Corporation, Santa Monica, California, 1965.
- [13] G. KOWALEWSKI, *Interpolation und genäherte Quadratur*, Teubner, Berlin, 1932.
- [14] J. L. LAGRANGE, *Sur les suites recurrentes*, Nouveau Mémoires de l'Académie Royale de Berlin, 6(1775), pp. 183-195.
- [15] E. LAGUERRE, *Théorie des équations numériques*, J. Math. Pures Appl., 3^e série, 9(1883), p. 99.
- [16] J. LEGRAS, *Précis d'analyse numérique*, Dunod, Paris, 1963.
- [17] N. MACON AND A. SPITZBART, *Inverses of Vandermonde matrices*, Amer. Math. Monthly, 65(1958), pp. 95-100.
- [18] M. MARTONE, *La Funzione Alef di Hoëne Wronski*, Catanzaro, 1891.
- [19] E. NETTO, *Vorlesungen über Algebra*, vol. I, Leipzig, 1896.
- [20] N. OBRESCHKOFF, *Verteilung und Berechnung der Nullstellen reeler Polynome*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1963.
- [21] F. W. J. OLVER, *The evaluation of zeros of high degree polynomials*, Philos. Trans. Roy. Soc. London. Ser. A, 244(1952), pp. 335-415.
- [22] A. OSTROWSKI, *On two problems in abstract algebra connected with Horner's rule*, Memorial to Von Mises, Academic Press, New York, 1954, pp. 40-48.

- [23] F. D. PARKER, *Inverses of Vandermonde matrices*, Amer. Math. Monthly, 71(1964), pp. 410–411.
- [24] E. SCHRÖDER, *Über unendlich viele Algorithmen zur Auflösung der Gleichungen*, Math. Ann., 2(1870), pp. 317–365.
- [25] M. STOJAKOVIĆ, *Solution du probleme d'inversion d'une classe importante de matrices*, C. R. Acad. Sci. Paris, 246(1958), pp. 1133–1135.
- [26] J. F. TRAUB, *Iterative Methods for the Solution of Equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.
- [27] ———, *Generalized sequences with applications to the discrete calculus*, Math. Comp., 19(1965), pp. 177–200.
- [28] ———, *Solution of linear difference and differential equations*, Bull. Amer. Math. Soc., 71(1965), pp. 538–541.
- [29] H. WEBER, *Lehrbuch der Algebra*, vol. I (English transl. 1898 ed.), Chelsea, New York.