EIGENVALUES AND EIGENFUNCTIONS OF FINITE-DIFFERENCE OPERATORS

By W. G. BICKLEY and JOHN McNAMEE

Communicated by E. T. Goodwin

Received 5 November 1958

1. *Introduction.* Numerical solution of differential and integral equations is concerned mainly with the determination of the wanted function at a finite number of discrete points which are, in general, uniformly spaced. A first approximation to the solution can be obtained if the given differential or integral system is replaced by a difference system. Any differential or integral operator can be expressed as an infinite series of difference operators and the difference system of the first approximation is obtained by neglecting all but the first few terms of the infinite expansions. We shall distinguish two processes for improving the approximation: the first uses a tabular interval of moderate length but the approximation to the given continuous system is improved by including in the difference system as many terms of the infinite expansions as are necessary or practicable; the second process uses an unvarying difference system of low accuracy, but the tabular interval is reduced in successive approximations, the process being continued until two successive approximations agree within the accuracy required. We regard these processes as essentially distinct. If the solutions obtained by the two processes approach limits, these limits need not coincide.

1·1. *Scope of the paper.* The principal aim of the present paper is to explain some simple formal relations which are latent in the first process; these relations establish a perspicuous correspondence between the tabular solution of the differential or integral system and the solution of the approximating difference system. A further consequence of these relations is that we are led to explore from a fresh viewpoint one of the simplest of interpolation techniques, namely, trigonometric interpolation. It transpires that trigonometric interpolation is quite closely related to current techniques for numerical solution of differential and integral equations.

We develop these concepts by a somewhat indirect route. We show first that the solution of a simple and familiar difference system can be resolved into components; the tabular values of these components define continuous periodic functions which may be integrated or differentiated. With due restrictions, these components prove useful in a wider context than the simple problems which we use to illustrate them.

In §§ 1·2 and 2 we envisage a generating function $g(x)$ which can be regarded as defined and bounded in the whole complex plane or at least along the whole real axis. For example, $g(x)$ may be defined as a one-valued analytic function in the whole plane; but it may be difficult to obtain numerical values of the function at more than a chosen

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

set of points and it is legitimate and useful to enquire whether a meaningful representation of $g(x)$ can be based on these function values. Again, as in § 3, the generating function may be defined as a bounded function in a given interval of the real axis by means of a differential equation with appropriate boundary conditions or by an integral equation; the continuation of the function along the real axis can be defined by periodicity. When a function is defined in this way, a numerical solution of the differential or integral equations will yield function values at a chosen set of points. It is then convenient to distinguish between the definition of a function and the explicit function values which are at our disposal or which we hope to obtain.

1·2. *Properties of discrete approximations to continuous functions.* The essential properties of discrete approximations to continuous functions have been elucidated by Whittaker (1). He considered a one-valued analytic function $g(x)$ whose values at the infinite set of points $x = \pm nh$ (n, an integer; h, the tabular interval) uniformly spaced along the real axis are $g(\pm nh)$. He showed that central difference interpolation expansions (based on these ordinates) yield values $C(x)$ which are identical with the values of $g(x)$ at tabular points; but the values of $C(x)$ and $g(x)$ at non-tabular points are not necessarily identical—indeed, they will in general be discrepant. The function $C(x)$ is then a well-defined function which he terms the *cardinal function* of the set of functions co-tabular with $g(x)$. It has the defining properties:

(i) it is a continuous function, co-tabular with the generating function;

(ii) it has no finite singularities in the complex plane;

(iii) it has no oscillations of period less than $2h$.

It is important to observe that short-period terms ($< 2h$) of the generating function are not simply suppressed in forming the cardinal function; they are converted into co-tabular long-period terms. An example of suppression is the truncation of a Fourier series; as examples of conversion, we may instance

$$\exp 2\pi i\left(\frac{N}{h}+m\right)x \quad \text{to} \quad \exp 2\pi imx, \qquad -\frac{1}{2h} < m \leqslant \frac{1}{2h};$$

$$\exp 2\pi i\left(\frac{N+\frac{1}{2}}{h}+m\right)x \quad \text{to} \quad \exp 2\pi i\left(-\frac{1}{2h}+m\right)x, \quad 0 < m \leqslant \frac{1}{2h};$$

$$\exp 2\pi i\left(\frac{N+\frac{1}{2}}{h}+m\right)x \quad \text{to} \quad \exp 2\pi i\left(\frac{1}{2h}+m\right)x, \quad -\frac{1}{2h} < m \leqslant 0,$$

where N is an integer.

We shall consider here a restricted set of cardinal functions, namely, those functions which are defined at a finite number of points uniformly spaced on a finite segment of the real axis; this segment is a half-period or an integral number of half-periods of any function of this set and the functions are continued along the real axis by periodicity. We shall show that this restricted set consists of eigenfunctions of the basic finite-difference operators and it suffices for many practical applications (e.g. those of § 3) in which we are required to determine a wanted function only at discrete points of a finite segment.

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

2. *Matrix representation of difference operators.*

It is convenient at the outset to employ matrix representations of difference operators and vector representations of functions $z(x)$ defined only at the tabular points $x = ph$, $p = 0, \ldots, n+1$, of the unit interval; this vector z has the components $z(x_p)$, $p = 1, \ldots, n$, the points x_0 and x_{n+1} being regarded as boundary points. Matrix representation of operators is convenient in numerical applications, and it is a suitable starting point for the considerations to be developed in the present paper.

The central-difference operators δ and μ and their powers may be defined in terms of the differential operator D $(= d/dx)$ by

$$\delta = 2\sinh(\tfrac{1}{2}hD), \quad \mu = \cosh(\tfrac{1}{2}hD). \tag{1}$$

It is known that the action of the second-difference operator δ^2 on the tabular values z_p can be expressed by

$$\delta^2 z_p = z_{p+1} + z_{p-1} - 2z_p.$$

An nth order matrix representation of the operator $-\delta^2$ (and implicitly of $4\mu^2 = 4 + \delta^2$), acting on such a table, is

$$\mathbf{D}^2 = 2I - S - S^*, \tag{2}$$

where I is the unit matrix, S is a matrix with units in the superdiagonal and zeros elsewhere, and S^* is the transpose of S. We choose to represent $-\delta^2$ rather than δ^2 in order that $\mathbf{D}^2$ may have positive eigenvalues; the bold character serves to distinguish the matrix representation from the differential operator. This matrix produces correct central second differences corresponding to z_1 and z_n only if z_0 and z_{n+1} vanish.

The eigenvalues and eigenfunctions of matrices of this type have been fully discussed by Rutherford (2) and Todd (3), using determinantal methods. We sketch briefly here a somewhat different method which emphasizes the dependence on boundary values, i.e. we consider the eigenvalue equation

$$\mathbf{D}^2 z = \lambda z, \tag{3}$$

z_0 and z_{n+1} being zero. Writing $z_p = A\rho^p$

(A constant), we find $\rho^2 - 2\rho\cos\theta + 1 = 0$, where $\lambda = 2(1-\cos\theta)$, and the boundary values then determine the eigenvalues and eigenvectors

$$\lambda_m = 4\sin^2(\tfrac{1}{2}m\pi h), \quad z_p = \sin(pm\pi h), \tag{4}$$

where $h = 1/(n+1)$; $p, m = 1, 2, \ldots, n$.

A plausible argument might be constructed thus: the eigenvalues of δ^2 are $-4\sin^2(\tfrac{1}{2}m\pi h)$, and we infer that the eigenvalues of δ may be

$$2i\sin(\tfrac{1}{2}m\pi h) = 2\sinh(\tfrac{1}{2}m\pi hi).$$

Remembering the defining relation, $\delta = 2\sinh(\tfrac{1}{2}hD)$, we surmise that the eigenvalues of $hD = hd/dx$ should be

$$m\pi hi$$

and the eigenfunctions of D will then be

$$\exp(pm\pi hi) = \exp(m\pi xi),$$

where $x = ph = p/(n+1)$.

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

The conclusion is certainly correct if it is permissible to pass from integral values of p to continuous values in the relation $x = ph$. It follows, however, from the properties cited in the Introduction that the cardinal function constructed from the tabular values of $\exp(m\pi xi)$ is identical with $\exp(m\pi xi)$ when m is restricted to the continuous range $0 \leqslant m < n+1$. In general, this continuous spectrum will be discretized by boundary conditions; for example, the boundary conditions of the matrix representation admitted only integral values of m in the specified range.

If this function is differentiated, the function is simply multiplied by a numerical factor. Equivalently, we may say that differentiation of the cardinal function constructed from the tabular values $\exp(pm\pi hi)$ leaves the function and hence the tabular values unchanged, save for a common factor of proportionality. In this sense, we may legitimately term the discrete function $\exp(pm\pi hi)$ an eigenfunction or eigenvector of the operator D when m is suitably restricted.

A direct proof of this useful property may serve to reinforce the general distinction between generating function and cardinal function and their identification in these particular instances. We shall make use of the known result

$$\sum_{-\infty}^{\infty} (-)^p \frac{\exp(2p\pi iv)}{t-p} = \pi \frac{\exp(2\pi ti\theta)}{\sin \pi t} \quad (|\theta| \neq \tfrac{1}{2}),$$

where θ is the difference between v and the integer nearest to v (Bromwich (4), p. 393). This result is valid only when t is not an integer; proceeding to the limit $t \to N$ (integral), we have

$$\left.\begin{aligned} \sum_{-\infty}^{\infty}{}' (-)^p \frac{\exp(2p\pi iv)}{N-p} &= (-)^N 2\pi iv \exp(2\pi Niv) \quad (0 \leqslant v < \tfrac{1}{2}) \\ &= 0 \qquad\qquad\qquad\qquad (v = \tfrac{1}{2}), \end{aligned}\right\} \tag{5}$$

and in general

$$\left.\begin{aligned} \sum_{-\infty}^{\infty}{}' (-)^p \frac{\exp(2p\pi iv)}{N-p} &= (-)^N 2\pi i\theta \exp(2\pi Ni\theta) \quad (v-[v] \neq \tfrac{1}{2}) \\ &= 0 \qquad\qquad\qquad\qquad (v-[v] = \tfrac{1}{2}), \end{aligned}\right\} \tag{6}$$

the dash on the summating sign indicating that the term $p = N$ is omitted.

If $C(x)$ is the cardinal function for the values $g(ph)$, we have by definition

$$C(x) = \sum_{-\infty}^{\infty} g(ph) \frac{\sin \pi(x-ph)/h}{\pi(x-ph)/h}$$

and

$$\left[\frac{dC(x)}{dx}\right]_{x=Nh} = \sum_{-\infty}^{\infty}{}' g(ph) \frac{\cos \pi(N-p)}{h(N-p)} = \frac{(-)^N}{h} \sum_{-\infty}^{\infty}{}' (-)^p \frac{g(ph)}{N-p}.$$

Putting

$$g(ph) = \exp(pm\pi hi),$$

we have

$$\left.\begin{aligned} \left[\frac{dC(x)}{dx}\right]_{x=Nh} &= m\pi i \exp(Nm\pi hi) \quad (0 \leqslant m < n+1) \\ &= 0 \qquad\qquad\qquad\qquad (m = n+1). \end{aligned}\right\}$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

We may differentiate repeatedly and show that all the derivatives of the cardinal function of $\exp(m\pi xi)$ are identical with the derivatives of the generating function when m is restricted as above. Again, using (6), the conversion of short-period terms into co-tabular long-period terms can be recognized from the derivative above.

The numerical applications of this result will be considered in the next section. It is convenient here to use the obviously correct eigenfunctions of D to deduce the corresponding functions of the difference operators. For the moment we ignore the possibility of matrix representations.

The following results are easily verified from the definitions.

$$\left.\begin{array}{lllll} \delta\text{:} & \text{eigenvalues} & 2i\sin(\frac{1}{2}m\pi h), & \text{eigenfunctions} & \exp(pm\pi hi); \\ \mu\text{:} & \text{eigenvalues} & \cos(\frac{1}{2}m\pi h), & \text{eigenfunctions} & \exp(pm\pi hi). \end{array}\right\} \tag{7}$$

Here, it is convenient to assume that ph denotes a tabular point, p being the location index. The value index m appears to be unrestricted. The restriction to long-period harmonics is, however, implicit if $\exp(pm\pi hi)$ is defined only at tabular points, since (i) short-period terms are indistinguishable from co-tabular long-period terms in the discrete definition, and (ii) the use of function values at points midway between tabular points (in forming δ or μ) implies appeal to a continuous function which here can only be the cardinal function.

The eigenvalues and eigenfunctions of δ^2, μ^2 and $2\mu\delta$, follow immediately; the eigenvalues of δ^2 are real and negative, μ^2 has real positive eigenvalues. It is readily verified that *either* the real *or* imaginary parts of the above *discrete* eigenfunctions will serve as eigenfunctions of μ, μ^2, δ^2, D^2.

For inverse operations we note that the trapezoidal and mid-ordinate approximations to the value of the integral of $g(x)$ over the range qh to ph are defined by

$$\begin{aligned} hT(g) &= h(\tfrac{1}{2}g_q + g_{q+1} + \ldots + g_{p-1} + \tfrac{1}{2}g_p) \\ &= \tfrac{1}{2}h\coth(\tfrac{1}{2}hD)[g_p - g_q] = h\left(\frac{\delta}{\mu}\right)^{-1}[g_p - g_q] \end{aligned} \tag{8}$$

and

$$\begin{aligned} hM(g) &= h(g_{q+\frac{1}{2}} + g_{q+\frac{3}{2}} + \ldots + g_{p-\frac{1}{2}}) \\ &= \tfrac{1}{2}h\,\mathrm{cosech}(\tfrac{1}{2}hD)[g_p - g_q] = h\delta^{-1}[g_p - g_q], \end{aligned} \tag{9}$$

also

$$\int_{qh}^{ph} g(x)\,dx = D^{-1}[g_p - g_q]. \tag{10}$$

If, for example, we form the mid-ordinate sum and the integral of the function $\exp(m\pi xi)$ over the range $x = qh$ to $x = ph$, we find

$$\int_{qh}^{ph} \exp(m\pi xi)\,dx - hM = \left[\frac{1}{m\pi i} - \frac{h}{2i\sin(\frac{1}{2}m\pi h)}\right][\exp(pm\pi hi) - \exp(qm\pi hi)],$$

and the denominators on the right side can be recognized as the eigenvalues of D and δ (equations (9) and (10)) corresponding to this eigenfunction.

3. *Applications.*

3·1. *Representation of an arbitrary discrete function by cardinal components.* We define a *discrete function* as one defined at $n+1$ equally-spaced ordinates in the interval $(-1, 1)$. Then a *complete set* of discrete functions in the given interval can be used to

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

represent linearly any discrete function of that interval. The preceding section has exhibited and described such a complete set of discrete eigenfunctions, having the property that differencing of any function merely multiplies it throughout by a constant factor of proportionality. Each function also has the property that the values at $n+1$ points in the interval $(-1, 1)$—provided only that $n+1$ exceeds the order m of the function—defines a unique continuous function, depending only on m and not on $n+1$. Differentiation of this continuous function, like differencing, merely multiplies it by a factor of proportionality throughout, so that it is an eigenfunction for differentiation. It is desirable to emphasize the dual role of these functions by using the term 'cardinal components'. Since the eigenvalues may be imaginary, it is convenient at this stage to separate the eigenfunctions into real and imaginary components. The spectrum of eigenvalues can be discretized in an infinity of ways by the assignment of appropriate boundary conditions, but the complete set of functions defined by integral values of the eigen-index m enjoys very simple orthogonality properties and is convenient in practical use; it is sufficient, here, to consider only positive m.

It is possible, alternatively, to use the semi-interval $(0, 1)$ in a similar way, by assuming symmetry and cosine functions only, or anti-symmetry and sine functions only. The representation of a function defined by its values at equidistant points of a given range is familiar, but we shall require also the representation of functions which are defined, not by tabular values, but by a differential equation within the range, and by boundary relations at the ends of the range; a similar problem arises in the numerical solution of integral equations. The remarks which follow refer particularly to the difficulties attending the representation of functions whose values are initially unknown. The range $(0, 1)$ suffices for the applications envisaged in §§ 3·2 and 3·3 and we shall consider only this range.

Components of the form $\cos(pm\pi h) = \cos(m\pi x)$ or $\sin(pm\pi h) = \sin(m\pi x)$ individually satisfy simple boundary conditions at the terminal points 0 and 1 of the chosen range. The prescribed boundary conditions of a given differential system need not even be compatible with the simple boundary conditions satisfied by the component functions; this is particularly evident when a function is represented by sine components only or cosine components only. It will be seen later that a problem involving the general linear boundary condition can generally be reduced to an equivalent problem with simpler boundary conditions and that representational difficulties arising from the nature of the prescribed boundary conditions can usually be evaded.

If the values of a function $g(x)$ at the tabular points of $(0, 1)$ are available (or can be determined), we may express $g(x)$ by the representation

$$C_1(x) = \Sigma\,(g'_m c_m + g'_m s_m), \tag{11}$$

where $$c_m = \cos(m\pi x), \quad s_m = \sin(m\pi x) \quad \left(m \leqslant \frac{1}{h} = n+1\right).$$

The coefficients g_m and g'_m are chosen so that $g(ph) = C_1(ph)$ for, say, $p = 0, 1, \ldots, n+1$, and the components c_m and s_m are chosen to suit the needs of a particular application,

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

subject however to the restriction that the array of the tabular values of $\cos(pm\pi h)$ and $\sin(pm\pi h)$ should constitute a non-singular matrix of order $n+2$. $C_1(x)$ is clearly a cardinal function and the differences of $g(x)$ and $C_1(x)$ are clearly identical so long as the differences involve only tabular points of $(0, 1)$. Limiting operations such as differentiation or integration can be expressed as an infinite series of difference operations and we may replace limiting operations on $g(x)$ at internal points of $(0, 1)$ by limiting operations on $C_1(x)$ if the series are rapidly convergent. When we are solving differential or integral equations by discrete techniques, the replacement of the tabular values $g(ph)$ by the continuous representation $C_1(x)$ is a convenient practical expedient since limiting operations are easily effected on the component functions c_m and s_m.

Near the ends of the interval, the discrepancy between the derivatives of $g(x)$ and of $C_1(x)$ will be greater than at points near the centre of the interval. This discrepancy can be exhibited simply when the cardinal function $C(x)$ determined by the tabular values $g(ph)$, $-\infty < p < +\infty$, exists. Writing $k_p = g(ph) - C_1(ph)$, the cardinal function

$$K(x) = \sum_{-\infty}^{\infty} k_p \frac{\sin \pi(x-ph)/h}{\pi(x-ph)/h}$$

and its derivatives measure the discrepancy between $C(x)$ and $C_1(x)$ and the derivatives of these functions. Since $C(x)$ and $C_1(x)$ are co-tabular in $(0, 1)$, the above summation is effected only over tabular points outside this interval, i.e. the discrepancy is due to the assumption of periodic continuation which is implicit in the use of the representation (11).

Periodic continuation can, however, be secured by a simple change of variable which is of practical value in numerical solutions. It has been pointed out by Lanczos (5) that any one of the changes of variable

$$x = \tfrac{1}{2}(1 \pm \cos \pi\theta), \quad x = \sin(\tfrac{1}{2}\pi\theta), \quad x = \cos(\tfrac{1}{2}\pi\theta)$$

converts $g(x)$ into a periodic function. This device considerably increases the scope of the present representation; for example, the change of variable $x = \cos \pi\theta$ converts the Chebychev polynomial $T_{n+1}(x)$ and the Legendre polynomial $P_{n+1}(x)$ into periodic cardinal functions of θ when $h_\theta < 1/(n+1)$.

It may be noted that a representation of the form (11) is closely related to a Fourier series representation.† It is well known that poor convergence of the derivatives of a Fourier series is due mainly to the destructive influence of the higher harmonics (see, e.g. Lanczos (5)). A similar phenomenon can be observed in the representation (11).

† The relation has been very clearly stated by de la Vallée Poussin (6), p. 384. He uses a fundamental interval of 2π and formulates the rule

'Quand on divise 2π en un nombre impair μ de parties, la formule d'interpolation se déduit de celle de Fourier en remplaçant, dans les lignes trigonométriques $\cos px$ et $\sin px$ qui y figûrent, les multiplicateurs p par leur résidu (mod μ), positif ou négatif, qui a la plus petite valeur absolue.

Si μ est pair et égal à $2n$, on devra commencer par annuler dans la formule de Fourier tous les coefficients b dont l'indice est un multiple de n (on pourrait dire aussi qu'on annule les sinus correspondants), et puis l'on pourra appliquer la règle précédente.'

This rule makes it possible to use known bounds of the Fourier representation to deduce corresponding bounds for the cardinal representation.

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

In the applications we envisage, $g(x)$ and $C(x)$ will not in general be available; but we may be able to obtain a second approximation $C_2(x)$ by using a smaller integral h'. If the higher harmonics of $C_2(x)$ (which appear only as co-tabular long-period terms in $C_1(x)$) do not decrease in any derivative, we can deduce the failure of $C_1(x)$ to yield that derivative. De la Vallée Poussin's rule (quoted in the footnote)† makes clear that, for a given tabular interval h, it is difficult to determine the effect of converting higher harmonics without reference to a smaller tabular interval; it would appear that a similar difficulty is encountered in most discrete approximations.

We might at this point be expected to emphasize the uncertainty which must be faced if we attempt to estimate the value of $g(x)$ at non-tabular points from its value at tabular points. This uncertainty scarcely requires emphasis for the difficulties attending the representation of a function by means of component functions in one of the central problems of analysis. Let us here anticipate the results of § 3·2 and assume that we employ two different uniform intervals h and h' and that the numerical solution of a differential equation yields cardinal functions $C_1(x)$ and $C_2(x)$ which we accept as approximations to the wanted function. If there are no serious discrepancies between $C_1(x)$ and $C_2(x)$, our confidence in the approximation $C_1(x)$ is somewhat strengthened; but an analyst would usually and reasonably require much stronger grounds for confidence.

3·2. *Numerical solution of linear differential equations in* $(0, 1)$. Cardinal components are convenient for numerical solutions in the interval $(0, 1)$. Sine components are particularly convenient since they all satisfy zero boundary conditions. At first sight this appears to be unduly restrictive, but a linear transformation of the dependent variable (cf. equation (22) below) enables us to transform a problem involving non-zero boundary values into an equivalent problem in which the boundary values are zero—the governing differential equation must, of course, be linear.

Let us suppose that we wish to determine z by a numerical solution of the differential equations

$$-\frac{d^2z}{dx^2}+q(x)\,z=f(x), \tag{12}$$

where boundary conditions are imposed at $x=0$ and $x=1$. The points $x_p=ph$ at which we seek to determine $z(x_p)$ are in general at our disposal and there are at least two convenient choices, viz. $p=0, 1, \ldots, n$ and $p=\frac{1}{2}, 1\frac{1}{2}, \ldots, n+\frac{1}{2}$. The sense in which we define the components of a function in terms of cardinal components depends on the choice of points x_p and is of practical importance. Here we shall indicate only very briefly the definitions which are pertinent to our present analysis.

Using an obvious extension of the notation of equations (8) and (9), we can write

$$T_n(f.g)=f_1g_1+f_2g_2+\ldots+f_{n-1}g_{n-1}+\tfrac{1}{2}(f_0g_0+f_{n+1}g_{n+1})$$

and

$$M_n(f.g)=f_{\frac{1}{2}}g_{\frac{1}{2}}+f_{1\frac{1}{2}}g_{1\frac{1}{2}}+\ldots+f_{n+\frac{1}{2}}g_{n+\frac{1}{2}}$$

as measures of the component of the function $f(x)$ in the direction of the function $g(x)$. It is easy to verify that

$$T_n[\sin m_1\pi x.\sin m_2\pi x]=M_n[\sin m_1\pi x.\sin m_2\pi x]=0$$

† For footnote see p. 538.

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

if m_1 and m_2 are different integers. We may then use $T_n[f.g] = 0$ as a definition of orthogonality if p assumes only integral values and $M_n[f.g] = 0$ if p is half an odd integer. Similar definitions may be used with cosine components or with mixed sine and cosine components. For example,

$$T_n[\sin m_1\pi x . \cos m_2\pi x] = M_n[\sin m_1\pi x . \cos m_2\pi x] = 0$$

if m_1 and m_2 are integers and $m_1 - m_2$ is even.

To avoid ambiguity between these definitions of component, we shall use only the T-definition in this section. We now consider the differential equation (12) and we shall examine first the simple conditions $z(0) = z(1) = 0$; sine components are then convenient for the representation of z. The terminal values of $f(x)$ will not in general vanish and this implies inaccuracy in the representation of $f(x)$ by sine components—but not necessarily an inaccuracy of the same magnitude in the determination of z. The terminal values of $q(x)$ are of less significance, since we shall require only the representation of $q(x)\,z$ and not of $q(x)$ alone.

We assume a discrete solution of the form

$$z = z_m s_m,$$

where

$$f = f_m s_m; \tag{13}$$

the z_m and f_m are here scalars and summation over the repeated index is intended, here and elsewhere. Representing $q(x)$ by a diagonal matrix Q, we may write

$$Qz = Qz_m s_m = \beta_m s_m;$$

the scalars β_m can be determined and eliminated if we notice that we can define a set of vectors σ_r such that

$$T(\sigma_r . s_m) = \delta_{rm}.$$

If we then multiply the preceding equation by σ_r, we have

$$\begin{aligned}\sigma_r Qz_m s_m &= z_m(\sigma_r Q s_m)\\ &= q_{rm} z_m = \beta_r.\end{aligned}$$

The scalars q_{ij} are the elements of the matrix Q in the basis s_m, σ_m; i.e.,

$$Q = q_{ij} s_i \sigma_j. \tag{14}$$

Equation (12) now becomes

$$(\lambda s)_m z_m + q_{mr} z_r s_m = f_m s_m, \quad \lambda_m = m^2\pi^2. \tag{15}$$

The equation pertaining to the pth component is selected on multiplying by σ_p:

$$(\lambda z)_p + q_{pr} z_r = f_p. \tag{16}$$

It is useful to compare the above solution with the cruder approximation obtained when the differential equation is replaced by the difference equation

$$-\frac{\delta^2 z_i}{h^2} + [q(x)\,z]_i = f(x_i). \tag{17}$$

We may write as before

$$Q = q_{mr} s_m \sigma_r, \quad f = f_m s_m,$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

but $z = z'_m s_m$. We find as above

$$\frac{(\lambda' z')_p}{h^2} + q_{pr} z'_r = f_p,$$

where

$$\frac{\lambda'_p}{h^2} = \frac{4}{h^2} \sin^2 (\tfrac{1}{2} p\pi h). \tag{18}$$

The difference between the two solutions can be exhibited by writing

$$z_m = z'_m + z''_m, \quad \lambda_m = \frac{\lambda'_m}{h^2} + \lambda''_m.$$

Subtracting (18) from (16), we find

$$\left[\left(\frac{\lambda'}{h^2} + \lambda''\right) z''\right]_p + q_{pr} z''_r = -(\lambda'' z')_p. \tag{19}$$

This equation is of the same form as that commonly used for the evaluation of the higher-difference correction. We recover the usual equation on multiplying (19) by s_p and summing over p.

The formal similarity is somewhat misleading. The inclusion of higher differences involves in general tabular values at points outside the range of determination. In the solution by cardinal components, these values are (implicitly) obtained by periodic continuation; in the normal use of the higher-difference correction, it may be possible to compute or estimate the extra values. The techniques are then closely related, but not completely equivalent.

The above discussion can be applied without essential modification to the more general equation

$$-\frac{d^2 z}{dx^2} + p(x)\frac{dz}{dx} + q(x)\, z = f(x); \tag{20}$$

the differential operator d/dx converts s_m into $m\pi c_m$, but the discussion is otherwise unchanged.

Prescribed boundary conditions are convenient in numerical work, and we may observe here that the solution of a problem involving boundary conditions of the form

$$\left.\begin{aligned} a_1 z(0) + a_2 z'(0) + a_3 z(1) + a_4 z'(1) &= a_5, \\ b_1 z(0) + b_2 z'(0) + b_3 z(1) + b_4 z'(1) &= b_5 \end{aligned}\right\} \tag{21}$$

can usually be achieved by adding a linear term to a solution satisfying zero boundary conditions. If

$$z(x) = y(x) + (1-x)\, z(0) + x z(1), \tag{22}$$

the terminal values $z(0)$ and $z(1)$ can be determined by using (21) and (22) if the matrix

$$\begin{bmatrix} a_1 - a_2 - a_4 & a_2 + a_3 + a_4 \\ b_1 - b_2 - b_4 & b_2 + b_3 + b_4 \end{bmatrix}$$

is not singular.

The above illustrations possessed two simplifying properties:

(i) the components were orthogonal,

(ii) each component individually satisfied the boundary conditions.

These properties are convenient but inessential, and a wider range of problems can be treated if they are discarded.

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

We illustrate this point by considering again equation (20), supposing that the unit interval is divided into an even number $(n+1)$ of tabular intervals. We seek a solution of the form

$$z = z_0 c_0 + z_m c_m + z'_m s_m,$$

where c_0 is a constant vector all of whose elements are unity and the components form an independent but not necessarily orthogonal set. Equation (20) becomes

$$z_m(\lambda c)_m + z'_m(\lambda s)_m + p(x)\,\nu_m[-zs+z'c]_m + q(x)\,[z_0 c_0 + z_m c_m + z'_m s_m] = f_0 c_0 + f_m c_m + f'_m s_m, \qquad (23)$$

where

$$\lambda_m = m^2\pi^2, \quad \nu_m = m\pi.$$

We assume that the boundary conditions are linear; from them, two relations between the $(n+2)$ scalars z_0, z_m, z'_m, can be obtained. The equations (23) are to be satisfied at n internal points and we thus possess sufficient equations for the determination of the scalars. The most convenient method of setting up the determining equations will often be suggested by the form of the boundary conditions. We may note, however, that the boundary conditions enable one to eliminate two of the coefficients z_0, z_m, z'_m, and thus reduce the number of independent vectors from $n+2$ to n. The numerical elements of these n vectors at the n internal points constitute a matrix A which is not in general an orthogonal matrix; the columns of A and the rows of the inverse A^{-1} can, however, be used as a basis in resolving the terms $p(x)\,dz/dx$ and $q(x)\,z$ into a linear sum of components as in (14). The formation of n equations for the scalars then proceeds as in (16). Higher-difference correction can be investigated as earlier.

A procedure somewhat different from the above has been used by Dennis and Poots (7) who seek a solution of the form

$$z = Y(x) + h(x),$$

$Y(x)$ being a Fourier series (in practice, truncated) and $h(x)$ a function introduced to standardize boundary conditions and to improve convergence.

Numerical examples of the techniques outlined above have been worked out and may be published later.

3·3. *Numerical solution of integral equations in* $(0, 1)$. The numerical solution of integral equations in terms of already given cardinal components is very similar to the solution for differential equations. Numerical integration is a simpler and in general more accurate process than numerical differentiation and at least two ways of resolving the kernel of the equation are available. We outline first a solution obtained by the type of resolution used in § 3·2; use of the inverse relations (8), (9), (10), brings to light formal relations similar to those we have already noted between differential and difference equations. The second way of resolving the kernel will sometimes yield a more accurate numerical solution of an integral equation; it is less useful in the solution of differential equations.

We shall consider the linear Fredholm equation

$$z(x) = f(x) + \int_0^1 K(x, y)\,z(y)\,dy, \qquad (24)$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

and we may suppose that $z(x)$, $f(x)$ and $K(x,y)$ all vanish at $x = 0$ and $x = 1$ (the reduction of the general linear Fredholm equation to this form can always be effected by elementary operations).

The equation involves the integral of a product and we interpolate some remarks on integrals of this type. Since there are two variables here, we modify the notation of the preceding section. We shall use a suffix r for the location index and a suffix m for the value index of the first variable x. The y variable is distinguished by Greek suffixes, ρ for the location index and μ or ν for the value index. We use the notation

$$s_{mp} = \sin(m\pi ph)$$

and similar notation for the orthogonal vectors σ.

For a fixed value of x, we may write

$$K(x,y)\,z(y) = K(x,y)\,z_\nu s_\nu(y) = G_\mu(x)\,s_\mu(y), \tag{25}$$

the integrand being expressed as a linear sum of sine components, *quâ* function of y. The coefficients $G_\mu(x)$ can be determined by forming the scalar product with respect to the y variable, i.e.

$$G_\mu(x) = \sigma_{\rho\mu}[K(x,y)\,z_\nu s_\nu(y)]_\rho = z_\nu \sigma_{\rho\mu}[K(x,y)\,s_\nu(y)]_\rho = z_\nu k_{\mu\nu}(x), \tag{26}$$

where
$$k_{\mu\nu}(x) = \sigma_{\rho\mu}[K(x,y)\,s_\nu(y)]_\rho.$$

In the last equation above, $[K(x,y)\,s_\nu(y)]_\rho$ denotes the single product of $K(x,y_\rho)$ by $s_\nu(y_\rho)$. Each of these products is then multiplied by $\sigma_{\rho\mu}$ and the triple product is then summed over ρ.

The auxiliary notation $G_\mu(x)$ can now be discarded and (25) becomes

$$K(x,y)\,z(y) = z_\nu k_{\mu\nu}(x)\,s_\mu(y). \tag{27}$$

It remains to express $K(x,y)\,z(y)$ in cardinal components *quâ* function of x. We have

$$K(x,y)\,z(y) = z_\nu k_{\mu\nu m} s_m(x)\,s_\mu(y), \tag{28}$$

where
$$k_{\mu\nu m} = \sigma_{mp} k_{\mu\nu}(x)_p.$$

Alternatively, the scalar $k_{\mu\nu m}$ can be expressed in the form

$$k_{\mu\nu m} = \sigma_{mp}[K(x_p,y)\,s_\nu]_\rho\,\sigma_{\rho\mu}. \tag{28a}$$

The integration in equation (25) can now be effected, using (10):

$$\begin{aligned}\int_0^1 K(x,y)\,z(y)\,dy &= z_\nu k_{\mu\nu m} s_m(x) \int_0^1 S_\mu(y)\,dy \\ &= \left.\begin{cases} 2\lambda_\mu^{-1} z_\nu k_{\mu\nu m} s_m(x) & (\mu \text{ odd}), \\ 0 & (\mu \text{ even}), \end{cases}\right\}\end{aligned} \tag{29}$$

where
$$\lambda_\mu = \mu\pi.$$

Writing $f(x) = f_m s_m$, equation (24) becomes

$$\left.\begin{aligned} z_m s_m &= f_m s_m + 2\lambda_\mu^{-1} z_\nu k_{\mu\nu m} s_m, \\ z_p &= f_p + 2\lambda_\mu^{-1} z_\nu k_{\mu\nu p}, \end{aligned}\right\} \tag{30}$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

the summation with respect to μ being over odd integers only. Trapezoidal or mid-ordinate evaluation of the integral would yield a similar result, the denominator λ_μ in the above equations being replaced by

$$\frac{2}{h}\sin(\tfrac{1}{2}\mu\pi h) \quad \text{(mid-ordinate summation)},$$

$$\frac{2}{h}\tan(\tfrac{1}{2}\mu\pi h) \quad \text{(trapezoidal summation)}.$$

The discussion of the higher-difference correction to these approximate integrations is similar to the preceding discussion for differential equations.

The second way of resolving the integrand in (24) is to express $K(x,y)$ and $z(y)$ *separately* in cardinal components, *quâ* functions of y. Here we may use either sine or cosine components; we illustrate by using cosine components. We write

$$K(x,y) = k_{m\mu}c_m(x)\,c_\mu(y), \tag{31}$$

where $$k_{m\mu} = \gamma_{mp}K(x_p,y_\rho)\gamma_{\mu\rho} \quad \text{and} \quad z(y) = z_\nu c_\nu(y).$$

The integrand can now be expressed as a linear sum of cosines in x and a product of cosines in y

$$K(x,y)\,z(y) = z_\nu k_{m\mu}c_m(x)\,c_\mu(y)\,c_\nu(y),$$

and

$$\int_0^1 K(x,y)\,z(y)\,dy = \tfrac{1}{2}z_\mu k'_{m\mu}c_m(x), \tag{32}$$

where

$$k'_{m\mu} = \begin{cases} k_{m\mu} & (\mu \neq 0), \\ 2k_{m\mu} & (\mu = 0). \end{cases}$$

In the last equation above, only the terms $\mu = \nu$ survive on integration; the equations to determine the scalars z_p are

$$z_p = f_p + \tfrac{1}{2}k'_{p\mu}z_\mu. \tag{33}$$

With this form of resolution, trapezoidal or mid-ordinate evaluation of the integral are very closely related to exact integration. Only the coefficients $k'_{m,n+1}$ differ; we have

$$k'_{m,n+1} = \begin{cases} k_{m,n+1} & \text{(exact integration)}, \\ 2k_{m,n+1} & \text{(trapezoidal summation)}, \\ 0 & \text{(mid-ordinate summation)}. \end{cases}$$

Occasionally, the coefficients $k_{m\mu}$ in (31) above assume a simpler form with consequent reduction of the arithmetical work. For example, if the kernel is of the form $h(x)\,k(y)$, the coefficients are of the form $h_m k_\mu$. Again, if the kernel is of the form $K(x-y)$ it may be advisable to resolve it into components of the type

$$\sin m\pi(x-y), \quad \cos m\pi(x-y).$$

We conclude this section with two examples, chosen to illustrate points of general interest rather than numerical subtlety.

Example 1.

$$z(x) = 2\sin 5\pi x \int_0^1 \sin 5\pi y\,z(y)\,dy.$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

The solution of the equation is $z(x) = \sin 5\pi x$. If we seek a numerical solution with $h = \frac{1}{2}$, $\sin 5\pi x$ is indistinguishable from the co-tabular long-period term $\sin \pi x$; and (for this interval) the above equation is equivalent to

$$z(x) = 2 \sin \pi x \int_0^1 \sin \pi y \, z(y) \, dy.$$

The solution of this equation is $z(x) = \sin \pi x$, and it is readily verified that this solution is in fact obtained by the second method of resolution above.

Again, with $h = \frac{1}{4}$, $\sin 5\pi x$ is converted into the co-tabular term $-\sin 3\pi x$, and the integral equation in effect reduces to

$$z(x) = 2 \sin 3\pi x \int_0^1 \sin 3\pi y \, z(y) \, dy;$$

the solution $z(x) = \sin 3\pi x$ is reproduced by resolving into cardinal components. With $h < \frac{1}{5}$, the numerical solution is able to yield the true solution of the integral equation. If the kernel is replaced by a sequence of Fourier components, we can recognize from the above simple example that a tabular solution yields the cardinal function of the solution of the integral equation.

Example 2.

$$z(x) = f(x) + \beta \int_0^1 \exp(x-y) \, z(y) \, dy$$

which has the solution

$$z(x) = f(x) + \frac{\beta}{1-\beta} \int_0^1 \exp(x-y) f(y) \, dy \quad (\beta \neq 1).$$

We denote the cosine representation of $\exp x$ and $\exp(-y)$ by $h_m c_m$ and $k_\mu c_\mu$. The second method of resolving the kernel leads to the equations

$$z_m = f_m + \tfrac{1}{2} \beta h_m k'_\mu z_\mu.$$

Since $k'_\mu z_\mu$ is a scalar, we easily find

$$z_m = f_m + \frac{\frac{1}{2} \beta f_\mu k'_\mu}{1 - \frac{1}{2} \beta h_\mu k'_\mu} h_m.$$

The exponential functions in this example are not particularly well adapted to cosine representation, and the verification of the singularity at $\beta = 1$ in the above discrete solution is an immediate test of the accuracy of the representation. Using $h = \frac{1}{4}$, we obtain

$$\tfrac{1}{2} h_m k'_m = 0{\cdot}99986.$$

A further test of the accuracy of the representation is the value of the constant by which h_m is multiplied in the above solution for z_m. Taking, for example,

$$f(x) = x(1-x) \quad \text{and} \quad f(x) = x^2,$$

(with $\beta = \frac{1}{2}$) the coefficients of h_m are

$$0{\cdot}0967 \quad \text{and} \quad 0{\cdot}1623.$$

The true values of these constants can be obtained from the analytical solution; they are

$$3e^{-1} - 1 = 0{\cdot}1036 \quad \text{and} \quad 2 - 5e^{-1} = 0{\cdot}1606.$$

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press

4. *Use and limitations of matrix representations.*

It may be useful to attempt some assessment of the value of matrix representations of the type used in § 2. A full discussion of this topic falls somewhat outside the scope of the present paper, since it would require a review and comparison of current methods of computation; but it is desirable to indicate, at least in outline, the scope of a representation which provides at once an expressive notation for difference equations and a convenient layout for numerical computations.

It would appear that matrix operators of the form $\mathbf{D}^2$ constitute a useful interpretation of central-difference operations on an *array* of variables, and analogous operators can be devised to interpret forward and backward operations. The concept of eigenfunctions as applied to differential and difference operators emerges naturally from the matrix representation and leads to the more fundamental concept of cardinal component. When cardinal components are available and practicable, the replacement of differential equations and boundary conditions is unnecessary.

Approximation to continuous systems by discrete functions is, however, a field of considerable difficulty, and the difference approximation remains the most widely applicable technique available, particularly in the linearized treatment of non-linear systems. Matrix formulations frequently enable us to devise concise and easily-checked algorithms for the solution of such difference systems.

REFERENCES

(1) WHITTAKER, E. T. *Proc. Roy. Soc. Edinb.* 35 (1915), 181.
(2) RUTHERFORD, D. E. *Proc. Roy. Soc. Edinb.* 62 (1947), 229; 63 (1952), 232.
(3) TODD, JOHN. *Proc. Camb. Phil. Soc.* 46 (1950), 116.
(4) BROMWICH, T. J. I'A. *An introduction to the theory of infinite series* (London, 1926).
(5) LANCZOS, C. *Applied analysis* (London, 1957).
(6) DE LA VALLÉE POUSSIN, CH.-J. *Acad. Roy. Belg. Bull. Cl. Sci.* (1908), 319.
(7) DENNIS, S. C. R. and POOTS, G. *Proc. Camb. Phil. Soc.* 51 (1955), 422.

IMPERIAL COLLEGE
LONDON

UNIVERSITY OF ALBERTA
EDMONTON, CANADA

https://doi.org/10.1017/S0305004100035593 Published online by Cambridge University Press