

requires some properties of the upper bound of a matrix A , denoted by $|A|$. These properties are also available in von NEUMANN and GOLDSTINE [9].

For a definite matrix A

$$|A| = \lambda_{\max}.$$

For any matrix A and a vector B

$$|AB| \leq |A| |B|.$$

Then, since

$$Y = C^{-1}B = (A + kI)^{-1}B = (I + kA^{-1})^{-1}A^{-1}B = (I + kA^{-1})^{-1}X$$

it follows that

$$\begin{aligned} |Y| &\leq |(I + kA^{-1})^{-1}| |X| = \max_i \left[\frac{1}{1 + \frac{k}{\lambda_i}} \right] |X| \\ &= \max_i \left[\frac{\lambda_i}{\lambda_i + k} \right] |X| = \frac{\lambda_{\max}}{\lambda_{\max} + k} |X| < |X|. \end{aligned}$$

JAMES D. RILEY

U. S. Naval Ordnance Lab.
White Oak, Silver Spring, Maryland
Now at Iowa State College, Ames, Iowa

¹ M. HERZBERGER, "The normal equations of the method of least squares and their solution," *Quart. of Appl. Math.*, v. 7, 1949, p. 217-223.

² F. S. SHAW, *An Introduction to Relaxation Methods*, Dover Publications, Inc., New York, 1953.

³ OLGA TAUSSKY, "Note on the condition of matrices," *MTAC*, v. 4, 1950, p. 111-112.

⁴ H. POLACHEK, "On the solution of systems of linear equations of high order," Naval Ordnance Laboratory, Memorandum 9522, 1948.

⁵ A. C. AITKEN, "On Bernoulli's numerical solution of algebraic equations," *Roy. Soc., Edinburgh, Proc.*, v. 46, 1926, p. 289-305.

⁶ A. C. AITKEN, "Studies in practical mathematics. V. On the iterative solution of a system of linear equations," *Roy. Soc., Edinburgh, Proc., Sec. A*, v. 63, 1950, p. 52-60.

⁷ DANIEL SHANKS, "An analogy between transients and mathematical sequences and some nonlinear sequence-to-sequence transforms suggested by it. Part I," Published in *Jn. Math. and Physics*, v. 34, 1955, p. 1-42, under the title, "Nonlinear transformations of divergent and slowly convergent sequences."

⁸ GEORGE E. FORSYTHE, "Solving linear algebraic equations can be interesting," *Amer. Math. Soc., Bull.*, v. 59, 1953, p. 299-329.

⁹ JOHN VON NEUMANN & H. H. GOLDSTINE, "Numerical inverting of matrices of high order," *Amer. Math. Soc., Bull.*, v. 53, 1947, p. 1021-1099.

¹⁰ G. H. HARDY, J. E. LITTLEWOOD, & G. PÓLYA, *Inequalities*, Cambridge, 1934.

¹¹ K. LEVENBERG, "A method for the solution of certain non-linear problems in least squares," *Quart. of Appl. Math.*, v. 2, 1944, p. 164-168.

On the Numerical Solution of Elliptic Difference Equations

1. Introduction. In recent years a considerable amount of progress has been made in improving rates of convergence of iterative methods for the solution of elliptic difference equations. The numerical solution of Laplace's equation in rectangular coordinates in a unit square may be obtained by replacing the differential system by a difference system over a rectangular network with mesh spacing h , and then solving the resulting difference equations by an iterative method. This problem is usually taken as a "model problem" in elliptic difference

equations because the rate of convergence of iterative methods may often be obtained precisely. Prior to about 1950 the number of iterations for given accuracy with various methods was about proportional to h^{-2} . RICHARDSON [1] proposed a method with potentially greater rate of convergence, but gave no very effective way for choosing the parameters required. In 1950 YOUNG [2] obtained a method called the method of "successive overrelaxation" and independently FRANKEL [3] obtained his "extrapolated Liebmann" method, a special case of successive overrelaxation, and his "second-order Richardson" method, later shown by RILEY [4] to be also a special case of successive overrelaxation, if appropriate modifications were made. SHORTLEY [5] obtained a method using Tchebycheff polynomials which was equivalent to Richardson's method except that the parameters required were obtained by a definite procedure. Young [6] has analysed this method under more general assumptions about the matrix of the system. In all these methods the number of iterations for given accuracy is proportional to h^{-1} .

It will be shown that further improvements in rate of convergence are possible. Under certain conditions there are methods in which the number of iterations for given accuracy is proportional to h^{-1} and even to h^0 . Specific examples of the use of the new methods will be given.

2. General Description and Theory. Let our equations to be solved be $C'V' = \rho'$. We normalize C' , transform our dependent variables, V' , and inhomogeneous part, ρ' , by obtaining $C = D^{-1/2}C'D^{-1/2}$, $V = D^{1/2}V'$, and $\rho = D^{-1/2}\rho'$, where D is the matrix of diagonal elements of C' and where the signs of our equations $C'V' = \rho'$ have been chosen so that the elements of D are positive. We assume in all that follows that C is symmetric and positive definite, that C has what Young calls property (A) [7], and that the rows and columns of C have been ordered as in the extrapolated Liebmann method [3] (Young's ordering σ_2 [8]). Self-adjoint linear second-order elliptic partial differential equations may always be replaced by difference equations in such a way that the preceding properties are obtained (Young [9]).

The iterative sequence for the method of successive overrelaxation is

$$(1) \quad V_i^{(n+1)} = V_i^{(n)} - \alpha \left[\sum_{j=1}^{i-1} c_{ij} V_j^{(n+1)} + \sum_{j=i}^N c_{ij} V_j^{(n)} - \rho_i \right];$$

$$(i = 1, 2, \dots, N; n > 0).$$

We define triangular matrices A and B such that $A + B = \alpha C$ and $B = A^T - (2 - \alpha)I$. In terms of A and B , (1) becomes

$$(2) \quad AV^{(n+1)} + BV^{(n)} = \alpha\rho.$$

In (2) A is the matrix which multiplies the "new" or "improved" variables in the extrapolated Liebmann method, B the "old" or "unimproved" variables. A^T is the transpose of A . The difference, $E^{(n)}$, between V and $V^{(n)}$, i.e., the error in iteration n , satisfies

$$(3) \quad AE^{(n+1)} + BE^{(n)} = 0,$$

whence

$$(4) \quad E^{(n+1)} = A^{-1}BE^{(n)}.$$

If the length of $E^{(n)}$ tends to zero as n tends to ∞ we say that the transformation $A^{-1}B$ is convergent [10]. The rate of convergence, R_K , is defined as $-\log K$, where K is the spectral norm (magnitude of largest magnitude characteristic value) of $A^{-1}B$ [10]. Let $\bar{\lambda}$ be the spectral norm of $C - I$, and let $\bar{\mu} = 1 - \bar{\lambda}$. $\bar{\lambda}$ lies in the range $(0, 1)$ and may be evaluated for simple problems such as the "model problem" mentioned in the introduction. Young [2] has shown that: (a) $K = |K_1|$ where $K_1 - \alpha\bar{\lambda}K_1^{\frac{1}{2}} + \alpha - 1 = 0$. (b) $K = \bar{\lambda}^2$ when $\alpha = 1$. (We call $\alpha = 1$ the Liebmann value of α .) If $\bar{\mu}$ is small $R_K \doteq 2\bar{\mu}$. For the model problem with h small, $\bar{\mu} \doteq \frac{1}{2}\pi^2h^2$ and $R_K \doteq \pi^2h^2$. (c) $K = K(\min) = \alpha_0 - 1$, when $\alpha = \alpha_0 \equiv 2[1 - (1 - \bar{\lambda}^2)^{\frac{1}{2}}] \div \bar{\lambda}^2$. (We call α_0 the "optimum" value of α .) If $\bar{\mu}$ is small $R_K \doteq 2\sqrt{2\bar{\mu}}$, and, for the model problem, $R \doteq 2\pi h$. (d) For $\alpha \geq \alpha_0$, all the characteristic roots, K_i , of $A^{-1}B$ have magnitude $\alpha - 1$.

The above results were also obtained by Frankel [3] for the model problem.

So much for the ordinary extrapolated Liebmann method. We next consider the following iteration process:

$$(5) \quad AV^{(2n+1)} + BV^{(2n)} = \alpha\rho,$$

$$(6) \quad A^T V^{(2n+2)} + B^T V^{(2n+1)} = \alpha\rho.$$

If we write (6) in "component" form we obtain

$$(7) \quad V_i^{(2n+2)} = V_i^{(2n+1)} - \alpha \left[\sum_{j=i+1}^N c_{ij} V_j^{(2n+2)} + \sum_{j=1}^i c_{ij} V_j^{(2n+1)} - \rho_i \right].$$

In order to solve (7) for $V_i^{(2n+2)}$ we must first solve for $V_N^{(2n+2)}$, then $V_{N-1}^{(2n+2)}$, and so on. Hence, if we think of (1) as representing the traversal of a network in a certain order (corresponding to the extrapolated Liebmann method) then (7) represents a traversal of the network in the opposite order (the ordering, however, still corresponds to the extrapolated Liebmann method). Thus in the iteration process (5), (6) we are going back and forth over a network. We will call (5), (6) the method of symmetric successive overrelaxation. The error $E^{(2n)}$, in (5), (6) satisfies

$$(8) \quad AE^{(2n+1)} + BE^{(2n)} = 0,$$

$$(9) \quad A^T E^{(2n+2)} + B^T E^{(2n+1)} = 0,$$

and hence

$$(10) \quad E^{(2n+2)} = A^{T^{-1}}B^T A^{-1}BE^{(2n)}.$$

Let $T = A^{T^{-1}}B^T A^{-1}B$. Then we have the following theorem:

THEOREM 1. *The characteristic roots of T are real, non-negative, and less than one for $0 < \alpha < 2$.*

Proof.

$$\begin{aligned}
 (11) \quad T &= A^{T^{-1}} B^T A^{-1} B \\
 &= A^{T^{-1}} [A - (2 - \alpha)I] A^{-1} [A^T - (2 - \alpha)I] \\
 &= A^{T^{-1}} \{ [I - (2 - \alpha)A^{-1}] [I - (2 - \alpha)A^{T^{-1}}] \} A^T \\
 &= A^{T^{-1}} \{ [I - (2 - \alpha)A^{-1}] [I - (2 - \alpha)A^{-1}]^T \} A^T.
 \end{aligned}$$

The quantity in braces in (11) is a semi-definite matrix (because it is a matrix times its transpose) and T is the transform of this semi-definite matrix by a transformation that leaves its roots invariant. Hence, the roots of T are real and non-negative. We may easily show from Young's results that $K < 1$ for $0 < \alpha < 2$. (10) may be regarded as composed of two successive linear transformations each with spectral norm K . As two successive convergent transformations give a convergent transformation, T is convergent for $0 < \alpha < 2$, and hence the roots of T are less than one.

The foregoing theorem is just a generalization to α in the range $(0, 2)$ of a result obtained by AITKEN [11] for $\alpha = 1$. The method of proof is similar to that used by Aitken.

Let the characteristic roots of T be λ_i , $i = 1, 2, \dots, N$ and let the spectral norm of T be λ_1 . We have:

THEOREM 2. *If $\alpha \geq \alpha_0$, $\lambda_1 \geq K^2$, equality obtaining if and only if the characteristic roots of T are all equal.*

Proof.

$$\begin{aligned}
 (12) \quad |A^{-1}B| \cdot |A^{-1}B| &= K^{2N} \\
 &= |A^{-1^T} B^T| \cdot |A^{-1}B| \\
 &= |T| \\
 &= \prod_i^N \lambda_i.
 \end{aligned}$$

Therefore, $\text{Max } \lambda_i \geq K^2$, equality occurring only if the λ_i are all equal, for otherwise we could not satisfy $K^{2N} = \prod_i \lambda_i$. The first equality in (12) follows from Young's result (d) above.

COROLLARY 1. *If $\alpha \geq \alpha_0$ the rate of convergence of T is less than twice the rate of $A^{-1}B$.*

Corollary 1 follows immediately from Theorem 2 and the definition of rate of convergence. Hence we have proved that per network traversal (5), (6) converges more slowly (except for the very special case where the λ_i are all equal) than (2), i.e., back and forth iteration is less efficient than going just one way.

Let $k = 1 - K$ and suppose that k is small. Then we expect that most of the roots λ_i will be grouped in a band about K^2 , for each root which is substantially less than K^2 requires many roots greater than K^2 in order that (12) may be satisfied. Hence we have the following rather loose corollary:

COROLLARY 2. *If $\alpha \geq \alpha_0$, if k is small, and if N is large, then almost all the λ_i lie in the neighborhood of K^2 .*

For the ordinary extrapolated Liebmann method, we find a simple relationship (Young's result "a" above) between K , the spectral norm of $A^{-1}B$, and $\bar{\lambda}$, the

spectral norm of $C - I$. We would like to obtain a simple relationship between λ_1 and $\bar{\lambda}$ for then we could make an exact comparison of the rate of convergence of the symmetric successive overrelaxation method with the ordinary successive overrelaxation method. However, using standard methods which work for the analysis of many other iteration methods, we find that we cannot even determine λ_1 for our model problem (section 4). Hence it is doubtful that λ_1 is related to $\bar{\lambda}$ through any simple equation. However, we can analyse the one-dimensional Laplace problem (section 3) and also we can perform experiments on a computer (section 3 and 4). This work leads us to the conclusion that the following assumption is plausible:

ASSUMPTION 1. *If we let our mesh-spacing, h , tend to zero, k and $\mu_1 = 1 - \lambda_1$ tend to zero at the same rate, i.e., k and μ_1 are of the same order of magnitude.*

We will assume that our assumption is true. Then we can construct an iteration process which, for sufficiently small h , is superior to the extrapolated Liebmann method. This iteration process is

$$(13) \quad A V^{(2n+1)} + B V^{(2n)} = \alpha \rho$$

$$(14) \quad A^T V'^{(2n+2)} + B^T V^{(2n+1)} = \alpha \rho$$

$$(15) \quad V^{(2n+2)} = V'^{(2n+2)} + \omega_{2n} [V'^{(2n+2)} - V^{(2n)}].$$

The ω_{2n} are numbers chosen according to a technique similar to that used by Shortley [5]. Let $E^{(2n)} = V - V^{(2n)}$, $E'^{(2n)} = V - V'^{(2n)}$. Then $E^{(2n)}$, the error at iteration n , satisfies

$$(16) \quad A E^{(2n+1)} + B E^{(2n)} = 0$$

$$(17) \quad A^T E'^{(2n+2)} + B^T E^{(2n+1)} = 0$$

$$(18) \quad E^{(2n+2)} = E'^{(2n+2)} + \omega_{2n} (E'^{(2n+2)} - E^{(2n)}).$$

Expand $E^{(2n)}$ in the characteristic vectors, U_i , of the homogeneous linear system (16), (17):

$$(19) \quad E_{2n} = \sum_i c_i^{(2n)} U_i.$$

Then

$$(20) \quad c_i^{(2m)} = [\lambda_i + \omega_{2(m-1)}(\lambda_i - 1)] c_i^{(2m-2)} \\ = \prod_{n=0}^{m-1} [\lambda_i + \omega_{2n}(\lambda_i - 1)] c_i^{(0)}.$$

Let

$$(21) \quad P_m(\lambda) = \prod_{n=0}^{m-1} [\lambda + \omega_{2n}(\lambda - 1)].$$

$P_m(\lambda) = 1$ for $\lambda = 1$. We may factor $P_m(\lambda)$ as follows:

$$(22) \quad P_m(\lambda) = \frac{\prod_{j=1}^m (\lambda - \Lambda_j)}{\prod_{j=1}^m (1 - \Lambda_j)},$$

where $\omega_{2j} = \Lambda_j / (1 - \Lambda_j)$, as may be verified by substituting into (21). If nothing is known about the distribution of the λ_i except that they lie between 0 and λ_1 , then we let (cf. Shortley [5] and Young [6])

$$(23) \quad P_m(\lambda) = \frac{T_m(a\lambda + b)}{T_m(a + b)},$$

$$(24) \quad T_m(x) = \cos(m \cos^{-1} x), \quad |x| \leq 1 \\ = \cosh(m \cosh^{-1} x), \quad |x| > 1.$$

$T_m(x)$ is Tchebycheff's polynomial of order m . In the interval between $a\lambda + b = -1$ and $a\lambda + b = 1$ the maximum value of $P_m(\lambda)$ will be $1/T_m(a + b)$. We let $a\lambda + b = -1$ when $\lambda = 0$ and $a\lambda + b = 1$ when $\lambda = \lambda_1$. We obtain $a = 2/\lambda_1$, $b = -1$. The roots of $T_m(x)$ lie at $x = \cos \frac{2j-1}{2m} \pi$, $j = 1, 2, \dots, m$. We obtain for Λ_j ,

$$(25) \quad \Lambda_j = \frac{\lambda_1}{2} \left(\cos \frac{2j-1}{2m} \pi + 1 \right).$$

Let $1/T_m(a + b) = r^m$. We call r the average spectral norm for the transformation involved in going from $E^{(2n)}$ to $E^{(2n+2)}$. The average rate of convergence is $R_r = -\log r$. Suppose that $1 - r$ is small and that m is large. Then

$$(26) \quad \frac{R_r}{2} = -\frac{1}{2} \log \left\{ \frac{1}{[T_m(a + b)]^{1/m}} \right\} = -\frac{1}{2} \log \left\{ \frac{1}{\left[T_m \left(\frac{1 + \mu_1}{1 - \mu_1} \right) \right]^{1/m}} \right\} \\ = -\frac{1}{2} \log \{ e^{-2\sqrt{\mu_1} m} \}^{1/m} \doteq \sqrt{\mu_1}.$$

We compute $R_r/2$ instead of R_r to get convergence rate per network traversal. Suppose that symmetric successive overrelaxation converged at the same rate per network traversal as ordinary successive overrelaxation. Then for our model problem we would have $R_r/2 = 2\sqrt{\pi h}$. We would not expect to obtain this rate but according to Assumption 1 we should have $R_r/2$ proportional to $h^{\frac{1}{2}}$.

Young [12] shows that for the ordinary successive overrelaxation (with consistent ordering and $\alpha = \alpha_0$) the error goes to zero as nK^n not K^n . The operator in (11) is diagonalizable so that in the new method the error goes to zero (on the average) as r^m not mr^m which gives the new method an additional advantage, especially when the accuracy to be obtained is not great.

From equations (8), (9) we obtain:

$$(27) \quad AA^T E^{(2n+2)} - B^T B E^{(2n)} = 0.$$

and for a characteristic vector, U_i ,

$$(28) \quad (AA^T \lambda_i - B^T B) U_i = 0$$

(28) is useful in analysing specific cases.

3. Laplace's Equation in One Dimension. We have

$$(29) \quad \left. \begin{aligned} V_0 &= c_1 \\ -\frac{1}{2}V_{p-1} + V_p - \frac{1}{2}V_{p+1} &= 0 \\ V_N &= c_2 \end{aligned} \right\}.$$

Equations (13), (14), and (15) become:

$$(30) \quad V_p^{(n+1)} - V_p^{(n)} - \alpha[\frac{1}{2}V_{p-1}^{(n+1)} - V_p^{(n)} + \frac{1}{2}V_{p+1}^{(n)}] = 0,$$

$$(31) \quad V_p^{(n+2)} - V_p^{(n+1)} - \alpha[\frac{1}{2}V_{p-1}^{(n+1)} - V_p^{(n+1)} + \frac{1}{2}V_{p+1}^{(n+2)}] = 0,$$

$$(32) \quad V_p^{(n+2)} = V_p^{(n+2)} + \omega_n[V_p^{(n+2)} - V_p^{(n)}],$$

with V_0 and V_N equal to c_1 and c_2 wherever they occur in (30), (31), and (32). When we work out equations (28) for this problem we obtain:

$$(33) \quad a_{p-1}U_{i,p-1} + b_pU_{i,p} + a_pU_{i,p+1} = 0,$$

where

$$(34) \quad a_0 = a_{N-1} = 0,$$

$$(35) \quad a = a_p = -\frac{\alpha}{2}\lambda_i + \frac{\alpha}{2}(\alpha - 1), \quad p = 1, 2, \dots, N - 2,$$

$$(36) \quad b = b_p = \left(1 + \frac{\alpha^2}{4}\right)\lambda_i - \left(1 + \frac{\alpha^2}{4}\right) + \alpha(2 - \alpha), \quad p = 2, 3, \dots, N - 1,$$

$$(37) \quad b_1 = \lambda_i - 1 + \alpha(2 - \alpha).$$

We try as a solution to (33):

$$(38) \quad U_{i,p} = \sin \frac{N-p}{N} \beta_i \pi i.$$

All the equations (33) have the same coefficients except the first and the last. If any one of the intermediary equations is satisfied by (38) then all will be and so will the last equation. We will also have:

$$(39) \quad aU_{i,0} + bU_{i,1} + aU_{i,2} = 0$$

which may be combined with the first of equations (33) to obtain:

$$(40) \quad aU_{i,0} + (b - b_1)U_{i,1} = 0.$$

We substitute (38) into the intermediary equations (33) and obtain for λ_i ,

$$(41) \quad \lambda_i = \frac{1 + \frac{\alpha^2}{4} - \alpha(2 - \alpha) - (\alpha - 1)\alpha t_i}{1 + \frac{\alpha^2}{4} - \alpha t_i},$$

where $t_i = \cos \frac{i\pi}{N} \beta_i$. It is convenient to introduce the following quantities:

$$(42) \quad \mu_i = 1 - \lambda_i$$

$$(43) \quad \gamma_i = 1 - t_i$$

$$(44) \quad \delta = 2 - \alpha$$

$$(45) \quad \xi_i = \frac{\delta^2}{4\alpha\gamma_i}.$$

With (42), (43), (44), and (45) in (41) we obtain

$$(46) \quad \mu_i = \frac{\delta}{1 + \xi_i}.$$

The boundary condition (40) becomes

$$(47) \quad \xi_i \sin \beta_i \pi i + \frac{\alpha}{2} \sin \frac{N-1}{N} \beta_i \pi i = 0.$$

We choose δ so that $\xi_1 = 1$. We can show *a posteriori* that this gives closely the optimum value of α when N is large for both (5), (6), and (2). We shall assume N large and obtain an asymptotic solution for the μ_i . First we consider when i is small. Let $\beta_i = 1 + \epsilon_i$. Then from (47)

$$(48) \quad -\left(\frac{\alpha}{2} + \xi_i\right) \epsilon_i + \frac{\alpha}{2N} \doteq 0.$$

We have $\gamma_i \doteq \frac{i^2 \pi^2}{2N^2}$ and so

$$(49) \quad \xi_i = \frac{\delta^2}{4\alpha\gamma_1} \frac{\gamma_1}{\gamma_i} \doteq \frac{1}{i^2},$$

$$(50) \quad \epsilon_i \doteq \frac{\alpha}{\alpha/2 + \xi_i} \cdot \frac{1}{2N}.$$

Hence ϵ_i is small for the first few values of i where the expansions used are valid.

We have $\mu_1 = \frac{\delta}{2}$ and the successive values of μ_i decrease as i increases, and ξ_i rapidly grows small. Next we obtain a solution when ξ_i is small. Let $\beta_i = \frac{N}{N-1} (1 + \epsilon_i)$. Then

$$(51) \quad \xi_i \sin \frac{N}{N-1} \pi i (1 + \epsilon_i) + \frac{\alpha}{2} \sin \pi i (1 + \epsilon_i) = 0$$

$$(52) \quad \sin \pi i (1 + \epsilon_i) \doteq (-1)^i \pi i \epsilon_i$$

$$(53) \quad \epsilon_i \doteq \frac{(-1)^{i+1} \xi_i \sin \frac{N}{N-1} \pi i}{i}, \quad i \neq N-1.$$

Hence β_i is close to $\frac{N}{N-1}$ for the larger values of i and so $\gamma_i \rightarrow 2$, $\xi_i \rightarrow 0$, and the μ_i approximate δ for large i so that we have a narrow band of roots from $1 - \delta$ to $1 - \frac{\delta}{2}$.

The approximation in (51) of neglecting ϵ_i in the first term to obtain (53) is not valid when $i = N - 1$, for then we have

$$(54) \quad \xi_{N-1} \sin [N\pi(1 + \epsilon_{N-1})] + \frac{\alpha}{2} \sin [(N - 1)\pi(1 + \epsilon_{N-1})] = 0$$

and there is no real solution with $|\epsilon_{N-1}| < 1$, when ξ_{N-1} is small. The last characteristic value is obtained from

$$(55) \quad U_{N-1,p} = \sinh (N - p)\beta_{N-1}.$$

For this characteristic vector, $t_{N-1} = \cosh \beta_{N-1}$ and γ_{N-1} is negative. (40) becomes

$$(56) \quad \delta^2 \sinh N\beta_{N-1} + 2\alpha^2(1 - \cosh \beta_{N-1}) \sinh (N - 1)\beta_{N-1} = 0$$

and if β_{N-1} is small,

$$(57) \quad (\delta^2 - \alpha^2\beta_{N-1}^2) \sinh N\beta_{N-1} + \alpha^2\beta_{N-1}^3 \cosh N\beta_{N-1} = 0.$$

Let

$$(58) \quad \beta_{N-1}^2 = \frac{\delta^2}{\alpha^2} (1 + \epsilon).$$

Then from (57), (58)

$$(59) \quad \epsilon \doteq \frac{\delta \cosh \pi}{\alpha \sinh \pi} \doteq \frac{\delta}{\alpha}$$

and when N is large so that δ is small we can show that

$$(60) \quad \xi_{N-1} \doteq 1 - \delta$$

and so $\mu_{N-1} \doteq 1$ and there is a single root close to zero.

We obtain for $\frac{R_r}{2}$, R_k , and $\frac{R_r}{2} \div R_k$:

$$(61) \quad \frac{R_r}{2} = \sqrt{\pi h}$$

$$(62) \quad R_k = 2\pi h$$

$$(63) \quad \frac{R_r}{2R_k} = \frac{1}{2\sqrt{\pi}} h^{-\frac{1}{2}}.$$

The factor of improvement is not large unless h is quite small, N quite large.

For this simple one-dimensional problem we can construct an iteration process in which the rate of convergence is independent of N . If we choose, for example, ω_n alternately equal to zero and $\frac{4}{3\delta} - 1$, then the first ω will damp out the error with characteristic value near zero and the second will damp every term in the band $\left(1 - \delta; 1 - \frac{\delta}{2}\right)$ by at least $\frac{1}{3}$. Further by choosing a smaller value of ξ_1 at the start, i.e., a larger value of α , we may increase the rate of convergence without limit, but the ω_n also increase without limit so that the whole process becomes unstable because of rounding error.

An example of this iteration process was tried on the IBM 701 computer [13]. N was chosen to be 100 and c_1 and c_2 were chosen to be zero. A starting approximation of $V^{(0)}$ equal to 10^9 at all interior points was chosen. The ω_n were chosen according to the scheme suggested in the preceding paragraph. Using the ordinary extrapolated Liebmann method approximately 300 iterations were required to damp the maximum error by 10^{-5} . Using the new method 16 iterations (32 passes over the network) damped the maximum error by about the same amount.

By using a polynomial derived from the Tchebycheff polynomials, as in section 2, to damp the amplitudes of the error characteristic vectors with roots in the band $\left(1 - \delta; 1 - \frac{\delta}{2}\right)$ we could obtain an even greater factor of improvement.

We make this analysis of Laplace's equation in one dimension in order to learn about the properties of symmetric successive overrelaxation, but not with the expectation that this iteration method would actually be used for one-dimensional problems. For one-dimensional problems direct Gaussian elimination methods are most efficient [14].

4. Laplace's Equation in Two Dimensions. A detailed analysis of the model problem mentioned in the introduction does not appear possible. We may obtain equations (28) in explicit form for the two-dimensional problem without much difficulty. We find that we can construct vectors composed of trigonometric and exponential functions which satisfy (28) at all mesh points not adjacent to boundaries, but then we find that these trigonometric and exponential functions may not be chosen in such a way that the λ_i are real and the boundary conditions are satisfied.

The two-dimensional problem was investigated by solving a numerical example on a computer [13]. A 30×30 network (including the boundary points) was chosen for the problem of the square for which the optimum value of α is $\alpha = 1.805$. A starting approximation of $V^{(0)} = 10^9$ was chosen at all interior points and the boundary condition $V = 0$ was applied over the entire boundary.

The ordinary extrapolated Liebmann method required 110 iterations at optimum α to damp the maximum error by 2.5×10^{-6} . By iterating with the process (1), (2) (i.e., $\omega_n = 0$) it was determined that $\lambda_1 \doteq .88$. A set of ω_n for $m = 20$ was constructed by the method suggested in section 2. Then the process (1), (2), (3) was applied for 20 iterations (40 passes over the network). At the end, the maximum error was damped by approximately 2.5×10^{-6} . This reduction in error corresponded quite closely with the theoretical value. Hence we

find that we have an improvement of a factor of about three for this two-dimensional example. A factor of three can be very important when a long series of problems is to be solved. However, we note that the method of symmetric successive overrelaxation requires storage for two iterates, $V^{(2n+2)}$ and $V^{(2n)}$, whereas ordinary successive overrelaxation only requires storage for one.

A distribution of ω_n similar to that chosen for the one-dimensional problem was tried for the two-dimensional problem. Rapid convergence was not obtained indicating that there are roots more or less evenly distributed between λ_1 and 0. However the results obtained lead us to conjecture that there is an iteration process of order h^0 even for the two dimensional problem. Label the corners of our square a, b, c, d in a clockwise direction. Suppose that our equations are ordered so that the first equation is centered at the meshpoint just interior to a , the last just interior to c . When we apply a distribution of the ω_n like that of section 3, we find that we have rapid convergence at mesh-points along the line ac but slow convergence (or even divergence) in the neighborhood of the corners b and d . Let $a \rightarrow c$ indicate that we pass over the network using the extrapolated Liebmann method starting at a and ending at c , and similarly for other letter combinations. (There are two ways of doing this but it should not make much difference which we choose so long as we are consistent.) Starting with $V^{(n)}$ we obtain $V^{(n+4)}$ by iterations $a \rightarrow c, c \rightarrow a, b \rightarrow d, d \rightarrow b$. Then we obtain $V^{(n+4)}$ from

$$V^{(n+4)} = V^{(n+4)} + \omega_n(V^{(n+4)} - V^{(n)}).$$

It seems plausible that a choice of the ω_n analogous to what was done for the one-dimensional problem would give convergence of order zero for this process.

5. Suggestions for the Use of the New Method for Multi-dimensional Problems. It is supposed that the method here presented would be used mostly in situations where we wish to solve a number of similar problems. In this case the following procedure is suggested :

- (1) For a typical problem obtain α and λ_1 by experimentation on a computer.
- (2) Estimate $\lambda_1(\max)$ for the series of problems.
- (3) Determine m and the ω_j from $\lambda_1(\max)$ so that r^m is less than the desired accuracy ϵ to be obtained.
- (4) Use the values of the ω_j obtained for the remaining problems.

$\lambda_1(\max)$ should be chosen conservatively so that it is sure to be large enough to cover all the cases treated. If λ_1 is chosen too small the new method becomes inefficient, so that it is better to overestimate λ_1 . It is not believed that the new method requires a very precise determination of α .

The new method is especially suited to a computer which has magnetic tapes which read "backward" as well as "forward."

The new method will perhaps also be useful for the numerical solution of elliptic difference equations in cases where relaxation methods converge very slowly; for example, for the equations of compressible fluid flow at Mach numbers close to but less than one.

JOHN W. SHELDON

Wilson Point
South Norwalk, Conn.

The research in this paper was performed in connection with consulting agreements between the author and Continental Oil Co., Ponca City, Oklahoma, and Curtiss-Wright Aircraft Co., Wright Aeronautical Division, Woodridge, N. J. Curtiss-Wright provided all computer time used.

¹ L. F. RICHARDSON, "The approximate arithmetical solution by finite differences of physical problems involving differential equations with an application to stresses in a masonry dam," Roy. Soc., London, *Phil. Trans.*, Sec. A, v. 210, 1910, p. 307-357.

² D. YOUNG, "Iterative methods for solving partial difference equations of elliptic type," Amer. Math. Soc., *Trans.*, v. 76, 1954, p. 92-111.

³ S. P. FRANKEL, "Convergence rates of iterative treatments of partial differential equations," *MTAC*, v. 4, 1950, p. 65-75.

⁴ J. D. RILEY, "Iteration procedures for the Dirichlet difference problem," *MTAC*, v. 8, 1954, p. 125-131.

⁵ G. SHORTLEY, "Use of Tchebycheff polynomial operators in the numerical solution of boundary value problems," *Jn. Appl. Phys.*, v. 24, 1953, p. 392-396.

⁶ D. YOUNG, "On Richardson's method for solving linear systems with positive definite matrices," *Jn. Math. and Physics*, v. 32, 1954, p. 243-255.

⁷ D. YOUNG, *op. cit.*, reference 2, p. 93.

⁸ D. YOUNG, *op. cit.*, p. 108.

⁹ D. YOUNG, *op. cit.*, section 4.

¹⁰ D. YOUNG, *op. cit.*, p. 96.

¹¹ A. C. AITKEN, "On the iterative solution of a system of linear equations," Roy. Soc., Edinburgh, *Proc.*, Sec. A, 63, 1950, p. 52-60.

¹² D. YOUNG, *op. cit.*, reference 2, p. 103.

¹³ The International Business Machines Company Type 701 Electronic Data Processing Machines located at the IBM Scientific Computing Service, 590 Madison Ave., New York City, N. Y.

¹⁴ JOHN SHELDON & L. H. THOMAS, "The use of large scale computing in physics," *Jn. Appl. Phys.*, v. 24, 1953, p. 235-242.

Solutions of a Nonlinear Differential Equation Arising in the Theory of Diffusion Flames

An extended theory of diffusion flames [1] leads to a nonlinear second order differential equation, which is, in its non-dimensional form [2]:

$$v'' = \frac{K}{x + x_1} \cdot v \cdot (v + x + x_1 - 1)$$

with the following two-point boundary conditions:

$$v = 1 - x_1 \quad \text{at} \quad x = 0$$

and

$$v \rightarrow 0 \quad \text{as} \quad x \rightarrow \infty.$$

This equation describes the reaction of sodium vapor diffusing into a halide, where $1 - x_1$ and K correspond effectively to a measure of the nozzle opening and the chemical rate of reaction, respectively. The equation has been solved numerically, with the aid of FERUT, to cover a wide range of experimental conditions. It presented many difficulties of a practical nature, which could hardly have been foreseen until considerable work had been done.

First of all, solutions for two cases, corresponding directly to specific experiments, were carried out by hand, using the equation in its original form [1] and employing MILNE's method of integration. Each of these required nearly a month of desk work by a proficient computer. One of them served as an independent check for the FERUT programs.

In order to adapt the problem for automatic solution it was decided to work