STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM
REKENAFDELING

MR 104

A method for solving elliptic difference equations

by

P.J. van der Houwen

April 1969

Contents

1. Introduction

In recent years a large number of methods have been developed to improve the rate of convergence of iterative processes for solving elliptic difference equations. In order to analyse these iterative methods one usually takes the discrete analogue of the Dirichlet problem for Poison's equation on a square of side π, which serves as a "model problem" in elliptic difference equations. For the model problem one often may obtain an analytic expression for the rate of convergence so that one can learn about the properties of the iterative method. Until 1950, the rates of convergence of the known iterative procedures were of order h^2, h being the mesh size of the grid used. Although Richardson had indicated as early as 1910 an iterative method with a potentially greater rate of convergence, he did not obtain a better method, because the parameters required for his method were not chosen optimally.

In 1950 Young, and independently Frankel, proposed a very powerful method with potentially a rate of convergence of order h. This method is known as the "method of successive overrelaxation of Young" (SOR method) or as the "extrapolated Liebmann method" as it was called by Frankel. In the paper of Frankel another method of the second order was described which also has a rate of convergence of order h. We shall call it "Frankel's method".

In 1953 Shortley applied Richardson's method with the optimal values of the required parameters. Asymptotically, the method has a rate of convergence of order h, however, for a large number of iterations the method turned out to be numerically unstable.

In 1955 Sheldon combined Richardson's method and the SOR method to obtain a process, the "method of symmetric successive overrelaxation", which experimentally proved to be of order $\sqrt{h}$ for the model problem. We shall call the method the SSOR method.

A new approach in accelerating iterative methods was given by Peaceman and Rachford in 1955 and by Douglas and Rachford in 1956. Asymptotically, their theories result in a still greater rate of convergence, namely of order $1/\ln h^{-1}$. However, it was shown in 1959 by Birkhoff and Varga that

the theory only holds for the model problem, so that the actual value of the method is doubtful.
In 1958 the disadvantage of Richardson's method, namely its instability, was overcome by Stiefel, who introduced a second order version of the method which can be proved to be stable.
In the following years a large number of contributions to elliptic difference equations were made. The greater part of these are modifications or generalizations of the methods already mentioned.

In this paper a method is described based on the second order Richardson method and the SOR method. We give a detailed analysis of the method for the model problem showing that the rate of convergence is of order $\sqrt{h}$. On a computer we got rates of convergence which agree with the theoretically predicted rates of convergence.

The author acknowledges Mr. P. Beertema for writing the computer program by which the numerical results were obtained.

2. Mathematical preliminaries

In this section we review Richardson's method of second degree for solving matrix equations of the type

$$Lu = f, \tag{2.1}$$

where L is a symmetric matrix with positive eigenvalues, f is a known vector and u is the unknown vector. The method is defined by the recurrence relations

$$\begin{aligned} u_1 &= u_0 - \omega_0(Lu_0 - f), \\ u_{k+1} &= \alpha_k u_k + (1 - \alpha_k)u_{k-1} - \omega_k(Lu_k - f), \quad k = 1, 2, \ldots, \\ \alpha_k &= 2y_0 \frac{T_k(y_0)}{T_{k+1}(y_0)}, \quad k = 1, 2, \ldots, \\ \omega_0 &= \frac{2}{b + a}, \quad \omega_k = \frac{4}{b - a} \frac{T_k(y_0)}{T_{k+1}(y_0)}, \quad k = 1, 2, \ldots, \end{aligned} \tag{2.2}$$

where u_0 is an initial guess for the solution u, $y_0 = (b + a)/(b - a)$, $[a,b]$ is a positive interval containing all eigenvalues λ of L, and $T_k(y)$ is the Chebyshev polynomial of degree k. $T_k(y)$ satisfies the recurrence relation

$$T_0(y) = 1, \; T_1(y) = y, \; T_{k+1}(y) = 2yT_k(y) - T_{k-1}(y), \; k = 1, 2, \ldots . \tag{2.3}$$

This iteration scheme was proposed by Stiefel [1958] and is called Richardson's method of second degree. A detailed discussion of Richardson's method and some accelerating procedures may be found in van der Houwen [1968], chapter IV. In the present investigation we are mainly interested in the average rate of convergence after K iterations of the iteration process for some operators L arising from the numerical solution of Poison's equation. The rate of convergence, denoted by R(K), is given by (cf. Forsythe and Wasow [1960], p. 231)

$$R(K) = \frac{1}{K} \ln T_K(y_0). \tag{2.4}$$

If $a \ll b$ it can be derived from the properties of $T_k(y)$ that

$$R(K) \sim 2\sqrt{\frac{a}{b}} - \frac{1}{K} \ln 2. \tag{2.4'}$$

The quotient b/a is called the P-condition number of the matrix L and will be denoted by P(L).

We now summarize the theory of the numerical solution of Poison's equation needed in the following sections. A detailed treatment of this theory may be found in van der Houwen [1967].

Consider Poison's equation

$$\Delta U + F = 0, \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{2.5}$$

on a square of side π with boundary conditions of the first kind. Let $X_{\pm}$ and $Y_{\pm}$ represent translations $\pm h$ along the x- and y-axis respectively. On a grid of square meshes of side h the following discrete analogue of (2.5) may be defined:

$$D(\gamma)u + f = 0, \tag{2.6}$$

where

$D(\gamma) = 1$ at the boundary points,

$D(\gamma) = L_1(X_+ + X_-)(Y_+ + Y_-) + L_2(X_+ + X_- + Y_+ + Y_-) + L_4$ at the internal points,

$$L_1 = \frac{2 - \gamma}{2h^2}, \quad L_2 = \frac{\gamma - 1}{h^2}, \quad L_4 = -\frac{2\gamma}{h^2}, \quad 1 \leq \gamma \leq 2,$$

and where u and f represent grid functions defined at the grid points. When -f assumes the boundary values of U at the boundary points and f assumes the values of F at the internal grid points, then it can be proved that the solution of the boundary value problem (2.5) and the solution of the discrete problem (2.6) differ by a term $O(h^2)$ (cf. Forsythe and Wasow [1960], section 23). Problem (2.6) will be called the model problem.

The operator $D(\gamma)$ can be represented by a symmetric matrix operator. The eigenvalues of $D(\gamma)$ appear to be negative so that we define

(2.7) $$L(\gamma) = -D(\gamma).$$

For $\gamma = 2$ this operator was considered by Frank [1960]. It is easily verified that

(2.8) $$P(L(2)) \sim 4h^{-2},$$

so that

(2.9) $$R(K) \sim h - \frac{\ln 2}{K} \sim h - \frac{0.693}{K}.$$

For $\gamma = 1$ it can be shown that (see van der Houwen [1967])

(2.10) $$R(K) \sim \sqrt{2}\, h - \frac{0.693}{K}.$$

At this point we remark that the iteration scheme with $\gamma = 1$ may be interpreted as the iteration scheme with $\gamma = 2$ applied to a square which is rotated over 45° (see figure 2.1 and 2.2).

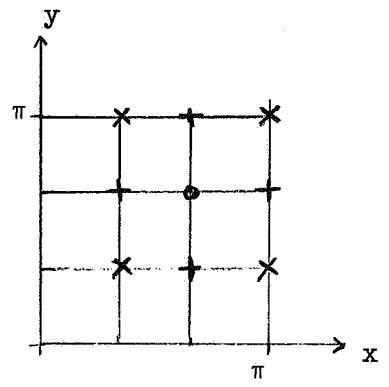

fig. 2.1

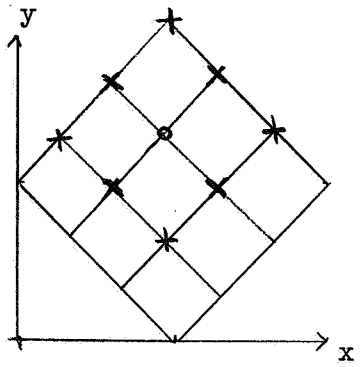

fig. 2.2

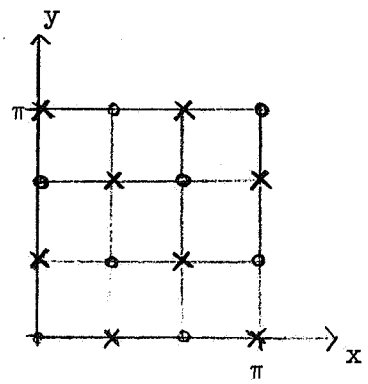

fig. 2.3

× corresponds to $\gamma = 1$, + corresponds to $\gamma = 2$

Further, we observe that using the operator $L(1)$ permits us to calculate u_k only in those net points (j,l) of figure 2.3 which are denoted by either a dot or a cross.

One point of departure in accelerating Richardson's method is to replace the matrix equation $L(\gamma)u = f$ by an equation

(2.11) $$L'(\gamma)u = f'$$

which has the same solution u, but in which $L'(\gamma)$ has a lower P-condition number than $L(\gamma)$. Such an approach was given in van der Houwen [1967] for the model problem mentioned above. The operator $L'(\gamma)$ and the vector f' were defined by

(2.12) $$L'(\gamma) = -(1 - D_1(\gamma))^{-1}D(\gamma),\quad f' = (1 - D_1(\gamma))^{-1}f,$$

where

(2.13) $$D_1(\gamma) = 0 \text{ at the boundary points,}$$
$$D_1(\gamma) = p(L_1(X_+Y_- + X_-Y_-) + L_2(X_- + Y_-)) \text{ at the internal points.}$$

For $\gamma = 1$ and $\gamma = 2$ it is possible to choose the parameter p such that the eigenvalues of $L'(\gamma)$ are real. The operator $L'(2)$ then reduces to the operator occurring in Liebmann's method (compare Forsythe and Wasow [1960]). The P-condition number of this operator appears to be $\frac{1}{4}$ of the value of $P(L(2))$:

(2.14) $$P(L'(2)) \sim h^{-2}.$$

Therefore, the asymptotic rate of convergence of Richardson's method with respect to $L'(2)$ is twice as large as the asymptotic rate of convergence of the scheme used by Frank. However, operators of type (2.12) are not symmetric and, in fact, the eigenvalues of $L'(1)$ and $L'(2)$ are very ill-conditioned, so that an arbitrary initial approximation u_0 will be a very poor approximation of the solution u. In Coolen and van der Houwen [1968] a method is given which eliminates the ill-conditioned eigenfunction components from the initial approximation. This preconditioning was found very successful, but the method explicitly uses the fact that the eigenvalues of the ill-conditioned components are known to be the greatest eigenvalues. In other cases, preconditioning may be less successful. Therefore, it is desirable to construct matrix representations with better conditioned eigenfunctions.

3. A symmetric matrix representation for the model problem

In this section a symmetric matrix representation of the boundary value problem is given for which the P-condition number is of order h^{-1}.

Let us define the operator

$$L'(\gamma) = -\frac{1}{2}\left[(1 - D_1(\gamma))^{-1} + (1 - D_2(\gamma))^{-1}\right]D, \tag{3.1}$$

where

$$\begin{aligned} &D_2(\gamma) = 0 \text{ at the boundary points,} \\ &D_2(\gamma) = p(L_1(X_-Y_+ + X_+Y_+) + L_2(X_+ + Y_+)) \text{ at the} \\ &\qquad\qquad\qquad\qquad\qquad\qquad \text{internal points.} \end{aligned} \tag{3.2}$$

This operator arises from averaging the operators which correspond to Gauss-Seidel's method starting from opposite corner points. We may expect that this averaging eliminates the ill-conditioning of the eigenfunctions.

Theorem 3.1.

Let $\gamma = 1$, then $L'(\gamma)$ is symmetric with eigenfunctions

$$e(n,m) = \sin njh \sin mlh,\ n, m = 1, 2, \ldots, \frac{\pi}{h} - 1 \tag{3.3}$$

and eigenvalues

$$\lambda(n,m) = 4h^{-2}\frac{(2 - \Omega\nu\mu)(1 - \nu\mu)}{\Omega^2\nu^2 - 4\Omega\nu\mu + 4},\ n, m = 1, 2, \ldots, \frac{\pi}{h} - 1. \tag{3.4}$$

Here we have $\nu = \cos nh$, $\mu = \cos mh$ and $\Omega = 2ph^{-2}$.

Proof

It is easily verified that $D(1)$ has the eigenfunctions $e(n,m)$ defined above with eigenvalues

$$\delta(n,m) = -2h^{-2}(1 - \cos nh \cos mh) = -2h^{-2}(1 - \nu\mu). \tag{3.5}$$

From this it follows that

$$L'(1)e(n,m) = -\frac{1}{2}\,\delta(n,m)(v_1 + v_2),$$

where

$$v_1 = (1 - D_1)^{-1}e(n,m) = e(n,m) + D_1 v_1,$$

$$v_2 = (1 - D_2)^{-1}e(n,m) = e(n,m) + D_2 v_2.$$

By substituting (3.3) we find $L'(1)e(n,m) = \lambda(n,m)e(n,m)$ where $\lambda(n,m)$ is given by (3.4).

The eigenfunctions $e(n,m)$ are orthogonal and the eigenvalues $\lambda(n,m)$ are real. Therefore, $L'(1)$ is a symmetric matrix representation of the boundary value problem.

It may be remarked that for other values of γ it appeared not possible to derive simple expressions for $e(n,m)$ and $\lambda(n,m)$ as given above.

We desire to determine the relaxation factor Ω such that the P-condition number of $L'(1)$ is as small as possible. At first sight, one should choose Ω according to the theory of Young, that is one minimizes the condition number of the operator $L'(1)$ defined by (2.12). In Coolen and van der Houwen it was shown that the theory of Young yields

$$\Omega = 2\gamma^{-1}(1 - h\sqrt{2\gamma^{-1}}),\ \gamma = 1,\ 2. \tag{3.6}$$

However, it will be shown here that for $\gamma = 1$ an exact analysis of (3.4) yields a slightly better value for Ω. Of course, when dealing with a more general problem than the model problem, one should take the value of Ω prescribed by the Young theory.

Let us suppose that Ω is close to but less than 2 and, temporally, let us assume that ν and μ are continuous variables.
It is readily seen that the stationary points of $\lambda(n,m)$ satisfy the equations

$$\mu(2 + \Omega - 2\Omega\nu\mu)(4 - 4\Omega\nu\mu + \Omega^2\nu^2) + 2\Omega(\Omega\nu - 2\mu)(1 - \nu\mu)(1 - \Omega\nu\mu) = 0,$$

$$\nu(2 + \Omega - 2\Omega\nu\mu)(4 - 4\Omega\nu\mu + \Omega^2\nu^2) - 4\Omega\nu(1 - \nu\mu)(1 - \Omega\nu\mu) = 0.$$

For $\nu \neq 0$ the stationary points of $\lambda(n,m)$ are situated on the line

$$\mu = \frac{1}{4}\Omega\nu. \tag{3.7}$$

Therefore, the extrema of $\lambda(n,m)$ are reached at points of the curve (3.7), the μ-axis, or at points of the boundary of the (ν,μ)-domain, i.e. the line segments $\nu = \pm \cos h$, $-\cos h \leq \mu \leq \cos h$ and $\mu = \pm \cos h$, $-\cos h \leq \nu \leq \cos h$ (see figure 3.1).

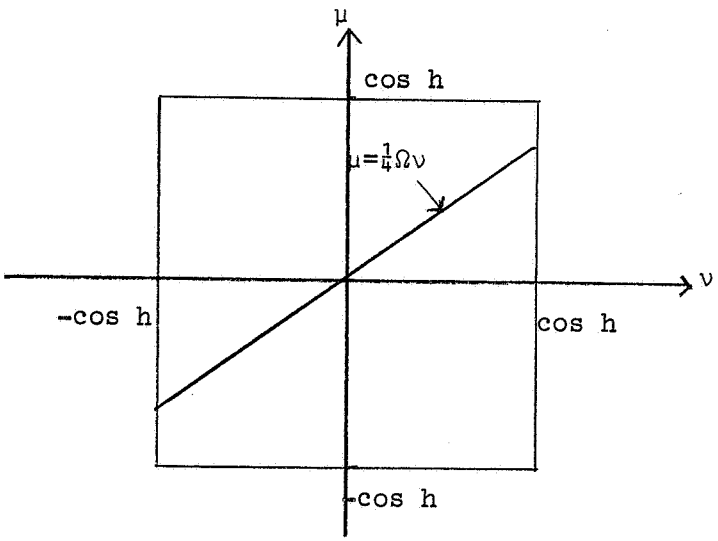

fig. 3.1

Along the curve $\mu = \frac{1}{4}\Omega\nu$ λ reaches a maximal value for $\nu = 0$ and along the μ-axis we have $\lambda = 2h^{-2}$. Therefore, we are only concerned with the values of λ at the boundary. Thus, returning to integer values for n and m, we have to consider the values:

$$\lambda(\tfrac{\pi}{2h},1),\ \lambda(n,1) = \lambda(\tfrac{\pi}{h} - n,\tfrac{\pi}{h} - 1),\ \lambda(1,m) = \lambda(\tfrac{\pi}{h} - 1,\tfrac{\pi}{h} - m), \tag{3.8}$$

where $n, m = 1, 2, \ldots, \frac{\pi}{h} - 1$.
In figure 3.2 the behaviour of the functions $\lambda(n,1)$ and $\lambda(1,m)$ is illustrated for $\Omega \sim 2$.

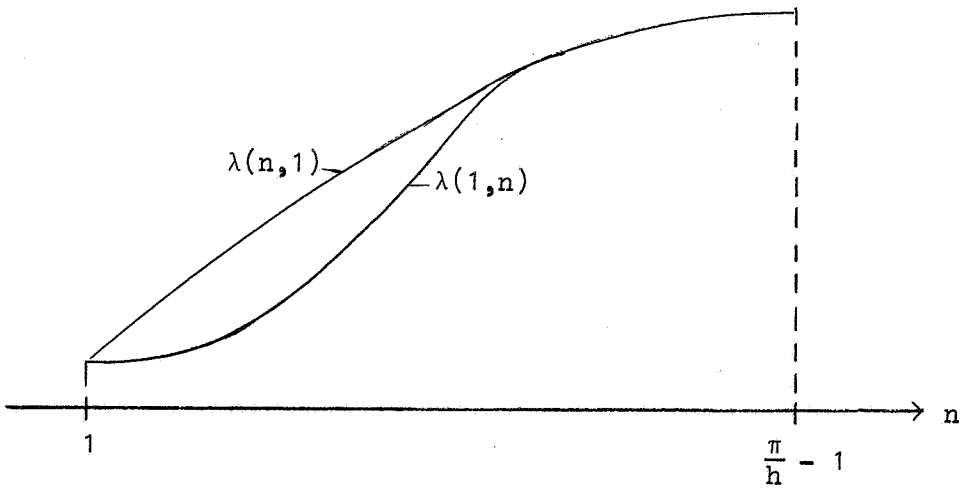

fig. 3.2

From this figure it may be concluded that the extrema of $\lambda(n,m)$ are bounded by the values of

$$\lambda(\tfrac{\pi}{2h},1) = 2h^{-2},$$

$$\lambda(1,1) = 2h^{-2}\,\frac{2(1-\cos^2 h(2-\Omega\cos^2 h)}{4-4\Omega\cos^2 h+\Omega^2\cos^2 h} \sim 2h^{-2}\,\frac{2\eta(\varepsilon+2\eta)}{\varepsilon^2+4\eta}, \tag{3.9}$$

$$\lambda(1,\tfrac{\pi}{h}-1) = 2h^{-2}\,\frac{2(1+\cos^2 h)(2+\Omega\cos^2 h)}{4-4\Omega\cos^2 h+\Omega^2\cos^2 h} \sim 2h^{-2},$$

where $\eta = 1-\cos^2 h \sim h^2$ and $\varepsilon = 2-\Omega$. The P-condition number is approximated by

$$P(L'(1)) \sim \frac{\varepsilon^2+4\eta}{2\eta(\varepsilon+2\eta)}. \tag{3.10}$$

This expression is minimized by $\varepsilon = 2\sqrt{\eta}$. Hence we find for Ω the approximate value (compare formula (3.6) for $\gamma = 1$)

$$\Omega \sim 2-\varepsilon = 2-2\sqrt{\eta} = 2-2\sin h \sim 2-2h. \tag{3.11}$$

This value of Ω yields the condition number

$$P(L'(1)) \sim \frac{2}{\sin h + \sin^2 h} \sim 2h^{-1}. \tag{3.12}$$

A slightly larger condition number is found for the SOR value of Ω defined by (3.6), namely

$$P(L'(1)) \sim \frac{3}{2}\sqrt{2}\, h^{-1} \sim 2.1\, h^{-1}. \tag{3.13}$$

The average rates of convergence of Richardson's method corresponding to (3.12) and (3.13) respectively, are

$$R(K) \sim \sqrt{2h} - \frac{\ln 2}{K} \sim 1.41\sqrt{h} - \frac{0.693}{K}, \tag{3.14}$$

$$R(K) \sim 2\sqrt{\frac{\sqrt{2}}{3}\, h} - \frac{\ln 2}{K} \sim 1.31\sqrt{h} - \frac{0.693}{K}. \tag{3.15}$$

In order to compare the new method with other iterative processes, for instance Young's method, one must bear in mind that the method described above is twice as laborious per iteration. Thus, comparing (3.14) with the rate of convergence of Young's SOR method, i.e. $2\sqrt{2}\, h$ for large values of K (see Coolen and van der Houwen [1968]), we may conclude that the new method becomes faster if

$$\frac{1}{2}\sqrt{2h} > 2\sqrt{2}\, h,$$

or equivalently

$$h < \frac{1}{16} = .0625, \quad N > 50. \tag{3.16}$$

Thus, only for rather small values of h the new method will be advantageous.

4. Numerical results for the model problem

In this section the results are given of a number of experiments with the model problem on the EL X8 computer at the Mathematical Centre at Amsterdam.

In order to check the accuracy of the numerical solution we have chosen

$$F(x,y) = -2\exp(x+y),\quad 0 < x < \pi,\ 0 < y < \pi, \tag{4.1}$$

$$U(x,y) = \exp(x+y),\qquad x = 0,\pi\ ,\ \ y = 0,\pi. \tag{4.2}$$

The inhomogeneous term (4.1) and the boundary function (4.2) give rise to the analytical solution

$$U(x,y) = \exp(x+y),\qquad 0 \leq x \leq \pi,\ 0 \leq y \leq \pi. \tag{4.3}$$

The iteration process was started by the following initial approximation:

$$u_0(j,1) = \left(\frac{1}{j} + \frac{1}{1} + \frac{1}{N-j} + \frac{1}{N-1}\right)^{-1} \left(\frac{u(j,0)}{1} + \frac{u(0,1)}{j} + \right. \tag{4.4}$$

$$\left. + \frac{u(N,1)}{N-j} + \frac{u(j,N)}{N-1}\right).$$

This net function equals at each point the average of its boundary neighbours (see figure 4.1).

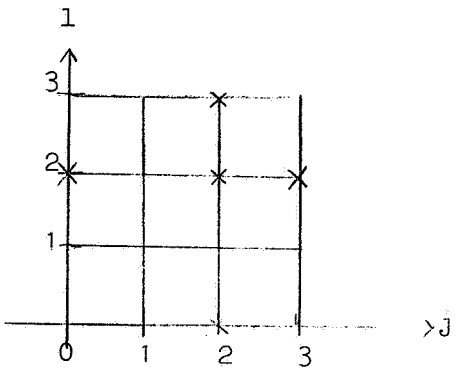

fig. 4.1

The method was applied for N = 10, N = 20 and N = 40. The following table contains the number of iterates K, the estimates a and b for the least and greatest eigenvalue of $L'(1)$, the theoretical average rate of convergence R(K) defined by formula (3.14), the experimental average rate of convergence $R^*(K)$ defined by

$$R^*(K) = -\frac{1}{K} \ln \left[\frac{||L'(1)u_k^*||}{||L'(1)u_0||} \right], \tag{4.5}$$

where u_k^* is the numerical net function and $||\ ||$ denotes the maximum norm. Further we have listed the relative precision $P_r(K)$ after K iterates, defined by

$$P_r(K) = \left|\left| \frac{U - u_K^*}{U} \right|\right| \tag{4.6}$$

and the asymptotic relative precision $P_r(\infty)$.

TABLE 4.1

N	a	b	K	R(K)	$R^*(K)$	$P_r(K)$	$P_r(\infty)$
10	3.9780	23.5009	3	.6453	.8469	91 %	10.65%
			6	.7591	.8828	13 %	
			9	.7976	.8757	10.31%	
			12	.8169	.8645	10.63%	
			15	.8284	.8843	10.65%	
20	7.1301	87.4078	3	.3662	.5241	120 %	2.63%
			6	.4722	.6196	22 %	
			9	.5105	.6199	5.4 %	
			12	.5298	.5877	2.81%	
			18	.5491	.6011	2.63%	
40	13.7313	337.0450	5	.2740	.4173	108 %	.66%
			10	.3400	.4164	17 %	
			15	.3631	.4226	3.2 %	
			20	.3746	.4148	.78%	
			30	.3862	.4157	.65%	

In practice it is important to determine the value of K beforehand; one would desire that K is such that the errors due to discretization and iteration are comparable, i.e.

$$||U - u|| \sim ||u - u_K||. \tag{4.7}$$

We can give a lower bound for K. According to Gerschgorin we have

$$||U - u|| = ch^2, \tag{4.8}$$

where c is a constant. Further we have

$$||u - u_K|| = 2\theta \exp(-2K\sqrt{\frac{a}{b}}) \cdot ||u - u_0||, \tag{4.9}$$

where $0 < \theta \leq 1$.
From (4.7), (4.8) and (4.9) it follows that

$$K \sim \sqrt{\frac{b}{a}} \left[\ln h^{-1} + \frac{1}{2} \ln\left(\frac{2\theta||u - u_0||}{c}\right)\right].$$

Thus, using (3.12), we obtain

$$K \geq 1.41 \frac{\ln h^{-1}}{\sqrt{h}}. \tag{4.10}$$

For very small values of h ($\ln h^{-1} >> \frac{1}{2} \ln(2\theta||u - u_0||/c)$) the equality sign approximately holds. For instance, the case N = 40 requires at least 14 iterations.

From table 4.1 one may conclude that the numerical rates of convergence are in agreement with the theoretical rates of convergence predicted by the theory of the preceding sections.

5. References

Coolen, T.M.T., Houwen, P.J. van der [1968] — On the acceleration of Richardson's method IV. A non-symmetrical case. Report TW 109, Math. Centre, Amsterdam.

Forsythe, G.E., Wasow, W.R. [1960] — Finite difference methods for partial differential equations. John Wiley & Sons, Inc., New York.

Houwen, P.J. van der [1967a] — On the acceleration of Richardson's method III. Applications. Report TW 108, Math. Centre, Amsterdam.

Houwen, P.J. van der [1968] — Finite difference methods for solving partial differential equations. MC Tract 20, Math. Centre, Amsterdam.

Stiefel, E.C. [1958] — Kernel polynomials in linear algebra and their numerical applications. National Bureau of Standards Applied Mathematics Series 49, "Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues", U.S. Government Printing Office, Washington, D.C.