Aplikace matematiky

Milan Práger; Jiří Taufer; Emil Vitásek

Overimplicit multistep methods

Aplikace matematiky, Vol. 18 (1973), No. 6, 399–421

Persistent URL: http://dml.cz/dmlcz/103497

OVERIMPLICIT MULTISTEP METHODS

MILAN PRÁGER, JIŘÍ TAUFER, EMIL VITÁSEK

(Received January 12, 1973)

1. INTRODUCTION

The efficient solution of many technical problems leading to initial-value problems for ordinary differential equations (typical examples are stiff problems) by multistep difference methods calls not only for high asymptotic accuracy but also for satisfying other requirements. One of such requirements is Dahlquist's A-stability which has often proved very reasonable. It is well-known, however, that in the class of basic methods for the numerical solution of initial-value problems (linear multistep methods, Runge-Kutta methods), A-stable methods of order higher than 2 do not exist (Dahlquist [1963]). This to a great extent negative result made us to seek a larger class of methods that would include A-stable methods of arbitrarily high order. Since it is also well-known that A-stable linear multistep methods are necessarily implicit (cf. again Dahlquist [1963]), the implicit character of our methods will be emphasized in such a way that instead of computing the approximate solution at one point from the (known) approximate solutions at l preceding points (as it is in the case of linear l-step methods) we shall compute the approximate solutions at k successive points simultaneously from some (generaly nonlinear) system of equations, supposing that the solution is known at l successive points. For this reason our methods will be called overimplicit methods.

In the paper, necessary and sufficient conditions for the convergence of overimplicit methods are given and the existence of A-stable methods of arbitrarily high order is studied.

2. OVERIMPLICIT MULTISTEP METHODS

In this section we define a general overimplicit multistep method. For the sake of simplicity, we shall treat only one differential equation of the first order

$$y' = f(x, y) \quad \text{in} \quad \langle a, b \rangle \tag{2.1}$$

(a, b being finite numbers) with the initial condition

$$y(a) = \eta\,. \tag{2.2}$$

Let us note, however, that all what follows is true also for systems of ordinary differential equations. The right-hand term of the given differential equation is assumed to be defined, continuous and satisfying the Lipschitz condition with respect to y (with a constant L independent of x) in the strip $a \leqq x \leqq b$, $-\infty < y < \infty$ so that the solution of the problem (2.1), (2.2) exists and is unique in the whole interval $\langle a, b\rangle$. The approximate solution will be sought at the points $x_i = a + ih$, $i = 0, 1, \ldots$ (or at some of them), where $h > 0$ is the mesh-size, and will be denoted by y_i. One step of the method under consideration consists – as it was already mentioned – in computing the values $y_{n+1}, \ldots, y_{n+k}$ of the approximate solution at the points $x_{n+1}, \ldots, x_{n+k}$ (assuming $y_{n-l+1}, \ldots, y_n$ to be known) simultaneously from the system

$$\begin{bmatrix} y_{n+1} \\ \vdots \\ y_{n+k} \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} y_{n-l+1} \\ \vdots \\ y_n \end{bmatrix} = h\boldsymbol{C} \begin{bmatrix} f_{n+1} \\ \vdots \\ f_{n+k} \end{bmatrix} + h\boldsymbol{D} \begin{bmatrix} f_{n-l+1} \\ \vdots \\ f_n \end{bmatrix} \tag{2.3}$$

where $f_j = f(x_j, y_j)$, $\boldsymbol{C}$ is a square matrix of order k and $\boldsymbol{B}$, $\boldsymbol{D}$ are $k \times l$ matrices.

The fact that the function $f(x, y)$ satisfies the Lipschitz condition guarantees the existence and the uniqueness of the solution of (2.3) for any sufficiently small h so that one step of our method is well-defined. In order to describe the whole method it is necessary, moreover, to indicate how to continue in the following step, i.e., how to choose l new initial values. The method will be practicable obviously only in the case when the new initial values will be chosen fom the values $y_{n-l+2}, \ldots, y_{n+k}$. Because this may be done in different ways, specify the new initial values as $y_{n-l+1+s}, \ldots$ $\ldots, y_{n+s}$ where s is an integer, $1 \leqq s \leqq k$. Hence our method is characterized not only by the matrices $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ but also by the parameter s. Let us note that if $s < k$ it is necessary to forget the values $y_{n+s+1}, \ldots, y_{n+k}$ just computed and to recompute them in the following step. To simplify the notation, we shall always denote the value of the approximate solution at the point x_j by only one symbol y_j even though this value need not be the same in different phases of the computation. This licence cannot cause any misunderstanding.

3. CONVERGENCE OF OVERIMPLICIT METHODS

Before formulating the main result of this section it is necessary to introduce some concepts and notations.

Definition 3.1. *The metod* (2.3) *given by the matrices* $\boldsymbol{B} = \{b_{ij}\}$, $\boldsymbol{C} = \{c_{ij}\}$, $\boldsymbol{D} = \{d_{ij}\}$ *and a parameter* s *is said to be of order* p (p *positive integer*) *if the*

following $k(p+1)$ *conditions are satisfied*:

$$1 + \sum_{j=1}^{l} b_{ij} = 0\,, \quad i - \sum_{j=1}^{l} b_{ij}(l-j) = \sum_{j=1}^{k} c_{ij} + \sum_{j=1}^{l} d_{ij}\,, \tag{3.1}$$

$$i^{\nu} + (-1)^{\nu} \sum_{j=1}^{l} b_{ij}(l-j)^{\nu} = \nu\Big[\sum_{j=1}^{k} c_{ij} j^{\nu-1} + (-1)^{\nu-1} \sum_{j=1}^{l} d_{ij}(l-j)^{\nu-1}\Big]\,,$$

$$\nu = 2, \ldots, p\,; \quad i = 1, \ldots, k\,.$$

Definition 3.2. *The method* (2.3) *is said to be consistent if it is of order at least one.*

Let us draw the reader's attention to the fact that both the consistence and the order of the method are local properties of the method, i.e., they depend only on the matrices $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$ and do not depend on the parameter s.

Definition 3.3. *Let* $y \in C^1$ *and put*

$$\begin{bmatrix} y(x+h) \\ \vdots \\ y(x+kh) \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} y(x-(l-1)h) \\ \vdots \\ y(x) \end{bmatrix} - h\boldsymbol{C} \begin{bmatrix} y'(x+h) \\ \vdots \\ y'(x+kh) \end{bmatrix} - \tag{3.2}$$

$$- h\boldsymbol{D} \begin{bmatrix} y'(x-(l-1)h) \\ \vdots \\ y'(x) \end{bmatrix} = \boldsymbol{L}(y(x); h)\,.$$

The vector $\boldsymbol{L}(y(x); h)$ *with components* $L_i(y(x); h)$ *is called the local error of the method.*

The conditions (3.1) express that the local error of the method (2.3) is of order h^{p+1}. More precisely, we have

Lemma 3.1. *Let the method* (2.3) *of order* p *be given; let* $y(x) \in C^{p+1}(\langle a, b\rangle)$ *and let* $Y = \max |y^{(p+1)}(x)|$. *Then there exists a constant* K *(depending only on the matrices* $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$*) such that*

$$|L_i(y(x); h)| \leqq KYh^{p+1} \tag{3.3}$$

for $i = 1, \ldots, k$ *and for any* $x \in (a, b)$ *for which* $\boldsymbol{L}(y(x); h)$ *has sense.*

Proof follows directly from Taylor's formula.

Since we are dealing with the multistep method it can be expected that the convergence will not be guaranteed by the single assumption that the local error is small but that some other conditions similar to Dahlquist's stability conditions will have to be fulfilled (cf., for example, Henrici [1962]). In order to be able to formulate them let us introduce some further notation.

Let the method (2.3) be given and let firstly $l \leqq s$. Define the matrix $\boldsymbol{R}$ by the equation

$$\boldsymbol{R} = [\boldsymbol{O}_{l,s-l}, \boldsymbol{I}_l, \boldsymbol{O}_{l,k-s}] \tag{3.4}$$

where $\boldsymbol{O}_{m,n}$ is $m \times n$ null matrix[1]) and $\boldsymbol{I}_l$ is the unit matrix of order l. Further, define the matrix $\boldsymbol{E}$ by the equation

$$\boldsymbol{E} = -\boldsymbol{R}\boldsymbol{B}\,. \tag{3.5}$$

Secondly, let $l > s$ and define the matrx $\boldsymbol{S}$ by

$$\boldsymbol{S} = [\boldsymbol{I}_s, \boldsymbol{O}_{s,k-s}]\,. \tag{3.6}$$

Further, let

$$i = \left[\frac{l-1}{s}\right], \tag{3.7}$$

[2])

construct the matrix

$$\boldsymbol{B}^{(1)} = [\boldsymbol{O}_{k,(i+1)s-l}, \boldsymbol{B}] \tag{3.8}$$

and divide the matrix $\boldsymbol{S}\boldsymbol{B}^{(1)}$ into $i + 1$ square blocks in such a way that

$$\boldsymbol{S}\boldsymbol{B}^{(1)} = [\boldsymbol{B}_0, \ldots, \boldsymbol{B}_i]\,. \tag{3.9}$$

Finally, construct the matrix

$$\boldsymbol{E} = \begin{bmatrix} \boldsymbol{O}_{s,s} & \boldsymbol{I}_s & \boldsymbol{O}_{s,s} & \ldots & \boldsymbol{O}_{s,s} \\ \vdots & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & \ddots & \boldsymbol{O}_{s,s} \\ \boldsymbol{O}_{s,s} & \ldots\ldots & & \boldsymbol{O}_{s,s} & \boldsymbol{I}_s \\ -\boldsymbol{B}_0 & \ldots\ldots\ldots & & & -\boldsymbol{B}_i \end{bmatrix}. \tag{3.10}$$

After introducing the matrix $\boldsymbol{E}$ we are able to define the stability of the overimplicit method (2.3).

---

[1]) If some index of the matrix $\boldsymbol{O}_{m,n}$ is zero then $\boldsymbol{O}_{m,n}$ does not occur in (3.4) at all.

[2]) The symbol $[a]$ denotes the integral part of the number a.

Definition 3.4. *The overimplicit method* (2.3) *is said to be stable if there exists a constant* Γ *such that for* $n = 0, 1, \ldots$

$$\|\boldsymbol{E}^n\| \leqq \Gamma \tag{3.11}$$

where $\boldsymbol{E}$ *is defined for* $l \leqq s$ *by* (3.5) *and for* $l > s$ *by* (3.10).

Remark 3.1. The condition (3.11) can be expressed alternatively in such a way that the spectral radius of $\boldsymbol{E}$ is less than or equal to 1 and that only linear elementary divisors correspond to eigenvalues of magnitude 1.

Let us note that only the matrix $\boldsymbol{B}$ and the parameter s are concerned in Definition 3.4.

Now we can formulate the basic theorems concerning the convergence of overimplicit methods.

Theorem 3.1. *A stable and consistent overimplicit method is convergent.*[4])

Before proving this theorem, we remind a useful lemma omitting its easy proof (cf., for example, Babuška, Práger, Vitásek [1966]).

Lemma 3.2. *Let* $\varphi(\nu), \psi(\nu), \chi(\nu)$ *be defined for* $\nu = 0, \ldots, n$ *and let* $\chi(\nu) \geqq 0$ *for* $\nu = 0, \ldots, n$. *Further, let*

$$\varphi(\nu) \leqq \psi(\nu) + \sum_{\mu=0}^{\nu-1} \chi(\mu)\, \varphi(\mu) \quad \text{for} \quad \nu = 0, \ldots, n\,. \tag{3.12}$$

Then

$$\varphi(\nu) \leqq \psi(\nu) + \sum_{\mu=0}^{\nu-1} \chi(\mu)\, \psi(\mu) \prod_{s=\mu+1}^{\nu-1} (1 + \chi(s)) \quad \text{for} \quad \nu = 0, \ldots, n\,. \tag{3.13}$$

Proof of Theorem 3.1. Let $e_j = y_j - y(x_j)$. Then we have for $n = rs + l - 1$, $r = 0, 1, \ldots$ according to (3.2)

$$\begin{bmatrix} e_{n+1} \\ \vdots \\ e_{n+k} \end{bmatrix} + \boldsymbol{B} \begin{bmatrix} e_{n-l+1} \\ \vdots \\ e_n \end{bmatrix} = h\boldsymbol{C} \begin{bmatrix} f(x_{n+1}, y_{n+1}) - f(x_{n+1}, y(x_{n+1})) \\ \vdots \\ f(x_{n+k}, y_{n+k}) - f(x_{n+k}, y(x_{n+k})) \end{bmatrix} + \tag{3.14}$$

$$+ h\boldsymbol{D} \begin{bmatrix} f(x_{n-l+1}, y_{n-l+1}) - f(x_{n-l+1}, y(x_{n-l+1})) \\ \vdots \\ f(x_n, y_n) - f(x_n, y(x_n)) \end{bmatrix} - \boldsymbol{L}(y(x_n); h)$$

[3]) Here one can take an arbitrary norm of the matrix as a linear mapping; for the definiteness let us consider the spectral norm.

[4]) The convergence is understood in the usual sense, cf., for example, Henrici [1962].

or

$$(3.15)\quad \begin{bmatrix} e_{n+1} \\ \vdots \\ e_{n+k} \end{bmatrix} + \mathbf{B} \begin{bmatrix} e_{n-l+1} \\ \vdots \\ e_n \end{bmatrix} = h\mathbf{C}\boldsymbol{\Phi}^{(k)}_{n+1} \begin{bmatrix} e_{n+1} \\ \vdots \\ e_{n+k} \end{bmatrix} + h\mathbf{D}\boldsymbol{\Phi}^{(l)}_{n-l+1} \begin{bmatrix} e_{n-l+1} \\ \vdots \\ e_n \end{bmatrix} - \mathbf{L}(y(x_n); h)$$

where

$$(3.16)\quad \boldsymbol{\Phi}^{(k)}_r = \begin{bmatrix} g_r & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & g_{r+k-1} \end{bmatrix}$$

with

$$(3.17)\quad g_r = \begin{cases} \dfrac{f(x_r, y_r) - f(x_r, y(x_r))}{e_r} & \text{for } e_r \neq 0 \\ 0 & \text{for } e_r = 0 \text{ or for } r < 0 \end{cases}$$

so that

$$(3.18)\quad |g_r| \leqq L$$

where L is the Lipschitz constant of the right-hand term of the given differential equation.

Our first task now is to find a bound of the norm of the vector $\mathbf{L}$. The solution $y(x)$ of (2.1) and (2.2) has in $\langle a, b\rangle$ continuous derivative so that we can define the function

$$(3.19)\quad \omega(\varepsilon) = \sup_{\substack{|x-x^*| \leqq \varepsilon \\ x, x^* \in \langle a,b\rangle}} |y'(x) - y'(x^*)|$$

and it holds

$$(3.20)\quad \lim_{\varepsilon \to 0} \omega(\varepsilon) = 0\,.$$

Obviously, there exist constants $\theta^{(1)}_{n,j}$, $\theta^{(2)}_{n,j}$, $\theta^{(3)}_{n,j}$ and $\theta^{(4)}_{n,j}$ smaller than or equal to 1 in the absolute value such that

$$\begin{aligned} y(x_{n+j}) &= y(x_n) + jh\, y'(x_n) + jh\theta^{(1)}_{n,j}\, \omega(jh)\,, \\ y(x_{n-(l-j)}) &= y(x_n) - (l-j)\, h\, y'(x_n) - (l-j)\, h\theta^{(2)}_{n,j}\, \omega((l-j)\, h)\,, \\ y'(x_{n+j}) &= y'(x_n) + \theta^{(3)}_{n,j}\, \omega(jh)\,, \\ y'(x_{n-(l-j)}) &= y'(x_n) + \theta^{(4)}_{n,j}\, \omega((l-j)\, h) \end{aligned}$$

and, consequently, by (3.2) we have

$$L_i(y(x_n); h) = (1 + \sum_{j=1}^{l} b_{ij})\, y(x_n) +$$

$$+ h[i - \sum_{j=1}^{l} (l-j)\, b_{ij} - \sum_{j=1}^{k} c_{ij} - \sum_{j=1}^{l} d_{ij}]\, y'(x_n) +$$

$$+ ih\theta_{n,i}^{(1)}\, \omega(ih) - h \sum_{j=1}^{l} (l-j)\, \theta_{n,j}^{(2)}\, \omega((l-j)\, h)\, b_{ij} -$$

$$- h \sum_{j=1}^{k} \theta_{n,j}^{(3)}\, \omega(jh)\, c_{ij} - h \sum_{j=1}^{l} \theta_{n,j}^{(4)}\, \omega((l-j)\, h)\, d_{ij}\,.$$

Owing to the consistency, the coefficients at $y(x_n)$ and $y'(x_n)$ vanish. Thus there exists a constant M (depending only on the given method) such that

$$|L_i(y(x_n); h)| \leqq Mh\omega(h \max(k, l-1)) \tag{3.21}$$

for $i = 1, \ldots, k$ and $n = l - 1, \ldots$

Further, let for any integer r

$$\mathbf{e}_r = (e_{(r-1)s+1}, \ldots, e_{(r-1)s+l+k-1})^T \tag{3.22}$$

equating the components with negative indices, if any, to zeros. Analogously, let for $r = 1, 2, \ldots$

$$\mathbf{t}_r = (-L_1(y(x_{(r-1)s+l-1}); h), \ldots, -L_k(y(x_{(r-1)s+l-1}); h))^T. \tag{3.23}$$

In further considerations two cases must be already distinguished: $l \leqq s$ and $l > s$.

Let firstly $l \leqq s$. Let us define the vector $\mathbf{e}_r^{(2)}$ in this case by the equation

$$\mathbf{e}_r^{(2)} = \mathbf{R}\mathbf{e}_r \tag{3.24}$$

where $\mathbf{R}$ is the matrix defined by (3.4). Using this notation, it is possible to rewrite (3.15) in the form

$$\mathbf{e}_{r+1} + \mathbf{B}\mathbf{e}_r^{(2)} = h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)}\mathbf{e}_{r+1} + h\mathbf{D}\boldsymbol{\Phi}_{rs}^{(l)}\mathbf{e}_r^{(2)} + \mathbf{t}_{r+1}\,, \quad r = 0, 1, \ldots \tag{3.25}$$

Since the matrix $\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)}$ is obviously regular for

$$h \leqq h_0 < \frac{1}{L\|\mathbf{C}\|} \tag{3.26}$$

it follows immediately from (3.25) that

$$\mathbf{e}_{rk+1} = -(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\, \mathbf{B}\mathbf{e}_r^{(2)} + h(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\, \mathbf{D}\boldsymbol{\Phi}_{rs}^{(l)}\mathbf{e}_r^{(2)} + \\ + (\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\, \mathbf{t}_{r+1} \tag{3.27}$$

for $h \leqq h_0$ and $r = 0, 1, \ldots$ Hence the error in a given step depends, as it can be expected, only on the component $\mathbf{e}_r^{(2)}$ of the error in the preceding step. Therefore, we shall be interested only in the behaviour of this component. From (3.27) we obtain

$$\mathbf{e}_{rs+l}^{(2)} = -\mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{B}\mathbf{e}_r^{(2)} + \tag{3.28}$$
$$+ h\mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{D}\boldsymbol{\Phi}_{rs}^{(l)}\mathbf{e}_r^{(2)} + \mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{t}_{r+1}$$

or

$$\mathbf{e}_{r+1}^{(2)} + \mathbf{R}\mathbf{B}\mathbf{e}_r^{(2)} = \mathbf{R}(\mathbf{I} - (\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1})\,\mathbf{B}\mathbf{e}_r^{(2)} + \tag{3.29}$$
$$+ h\mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{D}\boldsymbol{\Phi}_{rs}^{(l)}\mathbf{e}_r^{(2)} + \mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{t}_{r+1}\,, \quad r = 0, 1, \ldots$$

With the help of (3.5), the last equation can be rewritten in the form

$$\mathbf{e}_{r+1}^{(2)} = \mathbf{E}\mathbf{e}_r^{(2)} + \mathbf{v}_r\,, \quad r = 0, 1, \ldots \tag{3.30}$$

where

$$\mathbf{v}_r = \mathbf{R}[(\mathbf{I} - (\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1})\,\mathbf{B} + h(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{D}\boldsymbol{\Phi}_{rs}^{(l)}]\,\mathbf{e}_r^{(2)} + \tag{3.31}$$
$$+ \mathbf{R}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}\,\mathbf{t}_{r+1}\,.$$

Now realizing that the inequalities

$$\|(\mathbf{I} + \mathbf{V})^{-1}\| \leqq \frac{1}{1 - \|\mathbf{V}\|}\,, \quad \|\mathbf{I} - (\mathbf{I} + \mathbf{V})^{-1}\| \leqq \frac{\|\mathbf{V}\|}{1 - \|\mathbf{V}\|}$$

are obviously true for any matrix $\mathbf{V}$ with $\|\mathbf{V}\| < 1$ and using (3.21) we conclude from (3.31) for $h \leqq h_0$ that

$$\|\mathbf{v}_r\| \leqq h\alpha\|\mathbf{e}_r^{(2)}\| + Mh\beta\omega(h \max(k, l - 1)) \tag{3.32}$$

where

$$\alpha = \frac{\|\mathbf{B}\|\,\|\mathbf{C}\| + \|\mathbf{D}\|}{1 - hL\|\mathbf{C}\|}\,L\,, \quad \beta = \frac{1}{1 - hL\|\mathbf{C}\|}\,. \tag{3.33}$$

From (3.30) it follows

$$\|\mathbf{e}_r^{(2)}\| = \mathbf{E}^r\mathbf{e}_0^{(2)} + \sum_{\nu=0}^{r-1}\mathbf{E}^{r-1-\nu}\mathbf{v}_\nu\,, \quad r = 0, 1, \ldots \tag{3.34}$$

Since the method under consideration is stable, we get from (3.34)

$$\|\mathbf{e}_r^{(2)}\| \leqq \Gamma\|\mathbf{e}_0^{(2)}\| + \Gamma\sum_{\nu=0}^{r-1}\|\mathbf{v}_\nu\|\,, \quad r = 0, 1, \ldots$$

or, using (3.32),

$$\|\mathbf{e}_r^{(2)}\| \leqq \Gamma\|\mathbf{e}_0^{(2)}\| + \Gamma h\alpha \sum_{\nu=0}^{r-1} \|\mathbf{e}_\nu^{(2)}\| + r\Gamma M h\beta\omega(h \max(k, l-1)) . \tag{3.35}$$

Further, using this inequality and Lemma 3.2, we get

$$\|\mathbf{e}_r^{(2)}\| \leqq \Gamma\|\mathbf{e}_0^{(2)}\| + r\Gamma M h\beta\omega(h \max(k, l-1)) + \tag{3.36}$$

$$+ \Gamma h\alpha \sum_{\nu=0}^{r-1} [\Gamma\|\mathbf{e}_0^{(2)}\| + \nu\Gamma M h\beta\omega(h \max(k, l-1))] (1 + \Gamma h\alpha)^{r-1-\nu}$$

or, after an elementary modification

$$\|\mathbf{e}_r^{(2)}\| \leqq \Gamma(1 + \Gamma h\alpha)^r \|\mathbf{e}_0^{(2)}\| + \Gamma M\beta\omega(h \max(k, l-1)) \frac{(1 + \Gamma h\alpha)^r - 1}{\Gamma\alpha} . \tag{3.37}$$

Finally, with respect to obvious inequalities

$$(1 + \Gamma h\alpha)^r \leqq e^{\Gamma\alpha rh}, \quad \frac{(1 + \Gamma h\alpha)^r - 1}{\Gamma\alpha} \leqq rhe^{\Gamma\alpha rh},$$

we have from (3.37)

$$\|\mathbf{e}_r^{(2)}\| \leqq \Gamma e^{\Gamma\alpha rh}\|\mathbf{e}_0^{(2)}\| + \Gamma M\beta rhe^{\Gamma\alpha rh}\omega(h \max(k, l-1)) . \tag{3.38}$$

Choosing now the initial conditions so that $\|\mathbf{e}_0^{(2)}\| \to 0$ for $h \to 0$ and using (3.20) and (3.27), we get from (3.38) the assertion of our theorem in the case $l \leqq s$.

Consider now the case $l > s$. Let us define in this case the vector $\mathbf{e}_r^{(1)}$ by

$$\mathbf{e}_r^{(1)} = \boldsymbol{S}\mathbf{e}_r \tag{3.39}$$

where the matrix $\boldsymbol{S}$ is defined by (3.6). Now (3.15) can be rewritten in the form

$$\mathbf{e}_{r+1} + \boldsymbol{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} = h\boldsymbol{C}\boldsymbol{\Phi}_{rs+l}^{(k)}\mathbf{e}_{r+1} + h(\boldsymbol{D}\boldsymbol{\Phi}_{rs}^{(l)})^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \mathbf{t}_{r+1} \tag{3.40}$$

for $r = 0, 1, \ldots$, where i is defined by (3.7), $\boldsymbol{B}^{(1)}$ is the matrix (3.8) and

$$(\boldsymbol{D}\boldsymbol{\Phi}_{rs}^{(l)})^{(1)} = [\boldsymbol{O}_{k,(i+1)s-l}, \boldsymbol{D}\boldsymbol{\Phi}_{rs}^{(l)}] . \tag{3.41}$$

Since it is obviously

$$(\boldsymbol{D}\boldsymbol{\Phi}_{rs}^{(l)})^{(1)} = \boldsymbol{D}^{(1)}\boldsymbol{\Phi}_{(r-i-1)s+l}^{((i+1)s)}$$

where

$$\boldsymbol{D}^{(1)} = [\boldsymbol{O}_{k,(i+1)s-l}, \boldsymbol{D}] \tag{3.42}$$

it is possible to rewrite (3.40) in the form

$$\mathbf{e}_{r+1} + \mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(k)} \end{bmatrix} = h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)} \mathbf{e}_{r+1} + h\mathbf{D}^{(1)} \boldsymbol{\Phi}_{(r-i-1)s+l}^{((i+1)s)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \mathbf{t}_{r+1} \tag{3.43}$$

or, for h satisfying (3.26),

$$\mathbf{e}_{r+1} = -(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \tag{3.44}$$

$$+ h(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{D}^{(1)} \boldsymbol{\Phi}_{(r-i-1)s+l}^{((i+1)s)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + (\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{t}_{r+1},$$

$$r = 0, 1, \ldots$$

The error depends therefore only on the component $\mathbf{e}^{(1)}$ of the vector $\mathbf{e}$ and so we shall study only the behaviour of this component as we did in the preceding case. From (3.44) we have

$$\mathbf{e}_{r+1}^{(1)} = -\mathbf{S}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \tag{3.45}$$

$$+ h\mathbf{S}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{D}^{(1)} \boldsymbol{\Phi}_{(r-i-1)s+l}^{((i+1)s)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \mathbf{S}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{t}_{r+1}$$

or

$$\mathbf{e}_{r+1}^{(1)} + \mathbf{S}\mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} = \mathbf{S}[\mathbf{I} - (\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1}] \mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \tag{3.46}$$

$$+ h\mathbf{S}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{D}^{(1)} \boldsymbol{\Phi}_{(r-i-1)s+l}^{((i+1)s)} \begin{bmatrix} \mathbf{e}_{r-i}^{(1)} \\ \vdots \\ \mathbf{e}_r^{(1)} \end{bmatrix} + \mathbf{S}(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}_{rs+l}^{(k)})^{-1} \mathbf{t}_{r+1}.$$

Using now (3.9), we can write

$$\mathbf{e}_{r+1}^{(1)} + \mathbf{B}_0 \mathbf{e}_{r-i}^{(1)} + \ldots + \mathbf{B}_i \mathbf{e}_r^{(1)} = \mathbf{v}_r, \quad r = 0, 1, \ldots \tag{3.47}$$

where

$$\mathbf{v}_r = \mathbf{S}\left[\mathbf{I} - \left(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}^{(k)}_{rs+l}\right)^{-1}\right] \mathbf{B}^{(1)} \begin{bmatrix} \mathbf{e}^{(1)}_{r-i} \\ \vdots \\ \mathbf{e}^{(1)}_{r} \end{bmatrix} + \tag{3.48}$$

$$+ h\mathbf{S}\left(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}^{(k)}_{rs+l}\right)^{-1} \mathbf{D}^{(1)} \boldsymbol{\Phi}^{((i+1)s)}_{(r-i-1)s+l} \begin{bmatrix} \mathbf{e}^{(1)}_{r-i} \\ \vdots \\ \mathbf{e}^{(1)}_{r} \end{bmatrix} + \mathbf{S}\left(\mathbf{I} - h\mathbf{C}\boldsymbol{\Phi}^{(k)}_{rs+l}\right)^{-1} \mathbf{t}_{r+1} .$$

If we put

$$\boldsymbol{\eta}_v = \begin{bmatrix} \mathbf{e}^{(1)}_{v} \\ \vdots \\ \mathbf{e}^{(1)}_{v+i} \end{bmatrix}, \quad \mathbf{w}_v = \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \mathbf{v}_{v+i} \end{bmatrix} \tag{3.49}$$

for $v = -i, -i+1, \ldots$, we can rewrite (3.48) in the form

$$\boldsymbol{\eta}_{v+1} = \mathbf{E}\boldsymbol{\eta}_v + \mathbf{w}_v, \quad v = -i, -i+1, \ldots \tag{3.50}$$

where $\mathbf{E}$ is defined by (3.10), or

$$\boldsymbol{\eta}_v = \mathbf{E}^{v+i}\boldsymbol{\eta}_{-i} + \sum_{\mu=-i}^{v-1} \mathbf{E}^{v-1-\mu}\mathbf{w}_\mu, \quad v = -i, -i+1, \ldots \tag{3.51}$$

Using the conditions of stability, we have from (3.51)

$$\|\boldsymbol{\eta}_v\| \leqq \Gamma\|\boldsymbol{\eta}_{-i}\| + \Gamma \sum_{\mu=-i}^{v-1} \|\mathbf{w}_\mu\| . \tag{3.52}$$

It is obviously

$$\|\mathbf{w}_\mu\| \leqq h\alpha\|\boldsymbol{\eta}_\mu\| + Mh\beta\omega(h \max(k, l-1)) \tag{3.53}$$

where α and β are defined by (3.33). Hence we get from (3.52) after the same arrangement as in the case $l \leqq s$ the inequality

$$\|\boldsymbol{\eta}_v\| \leqq \Gamma e^{\Gamma\beta(v+i)h}\|\boldsymbol{\eta}_{-i}\| + \Gamma M\beta(v+i)\, h e^{\Gamma\alpha(v+i)h}\omega(h \max(k, l-1)) . \tag{3.54}$$

Using now the obvious inequality $\|\mathbf{e}^{(1)}_r\| \leqq \|\boldsymbol{\eta}_r\|$ we get immediately the assertion of our theorem even in the case $l > s$. The proof of Theorem 3.1 is complete.

Remark 3.2. Using the standard procedure consisting in the investigation of special differential equations it can be proved that the conditions of Theorem 3.1 are also necessary.

Theorem 3.2. *Let a stable overimplicit method of order $p \geqq 1$ be given. Let the solution of the problem* (2.1), (2.2) *have continuous derivatives up to order $p + 1$. Finally, let the initial conditions by which the approximate solution is determined be given with the accuracy of order $O(h^p)$. Then the discretization error is also of order $O(h^p)$.*

Proof. The proof of this theorem is a trivial modification of that of Theorem 3.1 with the use of Lemma 3.1.

4. A-STABILITY OF OVERIMPLICIT METHODS

The main goal of this section is to prove that there exist A-stable methods of arbitrarily high order in the class of overimplicit multistep methods. First of all, let us remind the definition of Dahlquist's A-stability (Dahlquist [1936]).

Definition 4.1. *Let α be a complex constant with negative real part. A numerical method for solving initial-value problems for ordinary differential equations is said to be A-stable if any solution of the difference equation which arises by aplying the given method to the differential equation $y' = \alpha y$ converges to zero for $n \to \infty$.*

In order to facilitate our task, we will seek A-stable methods of arbitrarily high orders in the subset of the class of overimplicit methods for which $l = 1$ and $\boldsymbol{B} = (-1, \ldots, -1)^T$ (which are consequently stable) and which are of order at least k. Thus, in what follows we shall deal only with the formulae of the form

$$\begin{bmatrix} y_{n+1} \\ \vdots \\ y_{n+k} \end{bmatrix} = \begin{bmatrix} y_n \\ \vdots \\ y_n \end{bmatrix} + h\boldsymbol{C} \begin{bmatrix} f_{n+1} \\ \vdots \\ f_{n+k} \end{bmatrix} + hf_n\boldsymbol{d} \tag{4.1}$$

where $\boldsymbol{C}$ is a square matrix of order k and $\boldsymbol{d}$ is a k-dimensional vector. Moreover, $\boldsymbol{C}$ and $\boldsymbol{d}$ are such that

$$\begin{bmatrix} y(x+h) \\ \vdots \\ y(x+kh) \end{bmatrix} = \begin{bmatrix} y(x) \\ \vdots \\ y(x) \end{bmatrix} + h\boldsymbol{C} \begin{bmatrix} y'(x+h) \\ \vdots \\ y'(x+kh) \end{bmatrix} + h\, y'(x)\, \boldsymbol{d} + \boldsymbol{O}(h^{p+1}) \tag{4.2}$$

holds for any sufficiently smooth function $y(x)$ and for some $p \geqq k$. Since these formulae do not need any starting procedures and since they are of order at least k we shall call them selfstarting overimplicit almost optimal methods.

We show first of all that the above class is not empty. If we recall the basic idea of the multistep methods of Adams type we are led naturally to the subset of the class of selfstarting methods of the form

$$y_{n+i} - y_n = \int_{x_n}^{x_{n+i}} P(x)\,\mathrm{d}x\,, \quad i = 1, \ldots, k \tag{4.3}$$

where $P(x)$ is the interpolating polynomial of degree k which has the values f_{n+i} at the points x_{n+i}, $i = 0, \ldots, k$. These methods will be called the selfstarting methods of Adams type. They can be obviously rewritten in the form

$$y_{n+i} - y_n = h \sum_{j=0}^{k} \gamma_{ij} f(x_{n+j}, y_{n+j})\,, \quad i = 1, \ldots, k \tag{4.4}$$

where

$$\gamma_{ij} = \int_0^i l_j(t)\,\mathrm{d}t \tag{4.5}$$

and $l_j(t)$ is the elementary Lagrange interpolating polynomial for the points $t = 0, \ldots, k$, i.e., the polynomial of degree k which has the value 0 at the points $t = 0, \ldots, k$, $t \neq j$ and the value 1 at the point $t = j$. It is obvious that a selfstarting formula of Adams type is even of order $k + 1$.

Selfstarting overimplicit almost optimal methods are defined as methods of the form (4.1) whose order is at least k. Let us examine the conditions on $\boldsymbol{C}$ and $\boldsymbol{d}$ which guarantee that the corresponding method will be of order k. If we substitute $l = 1$ and $b_{i1} = -1$, $d_{i1} = d_i$ ($\boldsymbol{d} = (d_1, \ldots, d_k)^T$) into (3.1) we get

$$\begin{aligned} &\sum_{j=1}^{k} c_{ij} + d_i = i\,, \quad i = 1, \ldots, k\,, \\ &\nu \sum_{j=1}^{k} j^{\nu-1} c_{ij} = i^{\nu}\,, \quad i = 1, \ldots, k\,, \quad \nu = 2, \ldots, k\,. \end{aligned} \tag{4.6}$$

If we introduce the notation

$$\mathbf{e} = \begin{bmatrix} 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix}, \qquad \boldsymbol{M} = \begin{bmatrix} 1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & k \end{bmatrix} \tag{4.7}$$

we can rewrite (4.6) in the form

$$\begin{aligned} \boldsymbol{C}\mathbf{e} + \boldsymbol{d} &= \boldsymbol{M}\mathbf{e} \\ 2\boldsymbol{C}\boldsymbol{M}\mathbf{e} &= \boldsymbol{M}^2\mathbf{e} \\ &\vdots \\ k\boldsymbol{C}\boldsymbol{M}^{k-1}\mathbf{e} &= \boldsymbol{M}^k\mathbf{e}\,. \end{aligned} \tag{4.8}$$

Thus, the equations (4.8) are necessary and sufficient conditions for the formula (4.1) to express a selfstarting overimplicit almost optimal method.

Now, let us examine the A-stability of the formula (4.1). If we use this formula for solving the differential equation $y' = \alpha y$, where α is a (complex) constant, we get

$$(\boldsymbol{I} - z\boldsymbol{C}) \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{n+k} \end{bmatrix} = y_n(\boldsymbol{e} + z\boldsymbol{d}) \tag{4.9}$$

where

$$z = \alpha h \,. \tag{4.10}$$

Since only the value y_{n+s} where s is an integer, $1 \leqq s \leqq k$, is used as the initial value in the following step of the method we are actually interested only in the values y_{rs}, $r = 0, 1, \ldots$ By Cramer's rule we obtain immediately

$$y_{(r+1)s} = \frac{P_s(z)}{Q(z)}\, y_{rs}\,, \quad r = 0, 1, \ldots \tag{4.11}$$

where

$$Q(z) = \det(\boldsymbol{I} - z\boldsymbol{C}) \tag{4.12}$$

and $P_s(z)$ is the determinant of the matrix which arises from the matrix $\boldsymbol{I} - z\boldsymbol{C}$ by replacing its s-th column by the vector $\boldsymbol{e} + z\boldsymbol{d}$. Hence we can formulate

Theorem 4.1. *The necessary and sufficient condition for the A-stability of the formula* (4.1) *is that*

$$\left|\frac{P_s(z)}{Q(z)}\right| < 1 \tag{4.13}$$

for any z *with* $\operatorname{Re} z < 0$.

Proof. The statement is obvious and it follows immediately from Definition 4.1 and from (4.11).

We can see that the A-stability of the given selfstarting overimplicit almost optimal method depends on the behaviour of the polynomials $P_s(z)$, $Q(z)$ and, consequently, in order to attain our main aim it will be substantial to be able to construct a selfstarting overimplicit almost optimal method with $\det(\boldsymbol{I} - z\boldsymbol{C})$ equal to any given polynomial. Thus, the following theorem will be of basic importance in the further investigation.

Theorem 4.2. *To every polynomial $Q(z)$ with the coefficient 1 at z^0 there exists one and only one selfstarting overimplicit almost optimal method, for which* (4.12) *holds.*[5])

Proof. First of all let us study the structure of the class of selfstarting overimplicit almost optimal methods in more detail. Let us note that if such a method is given and if a vector $\mathbf{t}$ is defined by

$$\mathbf{t} = \frac{1}{k+1} \boldsymbol{M}^{k+1}\mathbf{e} - \boldsymbol{C}\boldsymbol{M}^{k}\mathbf{e}\,, \tag{4.14}$$

then the corresponding matrix $\boldsymbol{C}$ and the vector $\boldsymbol{d}$ satisfy

$$\begin{aligned} \boldsymbol{M}^{\nu}\mathbf{e} &= \nu \boldsymbol{C}\boldsymbol{M}^{\nu-1}\mathbf{e}\,, \quad \nu = 2, \ldots, k\,, \\ \boldsymbol{M}^{k+1}\mathbf{e} &= (k+1)\,\boldsymbol{C}\boldsymbol{M}^{k}\mathbf{e} + (k+1)\,\mathbf{t} \end{aligned} \tag{4.15}$$

and

$$\boldsymbol{d} = \boldsymbol{M}\mathbf{e} - \boldsymbol{C}\mathbf{e}\,. \tag{4.16}$$

Conversely, let an arbitrary vector $\mathbf{t}$ be given. Let us define the matrix $\boldsymbol{C}$ by (4.15). This is possible because the system (4.15) can be rewritten in the form

$$\boldsymbol{M}^2\boldsymbol{V} = \boldsymbol{C}\boldsymbol{M}\boldsymbol{V}(\boldsymbol{I} + \boldsymbol{M}) + [\boldsymbol{0}, \ldots, \boldsymbol{0}, (k+1)\,\mathbf{t}] \tag{4.17}$$

where $\boldsymbol{V}$ is the Vandermonde matrix for numbers $1, 2, \ldots, k$, i.e.,

$$\boldsymbol{V} = \begin{bmatrix} 1^0 & 1^1 & \ldots & 1^{k-1} \\ 2^0 & 2^1 & \ldots & 2^{k-1} \\ \vdots & \vdots & & \vdots \\ k^0 & k^1 & \ldots & k^{k-1} \end{bmatrix} \tag{4.18}$$

and the matrix $\boldsymbol{M}\boldsymbol{V}(\boldsymbol{I} + \boldsymbol{M})$ is therefore regular. Further, let us define the vector $\boldsymbol{d}$ by (4.16). Then the matrix $\boldsymbol{C}$ and the vector $\boldsymbol{d}$ define obviously a selfstarting overimplicit almost optimal method. Thus, we have found that there exists a one-to-one correspondence between the class of selfstarting overimplicit almost optimal methods and the k-dimensional vector space.

---

[5]) The words "one and only one" refer here to the matrix $\boldsymbol{C}$ and the vector $\boldsymbol{d}$ from (4.1). Because an overimplicit method is given not only by $\boldsymbol{C}$ and $\boldsymbol{d}$ but also by the parameter s, there exist in fact exactly k methods having the property of Theorem 4.2 and differing only by the value of the parameter s.

Further, let be given a selfstarting overimplicit almost optimal method again and let

$$\sum_{i=0}^{k} a_i z^i = \det(\boldsymbol{I} - z\boldsymbol{C}). \tag{4.19}$$

Then the polynomial $\sum_{i=0}^{k} a_{k-i}\lambda^i$ is obviously the characteristic polynomial of the matrix $\boldsymbol{C}$ and by Cayley-Hamilton theorem one has

$$\sum_{i=0}^{k} a_{k-i}\boldsymbol{C}^i = \boldsymbol{O} \tag{4.20}$$

and, consequently, after multiplying by $\boldsymbol{Me}$,

$$\sum_{i=0}^{k} a_{k-i}\boldsymbol{C}^i\boldsymbol{Me} = \boldsymbol{0}. \tag{4.21}$$

From (4.8), since $\boldsymbol{C}$ satisfies (4.8), we have

$$\boldsymbol{C}^i\boldsymbol{Me} = \frac{1}{(i+1)!}\boldsymbol{M}^{i+1}\boldsymbol{e}, \quad i = 0, \ldots, k-1. \tag{4.22}$$

Using (4.22) we can rewrite (4.21) in the form

$$\sum_{i=0}^{k} a_{k-i}\frac{1}{(i+1)!}\boldsymbol{M}^{i+1}\boldsymbol{e} = \frac{1}{k!}\boldsymbol{t} \tag{4.23}$$

where $\boldsymbol{t}$ is defined by (4.14). It is easy to verify that (4.23) can be rewritten in the form

$$\boldsymbol{MVN}\begin{bmatrix} a_k \\ \vdots \\ a_1 \end{bmatrix} = \frac{1}{k!}\boldsymbol{t} - \frac{1}{(k+1)!}\boldsymbol{M}^{k+1}\boldsymbol{e} \tag{4.24}$$

where $\boldsymbol{V}$ is defined by (4.18) and $\boldsymbol{N}$ is given by

$$\boldsymbol{N} = \begin{bmatrix} \frac{1}{1!} & 0 & \ldots & 0 \\ & \ddots & & \vdots \\ 0 & \ddots & \ddots & \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{k!} \end{bmatrix}. \tag{4.25}$$

Since the matrix $\boldsymbol{MVN}$ is obviously regular we can verify conversely that $a'_k, \ldots, a'_1$ solving (4.24) (with a given $\mathbf{t}$) and $a'_0 = 1$ satisfy

$$\sum_{i=0}^{k} a'_i z^i = \det(\boldsymbol{I} - z\boldsymbol{C}) \tag{4.26}$$

where $\boldsymbol{C}$ is the matrix of the selfstarting overimplicit almost optimal method given by $\mathbf{t}$.

Now the proof of Theorem 4.2 is already easy. To a given polynomial $Q(z) = \sum_{i=0}^{k} a_i z^i$ with $a_0 = 1$, it is sufficient to compute the vector $\mathbf{t}$ by (4.23) and then with this $\mathbf{t}$ to compute $\boldsymbol{C}$ from (4.17) and $\boldsymbol{d}$ from (4.16). The above argument guarantees that $\boldsymbol{C}$ and $\boldsymbol{d}$ obtained in this way will give the required selfstarting overimplicit almost optimal method.

Remark 4.1. Let us note that for the selfstarting method of Adams type the vector $\mathbf{t}$ defined by (4.14) is the null vector because the method is of order $k + 1$.

If the selfstarting overimplicit almost optimal method is given by a polynomial $Q(z)$ in the sense of Theorem 4.2 it will be useful in what follows to express the coefficients of the polynomials $P_s(z)$ from (4.11) by the coefficients of $Q(z)$. This is the contents of the following theorem.

Theorem 4.3. *Let a polynomial* $Q(z) = \sum_{i=0}^{k} a_i z^i$ *with* $a_0 = 1$ *be given. Then for the polynomials* $P_s(z)$ *defined by* (4.11) *it holds*

$$P_s(z) = \sum_{j=0}^{k} \left(\sum_{i=0}^{j} \frac{s^{j-i}}{(j-i)!} a_i \right) z^j, \quad s = 1, \ldots, k. \tag{4.26}$$

Proof. By Cramer's rule, the vector $(y_{n+1}, \ldots, y_{n+k})^T$ defined by (4.9) satisfies

$$Q(z) \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{n+k} \end{bmatrix} = y_n \begin{bmatrix} P_1(z) \\ \vdots \\ P_k(z) \end{bmatrix}. \tag{4.27}$$

However, y_{n+i} is the approximate solution of the differential equation $y' = \alpha y$ by the overimplicit method of order k and, consequently, for any solution $y(x)$ of this differential equation

$$Q(z) \begin{bmatrix} y(x+h) \\ \vdots \\ y(x+kh) \end{bmatrix} - y(x) \begin{bmatrix} P_1(z) \\ \vdots \\ P_k(z) \end{bmatrix} = \boldsymbol{O}(h^{k+1}). \tag{4.28}$$

In particular, for $\alpha = 1$

$$Q(h)\, e^{sh} - P_s(h) = O(h^{k+1}), \quad s = 1, \ldots, k. \tag{4.29}$$

Comparing the coefficients at the same powers of h on the left- and the right-hand sides of (4.29) we obtain (4.26) immediately.

Since by means of Theorem 4.2 we are able to construct a self-starting overimplicit almost optimal method such that the polynomial $Q(z)$ given by (4.12) is an arbitrary polynomial given in advance, the further course of our study will be to choose this polynomial to have some convenient properties. The choice of this polynomial will be closely connected with the Padé approximation of the exponential function.

Thus, let

$$R(z) = \sum_{i=0}^{k} r_i z^i \tag{4.30}$$

where

$$r_i = \frac{(2k - i)!\, k!}{i!\,(k - i)!\,(2k)!}. \tag{4.31}$$

Then $R(z)/R(-z)$ is the Padé approximation of e^z, i.e., it holds

$$e^z = \frac{R(z)}{R(-z)} + O(z^{2k+1}) \quad \text{for} \quad z \to 0 \tag{4.32}$$

cf., for example, Varga [1962]). From (4.32) it can be easily obtained

$$\sum_{i=0}^{j} \frac{(-1)^i r_i}{(j - i)!} = r_j, \qquad j = 0, \ldots, k, \tag{4.33}$$

$$\sum_{i=0}^{k} \frac{(-1)^i r_i}{(j - i)!} = 0, \quad j = k + 1, \ldots, 2k.$$

The following lemmas will be useful in the proof of the basic statement of this section.

Lemma 4.1. *All the zeros of the polynomial $R(z)$ given by* (4.30) *and* (4.31) *have negative real parts* (*cf. Birkhoff, Varga* [1965]).

Proof. To prove this lemma the so-called Routh criterion will be used. We shall formulate it for our purpose in this way (cf. Gantmacher [1966]):

Routh criterion. *Let a polynomial* $S(z) = \sum_{i=0}^{k} s_i z^i$ *with real coefficients be given. Let us define the polynomials* $S_1(z)$ *and* $S_2(z)$ *by*

$$S_1(z) = \sum_{j=0}^{[k/2]} (-1)^j s_{k-2j} z^{k-2j}, \tag{4.34}$$

$$S_2(z) = \sum_{j=0}^{[(k-1)/2]} (-1)^j s_{k-2j-1} z^{k-2j-1}.$$

Further, let us suppose that there exist nonzero polynomials $S_3(z), \ldots, S_{k+1}(z)$ *with decreasing degrees and numbers* α_i, $i = 1, \ldots, k$ *such that*

$$S_i(z) = \alpha_i z\, S_{i+1}(z) - S_{i+2}(z), \quad i = 1, \ldots, k-1, \tag{4.35}$$

$$S_k(z) = \alpha_k z\, S_{k+1}(z).$$

Finally, let all α_i, $i = 1, \ldots, k$ *have the same signs. Then all the zeros of the polynomial* $S(z)$ *have negative real parts.*

In the proof of our lemma we shall investigate the polynomial

$$S(z) = \frac{(2k)!}{k!} z^k R\left(\frac{1}{z}\right) \tag{4.36}$$

instead of $R(z)$ so that $s_j = (k+j)!/(j!\,(k-j)!)$. We shall show that α_i's and the polynomials $S_i(z)$ from the Routh criterion are given by

$$\alpha_i = 2(2i-1), \quad i = 1, \ldots, k \tag{4.37}$$

and

(4.38)

$$S_i(z) = \sum_{j=0}^{[(k-i+1)/2]} (-1)^j \frac{(2k-2j)!}{(k-j)!} \frac{(j+i-1)!}{(2j+2i-2)!} \frac{(k-j-i+1)!}{j!\,(k-2j-i+1)!} z^{k-2j-i+1},$$

respectively. Indeed, $S_k(z) = \alpha_k z\, S_{k+1}(z)$ and for $i < k$

(4.39)

$$\alpha_i z\, S_{i+1}(z) - S_{i+2}(z) =$$

$$= \sum_{j=0}^{[(k-i)/2]} (-1)^j \frac{(2k-2j)!}{(k-j)!} \frac{(j+i)!}{(2j+2i)!} \frac{(k-j-i)!}{j!\,(k-2j-i)!} 2(2i-1)\, z^{k-2j-i+1} -$$

$$- \sum_{j=0}^{[(k-i-1)/2]} (-1)^j \frac{(2k-2j)!}{(k-j)!} \frac{(j+i+1)!}{(2j+2i+2)!} \frac{(k-j-i-1)!}{j!\,(k-2j-i-1)!} z^{k-2j-i-1}.$$

Substituting $j - 1$ for j into the second sum in (4.39) we get

$$\alpha_i z\, S_{i+1}(z) - S_{i+2}(z) = \frac{(2k)!}{k!}\,\frac{i!}{(2i)!}\,2(2i-1)\,z^{k-i+1} + \tag{4.40}$$

$$+ \sum_{j=1}^{[(k-i)/2]} (-1)^j \frac{(2k-2j)!}{(k-j)!}\,\frac{(j+i)!}{(2j+2i)!}\,\frac{(k-j-i)!}{(j-1)!\,(k-2j-i)!}\cdot$$

$$\cdot\left(\frac{2(2i-1)}{j} + \frac{2(2k-2j+1)}{k-2j-i+1}\right) z^{k-2j-i+1} + \sigma$$

where σ is equal to zero for $k - i$ even and equal to $(-1)^{(k-i+1)/2}$ for $k - i$ odd. But now it is already easy to see that the right-hand term of (4.40) is exactly $S_i(z)$.

Lemma 4.2. *Let $U(z)$ be a polynomial with real coefficients and of degree at least one. Further, let all the zeros of $U(z)$ have positive real parts. Then*

$$\left|\frac{U(-z)}{U(z)}\right| < 1 \tag{4.41}$$

for any complex z such that Re $z < 0$.

Proof. The statement is obvious when one realizes that $|(z + z_1)/(z - \bar{z}_1)| < 1$ for Re $z_1 > 0$ and Re $z < 0$ and that $U(z)$ has real coefficients.

Now we have all ready for the proof of the basic theorem of this section.

Theorem 4.4. *In the class of overimplicit methods there exist A-stable methods of arbitrarily high orders.*

Proof. Let be given an integer k, $k \geqq 1$, and an integer s, $1 \leqq s \leqq k$. Let

$$Q(z) = R(-sz) = \sum_{i=0}^{k} (-1)^i r_i s^i z^i \tag{4.42}$$

where $R(z)$ is defined by (4.30). By Theorem 4.2 construct for this polynomial $Q(z)$ the corresponding selfstarting overimplicit almost optimal method, which is consequently of order k. We shall prove that with s given above this method is A-stable. To show this let us compute the corresponding polynomial $P_s(z)$ (cf. (4.11)). By Theorem 4.3 and by (4.33), the coefficients p_j of $P_s(z)$ satisfy

$$p_j = \sum_{i=0}^{j} \frac{s^{j-i}}{(j-i)!}\,(-1)^i s^i r_i = s^j \sum_{i=0}^{j} \frac{(-1)^i r_i}{(j-i)!} = s^j r_j \tag{4.43}$$

for $j = 0, \ldots, k$ and, consequently,

$$P_s(z) = R(sz) = Q(-z)\,. \tag{4.44}$$

However, (4.44), Lemmas 4.1 and 4.2 and Theorem 4.1 imply our statement immediately. Theorem 4.4 is proved.

It can be seen from the proof of Theorem 4.4 that the selfstarting overimplicit almost optimal method constructed gives in the case of linear differential equation with constant coefficients the Padé approximation of the corresponding exponential function with the numerator and denominator of degree k and thus a much more accurate approximation of this function than it was guaranteed apriori by the corresponding method. On the other hand, the selfstarting methods of Adams type are of order $k + 1$ and therefore they are optimal in the class of selfstarting overimplicit methods from the point of view of asymptotic accuracy. From this reason it is natural to be interested in their A-stability, too. We shall study this problem only in the case $s = k$.

Theorem 4.5. *In order that a selfstarting method of Adams type with $s = k$ may be A-stable, it is necessary and sufficient for all roots of the polynomial $Q(z) = \det(\boldsymbol{I} - z\boldsymbol{C})$ where $\boldsymbol{C} = \{\gamma_{ij}\}$ and γ_{ij} are defined by (4.5) to have positive real parts.*

Proof. Let $Q(z) = \sum_{i=0}^{k} q_i z^i$ $(q_0 = 1)$ and let $P_k(z) = \sum_{i=0}^{k} p_i z^i$ be the polynomials from (4.11). Then with respect to Theorem 4.1 and Lemma 4.2 it is obviously sufficient to prove only that $P_s(z) = Q(-z)$, or

$$p_j = (-1)^j q_j, \quad j = 0, \ldots, k \tag{4.45}$$

where

$$p_j = \sum_{i=0}^{j} \frac{k^{j-i}}{(j-i)!} q_i \tag{4.46}$$

as it follows from Theorem 4.3. Hence $p_0 = 1$ and (4.45) is true for $j = 0$. From Theorem 4.2 and Remark 4.1 we can see that $\boldsymbol{C}$ is uniquely determined by the choice $\boldsymbol{t} = \boldsymbol{0}$. The coefficients $q_1, \ldots, q_k$ are, consequently, determined uniquely by the system

$$\sum_{j=0}^{k-1} q_{k-j} \frac{v^{j+1}}{(j+1)!} = -\frac{v^{k+1}}{(k+1)!}, \quad v = 1, \ldots, k \tag{4.47}$$

(cf. the proof of Theorem 4.2). Therefore, if we prove that

$$\sum_{j=0}^{k-1} (-1)^{k-j} p_{k-j} \frac{v^{j+1}}{(j+1)!} = -\frac{v^{k+1}}{(k+1)!} \tag{4.48}$$

for $v = 1, \ldots, k$ where p_j are defined by (4.46), the validity of (4.45) will be proved even for $j = 1, \ldots, k$ since the matrix of the system (4.47) is regular. In order to prove

(4.48) let us investigate the sum

$$(4.49)\quad \sum_{j=0}^{k}(-1)^{k-j}\,p_{k-j}\,\frac{v^{j+1}}{(j+1)!}=\sum_{j=0}^{k}(-1)^{j}\,\frac{v^{k+1-j}}{(k+1-j)!}\sum_{i=0}^{j}\frac{k^{j-i}}{(j-i)!}\,q_i=$$

$$=\sum_{i=0}^{k}q_i\sum_{j=i}^{k}(-1)^{j}\,\frac{k^{j-i}v^{k+1-j}}{(j-i)!\,(k+1-j)!}=$$

$$=\sum_{i=0}^{k}q_{k-i}\sum_{j=k-i}^{k}(-1)^{j}\,\frac{k^{j+i-k}v^{k+1-j}}{(j+i-k)!\,(k+1-j)!}=$$

$$=\sum_{i=0}^{k}q_{k-i}\sum_{j=0}^{i}(-1)^{j+k-i}\,\frac{k^{j}v^{i+1-j}}{j!(i+1-j)!}=$$

$$=(-1)^{k-1}\sum_{i=0}^{k}\frac{q_{k-i}}{(i+1)!}\sum_{j=0}^{i}(-1)^{i+1-j}\binom{i+1}{j}k^{j}v^{i+1-j}=$$

$$=(-1)^{k-1}\sum_{i=0}^{k}\frac{q_{k-i}}{(i+1)!}\left[(k-v)^{i+1}-k^{i+1}\right].$$

It follows from (4.47) that the last term is equal to zero for $v = 1, \ldots, k$. From this statement and with regard to $p_0 = 1$ we obtain (4.48), Theorem 4.5 is proved.

Theorem 4.5 enables us to determine whether a particular method of Adams type is A-stable or not. It has been shown by direct computation that the Adams methods up to $k = 8$ are A-stable, for $k = 9, 10$ they are not A-stable, but we have not yet the general result for general k.

Finally, we should like to mention that the problems presented above are far from exhausting all aspects of the overimplicit methods. Many problems remain open, for example, the properties of the overimplicit methods whose local errors in different lines are different, the above mentioned general discussion of A-stability of Adams-type methods etc. It is also clear that the overimplicit methods will prove very useful in the construction of numerical methods for solving partial differential equations of parabolic type which are of arbitrarily high order of accuracy with respect to time mesh-size. As a matter of fact, when solving a parabolic equation by transforming it to a system of ordinary differential equations (discretizating only the space variables) the resulting system of ordinary differential equations is the stiffer the finer is the space mesh (cf. Taufer [1972]).

References

[1] *I. Babuška M. Práger* and *E. Vitásek:* Numerical processes in differential equations, Interscience publishers, London, New York, Sydney (1966).

[2] *G. Birkhoff* and *R. S. Varga:* Discretization errors for well-set Cauchy problems I, J. Math. and Phys. 44 (1965), 1—23.

[3] *G. Dahlquist:* A special stability problem for linear multistep methods, BIT 3 (1963), 27—43.

[4] *F. R. Gantmacher* (Ф. Р. Гантмахер): Теория матриц, Наука, Москва (1966).
[5] *P. Henrici:* Discrete variable methods in ordinary differential equations, J. Wiley & Sons, Inc., New York, London (1962).
[6] *J. Taufer* (Й. Тауфер): Об одном обобщенном многошаговом методе, сб. Применение функциональных методов к краевым задачам математической физики, Новосибирск (1972).
[7] *R. S. Varga:* Matrix iterative analysis, Prentice-Hall, Inc., Englewood Cliffs, New Jersey (1962).
[8] *E. Vitásek* (E. Витасек): Строго неявные методы для решения дифференциальных уравнений, сб. Применение функциональных методов к краевым задачам математической физики, Новосибирск (1972).

Souhrn

SILNĚ IMPLICITNÍ MNOHOKROKOVÉ METODY

MILAN PRÁGER, JIŘÍ TAUFER, EMIL VITÁSEK

Článek se zabývá numerickým řešením obyčejných diferenciálních rovnic pomocí nové třídy metod nazvaných silně implicitní mnohokrokové metody. Základní myšlenka těchto metod spočívá v tom, že ze známých hodnot řešení v l bodech se podle předpisu (2.3) vypočítává k hodnot nových najednou. Asi polovina práce je věnována podrobnému studiu konvergence těchto silně implicitních metod.

Zbývající část práce je věnována studiu A-stability zavedených metod (viz definice 4.1). V této části je ukázáno, že ve třídě silně implicitních mnohokrokových metod existují A-stabilní metody libovolně vysokého řádu. Tím je vlastně ukázána vhodnost použití této nové třídy pro řešení diferenciálních rovnic se silným tlumením a možnost aplikací na rovnice parabolického typu.

Authors' addresses: Dr. *Milan Práger* CSc., Dr. *Jiří Taufer* CSc., Dr. *Emil Vitásek* CSc., Matematický ústav ČSAV v Praze, Žitná 25, 115 67 Praha 1.