Accelerating the Jacobi method for solving simultaneous equations by Chebyshev extrapolation when the eigenvalues of the iteration matrix are complex

By H. E. Wrigley

Chebyshev extrapolation has been applied successfully to accelerate the convergence of iterative solutions of simultaneous equations which arise in the numerical solution of partial differential equations. Its use is based upon the assumption that the eigenvalues of the iteration matrix are real. In this paper, the analysis of Chebyshev extrapolation is extended to the case when the eigenvalues of the iteration matrix are complex.

1. Introduction

The numerical solution of many elliptic and parabolic partial differential equations reduces to the problem of solving a set of simultaneous equations of the form

$$\boldsymbol{x} = M\boldsymbol{x} + \boldsymbol{b} \tag{1.1}$$

where $\boldsymbol{x}$ is an unknown vector of length N (say), $\boldsymbol{b}$ is a constant vector of equal length, and M is an $N \times N$ matrix with real elements. If the form of the matrix M is such that a direct solution, $\boldsymbol{x} = (I - M)^{-1}\boldsymbol{b}$ would be too time-consuming, then, provided the eigenvalues of M are bounded by the unit circle, the equations are likely to be solved by the iterative process

$$\boldsymbol{x}^{(n+1)} = M\boldsymbol{x}^{(n)} + \boldsymbol{b}. \tag{1.2}$$

Several schemes have been devised to accelerate the convergence of this iterative process. One of the most powerful is the so-called "Chebyshev Extrapolation" scheme (Stiefel, 1958), which takes the form

$$\tilde{\boldsymbol{x}}^{(n+1)} = M\boldsymbol{x}^{(n)} + \boldsymbol{b}$$

$$\boldsymbol{x}^{(n+1)} = \boldsymbol{x}^{(n-1)} + \alpha_n\{\boldsymbol{x}^{(n)} - \boldsymbol{x}^{(n-1)} + \beta(\tilde{\boldsymbol{x}}^{(n+1)} - \boldsymbol{x}^{(n)})\} \tag{1.3}$$

where α_n and β are extrapolation parameters, the first of which, as the notation implies, varies with iteration number. This formulation of the extrapolation procedure is preferred since, at each iteration, only one parameter, α_n, need be calculated. Also, α_n can be determined from α_{n-1} by means of a simple recurrence relation. It is assumed in applying Chebyshev extrapolation that the eigenvectors $\boldsymbol{u}_j$ of M form a complete set, and that the corresponding eigenvalues, λ_j, are real. Although these properties have been shown to hold (Varga, 1961) for only a limited number of idealized problems, the success of the method in application confirms the belief that they hold for a wide variety of practical cases also.

There are problems expressable in the form of equation (1.1) for which the eigenvalues of the matrix M can be complex. The author had applied Chebyshev extrapolation to a number of such cases without appreciating that the eigenvalues were complex. The convergence of the iteration scheme had been accelerated, although in a number of cases, subsequently identified as those in which the imaginary parts of the eigenvalues were large, the acceleration achieved was not great. Several questions arose naturally from these observations.

1. Why does Chebyshev extrapolation produce an acceleration in the complex eigenvalue case?
2. Is Chebyshev extrapolation the best acceleration scheme in these circumstances?
3. How does its effectiveness depend on the imaginary parts of the eigenvalues?
4. What information on the eigenvalue distribution is required in order to make the best use of the extrapolation procedure?

In Section 3, question 1 is answered by showing that the answer to question 2 is in the affirmative provided the eigenvalues lie within a particular region of the complex plane. It is also shown that question 3 can be answered by considering the effect on this region of a change in the imaginary parts of the eigenvalues. In the same Section, the theory of Chebyshev extrapolation is extended to consider complex eigenvalues, and the relationship between the convergence rate and certain acceleration parameters is determined. The dependence of the optimum values of these parameters upon certain eigenvalues of the matrix M is established. In Section 4, it is shown how the required eigenvalues may be determined during the course of the calculations. Iterations begun with non-optimum values of the acceleration parameters can be interrupted and the extrapolation process re-started using the optimum values to give an improvement in the convergence rate. In Section 5, a strategy is described which can be applied in certain cases to increase the effectiveness of Chebyshev extrapolation when the imaginary parts of the eigenvalues are large. Finally, in Section 6, there is a discussion on the types of problem to which the method described in this paper can be applied.

Before these points are considered, however, a brief description of Chebyshev extrapolation as applied in the real-eigenvalue case is presented, together with formulae which are needed later for comparison with the complex-eigenvalue case.

2. Real eigenvalues

Application of Chebyshev extrapolation in the real-eigenvalue case requires a knowledge of two parameters, a and b, which are respectively the upper and lower bounds of the eigenvalues of M. The extrapolation parameters α_n and β are then given in terms of a and b as follows:

$$\begin{aligned} \beta &= 2/(2-a-b) \\ \alpha_0 &= 1 \\ \alpha_n &= 2\gamma T_n(\gamma)/T_{n+1}(\gamma) \text{ for } n \geqslant 1 \end{aligned} \tag{2.1}$$

where $\gamma = (2-a-b)/(a-b)$ and $T_n(\gamma)$ is the Chebyshev polynomial of degree n in γ. For $\gamma < 1$, $T_n(\gamma) = \cos(n \cos^{-1}\gamma)$ while, for $\gamma > 1$,

$$T_n(\gamma) = \cosh(n \cosh^{-1}\gamma).$$

For $n > 1$, α_n can be determined from α_{n-1} by means of the recurrence relation

$$\alpha_n = [1 - (\alpha_{n-1}/4\gamma^2)]^{-1}.$$

Let $\epsilon^{(n)}$ be the error in $x^{(n)}$ so that $x^{(n)} = x + \epsilon^{(n)}$, where x is the solution of equation (1.1). Then, if $\epsilon^{(0)}$ is expressed in terms of the eigenvectors of M as follows:

$$\epsilon^{(0)} = \sum_{j=1}^{N} a_j u_j$$

it is a simple matter to show, using equations (2.1) and (1.3) that the error $\epsilon^{(n)}$ is given by

$$\epsilon^{(n)} = \sum_{j=1}^{N} a_j \left\{\frac{T_n(\gamma_j)}{T_n(\gamma)}\right\} u_j \tag{2.2}$$

where $\gamma_j = (2\lambda_j - a - b)/(a-b)$. As λ_j varies from b to a, γ_j varies from -1 to $+1$. Thus over the range of the eigenvalues, $T_n\{(2\lambda - a - b)/(a-b)\}/T_n(\gamma)$ is the smallest polynomial, of degree n in λ, which is equal to unity when $\lambda = 1$, in the sense that its maximum absolute value is smaller than that of any other polynomial of the same degree. For this reason, Chebyshev extrapolation is considered to optimize the convergence rate, and the optimum values of a and b are defined as those for which $a = \lambda_1$, the largest eigenvalue, and $b = \lambda_N$, the smallest eigenvalue.

In general, the eigenvalues of the matrix M will not be known, and the extrapolation process is likely to be applied with non-optimum values of a and b. If $a > \lambda_1$ and $b < \lambda_N$, the reduction in convergence rate need not be serious, since γ_j, and hence $T_n(\gamma_j)$ remains bounded by unity. However, convergence is retarded because the quantity γ is reduced by increasing the value of $(a-b)$. On the other hand, if any of the eigenvalues lie outside the range (b, a), the convergence rate can be seriously affected and the extrapolation can cause the iteration process to diverge. For eigenvalues outside the range (b, a), $|\gamma_j| > 1$. Thus

$$T_n(\gamma_j) = \tfrac{1}{2}\exp[n \cosh^{-1}\gamma_j] + \tfrac{1}{2}\exp[-n \cosh^{-1}\gamma_j]$$

and the ratio $|T_n(\gamma_j)/T_n(\gamma)|$ is asymptotically equal to $\exp\{n[\cosh^{-1}\gamma_j - \cosh^{-1}\gamma]\}$. If $|\gamma| > |\gamma_j| > 1$, $|T_n(\gamma_j)/T_n(\gamma)| \to 0$ as $n \to \infty$, and the extrapolation process is convergent. If $|\gamma| = |\gamma_j| > 1$, $|T_n(\gamma_j)/T_n(\gamma)| \to 1$ as $n \to \infty$, and the process will not converge for general $\epsilon^{(0)}$. If $|\gamma_j| > |\gamma| > 1$, $|T_n(\gamma_j)/T_n(\gamma)| \to \infty$ as $n \to \infty$, and the process will, in general, diverge. $|\gamma_j| > |\gamma|$ if $\lambda_j > 1$ or $\lambda_j < a + b - 1$. The first condition can be discounted, but the second can be satisfied if care is not exercised in the choice of b.

The reduction factor $\rho_j^{(n)}$ for eigenvalue λ_j over iteration n is seen from equation (2.2) to be

$$\rho_j^{(n)} = \{T_{n-1}(\gamma_j)\, T_n(\gamma)\}/\{T_n(\gamma_j) T_{n-1}(\gamma)\}.$$

For those eigenvalues such that $|\gamma_j| \geqslant 1$, $\rho_j^{(n)}$ tends to a limit ρ_j independent of n for n sufficiently large. The reciprocal of ρ_j is termed the *asymptotic convergence rate*, μ_j, of the jth eigenvector contribution, and is defined by the equation

$$\mu_j = \exp\{\cosh^{-1}\gamma_j - \cosh^{-1}\gamma\}. \tag{2.3}$$

The asymptotic convergence rate, μ, of the extrapolation process is defined by $\mu = \max_{j \in J} \mu_j$ where $J = \{j;\ |\gamma_j| \geqslant 1\}$. No asymptotic convergence rate, in the sense defined above, exists for eigenvectors such that $|\gamma_j| < 1$, although it would not be correct to infer from this that nothing is known about the convergence rate of the extrapolation process in the case when all eigenvalues are real and $|\gamma_j| < 1$ for all j. (See, for example, Golub and Varga, 1961.)

3. Complex eigenvalues

Before the optimum covergence parameters a and b are defined for the complex-eigenvalue case, some of the fundamental properties of Chebyshev polynomials of complex argument will be established. The following definition is taken for any value of the complex argument z:

$$T_n(z) = \cosh(n \cosh^{-1} z). \tag{3.1}$$

The alternative definition, $T_n(z) = \cos(n \cos^{-1} z)$, could have been taken; only the detail of the following analysis would have been changed. It is convenient to introduce the variables u and v defined by the equation

$$z = x + iy = \cosh(u + iv) \tag{3.2}$$

so that u and v can be determined from the equations

$$x = \cosh u . \cos v, \quad y = \sinh u . \sin v. \tag{3.3}$$

Since cosh and cos are both even functions, while sinh and sin are both odd, u and v will be of indeterminate sign. The u-positive solution can be taken without loss of generality.

Consider the special case of z real. Then $y = 0$ and either $u = 0$ or $v = 0$ or π. If $x > 1$, $v = 0$ and $T_n(z) = \cosh(nu) = \cosh(n \cosh^{-1} x)$. If $x < -1$, $v = \pi$ and $u = \cosh^{-1}(-x)$. Thus

$$\begin{aligned} T_n(z) &= \cosh(n \cosh^{-1}(-x) + inv) \\ &= (-1)^n \cosh(n \cosh^{-1}(-x)). \end{aligned}$$

Finally, if $|x| < 1$, $u = 0$ and

$$T_n(z) = \cosh(inv) = \cos nv = \cos(n \cos^{-1} x).$$

Thus equation (3.1) is consistent with the usual definitions of Chebyshev polynomials of real argument. Further, it is a simple matter to show that Chebyshev polynomials of complex argument satisfy the well-known recurrence relationship $T_{n+1}(z) = 2zT_n(z) - T_{n-1}(z)$. Thus the analysis of Chebyshev extrapolation for real eigenvalues applies in the case of complex eigenvalues also. For α_n and β defined as in equation (2.1) with real a and b, the error after n iterations is given by equation (2.2) in which the numerator polynomials for complex eigenvalues are defined by equation (3.1).

Two further properties of Chebyshev polynomials of complex argument are established in the following lemmas:

Lemma 1: Whereas $|T_n(x)| < 1$ for all n when x is real and $|x| < 1$, $|T_n(z)| \to \infty$ as $n \to \infty$ when z is complex, *even though* $|z| < 1$.

Proof: From equations (3.1) and (3.2)

$$|T_n(z)|^2 = \cosh^2(nu)\cos^2(nv) + \sinh^2(nu).\sin^2(nv)$$
$$= \cosh^2(nu) - \sin^2(nv).$$

$\cosh^2(nu)$ is an increasing function of n while $\sin^2(nv)$ remains bounded by unity. Thus, unless $u = 0$ (z real, $|z| < 1$), $|T_n(z)| \to \infty$ as $n \to \infty$.

Lemma 2: If x is real, the ratio $T_{n+1}(x)/T_n(x)$ tends to the limit $\exp(\cosh^{-1} x)$ for large n only if $x > 1$. On the other hand, if z is complex, the ratio $T_{n+1}(z)/T_n(z)$ tends to the limit $\exp(\cosh^{-1} z)$ for *all* values of z.

Proof: From equations (3.1) and (3.2)

$T_n(z) = \cosh(nu + inv) \approx \frac{1}{2}\exp(nu + inv)$ for large n, since the u-positive solution of equations (3.3) has been chosen. Thus

$T_{n+1}(z)/T_n(z) \approx \exp(u + iv) = \exp(\cosh^{-1} z)$ for large n.

The first lemma implies that, after a sufficient number of iterations, the error will be dominated by contributions from eigenvectors whose eigenvalues are either complex or real and lie outside the interval (b, a). The second lemma implies that, after a sufficient number of iterations, these dominant error contributions decay at a steady rate given by the μ_j of equation (2.3).

Thus, after a sufficient number of iterations, and considering only those eigenvectors which at that stage are likely to dominate the error, the effect of the extrapolation process is to change the eigenvalues of the problem from the λ_j of matrix M to the μ_j of equation (2.3). Taking the logarithmic form of $\cosh^{-1}$, equation (2.3) can be written

$$\mu_j = \frac{[(\lambda_j - a)^{1/2} + (\lambda_j - b)^{1/2}]^2}{[(1 - a)^{1/2} + (1 - b)^{1/2}]^2} \tag{3.4}$$

or, expressing λ_j in terms of μ_j and deleting the suffix,

$$\frac{2\lambda - a - b}{a - b} = \frac{1}{2}\left\{\frac{c\mu}{a - b} + \frac{a - b}{c\mu}\right\} \tag{3.5}$$

where $c = [(1 - a)^{1/2} + (1 - b)^{1/2}]^2$. Define the variables t and w by means of the equations

$$t = (2\lambda - a - b)/(a - b), \quad \mathrm{w} = c\mu/(a - b).$$

Then equation (3.5) becomes a special case of the Joukowski transformation

$$t = \frac{1}{2}\left(\mathrm{w} + \frac{1}{\mathrm{w}}\right). \tag{3.6}$$

Consider the ellipse in t-space with semi-axes $\frac{1}{2}(k \pm 1/k)$ where k is some parameter. Then, according to equation (3.6), the ellipse is mapped into circles in w-space of radii k and $1/k$. The interior of the ellipse is mapped into the space between the circles. From the point of view of the iteration process, if $\lambda = x + iy$, then the ellipse in λ-space with the equation

$$\frac{[(2x - a - b)/(a - b)]^2}{[\frac{1}{2}(k + 1/k)]^2} + \frac{[2y/(a - b)]^2}{[\frac{1}{2}(k - 1/k)]^2} = 1 \tag{3.7}$$

is mapped into concentric circles in μ-space with centres at the origin and radii μ_0 and $1/\mu_0$ where

$$\mu_0 = k(a - b)/[(1 - a)^{1/2} + (1 - b)^{1/2}]^2. \tag{3.8}$$

The interior of the ellipse is mapped into the annulus between these two circles.

Before these results can be applied, it is necessary to discuss what is meant by optimum convergence. In the real-eigenvalue case, optimum convergence is obtained by setting a and b equal to the largest and smallest eigenvalues, respectively. Consequently, all eigenvector contributions to the error decay at approximately the same rate. In the complex-eigenvalue case, the situation is different. Eigenvalues, λ, which lie on the ellipse defined by equation (3.7) are transformed into eigenvalues, μ, which lie on the circle of radius μ_0 given by equation (3.8). Since it is impossible to find an ellipse which passes through all the eigenvalues, λ, it is impossible to find values of a and b in the complex case which make all eigenvector contributions decay at the same rate. However, the ellipse defined by equation (3.7) can be made to pass through any two points in the complex plane. In particular, it can be made to pass through the eigenvalues corresponding to the two largest error contributions. Values of a and b can then be found which minimize the radius, μ_0, of the circle into which this ellipse is transformed. Thus the two largest error contributions can be made to decay at the same rate which will be, moreover, the fastest rate for these two eigenvectors considered together. In the present paper, this is taken as the definition of optimum convergence in the complex-eigenvalue case. Eigenvector contributions whose eigenvalues lie outside the ellipse will decay at a slower rate, but, since the two largest error contributions have been considered explicitly, it is often found that such contributions are small. Their slower convergence rate does not necessarily delay the overall convergence of the iteration process.

Let λ_d be the eigenvalue of the dominant error contribution, and λ_s the eigenvalue of the sub-dominant contribution. Assume λ_d and λ_s to be known. The ellipse in λ-space defined by equation (3.7) is to pass through these two points. Thus, given k, substitution in equation (3.7) enables corresponding values of a and b to be determined. Substitution of these values in equation (3.8) gives the asymptotic convergence rate which would result for the two largest error contributions if Chebyshev extrapolation were to be applied using the derived values of a and b. It is desired to minimize μ_0, i.e. to find the value of k, and hence of a and b, for which $\mu_0' = 0$ where the prime indicates differentiation with respect to k. It is convenient at this point to introduce the variables $\theta = (2 - a - b)$ and $\eta = (a - b)$, so that $\gamma = \theta/\eta$. Then

$$\mu_0 = k/[\gamma + (\gamma^2 - 1)^{1/2}]$$

and

$$\mu_0' = \frac{\mu_0}{k}\{1 - [k\gamma'/(\gamma^2 - 1)^{1/2}]\}.$$

In equation (3.7), a and b can be expressed in terms of θ and η. Substitution of λ_d and λ_s gives two equations from which expressions for θ and η in terms of λ_d, λ_s and k can be obtained. Differentiation with respect to k gives expressions for θ' and η'. Since $\eta^2\gamma' = \eta\theta' - \theta\eta'$, the values of γ and γ' and hence of μ_0 and μ_0' corresponding to any value of k can be obtained. The iterative method of Muller (1956) for solving algebraic equations can then be applied systematically to find the required zero of μ_0'.

In the next Section, a method of determining the required eigenvalues, λ_d and λ_s, is described. Before this is undertaken, however, consideration should be given to the question of whether Chebyshev extrapolation is the optimum acceleration scheme in the complex-eigenvalue case.

In Section 2, Chebyshev extrapolation was termed optimum because the polynomials on the right of equation (2.2) were in some sense the smallest over the range of the eigenvalues. In the complex-eigenvalue case, one would expect these to be replaced by polynomials which possess, over the domain in the complex plane containing the eigenvalues, the mini-max property associated with Chebyshev polynomials of real argument in the range $(-1, +1)$. However, the eigenvalues are a set of points in the complex plane. They do not form a continuum and hence can be considered to lie within an infinity of different domains. Moreover, the polynomial possessing the required mini-max property will depend upon the choice of domain. Clayton (1963) has shown that if $P_n(z)$ is the class of nth degree polynomials with real coefficients and equal to unity when $z = 1$, then that member, $P_n^*(z)$, which over the *ellipse* of semi-major axis A (< 1) and eccentricity E has the smallest maximum modulus value is given by

$$P_n^*(z) = T_n(z/AE)/T_n(1/AE). \qquad (3.9)$$

If the centre of the ellipse, still assumed to lie within the unit circle, is transferred to the point $(C, 0)$ the mini-max polynomial takes the form

$$P_n^*(z) = T_n\{(z - C)/AE\}/T_n\{(1 - C)/AE\}. \quad (3.10)$$

In applying Chebyshev extrapolation, it is assumed that the eigenvalues are contained within some member of the family of ellipses defined by equation (3.7). Writing this equation in the form

$$\frac{\{x - \frac{1}{2}(a + b)\}^2}{\{\frac{1}{4}(a - b)(k + 1/k)\}^2} + \frac{y^2}{\{\frac{1}{4}(a - b)(k - 1/k)\}^2} = 1,$$

it can be seen that these ellipses have their centres at the point $(\frac{1}{2}(a + b), 0)$ and are such that the product $AE = \frac{1}{2}(a - b)$. The mini-max polynomial over their interior is, from the equation (3.10),

$$P_n^*(z) = T_n\{(2z - a - b)/(a - b)\}/T_n\{(2 - a - b)/(a - b)\}.$$

Comparison with equation (2.2) indicates that, over the ellipse assumed to contain the eigenvalues, Chebyshev extrapolation as described in Section 2 is the optimum acceleration scheme.

This conclusion will no longer hold if the shape of the domain is changed. For example, Zarantonello (1957) has shown that over a circle of radius $A < 1$, the mini-max polynomial, $P_n^*(z)$, is simply z^n. That is, if the domain containing the eigenvalues is assumed to be circular, no extrapolation scheme can be devised which will improve the convergence rate of the iteration process of equation (1.2). Determination of the bounding ellipse in the above analysis was seen to require a knowledge of at least two eigenvalues. To delineate a more irregularly shaped domain would require further information on the eigenvalue distribution. If the situation is to be avoided in which more time is spent in determining the eigenvalues than in solving the simultaneous equations, it would seem that, except in special circumstances, Chebyshev extrapolation will be the best acceleration scheme to apply.

To answer question 1 of the introduction, it is necessary only to note that z^n belongs to the class of polynomials of which $P_n^*(z)$ defined by equation (3.9) is the member possessing the mini-max property over the ellipse. Question 3 can be answered by noting that the eccentricity of the ellipse will be small when the imaginary parts of the eigenvalues are large. For small E,

$$T_n(z/AE) \approx \tfrac{1}{2}\exp(n\cosh^{-1}(z/AE))$$

$$= \tfrac{1}{2}\left(\frac{z}{AE} + \left[\left(\frac{z}{AE}\right)^2 - 1\right]^{1/2}\right)^n$$

$$\approx 2^{n-1}(z/AE)^n.$$

Hence $$\lim_{E \to 0}\left\{\frac{T_n(z/AE)}{T_n(1/AE)}\right\} = z^n.$$

Thus, as the ellipse containing the eigenvalues tends to a

circle, the mini-max polynomial degenerates into Z^n, and the acceleration achieved by Chebyshev extrapolation decreases to zero.

4. An example

In general, when applying Chebyshev extrapolation to the complex-eigenvalue case, neither the eigenvectors which dominate the error nor the corresponding eigenvalues are likely to be known. Thus the extrapolation process must be started with estimated values of a and b, and an attempt made to extract from the calculations the information required to compute their correct, i.e. optimum, values. For any a and b, the asymptotic convergence rate of the error contributions which ultimately dominate will be given by equation (3.4). If μ_d and μ_s can be determined from the iteration process, the eigenvalues λ_d and λ_s can be deduced from equation (3.5). The optimum value of k, and hence the optimum values of a and b, can be found from equations (3.7) and (3.8). The *extrapolation* process can then be restarted with consequent improvement in the convergence rate.

As an illustration of this technique, consider the following example. The solution of equation (1.1) is required for the case when

$$M = \begin{bmatrix} 0{\cdot}9205 & -0{\cdot}8526 & 0{\cdot}3265 & 0{\cdot}3054 \\ 1{\cdot}0961 & -0{\cdot}6522 & 0{\cdot}8152 & 0{\cdot}3284 \\ 0{\cdot}0677 & 0{\cdot}2922 & 0{\cdot}8561 & -0{\cdot}1328 \\ -0{\cdot}9395 & 0{\cdot}8977 & -0{\cdot}5330 & 0{\cdot}6556 \end{bmatrix}$$

and $b' = (0{\cdot}4999, -0{\cdot}5573, -0{\cdot}0876, 1{\cdot}0876)$.

The eigenvalues of M are

$$\lambda_1 = 0{\cdot}9612,\ \lambda_2 = 0{\cdot}8018,\ \lambda_3 = 0{\cdot}0064 + 0{\cdot}3981i \text{ and } \bar{\lambda}_3.$$

The corresponding eigenvectors are:

$$u_1' = (-0{\cdot}1798, 0{\cdot}3188, 0{\cdot}9228, -0{\cdot}1204)$$
$$u_2' = (0{\cdot}1890, 0{\cdot}3645, 0{\cdot}0309, 0{\cdot}9113)$$
$$u_3' = (0{\cdot}3732, 0{\cdot}5133, -0{\cdot}1410, -0{\cdot}2782) + i(0{\cdot}5715, 0{\cdot}3454, -0{\cdot}2322, -0{\cdot}0118)$$

and $\bar{u}_3$.

The solution of the equations is $x' = (1, 2, 3, 1)$.

The iteration process is started with an estimate, $x^{(0)}$, of x given by $x^{(0)} = (1{\cdot}7556, 3{\cdot}7099, 3{\cdot}6717, 1{\cdot}2344)$. By this choice,

$$\epsilon^{(0)} = x^{(0)} - x = (0{\cdot}7566, 1{\cdot}7099, 0{\cdot}6717, 0{\cdot}2344) = u_1 + u_2 + u_3 + \bar{u}_3.$$

All eigenvectors contribute equally to the initial error which is therefore in its most general form. The initial values of a and b are 0·75 and zero, respectively. These values are quite arbitrary and it remains to be seen whether or not they are satisfactory. The results of the calculations are summarized in Table 1.

It is seen that a poor choice of extrapolation parameters has been made since the estimate of x after 17 iterations is not markedly closer to the true solution that it was initially. Before the convergence rate can be improved, however, it is necessary to determine from Table 1 the asymptotic convergence rates μ_d and μ_s of the dominant and sub-dominant error contributions. These could be determined from the vector sequence $\epsilon^{(n)}$, $\epsilon^{(n+1)}$, etc., if the absolute error in each estimate of the solution were known. In the absence of such information, it is necessary to consider the quantity

$$y^{(n)} = x^{(n+1)} - x^{(n)} = \epsilon^{(n+1)} - \epsilon^{(n)}.$$

The vectors $y^{(n)}$, $y^{(n+1)}$, etc., form a sequence in which the contributions from the dominant eigenvectors decay asymptotically at the same rate as in the vector sequence $\epsilon^{(n)}$, $\epsilon^{(n+1)}$, etc. Thus both μ_d and μ_s can be determined from $y^{(n)}$. Values of $y^{(n)}$ derived from Table 1 are given in Table 2.

Table 1

ITERATION NUMBER n	ESTIMATE OF SOLUTION, x			
	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$x_4^{(n)}$
0	1·7556	3·7099	3·6717	1·2344
1	−0·2078	1·5025	4·3486	1·8526
2	1·7435	2·7345	3·4543	1·3552
4	−0·3292	1·0704	4·1466	1·4619
6	0·7731	2·5601	3·6317	0·5550
8	2·1577	3·1209	2·9797	0·7358
10	0·3158	1·2485	3·6339	1·4580
11	2·1604	3·2177	2·8871	0·6055
12	0·0759	1·7177	3·6984	0·8646
13	0·7347	1·5148	3·3839	1·4074
14	2·0249	3·1984	2·8705	0·5233
15	−0·1270	1·3875	3·7051	1·0404
16	1·1527	1·8393	3·1540	1·3099
17	1·7745	3·0644	2·9151	0·4926

Table 2

ITERATION NUMBER n	$y^{(n)} = x^{(n+1)} - x^{(n)}$			
	$y_1^{(n)}$	$y_2^{(n)}$	$y_3^{(n)}$	$y_4^{(n)}$
10	1·8446	1·9692	−0·7469	−0·8525
11	−2·0845	−1·4999	0·8113	0·2591
12	0·6588	−0·2029	−0·3145	0·5428
13	1·2902	1·6836	−0·5134	−0·8842
14	−2·1519	−1·8108	0·8346	0·5171
15	1·2798	0·4518	−0·5511	0·2695
16	0·6218	1·2251	−0·2389	−0·8173
17	−1·9786	−1·9085	0·7612	0·7054

The variation in sign and magnitude of the elements of $y^{(n)}$ indicates the presence of a complex dominant eigenvalue, μ_d. This can be determined by the method described by Wilkinson (1954). If μ_d and $\bar{\mu}_d$ are the roots of the equation $\mu^2 + p\mu + q = 0$, then for sufficiently large n

$$y^{(n+2)} + py^{(n+1)} + qy^{(n)} = 0.$$

This relationship will be satisfied in particular for any two components of the vectors $y^{(n)}$, $y^{(n+1)}$ and $y^{(n+2)}$. Thus

$$y_i^{(n+2)} + py_i^{(n+1)} + qy_i^{(n)} = 0$$
$$y_j^{(n+2)} + py_j^{(n+1)} + qy_j^{(n)} = 0$$

giving

$$p = \frac{y_i^{(n+2)}y_j^{(n)} - y_i^{(n)}y_j^{(n+2)}}{y_i^{(n)}y_j^{(n+1)} - y_i^{(n+1)}y_j^{(n)}},$$

$$q = \frac{y_i^{(n+1)}y_j^{(n+2)} - y_i^{(n+2)}y_j^{(n+1)}}{y_i^{(n)}y_j^{(n+1)} - y_i^{(n+1)}y_j^{(n)}}.$$

The values of p and q estimated from the ith and jth elements of the vectors $y^{(n)}$, $y^{(n+1)}$ and $y^{(n+2)}$ are denoted by $p_n(i,j)$ and $q_n(i,j)$, respectively. In Table 3, the values of p_n and q_n are recorded for $n = 14, 15$ for all combinations of i and j.

In a practical problem, it would not be possible to consider all estimates of p and q. A number of element combinations would be selected, and the values of p and q computed for these points only. However, some of the elements chosen may retain, except for very large n, significant components of eigenvectors other than the dominant one. Such elements give rise to poor estimates of p and q as is shown in Table 3 for i or $j = 3$. To overcome this, a further selection process is required. The following method has worked well in practice. Continue the calculation until successive estimates of p for some element combination agree to within a certain prescribed accuracy. Let this be P, and let the value of q_n for this same element combination be Q. Then form the average of those p-values which differ from P by less than (say) 4%. Estimate q by forming the average of those q-values which differ from Q by the same amount. The roles of P and Q can be interchanged if desired. For example, in Table 3, $q_{14}(1,2)$ and $q_{15}(1,2)$ agree to two decimal places, i.e. $Q = 0{\cdot}96$, $P = 1{\cdot}20$. All $q_{15}(i,j)$, except those in which either i or $j = 3$, differ from Q by less than 4%. Their mean is 0·98. All $p_{15}(i,j)$, except those in which $j = 3$, differ from P by less than 4%. Their mean is 1·18. Thus μ_d and $\bar{\mu}_d$ are the roots of the equation

$$\mu^2 + 1{\cdot}18\mu + 0{\cdot}98 = 0$$

i.e. $\mu = -0{\cdot}59 \pm 0{\cdot}79i$ and $|\mu| = 0{\cdot}986$. Substitution in equation (3·5) gives $\lambda_d = 0{\cdot}006 \pm 0{\cdot}394i$. Thus the error is dominated by contributions from the complex eigenvectors u_3 and u_3. Further, their asymptotic convergence rate is of modulus 0·986, so that after 17 iterations their error contribution is reduced by a factor of only 1·25, which explains why the error was still large when the iterations were discontinued. The asymptotic convergence rate, μ_s, of the sub-dominant error contribution is determined by forming the vector sequence $z^{(n)}$ defined by the equation

$$z^{(n)} = y^{(n+2)} + py^{(n+1)} + qy^{(n)}. \qquad (4.1)$$

The dominant contribution to $z^{(n)}$ is made by the eigenvector which provides the sub-dominant contribution to $y^{(n)}$. The values of $z^{(n)}$ derived from Table 2 by means of equation (4.1), in which $p = 1{\cdot}18$ and $q = 0{\cdot}98$, are given in Table 4.

The elements of $z^{(n)}$ are of constant sign and are generally decreasing. This implies that μ_s is real, positive and less than unity, and the fact that the largest element is $z_3^{(n)}$ identifies the sub-dominant error contribution with the eigenvector u_1. Successive estimates of μ_s are defined by the relation $\mu_s^{(n)} = (z^{(n+1)}, z^{(n)})/(z^{(n)}, z^{(n)})$, and their values are recorded in Table 5.

Differences between successive estimates decrease in magnitude up to $n = 13$, and thereafter increase. This is because the values of p and q deduced from Table 3

Table 3

i \ j	p_{14}			q_{14}		
	2	3	4	2	3	4
1	1·12	0·04	1·16	0·96	0·31	0·98
2		0·95	1·17		0·91	0·97
3			1·10			1·01
	p_{15}			q_{15}		
	2	3	4	2	3	4
1	1·20	3·15	1·18	0·96	0·02	0·97
2		1·25	1·19		0·84	1·00
3			1·16			0·88

Table 4

ITERATION NUMBER n	$z^{(n)} = y^{(n+2)} + py^{(n+1)} + qy^{(n)}$			
	$z_1^{(n)}$	$z_2^{(n)}$	$z_3^{(n)}$	$z_4^{(n)}$
10	0·0068	−0·0431	−0·0891	0·0132
11	0·0248	−0·0258	−0·0894	0·0103
12	0·0162	−0·0231	−0·0794	0·0058
13	0·0049	−0·0351	−0·0694	0·0132
14	0·0230	−0·0164	−0·0713	0·0075
15	0·0093	−0·0201	−0·0608	0·0052

Table 5

ITERATION NUMBER (n)	10	11	12	13	14
$\mu_s^{(n)}$	0·937	0·870	0·907	0·919	0·828

are only approximate, so that the sequence $z^{(n)}$ retains contributions from the complex eigenvectors u_3 and $\bar{u}_3$. Moreover, since these contributions decay more slowly than those from u_1, the convergence rate of the sequence $z^{(n)}$ will eventually become that of the complex eigenvectors. Only by improving the estimates of p and q can this difficulty be avoided. However, for practical purposes, it is generally adequate to truncate the sequence $\mu_s^{(n)}$ as soon as differences between successive estimates start to increase, and to determine the limit of the sequence from the last few reliable values. If these values oscillate, the mean of the last two generally gives an improved estimate of the limit. If the last two differences are of the same sign, Aitken's δ^2-process can generally be applied. This is the case in the present example where extrapolation using $n = 11 - 13$ gives $\mu_s = 0{\cdot}921$. Substitution in equation (3.5) gives $\lambda_s = 0{\cdot}961$. With $\lambda_d = 0{\cdot}006 + 0{\cdot}394i$ and $\lambda_s = 0{\cdot}961$, equations (3.7) and (3.8) give an optimum value of $k = 1{\cdot}627$. The corresponding values of a and b are 0·867 and −0·697, respectively. The value of μ_0 for the two dominant error modes is 0·915. For the correct values, $\lambda_d = 0{\cdot}0064 + 0{\cdot}3981i$ and $\lambda_s = 0{\cdot}9612$, optimum $k = 1{\cdot}635$ giving $a = 0{\cdot}865$, $b = -0{\cdot}697$ and $\mu_0 = 0{\cdot}915$. Note that the largest change from the initial estimate is in the value of b. This often proves to be the case. If the extrapolation process is divergent with the initial estimates of a and b, it is likely to be due to a poor choice of b. Also, the above technique is likely to break down if μ_d and μ_s are of equal modulus. If this is the case, it will be necessary to interrupt the calculations and re-start the extrapolation process using different values of a and b. The accuracy with which the optimum values of a and b are computed in this example is due, in some measure, to the simple nature of the problem. Nevertheless, it serves to illustrate the feasibility of the method for large-scale calculations.

From iteration 18 onwards, the optimum convergence parameters can be used, although a further 78 iterations will be required to reduce the error by a factor of $1/(0{\cdot}915)^{78} \approx 10^3$. In the absence of Chebyshev extrapolation, a total of 174 iterations will be required to reduce the error by a factor of $1/(0{\cdot}961)^{174} \approx 10^3$. If the expected convergence rate is not realized when the extrapolation process is re-started, it implies that eigenvector contributions other than those originally considered have come to dominate the error. It has then to be decided whether to accept the new asymptotic convergence rate or to repeat the above calculations to deduce the new optimum values of a and b.

5. A strategy

The optimum asymptotic convergence rate for the above problem is 0·915. The convergence rate of the unextrapolated scheme is 0·961. Thus extrapolation doubles the rate at which the error is reduced. Chebyshev extrapolation applied to a real-eigenvalue case would generally achieve a greater improvement over the unextrapolated scheme. The reason for the relatively poor improvement in the above example is the existence of a complex eigenvalue, λ_3, with a substantial imaginary part. A strategy is described below which overcomes this difficulty in the present example by effectively reducing the imaginary parts of the complex eigenvalues. When no such technique is available, Chebyshev extrapolation is likely to achieve a small, though none the less welcome, acceleration of the iteration process.

The following strategy is applicable whenever the squares of the eigenvalues have smaller imaginary parts than the eigenvalues themselves, i.e. whenever the complex eigenvalues have real parts of modulus less than $\frac{1}{2}$. This occurs in many physical problems where the high-order modes, and hence the smaller eigenvalues, are complex due to inadequacies in either the mathematical or numerical treatment. As an example, consider the matrix formed by squaring the 4×4 matrix introduced in Section 4. This will have eigenvalues $\lambda_1^2 = 0{\cdot}9238$, $\lambda_2^2 = 0{\cdot}6429$, $\lambda_3^2, \bar{\lambda}_3^2 = -0{\cdot}1584 \pm 0{\cdot}0051i$. With $\lambda_d = \lambda_3^2$ and $\lambda_s = \lambda_{12}^2$ the optimum value of $k = 1{\cdot}02$, giving $a = 0{\cdot}924$, $b = -0{\cdot}209$ and $|\mu| = 0{\cdot}610$. Thus Chebyshev extrapolation applied to every other iterate with these values of a and b results in a mean convergence rate per iteration of $(0{\cdot}61)^{1/2} = 0{\cdot}78$. Further, since $(0{\cdot}96)^6 \approx (0{\cdot}92)^3 \approx 0{\cdot}78$, Chebyshev extrapolation every other iteration is three times faster than Chebyshev extrapolation every iteration, and six times faster than no extrapolation at all. The calculations reported in the previous Section were repeated with Chebyshev extrapolation applied every other iteration using values of $a = 0{\cdot}924$ and $b = -0{\cdot}209$. The results of the calculation are summarized in Table 6. Note that the error is reduced by a factor of about ten every nine iterations, so that the solution has been achieved to the required accuracy after a total of 28 iterations.

Table 6

ITERATION NUMBER (n)	ESTIMATE OF SOLUTION			
	$x_1^{(n)}$	$x_2^{(n)}$	$x_3^{(n)}$	$x_4^{(n)}$
0	1·7556	3·7099	3·6717	1·2344
10	1·0318	2·1500	3·1262	1·0036
18	1·0072	2·0166	3·0117	0·9709
28	1·0006	2·0017	3·0010	0·9984

6. Concluding remarks

The method outlined in Section 4 for determining the optimum convergence parameters is too complicated to be worth while if the equations (1.1) are to be solved only once. If this is the case, and complex eigenvalues are known to be present, it is probably better not to apply Chebyshev extrapolation unless one has a reasonable approximation to the optimum values of the convergence parameters. For many problems, however, equations of the type (1.1) have to be solved repeatedly with little or no change in either the matrix M or its eigenvalues and eigenvectors.

Typical of such cases is the problem of solving the non-linear reactor-kinetics equations in three-space dimensions (Curtis, Tyror and Wrigley, 1961). When finite-differenced in both space and time, the neutron diffusion equation for advancing the solution one time step assumes the form of equation (1.1). The coefficients of matrix M depend upon control-rod penetration and temperatures which are themselves dependent upon the neutron flux x. This non-linearity introduces a coupling between points widely separated on the finite-difference mesh, which is reflected in the form of the matrix M, and it is not unusual for some of the eigenvalues of M to be complex. As reactor conditions are controlled to change slowly with time, the optimum convergence parameters will also change slowly, and need not be determined for each time step.

A second example derives from the same problem. It is desirable to distribute the mesh points unevenly over the reactor and to concentrate them in regions where the neutron flux, x, is rapidly varying, e.g. in the neighbourhood of control rods. If the mesh lines run from boundary to boundary, there will be unimportant regions of the reactor where the mesh-point distribution is dense. On the other hand, if the mesh lines terminate at points within the reactor, the solution will be inaccurate unless many-point finite-difference formulae are used where the mesh lines end. Again, a complicated coupling between mesh points is produced which can cause the finite-difference coefficient matrix to have complex eigenvalues.

A further example is provided by the Carlson (1959) SNG method for solving the neutron transport equations. In this case, parameteric studies require the solution to be obtained for a succession of slightly varying matrices M. Chebyshev extrapolation has already been applied to this type of calculation by Blue and Flatt (1959). Since they assume parameter b to be always zero and make no acknowledgement of the fact that the eigenvalues can be complex, their method can, on occasion, seriously affect the convergence. Moreover, in the complex-eigenvalue case, their method is seldom likely to optimize the convergence rate. Equations of the type (1.1) in which the matrix M has complex eigenvalues can also be set up in multigroup collision-probability studies.

That these examples all derive from problems in the field of nuclear energy, is due solely to the author's preoccupation with work of this kind. Examples of complex eigenvalues in repetitive calculations can doubtless be found in other fields of study (see, for example, Kjellberg, 1958). Experience has shown that substantial economies can be effected by devoting time at the start of the calculations to the determination of optimum convergence parameters.

7. Acknowledgement

The author wishes to thank Dr. G. N. Lance for his advice on the preparation of this paper.

References

BLUE, E., and FLATT, H. P. (1960). "Convergence of the SN Method for Thermal Systems," *Nuclear Science and Engineering*, Vol. 7, p. 127.

CARLSON, B. (1959). "Numerical Solution of Transient and Steady State Neutron Transport Equations," *Report LA-2260*. Los Alamos Scientific Laboratory, University of California.

CLAYTON, A. J. (1963). "Further Results on Polynomials having Least Maximum Modulus over an Ellipse in the Complex Plane," U.K.A.E.A. Memorandum—AEEW-M.348.

CURTIS, A. R., TYROR, J. G., and WRIGLEY, H. E. (1961). "STAB: A Kinetic, Three-dimensional, One-group, Digital Computer Programme," *U.K.A.E.A. Report*: AEEW-R.77. H.M.S.O.

GOLUB, G. H., and VARGA, R. S. (1961). "Chebyshev Semi-iterative Methods, Successive Over-relaxation Iterative Methods and Second Order Richardson Iterative Methods," *Num. Math.*, Vol. 3, p. 147.

KJELLBERG, G. (1958). "On the Convergence of Successive Over-relaxation Applied to a Class of Linear Systems of Equations with Complex Eigenvalues," *Ericsson Technics*, No. 2, p. 245.

MULLER, D. E. (1956). "A Method for solving Algebraic Equations using an Automatic Computer," *Math. Tables Aids Comput.*, Vol. 10, p. 208.

STIEFEL, E. L. (1958). "Kernel Polynomials in Linear Algebra and their Numerical Applications," *Nat. Bur. Stand. Appl. Math. Series* 49, p. 1. U.S. Government Printing Office, Washington, D.C.

VARGA, R. S. (1961). "Proceedings of Symposia in Applied Mathematics". Vol. XI—*Nuclear Reactor Theory*, Providence, Rhode Island: American Mathematical Society.

WILKINSON, J. H. (1954). "The Calculation of the Latent Roots and Vectors of Matrices on the Pilot Model of the A.C.E.," *Proc. Camb. Phil. Soc.*, Vol. 50, Pt. 4, p. 536.

ZARANTONELLO, E. H. (1957). *See* R. S. Varga—"A Comparison of the Successive Over-relaxation Method and Semi-iterative Methods using Chebyshev Polynomials," *J. Soc. Indust. Appl. Maths.*, Vol. 5, p. 39.