ON A HIGH-ACCURACY DIFFERENCE SCHEME FOR AN ELLIPTIC EQUATION WITH SEVERAL SPACE VARIABLES*

A. A. SAMARSKII and V. B. ANDREYEV

(Moscow)

(Received 21 June 1963)

1. Suppose that in the region \( D_p = \{0 < x_\alpha < 1, \alpha = 1, \ldots, p\} \) we are looking for a solution to the differential equation

\[
Lu = \sum_{\alpha=1}^{p} L_\alpha u = -f(x), \quad L_\alpha u = \frac{\partial^2 u}{\partial x_\alpha^2},
\]

which satisfies the condition

\[
u |_{\Gamma} = g(x)
\]

on the boundary \( \Gamma \). Let \( \bar{\omega}_h = \{x_i = (i_1 h, i_2 h, \ldots, i_p h) \in \bar{D}_p\} \) be a square net with step \( h = 1/N \); and let \( \gamma \) be the boundary of the net \( \bar{\omega}_h \).

The numerical solution of the problem (1)-(2) is usually found with the use of the difference scheme

\[
\Lambda y + f(x) = 0, \quad y |_{\gamma} = g(x),
\]

where

\[
\Lambda = \sum_{\alpha=1}^{p} \Lambda_\alpha, \quad \Lambda_\alpha y = y_{x_\alpha} x_\alpha
\]

(see [1] for the notation). This scheme gives second order accuracy.

There are many iterative methods for solving the problem (3), and of these we have picked out those used in [2]-[8] which give the fastest rate of convergence. Without going into detail about any one method we

* Zh. vych. mat. 3, No. 6, 1006-1013, 1963.

1373

2. To find the numerical solution of the problem (1)-(2) we use the scheme

\[
\Lambda' y = \Lambda y + \frac{h^2}{6} \sum_{a=1}^{p} \sum_{\beta > a} \Lambda_{a \beta} y = - \varphi(x), \quad y|_{\Gamma} = g,
\]

(5)

where

\[
\varphi(x) = f(x) + \frac{h^2}{12}.
\]

(6)

This scheme has fourth order approximation in the class of sufficiently smooth solutions of (1), so that

\[
\psi = \Lambda'u + \varphi = O(h^4).
\]

(7)

It is not difficult to show that the scheme (5) has fourth order accuracy. Let us introduce the scalar products (see [11]):

\[
(\eta, y) = \sum_{\omega} y_{\omega} \eta_{\omega h}, \quad (y, \eta)_{a} = \sum_{\omega_{a} + 1} y_{\omega} \eta_{\omega h},
\]

(8)

and the norms:

\[
\|\eta\| = \sqrt{(\eta, \eta)}, \quad \|\eta_{x_a}\| = \sqrt{(\eta_{x_a}, \eta_{x_a})_{a}}.
\]

(9)

Let \( u \) be the solution of the problem (1)-(2), and \( y \) the solution of problem (5). For their difference \( z = y - u \) we obtain

\[
\Lambda' z = - \psi, \quad z|_{\Gamma} = 0.
\]

(10)

Making a scalar multiplication of this equation by \( z \) we write down the energy identity (see [11]):

\[
I = \frac{h^2}{6} \sum_{a=1}^{p} \sum_{\beta > a} \|z_{x_a x_{\beta}}\|^2 + (\psi, z), \quad I = \sum_{a=1}^{p} \|z_{x_a}\|^2.
\]

(11)

We use the obvious inequalities:

\[
\|z\|^2 \leq \frac{1}{4p} I, \quad \frac{h^2}{6} \sum_{a=1}^{p} \sum_{\beta > a} \|z_{x_a x_{\beta}}\|^2 \leq \frac{p - 1}{3} I,
\]

(12)

\[
(\psi, z) \leq \|z\| \|\psi\| \leq \frac{1}{\sqrt{4p}} I^{1/2} \|\psi\| \leq c_0\|z\| + \frac{1}{16c_0} \|\psi\|^2.
\]
where \( c_0 \) is an arbitrary positive constant. We insert these estimates in (11) and choose \( c_0 \) correspondingly. We then obtain

\[
\| z \| \leq M_p \| \psi \|, \quad \text{where} \quad M_p = \frac{3}{4p(4-p)}, \quad p \leq 3.
\]  

(13)

We have thus proved the following theorem.

**Theorem 1.** If the condition

\[
\| \psi \| \leq M h^4,
\]  

(14)

is satisfied then the difference scheme (5) for \( p \leq 3 \) converges in the mean at a rate \( O(h^4) \) so that

\[
\| y - u \| \leq M' h^4, \quad M' = M \cdot M_p,
\]  

(15)

where \( M \) is a positive constant which does not depend on \( h \).

**Note 1.** If instead of (1) we consider the equation

\[
\tilde{L} u = L u - q(x) u = - f(x), \quad 0 < c_1 < q(x), \quad u_{|\Gamma} = g(x),
\]  

(1')

then it is easy to see that the solution of the problem

\[
A' y - d y + \varphi(x) = 0, \quad y_{|\Gamma} = g(x),
\]  

(5')

where

\[
d(x) = q(x) + \frac{h^2}{12} A q(x),
\]

converges in the mean at a rate \( O(h^4) \) to the solution of the problem (1') for \( p = 4 \) also.

3. Let us examine the following iterative scheme for the approximate solution of problem (5) for \( p = 2, 3 \):

\[
v_{\tau} = \Lambda v + \frac{h^3}{6} \sum_{\alpha=1}^{p} \sum_{\beta>\alpha} \Lambda_{\alpha} \Lambda_{\beta} \tilde{\psi} + \varphi, \quad v_{|\tau} = g(x), \quad v(x, 0) = v^0(x),
\]  

(16)

where \( v = v^{n+1} \) is the \((n + 1)\)-th iteration, \( \tilde{\psi} = v^{(n)} \), \( v_{\tau} = (v - \tilde{v})/\tau_n \), \( \tau_n > 0 \) is an iterative parameter to be chosen later. The initial value \( v(x, 0) = v^{(0)}(x) \) is determined by the choice of the zero iteration. Let us construct two one-dimensional alternating direction algorithms for the numerical solution of problem (16).

A. We insert \( \Lambda v = \Lambda \tilde{v} + \tau \Lambda v_{\tau} \) in (16) and, following [6], replace
the operator \((E - \tau A)\) where \(E\) is the unit operator by the operator \(A\), where

\[
A = \prod_{\alpha=1}^{p} A_{\alpha}, \quad A_{\alpha} = E - \tau \Lambda_{\alpha}.
\]

Then instead of (16) we have the scheme

\[
Av = [A + \tau \Lambda'] \bar{v} + \tau \varphi, \quad v|_{\gamma} = g, \quad v(x, 0) = v^{(0)}(x), \tag{17}
\]

which we shall call the generating scheme. Introducing intermediate values \(v(1), \ldots, v(p) = v\) we reduce the solution of problem (5) to the solution of \(p\) one-dimensional problems:

\[
A_{1}v_{(1)} = [A + \tau \Lambda'] \bar{v} + \tau \varphi, \tag{18}
\]

\[
A_{\alpha}v_{(\alpha)} = v_{(\alpha-1)}, \quad \alpha = 2, \ldots, p; \quad v_{(\alpha)} = A_{\alpha+1} \ldots A_{p}g \text{ for } x_{\alpha} = 0, 1.
\]

B. Putting \(w = v_{\gamma}\), we rewrite the generating scheme in the form

\[
Aw = \Lambda' \bar{v} + \varphi, \quad w|_{\gamma} = 0. \tag{19}
\]

From this we have the alternating direction algorithm

\[
A_{1}w_{(1)} = \Lambda' \bar{v} + \varphi, \tag{20}
\]

\[
A_{\alpha}w_{(\alpha)} = w_{(\alpha-1)}, \quad \alpha = 2, \ldots, p; \quad w_{(\alpha)} = 0 \text{ for } x_{\alpha} = 0, 1,
\]

\[
v = \bar{v} + \tau w_{(p)}.
\]

To go from \(\bar{v}\) to \(v\) during the computation we must store the two layers: \(\bar{v}\) and \(w_{(\alpha)}, \alpha = 1, 2, \ldots, p\). However this algorithm requires fewer operations than (18) (thus it is not necessary to calculate \(A\bar{v}\)) and, furthermore, the functions \(w_{(\alpha)}\) always satisfy zero boundary conditions. For \(p = 2\) by analogy with [2] we can use an algorithm which does not contain the product \(\Lambda_{1}\Lambda_{2}\bar{v}\):

\[
A_{1}w_{(1)} = \Lambda_{1} \bar{v} + \left(1 + \frac{h_{2}^{2}}{6\tau}\right) \Lambda_{2}\bar{v} + \varphi, \tag{21}
\]

\[
A_{2}w_{(2)} = w_{(1)} - \frac{h_{2}^{2}}{6\tau} \Lambda_{2}\bar{v}, \quad v = \bar{v} + \tau w_{(2)}; \quad w_{(0)} = 0, \quad x_{2} = 0, 1.
\]

Each of the equations \(A_{\alpha}w_{(\alpha)} = \varphi_{\alpha}\) where \(\varphi_{\alpha}\) is a given function can be solved using the formulae of one-dimensional successive substitution (see [9], pp. 283-309). All the computing algorithms which we have
On a high-accuracy difference scheme

4. We show that the iterations defined on scheme (17) converge whatever the choice of the zero iteration \( v^{(0)}(x) \) and of the sequence \( \{\tau_n\} \) satisfying the condition \( 0 < c_1 < \tau_n < c_2 \), where \( c_1 \) and \( c_2 \) are constants which do not depend on the iteration number \( n \). Following [3] we give a method of choosing \( \{\tau_n\} \) for which the rate of convergence of the iterations will be "sufficiently fast". We obtain the following conditions for the difference \( z = v - y \), where \( y \) is the solution of the initial problem (5), \( v = v^{(n)} \) is the solution of problem (17):

\[
A z = A^* \tilde{z}, \quad z|_\gamma = 0, \quad z(x, 0) = z^{(0)}(x) = v^{(0)} - y(x) \tag{22}
\]

Let us expand \( z \) and \( \tilde{z} \) in terms of the eigenfunctions

\[
\mu_k = \prod_{\alpha=1}^{p} \sin k_\alpha \pi x_\alpha, \quad k_\alpha = 1, \ldots, N - 1, \quad k = \{k_1, \ldots, k_p\}, \quad x_\alpha = i_\alpha h, \tag{23}
\]

of the operators \( A_{\alpha'} \):

\[
z = z^{(n+1)} = \sum_k a_k^{(n+1)} \mu_k, \quad \tilde{z} = z^{(n)} = \sum_k a_k^{(n)} \mu_k. \tag{24}
\]

Inserting (24) in (22) and using the linear independence of \( \{\mu_k\} \) we obtain

\[
a_k^{(n+1)} = \rho_k^{(n+1)} a_k^{(n)}, \tag{25}
\]

\[
\rho_k^{(n+1)} = 1 - \lambda \left[ \sum_{\alpha=1}^{p} \xi_\alpha - \frac{2}{3} \sum_{\alpha=1}^{p} \sum_{\beta > \alpha} \xi_\alpha \xi_\beta \right] \prod_{\alpha=1}^{p} (1 + \lambda \xi_\alpha)^{-1}, \tag{26}
\]

\[
\lambda = \lambda_{n+1} = \frac{4 \xi_{n+1}}{h^3}, \quad \xi_\alpha = \xi_{k_\alpha} = \sin^3 \frac{k_\alpha \pi h}{2}.
\]

Theorem 2. The iterative method (17) for \( p = 2, 3 \) converges in the metric \( L^2(\omega_h) \) whatever parameters \( \tau_n \) satisfying the condition \( 0 < c_1 < \tau_n < c_2 \) are chosen.

Thus using (25) we can write

\[
a_k^{(n+1)} = \prod_{\alpha=1}^{n+1} \rho_k^{(s)} a_k^{(0)}, \tag{27}
\]

and it follows from (24) and (27) that

mentioned give the same generating scheme (17) as we can see by eliminating the intermediate values of \( v(x) \) or \( w(x), \quad \alpha = 1, \ldots, p - 1 \).
\[ z^{(n+1)} = \sum_k a_k^{(0)} \prod_{s=1}^{n+1} \rho_k^{(s)} \mu_k(i). \] (28)

Hence

\[ \| z^{(n+1)} \| \leq \left[ \sum_{k=0}^{N-1} h^p \left( \sum_k a_k^{(0)} \prod_{s=1}^{n+1} \rho_k^{(s)} \mu_k(i) \right)^2 \right]^{1/2} \leq \max_k \prod_{s=1}^{n+1} \rho_k^{(s)} \| z^0 \|. \] (29)

where \( z^{(0)} = v^{(0)} - y \) is the difference between the zero approximation and the exact solution of (5). We have to show that

\[ \max_k \prod_{s=1}^{n+1} \rho_k^{(s)} \to 0 \text{ as } n \to \infty. \]

Let us first estimate \( \rho_k^{(s)} \). Since \( 1 \leq k_x \leq N - 1 \) we have

\[ \sin^2 \frac{\pi h}{2} \leq \sum_k \leq 1 \]

and therefore

\[ 2 \xi_\alpha \xi_\beta \leq \sum_k \leq \xi_\alpha + \xi_\beta. \] (31)

Using (26) and (31) we obtain

\[ 0 < \rho_k^{(s)} < 1 \]

\[ \left( 1 - \frac{p-1}{3} \right) \sum_a \xi_a \]

\[ \Pi \frac{p}{(1 + \lambda \xi_a)} \] (32)

It follows from (32) that

\[ \rho_k^{(p)} \leq \rho < 1 \] (33)

for \( 0 < c_1 \leq \lambda \leq c_2, p = 2, 3 \), where \( c_1, c_2 \) and \( \rho \) do not depend on the number of the iteration.

Using (29) and (33) we obtain

\[ \| z^{(n+1)} \| \leq \rho^{n+1} \| z^{(0)} \| \to 0 \text{ as } n \to \infty. \]

Note 2. For equation (1') and the corresponding difference scheme (5') (see Note 1) the generating scheme is

\[ Av = [A + \tau \Lambda' - \tau d] \varphi \] (17')
It is not difficult to see that Theorem 2 remains valid for (17') for \( p = 4 \) also.

5. **Lemma 1.** If \( 0 < q < 1 < M \) and

\[
\rho(a, b) = 1 - \frac{2(a + b)}{3(1 + a)(1 + b)},
\]

then

\[
\rho^2 = \max_{m < a, b < M} \rho(a, b) = \max \left[ 1 - \frac{4M}{3(1 + m)^2}, \ 1 - \frac{4M}{3(1 + M)^2} \right].
\]

In fact it is not difficult to see by a direct check that

\[
\rho^2(a, b) \leq \rho(a, a) \rho(b, b).
\]

Since the region of definition of \( \rho(a, b) \) together with the point \((a, b)\) also contains the points \((a, a), (b, b)\) and conversely, on the basis of (36) we can state that \( \max_{m < a, b < M} \rho(a, b) \) is attained when \( a = b \). Let us examine the behaviour of the function

\[
\overline{\rho}(a) = 1 - \frac{4a}{3(1 + a)^2},
\]

\[
\frac{d\overline{\rho}}{da} = -\frac{4(1 - a)}{3(1 + a)^3} = \begin{cases} > 0 & \text{for } a > 1, \\ < 0 & \text{for } a < 1. \end{cases}
\]

It follows from (37) that \( \max \overline{\rho}(a) \) is attained either when \( a = M \) or when \( a = \frac{a}{2} \), which the following lemma proves.

**Lemma 2.** If \( 0 < q < M < \frac{a}{2} < M \) and

\[
\rho(a, b, c) = 1 - \frac{1}{3} \theta(a, b, c),
\]

where

\[
\theta(a, b, c) = \frac{a + b + c}{(1 + a)(1 + b)(1 + c)},
\]

then

\[
\rho^2 = \max_{m < a, b, c < M} \rho(a, b, c) = \max \left[ 1 - \frac{m}{(1 + m)^2}, \ 1 - \frac{M}{(1 + M)^2} \right].
\]

The lemma will be proved if we can show that the minimum of the function \( \theta(a, b, c) \) is attained either when \( a = b = c = m \), or when \( a = b = c = M \). Let us fix \( a \) and examine the behaviour of \( \theta(a, b, c) \) depending on the change in \( b \) and \( c \). By a direct check we see that
\[ \theta^2 (a, b, c) \geq \theta (a, b, b) \theta (a, c, c). \]  

(40)

Noting that
\[ \frac{\partial \theta (a, b, c)}{\partial b} = 2 \frac{1 - a - b}{(1 + a)(1 + b)^2} = \begin{cases} 0 & \text{for } b \leq 1 - a, \\ > 0 & \text{for } b > 1 - a, \end{cases} \]

(41)

and using (40) we obtain
\[ \bar{\theta} (a) = \min_{m < b, c < M} \theta (a, b, c) = \min_{m < b < M} \theta (a, b, b) = \]

\[ = \min \left[ \frac{a + 2m}{(1 + a)(1 + m)^2}, \frac{a + 2M}{(1 + a)(1 + M)^2} \right]. \]

(42)

Further,
\[ \frac{d \bar{\theta} (a)}{da} = \begin{cases} \text{either } \frac{1 - 2m}{(1 + a)^3(1 + m)^2} > 0, \\ \text{or } \frac{1 - 2M}{(1 + a)(1 + M)^2} < 0. \end{cases} \]

(43)

On the basis of (42) and (43) we conclude:

\[ \min_{m < a < M} \theta (a) = \min_{m < a, b, c < M} \theta (a, b, c) = \min_{m < a < M} \left( \frac{3m}{(1 + m)^2}, \frac{3M}{(1 + M)^2} \right). \]

(44)

6. Now let us choose the sequence \( \{\lambda_n\} \) so that it satisfies the conditions

\[ \lambda_n \xi^{(n)} = m, \quad \lambda_n \xi^{(n+1)} = M, \quad \xi^{(1)} = \sin^2 \frac{\pi h}{2}, \]

(45)

and the number of iterations \( n = n_0 \) so that

\[ \xi^{(n_0)} < 1, \quad \xi^{(n_0+1)} > 1. \]

(46)

It follows that

\[ \lambda_m = mq^{n-1} \sin^{-\frac{\pi h}{2}}, \quad \xi^{(n)} = q^{-n+1} \sin^2 \frac{\pi h}{2}, \]

(47)

\[ 2 \ln \sin \frac{\pi h}{2} \cdot \ln^{-1} q \leq n_0 \leq 2 \ln \sin \frac{\pi h}{2} \ln^{-1} q + 1, \]

(48)

where \( q = m/M \).

Lemma 3. If a cycle of \( n_0 \) iterations using the method (17) is carried out with the set of parameters \( \{\lambda_n\} \) defined in (47) then

\[ \| z^{(n_0)} \| \leq \rho_p \| z^{(0)} \|. \]

(49)
where \( p_p \) is given by formulae (35) and (39).

Thus when

\[
\xi^{(n)} \leq \xi_\alpha \leq \xi^{(n+1)}, \tag{50}
\]
\[
m \leq \lambda_n \xi_\alpha \leq M, \tag{51}
\]

and the intervals \([\xi^n, \xi^{n+1}]\) cover the whole region of values of \( \xi_\alpha \) for each value of \( \xi_\alpha \), there exists in consequence at least one value of \( n \) for which

\[
p_k^{(n)} < p_p. \tag{52}
\]

The inequality (43) and Theorem 2 prove the lemma.

**Note 3.** Bearing (37) and (41) in mind it is easy to see that \( \max(M/m) \) (or \( \min n_0 \), which is equivalent) is attained for fixed \( p_p \) when the first and second terms on the right-hand side of (35) and (39) are equal:

\[
1 - \frac{4m}{3(1 + m)^3} = 1 - \frac{4M}{3(1 + M)^3}; \quad 1 - \frac{m}{(1 + m)^3} = 1 - \frac{M}{(1 + M)^3}. \tag{53}
\]

Then when \( p = 2 \)

\[
M = \frac{1}{m}, \tag{54}
\]

when \( p = 3 \)

\[
M = \frac{\sqrt{(3 + m)^{3/2} + 4/m} - (3 + m)}{2}. \tag{55}
\]

**Theorem 3.** In order to reduce \( L_2 \), the norm of the error \( \|z^0\| \), \( 1/\varepsilon \) times using method (17) it is sufficient to carry out \( k_0 \) cycles of \( n_0 \) iterations with the set of parameters \( \{\lambda_n\} \) given by (47), where \( n_0 \) is defined by (48) and \( k_0 \) by (56):

\[
k_0 \geq \frac{\ln(1/\varepsilon)}{\ln(1/p_p)}. \tag{56}
\]

The proof of the theorem follows from Lemma 3.

**Note 4.** It follows from Theorem 3 that the total number of iterations required to reduce the error \( \|z^0\| \) \( 1/\varepsilon \) times is

\[
\nu \leq \frac{2 \ln \sin \frac{\pi h}{2} \ln \varepsilon}{\ln q \ln p_p}. \tag{57}
\]

Using Note 3 and optimising \( \nu \) w.r.t. \( m \) we obtain:
for $p = 2$

$$v_{opt} = 3.00 \ln \frac{1}{\sin (\pi h/2)} \ln \frac{1}{\epsilon},$$

$$m_{opt} = 0.24, \quad \rho_{\text{opt}} = 0.79, \quad q = \frac{m}{M} = 0.058;$$

for $p = 3$

$$v_{opt} = 8.40 \ln \frac{1}{\sin (\pi h/2)} \ln \frac{1}{\epsilon},$$

$$m_{opt} = 0.13, \quad \rho_{\text{opt}} = 0.91, \quad q = \frac{m}{M} = 0.080.$$

REFERENCES


