

AN ECONOMICAL ALGORITHM FOR THE NUMERICAL SOLUTION OF SYSTEMS OF DIFFERENTIAL AND ALGEBRAIC EQUATIONS*

A. A. SAMARSKII

(Moscow)

(Received 9 January 1964)

By analogy with [1], [2], [3], an economical (in the sense of number of operations) difference scheme of the second order of accuracy (alternating triangular scheme) is proposed for the solution of a system of ordinary differential equations of the first order. The scheme can be used as an iterative process for the solution of a system of linear algebraic equations. All the results can be extended to the case of linear operator equations.

1. We take Cauchy's problem for the system of equations

$$\frac{du}{dt} + Au = f(t), \quad 0 < t \leq T, \quad u(0) = u_0, \quad (1)$$

where $u = u(t) = (u_1(t), \dots, u_n(t))$, $f(t) = (f_1(t), \dots, f_n(t))$ are vectors, $A = A(t) = (a_{ik}(t))$ is a square $n \times n$ matrix.

Let $\omega_\tau = \{t_j = j\tau\}$ be a net with interval τ on the closed interval $[0, T]$. Several difference schemes can be used for the solution of problem (1). We shall compare them as regards the number of arithmetical operations q , expended in passing from the layer t_j to the layer t_{j+1} , i.e. in determining the vector y^{j+1} from a given y^j . For the explicit scheme $y^{j+1} = y^j + \tau(f - Ay)^j$ we have $q = 2n^2 + 2n$; but in accuracy it is of first order. The implicit scheme $y^{j+1} + 0.5\tau(Ay)^{j+1} = y^j - 0.5\tau(Ay)^j + 0.5\tau(f^j + f^{j+1})$ has the second order of accuracy, but for this $q = O(n^3)$. We shall describe a scheme as economic when $q = O(n^2)$ for it, as in the case of an explicit scheme.

* Zh. vych. mat., 4, No. 3, 580-585, 1964.

2. We shall assume that the matrix A can be represented as the sum of two triangular positive definite matrices:

$$A = A_1 + A_2, \quad A_1 = (a_{ik}^-), \quad A_2 = (a_{ik}^+), \quad a_{ik}^- = 0, \quad k > i; \quad a_{ii}^- + a_{ii}^+ = a_{ii}, \quad (2)$$

$$(A_\alpha \xi, \xi) \geq c_1 \|\xi\|^2, \quad c_1 = \text{const} > 0, \quad \alpha = 1, 2, \quad (3)$$

where $(v, \xi) = \sum_{i=1}^n v_i \xi_i$, $\|\xi\| = \sqrt{(\xi, \xi)}$ are the scalar product and norm, ξ is an arbitrary real vector. If A is a symmetric matrix, we naturally put $a_{ii}^- = a_{ii}^+ = a_{ii}/2$; the symmetry conditions now yield

$$(\xi, A_1 v) = (A_2 \xi, v), \text{ i. e. } A_1^* = A_2, \quad A_2^* = A_1. \quad (4)$$

We consider the following two-interval difference scheme:

$$\left. \begin{aligned} \frac{y - \check{y}}{\tau} + A_1 y + \check{A}_2 \check{y} &= f(t), & y(0) &= u_0, \\ \frac{\hat{y} - y}{\tau} + A_1 y + \hat{A}_2 \hat{y} &= f(t), \end{aligned} \right\} \quad (5)$$

where $t = t_{2j+1}$, $\check{t} = t_{2j}$, $\hat{t} = t_{2j+2}$, $y = y(t)$, $\check{y} = y(\check{t})$,

$$\hat{y} = y(\hat{t}), \quad A_1 = A_1(t), \quad \check{A}_2 = A_2(\check{t}), \quad \hat{A}_2 = A_2(\hat{t}).$$

To find y , given \check{y} , we must invert the triangular matrix $E + \tau A_1$, and to find \hat{y} we must invert the triangular matrix $E + \tau A_2$. This is possible for any $\tau > 0$, if $a_{ii}^\pm \geq 0$, and $O(n^2)$ operations are required, i.e. scheme (5) is economic. We write down the computation formulae as

$$\left. \begin{aligned} y_i &= \frac{\check{y}_i + \tau(f_i - \check{v}_i^+ - a_{ii}^+ \check{y}_i - v_i^-)}{1 + \tau a_{ii}^-}, & \check{v}_i^+ &= \sum_{k=i+1}^n \check{a}_{ik} \check{y}_k, & i &= 1, \dots, n, \\ \hat{y}_i &= \frac{y_i + \tau(f_i - v_i^- - a_{ii}^- y_i - \hat{v}_i^+)}{1 + \tau a_{ii}^+}, & v_i^- &= \sum_{k=1}^{i-1} a_{ik} y_k, & y_i(0) &= u_{0i}. \end{aligned} \right\} \quad (6)$$

If, in addition to the vector \check{y}_i , we store the vector \check{v}_i^+ , to find y_i (in one step) we have to use $q = n^2 + 7n$ arithmetical operations; on storing v_i^- and y_i , we can also find \hat{y}_i by using only $q = n^2 + 7n$ operations. In the explicit scheme, one step requires $q = 2n^2 + 2n$ operations. Thus, with $n > 5$ the proposed two-step scheme requires fewer operations than even the explicit scheme with the interval τ . In addition, as will be shown below, scheme (5) has the second order of accuracy.

3. Let us find the error of approximation of the scheme. Let u be the solution of problem (1), y the solution of problem (5). We have for $z = y - u$:

$$z + \tau A_1 z = \check{z} - \tau \check{A}_2 \check{z} - \tau \psi_1, \quad \hat{z} + \tau \hat{A}_2 \hat{z} = z - \tau A_1 z - \tau \psi_2, \quad z(0) = 0, \quad (7)$$

where ψ_1, ψ_2 are local approximation errors. The total approximation error of scheme (5) is

$$\psi = \psi_1 + \psi_2 = (\hat{A}_2 \hat{u} - 2A_2 u + \check{A}_2 \check{u}) + 2 \left(\frac{\hat{u} - \check{u}}{2\tau} - \frac{du}{dt} \right).$$

$$\text{If } A_\alpha(t) u, \frac{du}{dt} \in C^{(1,1)}[0, T], \text{ then } \psi = O(\tau^2). \quad (8)$$

We can write $\psi_\alpha, \alpha = 1, 2$, as

$$\left. \begin{aligned} \psi_\alpha &= \dot{\psi}_\alpha + \psi_\alpha^*, & \dot{\psi}_\alpha &= O(\tau), & \psi_\alpha^* &= O(\tau^2), & \alpha &= 1, 2, \\ \dot{\psi}_1 + \dot{\psi}_2 &= 0, & \dot{\psi}_2 &= \tau \left[\frac{d}{dt} (A_2 u) + \frac{1}{2} \frac{d^2 u}{dt^2} \right]. \end{aligned} \right\} \quad (8')$$

We shall assume in future that conditions (8) are fulfilled.

4. We turn to the derivation of the *a priori* inequalities for the solution of problem (7). Let (z, ξ) be the scalar product and $\|z\| = \sqrt{(z, z)}$ the related norm. We introduce the norm

$$\|z\|_\alpha^2 = \|z\|^2 + \tau^2 \|A_\alpha z\|^2, \quad \alpha = 1, 2. \quad (9)$$

Lemma 1. We have the inequalities

$$2\tau (A_\alpha z, z) \geq \sigma_\alpha \|z\|_\alpha^2, \quad \sigma_\alpha = \frac{2c_1 \tau}{1 + \tau^2 \|A_\alpha\|^2}, \quad (10)$$

$$\|z + \tau A_\alpha z\|^2 \geq (1 + \sigma_\alpha) \|z\|_\alpha^2, \quad \|z - \tau A_\alpha z\|^2 \leq (1 - \sigma_\alpha) \|z\|_\alpha^2, \quad \alpha = 1, 2. \quad (11)$$

It is sufficient to prove (10). It follows from (3) and (9) that

$$\|z\|_\alpha^2 \leq (1 + \tau^2 \|A_\alpha\|^2) \|z\|^2 \leq (1 + \tau^2 \|A_\alpha\|^2) \frac{1}{c_1} (A_\alpha z, z).$$

On rewriting the first of equations (7) in the form $z + \tau A_1 z + \tau \psi_1 = \check{z} - \tau \check{A}_2 \check{z}$, evaluating the squares of the norms of both sides and using Lemma 1, we obtain

$$(1 + \sigma_1) \|z\|_1^2 \leq (1 - \sigma_2) \|\check{z}\|_2^2 - \tau^2 \|\psi_1\|^2 - 2\tau (\psi_1, z + \tau A_1 z), \quad (12)$$

$$(1 + \sigma_2) \|\hat{z}\|_2^2 \leq (1 - \sigma_1) \|z\|_1^2 + \tau^2 \|\Psi\|^2 - 2\tau (\Psi_2, z - \tau A_1 z). \quad (13)$$

Let A_1^* be the operator adjoint to A_1 ; now

$$(A_1 z, \Psi_2) = (z, A_1^* \Psi_2) \leq \|z\| \|A_1^* \Psi_2\|. \quad (14)$$

We sum (12) and (13) and use the inequality

$$\begin{aligned} -2\tau (z, \Psi_1 + \Psi_2) &\leq c_0 \tau \|z\|^2 + \frac{1}{c_0} \tau \|\Psi_1 + \Psi_2\|^2, & 2\tau^2 (z, A_1^* (\Psi_2 - \Psi_1)) &\leq \\ &\leq c_0 \tau \|z\|^2 + \frac{1}{c_0} \tau^2 \|A_1^* (\Psi_2 - \Psi_1)\|^2, \end{aligned}$$

where $c_0 = 2c_1/(1 + \tau^2 \|A_1\|^2) = \sigma_1/\tau$. We now obtain

$$\|\hat{z}\|_2^2 \leq \frac{1 - \sigma_2}{1 + \sigma_2} \|\check{z}\|_2^2 + \frac{\tau}{1 + \sigma_2} \|\Psi\|^2 \leq \|\check{z}\|_2^2 + \tau \|\Psi\|^2, \quad (15)$$

$$\|\Psi\|^2 = [\tau \|\Psi_1 + \Psi_2\|^2 + \tau^2 \|A_1^* (\Psi_2 - \Psi_1)\|^2] \frac{1 + \tau^2 \|A_1\|^2}{c_1} + \tau \|\Psi_2\|^2 - \|\Psi_1\|^2. \quad (16)$$

It follows from (15) and the initial condition $z(0) = 0$ that

$$\|z^{2j}\|_2 \leq \|\overline{\Psi^{2j}}\|, \quad j = 0, 1, \dots, \|\overline{\Psi^j}\| = \left(\sum_{j'=1}^j \tau \|\Psi^{j'}\|^2 \right)^{1/2}. \quad (17)$$

We have thus proved

Theorem 1. Given the auxiliary condition (3), inequality (17) holds for the solution of problem (7).

Theorem 2. If conditions (3) and (8) are fulfilled, scheme (5) is of the second order of accuracy:

$$\|z^j\| \leq M\tau^2, \quad j=1, 2, \dots,$$

where M is a positive constant independent of τ .

To prove Theorem 2, it is sufficient to show that conditions (8') imply $\|\Psi\| = O(\tau^2)$, then use Theorem 1.

5. All the above results (except for formulae (6)) retain their force if A_1, A_2 are arbitrary linear operators in Hilbert space H satisfying (3). If A_α are self-adjoint operators, scheme (5) is a generalization of the familiar algorithm of alternating directions for the two-dimensional

equation of heat conduction. With a slight modification of the arguments, the method described above can be used to show that the algorithm [4] is convergent for an arbitrary region at the rate $O(h^2 + \tau^2 h^{-1/2})$, where h is the interval of the spatial net.

It can be shown in particular, by using Theorem 1, that the alternating method [5] for the one-dimensional equation of heat conduction is convergent at a rate $O(\tau^2/h^2 + h^2)$.

An economic method for solving a system of equations of the parabolic type:

$$\frac{\partial u}{\partial t} = \frac{\partial}{\partial x} \left(k \frac{\partial u}{\partial x} \right), \quad u = (u_1, \dots, u_n), \quad k = (k_{ij}), \quad (18)$$

$$\frac{y - \check{y}}{\tau} + A_1 y + A_2 \check{y} = 0, \quad A_1 y = - (a^- y_x^-)_x, \quad A_2 y = - (a^+ y_x^+)_x.$$

was proposed in [1]. In this case the algorithm (5) yields a higher order of accuracy $O(\tau^2 h^{-1/2} + h^2)$.

Problem (1) can also be solved by using the scheme

$$\frac{y - \check{y}}{\tau} + 2A_1 y = f, \quad \frac{\check{y} - y}{\tau} + 2A_2 \check{y} = f, \quad y(0) = u_0, \quad (19)$$

which is an analogue of the locally one-dimensional scheme [6]. This scheme is of the first order of accuracy and is economic.

It should be noted that condition (3) implies no loss of generality, if we bear in mind the requirement $u = u e^{\mu t}$, where $\mu = \text{const.} > 0$ is arbitrary. We only need to require that $(A_\alpha \xi, \xi) \geq -\text{const.} \|\xi\|^2$, $\alpha = 1, 2$, $\text{const.} > 0$.

Condition (3) is necessary for Sections 6 and 7.

6. Given the system of n linear algebraic equations

$$Au = A_1 u + A_2 u = f, \quad A = (a_{ik}), \quad (20)$$

where A_1, A_2 are triangular positive definite matrices.

Scheme (5) can be used as an iterative process for solving equations (20), taking y^j as the iteration number j .

Let us take the following iteration scheme:

$$(D + A_1) y = (D - A_2) \check{y} + f, \quad y(0) = y_0, \quad (21)$$

$$(D + A_2) y = (D - A_1) y + f \quad D = \left(\frac{1}{\tau_{ik}} \right), \quad \tau_{ik} > 0, \quad (22)$$

where y_0 is the arbitrary initial approximation, D is a diagonal matrix, $\check{y} = y^{2j}$ is the iteration number $2j$, $y = y^{2j+1}$, $\hat{y} = y^{2j+2}$. If the iterations are found in accordance with scheme (21), taking $y = y^{j+1}$, $\check{y} = y^j$, $j = 0, 1, \dots$, in the case $1/\tau_{ik} = a_{ik}/2$ we obtain the Nikrasov-Seidel method [7] (the "downwards" algorithm). Along with (21), we can consider the "upwards" algorithm (22) (taking $\hat{y} = y^{j+1}$, $y = y^j$, $j = 0, 1, \dots$). It follows from the analogue of formula (6) that the alternating triangular algorithm (21)-(22) is more economic than each of algorithms (21) and (22) separately.

If A_1, A_2 are linear self-adjoint operators in Hilbert space H , the method (21)-(22) is a generalization of the iterative method [4] for solving the Dirichlet difference problem.

7. To prove the convergence of the iterations, we need to obtain an upper bound for the solution of the Cauchy problem

$$(D + A_1) z = (D - A_2) \check{z}, \quad (D + A_2) \hat{z} = (D - A_1) z, \quad z(0) = z_0 \quad (23)$$

for the error $z = y - u$, where u is the exact solution of (20), \hat{y}, y, \check{y} are the iteration numbers $2j + 2$, $2j + 1$ and $2j$ ($j = 0, 1, \dots$) respectively. If $\tau_{ik} = \text{const.} = \tau$, we shall utilize the norm (9) for the upper bound of z ; in the general case, when $D = (1/\tau_{ik})$ is a diagonal matrix, we shall understand by $\|z\|_\alpha^2$ the expression

$$\|z\|_\alpha^2 = \|D^{1/2} z\|^2 + \|D^{-1/2} A_\alpha z\|^2, \quad \alpha = 1, 2. \quad (24)$$

Theorem 3. If A_1, A_2 satisfy condition (3), the alternating triangular scheme (21)-(22) is convergent at the rate of a geometric progression

$$\|y^{2j} - u\|_2 \leq \bar{\rho} \|y^{2(j-1)} - u\|_2, \quad 0 < \bar{\rho} < 1, \quad j = 1, 2, \dots, \quad (25)$$

with any positive definite diagonal matrix D .

It is sufficient to perform the proof for $\tau_{ik} = \text{const.} = \tau$. By Lemma 1, we obtain from (23):

$$(1 + \sigma_1) \|z\|_1^2 \leq (1 - \sigma_2) \|\check{z}\|_2^2, \quad (1 + \sigma_2) \|z\|_2^2 \leq (1 - \sigma_1) \|\hat{z}\|_1^2, \quad (26)$$

where σ_k is given by (10). Hence we eliminate $\|z\|_1^2$:

$$\|\hat{z}\|_2^2 \leq \rho^2 \|\check{z}\|_2^2, \quad \rho^2 = \frac{(1 - \sigma_1)(1 - \sigma_2)}{(1 + \sigma_1)(1 + \sigma_2)}. \quad (27)$$

Since $0 < \sigma_k \leq 1$ for $\tau > 0$, we always have $\rho^2 < 1$. On introducing $c_2 = \max(\|A_1\|, \|A_2\|)$, we obtain

$$\rho \leq \bar{\rho} = \frac{1 - \sigma}{1 + \sigma}, \quad \text{где } \sigma = \frac{2c_1\tau}{1 + c_2^2\tau^2}. \quad (28)$$

The minimum of $\bar{\rho}$ is obtained with

$$\tau = \tau_0 = 1/c_2 \quad (29)$$

and is equal to $\bar{\rho}_{\min} = (1 - \eta)/(1 + \eta)$, where $\eta = c_1/c_2$. Condition (29) may conveniently be employed in practice. In the general case $\sigma = 2c_1\tau_*/(1 + \tau^*)^2 c_2^2$, where $\tau_* = \min_i \tau_{ii}$, $\tau^* = \max_i \tau_{ii}$; for the proof, we write equations (23) in the form

$$\begin{aligned} \frac{1}{\sqrt{\tau_{ii}}} z_i + \sqrt{\tau_{ii}} (A_1 z)_i &= \frac{1}{\sqrt{\tau_{ii}}} \check{z}_i - \sqrt{\tau_{ii}} (A_2 \check{z})_i, \\ \frac{1}{\sqrt{\tau_{ii}}} \hat{z}_i + \sqrt{\tau_{ii}} (A_2 \hat{z})_i &= \frac{1}{\sqrt{\tau_{ii}}} z_i - \sqrt{\tau_{ii}} (A_1 z)_i, \quad i = 1, \dots, n, \end{aligned}$$

the square of the usual norm of both sides is evaluated, and the analogue of Lemma 1 is applied:

$$2(z, A_\alpha z) \geq \sigma_\alpha \|z\|_\alpha^2, \quad \sigma_\alpha = 2c_1 \quad A_\alpha \text{ в } (\tau^*), \quad \alpha = 1, 2.$$

8. *Note 1.* For each of the algorithms (21) and (22) individually, an upper bound of the form $\|z\|_\alpha^2 \leq \rho_\alpha \|\check{z}\|_\alpha^2$, is obtained, where $\rho_\alpha = (1 - \sigma_\alpha)/(1 + \sigma_\alpha)$, $\alpha = 1, 2$. This implies, in particular, the convergence of the Nekrasov-Seidel method for any matrix A , represented as the sum of two positive-definite triangular matrices for which $a_{ii}^- = a_{ii}^+ = a_{ii}/2$. The Nekrasov method is generally speaking convergent at the same rate as the corresponding alternating-triangular algorithm (21)-(22) (cf. [7], Chap. II, Section 34), but method (21)-(22) is more economic, since $n^2 + 7n$ operations instead of $2n^2 + 5n$ are required in this case for the evaluation of one iteration.

Note 2. Theorem 3 holds if A_1, A_2 are arbitrary positive-definite linear operators in Hilbert space H .

Note 3. Let A_α ($\alpha = 1, 2$) be a self-adjoint finite-dimensional linear operator in H , $\lambda_1^{(\alpha)}$, $\lambda_n^{(\alpha)}$ its minimum and maximum eigenvalues,

$\lambda_1 = \min(\lambda_2^{(1)}, \lambda_1^{(2)}), \lambda_n = \max(\lambda_n^{(1)}, \lambda_n^{(2)})$. Now $\lambda_1 \| \xi \| \leq (\xi, A_\alpha \xi) \leq \lambda_n \| \xi \|^2$, and we obtain $\sigma = 2\lambda_1 \tau / (1 + \tau^2 \lambda_1 \lambda_n)$, $\max \sigma = \sqrt{\bar{\eta}}$, $\eta = \lambda_1 / \lambda_n$, $\min \bar{\rho} = (1 - \sqrt{\bar{\eta}}) / (1 + \sqrt{\bar{\eta}})$.

In particular, for the Dirichlet difference problem in the square ($0 \leq x_\alpha \leq 1, \alpha = 1, 2$)

$$\lambda_1 = 4 \sin^2 \frac{\pi h}{2} / h^2, \quad \lambda_n = 4 \cos^2 \frac{\pi h}{2} / h^2$$

and the iterations [4] are convergent at the rate

$$\bar{\rho} = \left(1 - \tan \frac{\pi h}{2}\right) / \left(1 + \tan \frac{\pi h}{2}\right),$$

so that the number of iterations $\sim 1/h$. This is the same as the formula [8], obtained by another method. It can be shown that the iterative scheme [4] can be extended to an arbitrary region.

9. In the case of the system of second order equations

$$\frac{d^2 \mathbf{u}}{dt^2} + A(t) \mathbf{u} = f(t), \quad \mathbf{u}(0) = \mathbf{u}_0, \quad \frac{d\mathbf{u}}{dt}(0) = \bar{\mathbf{u}}_0 \quad (30)$$

the alternating-triangular scheme has the form

$$\left. \begin{aligned} \check{y}_{\check{ii}} + \check{A}_1 \check{y} + \check{A}_2 \check{y} &= \check{f}, & \hat{y}_{\hat{ii}} + A_1 \hat{y} + A_2 \hat{y} &= f, \\ \mathbf{y}(0) &= \mathbf{u}_0, & \mathbf{y}(\tau) &= \bar{\mathbf{u}}_0, \end{aligned} \right\} \quad (31)$$

where

$$\begin{aligned} y_{\check{ii}} &= y_{\check{ii}}^{2j+1}, & \hat{y}_{\hat{ii}} &= y_{\hat{ii}}^{2j+2}, & y_{\hat{ii}}^{j+1} &= (y^{j+1} - 2y^j + y^{j-1})/\tau^2, \\ \check{y} &= y^{2j-1}, & \bar{\mathbf{u}}_0 &= \mathbf{u}_0 + \tau \bar{\mathbf{u}}_0 + 0.5 \tau^2 (A(0) \mathbf{u}_0 - f(0)). \end{aligned}$$

If A is a symmetric matrix, $A_1^* = A_2$, the scheme has the second order of accuracy.

Translated by D.E. Brown

REFERENCES

1. Samarskii, A.A., Homogeneous difference schemes using non-uniform nets for equations of the parabolic type. *Zh. vych. mat.*, 3, No.2, 266-298, 1963.

An economical algorithm

2. Tikhonov, A.N. and Samarskii, A.A., On the theory of homogeneous difference schemes. Outlines of the Joint Soviet-American Symposium on Partial Differential Equations, 266-274, Novosibirsk, August, 1963.
3. Samarskii, A.A., On numerical methods of solving multi-dimensional problems. Outlines of the Joint Soviet-American Symposium on Partial Differential Equations, 236-240. Novosibirsk, August, 1963.
4. Peaceman, D.W. and Rachford, H.H., The numerical solution of parabolic and elliptic differential equations. *J. Industr. Math. Soc.*, 3, No. 1, 28-41, 1955.
5. Saul'ev, V.K., *Integrirovaniye uravnenii parabolicheskogo tipa metodom setok* (Integration of equations of the parabolic type by the net method). Moscow, Fizmatgiz, 1960.
6. Samarskii, A.A., An economic difference method for solving a multi-dimensional parabolic equation in an arbitrary region. *Zh. vych. mat.*, 2, No. 5, 787-811, 1962.
7. Faddeev, D.K. and Faddeeva, V.N., *Vychislitel'nye metody lineinoy algebry* (Computational methods of linear algebra). Moscow-Leningrad, Gost., 1963.
8. Lees, M., A note on the convergence of alternating direction methods. *Math. Comput.*, 16, No. 7; 70-75, 1962.