SIAM J. NUMER. ANAL.
Vol. 21, No. 4, August 1984

002

ITERATED SPLITTING METHODS OF HIGH ORDER FOR TIME-DEPENDENT PARTIAL DIFFERENTIAL EQUATIONS*

P. J. VAN DER HOUWEN†

Abstract. Nonlinear Chebyshev iteration is applied for solving the implicit relations which arise when an implicit linear multistep method is used in order to integrate semi-discrete initial value problems for partial differential equations. The approximate inverse occurring in the defect correction process is obtained by employing splitting methods. In order to accelerate convergence the Chebyshev iteration process is tuned in such a way that the lower frequencies in the iteration error are strongly damped without using a large number of iterations. For moderate accuracies this method is already markedly more efficient then conventional ADI methods.

Key words. numerical analysis, method of lines, initial-boundary value problems, defect correction, Chebyshev iteration, splitting methods

1. Introduction. Consider an initial-boundary value problem in two space dimensions and assume that this problem can be semi-discretized (by finite differences or finite element methods) into an explicit system of ordinary differential equations of the form

$$\frac{d^\nu y}{dt^\nu} = f(t, y), \qquad \nu \geqq 1, \tag{1.1}$$

where the boundary conditions are lumped into the right-hand side and the initial condition is of the form

$$\frac{d^i y}{dt^i}(t_0) = y_0^{(i)}, \qquad i = 0, \cdots, \nu - 1. \tag{1.2}$$

We assume that the Jacobian matrix $\partial f/\partial y$ has *negative* eigenvalues.

Suppose that a linear multistep method is chosen for the integration of (1.1). Then in each integration step we have to solve a, usually nonlinear, system of equations of the form

$$y - b_0 \tau^\nu f(t_{n+1}, y) = \sum_{l=1}^{k} [a_l y_{n+1-l} + b_l \tau^\nu f(t_{n+1-l}, y_{n+1-l})], \tag{1.3}$$

where τ is the stepsize $t_{n+1} - t_n$, y_n is the numerical approximation to $y(t_n)$ and $\{a_l, b_l\}$ are coefficients specifying the k-step method chosen. The solution of (1.3) is denoted by η, the approximation to η obtained in actual computation by y_{n+1}. We will write (1.3) in the compact form

$$Ly = \Sigma. \tag{1.3'}$$

In this paper we analyse a special class of nonlinear Chebyshev iteration methods for solving (1.3'). The special features of this iteration process are (i) the application of a three-term Chebyshev recursion, (ii) the use of splitting functions in the definition of the approximate inverse of L, (iii) the strong damping of the lower frequencies by the amplification operator.

Since for computational reasons one wishes a relatively low number of iterations, the iteration result may differ considerably from the solution of (1.3). Therefore, we will also consider the stability of the iteration result for a class of model problems.

* Received by the editors November 25, 1981, and in final revised form November 22, 1983.
† Centrum voor Wiskunde en Informatica, Kruislaan 413, Amsterdam, The Netherlands.

Finally, a few numerical experiments will be reported comparing the method proposed in this paper with conventional splitting methods.

2. The iteration error. Suppose we want to solve the problem

$$Ly = \Sigma, \tag{2.1}$$

where L is a (nonlinear) operator in $\mathbb{R}_r$ and Σ a given vector. We will assume that L has an inverse L^{-1}. For such problems one may define the iteration process (cf. Stetter [7])

$$y^{(j+1)} = y^{(j)} + \tilde{L}_j^{-1}(\Sigma + \tilde{\Sigma}_j - Ly^{(j)}) - \tilde{L}_j^{-1}\tilde{\Sigma}_j, \qquad j = 0, 1, \cdots, \tag{2.2}$$

where $y^{(0)}$ is an approximation to the solution η of (2.1), $\tilde{\Sigma}_j$ are approximations to Σ and $\tilde{L}_j^{-1}$ are approximations to L^{-1}.

In this paper we will consider the *two-step version* of (2.2),

$$y^{(j+1)} = \mu_j y^{(j)} + (1-\mu_j)y^{(j-1)} + \lambda_j[\tilde{L}_j^{-1}(\Sigma + \tilde{\Sigma}_j - Ly^{(j)}) - \tilde{L}_j^{-1}\tilde{\Sigma}_j], \qquad j = 0, 1, 2, \cdots, \tag{2.3}$$

where $\mu_0 = 1$, and μ_j and λ_j are parameters which will be used in order to accelerate the convergence.

If the operators $\tilde{L}_j^{-1}$ and L are differentiable then the iteration error $\varepsilon_j = y^{(j)} - \eta$ of (2.3) satisfies a relation of the form

$$\varepsilon_{j+1} = [\mu_j - \lambda_j(\tilde{L}_j^{-1})'L']\varepsilon_j + (1-\mu_j)\varepsilon_{j-1} + O(\|\varepsilon_j\|^2), \tag{2.4}$$

where $(\tilde{L}_j^{-1})'$ and L' denote the derivatives (Jacobian matrices) of the operators $\tilde{L}_j^{-1}$ and L evaluated at $\tilde{\Sigma}_j$ and η, respectively. We remark that the second order term in (2.4) vanishes if $\tilde{L}_j^{-1}$ and L are affine operators. Furthermore, since L and $\tilde{L}_j$ are supposed to be the left-hand side operator in (1.3) and its approximation, respectively, we expect $\tilde{L}_j y$ to be of the form $I - b_0\tau^\nu \tilde{f}(t_{n+1,y})$ with $\tilde{f} \approx f$. In such cases the order constant in (2.4) will contain a factor τ^ν. Finally, we observe that by defining

$$\tilde{\Sigma}_j := \tilde{L}_j y^{(j)} \tag{2.5}$$

and by writing (according to the inverse mapping theorem)

$$(\tilde{L}_j')^{-1} = (\tilde{L}_j^{-1})',$$

the recurrence relation (2.4) may be written as

$$\varepsilon_{j+1} = [\mu_j - \lambda_j(\tilde{L}_j')^{-1}L']\varepsilon_j + (1-\mu_j)\varepsilon_{j-1} + O(\|\varepsilon_j\|^2), \tag{2.4'}$$

where $\tilde{L}_j'$ is evaluated at $y^{(j)}$. In many cases this error equation is more convenient than (2.4) because we often cannot explicitly derive the matrix $(\tilde{L}_j^{-1})'$ whereas the matrix $(\tilde{L}_j')^{-1}$ is rather easily obtained.

In this paper it will be assumed that $\tilde{\Sigma}_j$ is defined by (2.5).

2.1. Chebyshev iteration. In the special case where $\tilde{L}_j'$ *does not depend on* j *and* $\tilde{L}_j^{-1}$, L *are affine*, the process (2.3) reduces to the familiar polynomial iteration method [12]. We find

$$\varepsilon_{j+1} = P_{j+1}(\tilde{L}'^{-1}L')\varepsilon_0, \qquad j = 0, 1, \cdots, \tag{2.4''}$$

where P_j is a polynomial of degree j in $\tilde{L}'^{-1}L'$ generated by the recurrence relation

$$P_0(\alpha) = 1, \qquad P_{j+1}(\alpha) = (\mu_j - \lambda_j\alpha)P_j(\alpha) + (1-\mu_j)P_{j-1}(\alpha), \qquad j = 0, 1, \cdots. \tag{2.6}$$

We will assume that the *iteration matrix* $\tilde{L}'^{-1}L'$ has its eigenvalues α in a positive interval. Then all eigenvector components in the iteration error corresponding to eigenvalues in the interval $[a, b]$ are maximally damped if we choose

$$P_j(\alpha)=\frac{T_j(w_0+w_1\alpha)}{T_j(w_0)}, \quad w_0=\frac{b+a}{b-a}, \quad w_1=\frac{2}{a-b}. \tag{2.7}$$

Since T_j satisfies a recurrence relation of the form (2.6) we can explicitly derive the parameters μ_j and λ_j. The resulting method is the well-known *Richardson method* [12] (or *Chebyshev iteration method*) applied to the preconditioned problem ($\tilde{L}_j^{-1}$ does not depend on j because it is affine)

$$\tilde{L}^{-1}Ly=\tilde{L}^{-1}\Sigma. \tag{2.1'}$$

If $\tilde{L}_j'^{-1}$ *does not depend on* j *but* $\tilde{L}_j^{-1}$, $\tilde{L}$ *are nonlinear*, we formally may define the parameters μ_j and λ_j by (2.6) and (2.7). Then, neglecting second order terms, (2.4″) presents a first order approximation to the error equations. Thus, for *sufficiently close initial approximations* $y^{(0)}$ the error equation (2.4″) can be used in the analysis of the iteration process.

In this paper it will be assumed that $\tilde{L}_j'$ does not depend on j.

2.2. Damping of low frequencies and consistency. In the usual application of iteration processes of the form (2.3), one chooses the parameters such that all frequencies in the initial error $\varepsilon_0=y^{(0)}-\eta$ are damped by roughly the same factor. However, if the problem (2.1) originates from a partial differential equation, then often the solution mainly consists of low frequency modes so that one chooses the discrete problem (2.1) such that its solution η does not contain high frequencies (for example, the backward differentiation formulas). Thus, if the initial approximation $y^{(0)}$ does not contain high frequencies (e.g. if $y^{(0)}$ is obtained by extrapolation of preceding y_n values) then ε_0 will also be free of high frequencies. In such cases only the low frequencies should be strongly damped whereas the high frequencies need only marginal damping. If the low frequencies correspond to large eigenvalues of the iteration matrix this can be achieved by choosing $a \gg \bar{a}$ and $b=\bar{b}$ where $[\bar{a}, \bar{b}]$ denotes the (positive) spectrum of the iteration matrix. As a consequence, the damping of the low frequencies increases considerably as is immediately clear from (2.7) which yields after m iterations the *damping factor*

$$D:=\max_{a\leqq\alpha\leqq b}|P_m(\alpha)|=T_m^{-1}\left(\frac{b+a}{b-a}\right)\approx\left\{\cosh\left(2m\left[\sqrt{\frac{a}{b-a}}+O\left(\frac{a}{b-a}\right)\right]\right)\right\}^{-1} \tag{2.8}$$

as $a/(b-a)\ll 1$ ($<.025$ say). It turns out that in most applications $a/(b-a)$ is rather small so that for prescribed damping D the number of iterations m can be found from the approximate expression for D, i.e.

$$m\approx\frac{\operatorname{arccosh}(1/D)}{\operatorname{arccosh}[(b+a)/(b-a)]}\approx\frac{1}{2}\sqrt{\frac{b-a}{a}}\operatorname{arccosh}\left(\frac{1}{D}\right). \tag{2.8'}$$

It is the purpose of this paper to derive iteration processes of the form (2.3) which strongly damp the low frequency modes and which have a modest damping of the higher frequencies. In the analysis we assume that only a few iterations are performed; otherwise the method becomes too expensive. As a consequence, $y^{(m)}$ may differ considerably from the solution of (1.3). This implies that one should consider the

consistency of the result $y^{(m)}$ as $\tau \to 0$. Evidently, the local error $y^{(m)} - y(t_{n+1})$ at t_{n+1} consists of the local error $\eta - y(t_{n+1})$ of the generating multistep formula and the iteration error ε_m, approximately given by (2.4″).

In our applications the matrix $\tilde{L}'^{-1}L'$ converges to the matrix $\alpha_0 I$ as $\tau \to 0$, i.e.

$$\tilde{L}'^{-1}L' = \alpha_0 I + \tau^r B(\tau), \qquad r \geqq 1, \tag{2.9}$$

where $B(\tau)$ is a nonvanishing, uniformly bounded matrix as $\tau \to 0$. Then

$$\varepsilon_m = P_m(\tilde{L}'^{-1}L)\varepsilon_0 = [P_m(\alpha_0)I + \tau^r P'_m(\alpha_0)B(\tau) + \tfrac{1}{2}\tau^{2r}P''_m(\alpha_0)B^2(\tau) + \cdots]\varepsilon_0. \tag{2.10}$$

From the definition (2.7) of $P_m(\alpha)$ we derive that

$$\begin{aligned}\varepsilon_m = T_m^{-1}(w_0)[&T_m(w_0 + w_1\alpha_0) + w_1 T'_m(w_0 + w_1\alpha_0)\tau^r B(\tau) \\ &+ \tfrac{1}{2}w_1^2 T''_m(w_0 + w_1\alpha_0)\tau^{2r}B^2(\tau) + \cdots]\varepsilon_0.\end{aligned} \tag{2.10′}$$

Introducing the damping factor D and observing that

$$w_1 = -\frac{2}{b-a} > -\frac{2}{[\sqrt{b}+\sqrt{a}]^2}\sqrt[m]{\frac{2}{D}}$$

we obtain

$$\begin{aligned}\|\varepsilon_m\| \leqq D\Bigg[&|T_m(w_0 + w_1\alpha_0)| + \frac{\tau^r}{\sqrt[m]{D/2}}|T'_m(w_0 + w_1\alpha_0)|\,\|\bar{B}(\tau)\| \\ &+ \frac{1}{2}\left(\frac{\tau^r}{\sqrt[m]{D/2}}\right)^2 |T''_m(w_0 + w_1\alpha_0)|\,\|\bar{B}(\tau)\|^2 + \cdots\Bigg]\|\varepsilon_0\|,\end{aligned} \tag{2.11}$$

where $\bar{B}(\tau)$ denotes the "normalized" matrix $2B(\tau)/[\sqrt{a}+\sqrt{b}]^2$ and D is assumed to be a given number independent of τ (e.g. $D = 1/10$).

The estimate (2.11) is suitable for practical use if τ is sufficiently small, i.e.

$$\tau^r \ll \frac{2\sqrt[m]{D/2}}{\|\bar{B}(\tau)\|}. \tag{2.12}$$

In this range of integration steps the iteration error ε_m can be decreased if we are able to choose $T_m(w_0 + w_1\alpha_0) = 0$, i.e.

$$w_0 + w_1\alpha_0 = \cos\left(\frac{2l+1}{2m}\pi\right), \qquad l \in \{0, 1, \cdots, m-1\}, \tag{2.13}$$

or equivalently

$$a = \frac{2\alpha_0 + b[\cos((2l+1)\pi/2m) - 1]}{\cos((2l+1)\pi/2m) + 1}, \qquad l \in \{0, 1, \cdots, m-1\}, \tag{2.13′}$$

where we assume

$$b > \alpha_0 > \frac{1}{2}b\left(1 - \cos\left(\frac{2l+1}{2m}\pi\right)\right). \tag{2.14}$$

Substitution of (2.13) into (2.11) and using the relation

$$T'_m(w) = \sqrt[m]{\frac{1 - T_m^2(w)}{1 - w^2}}$$

yields

$$\|\varepsilon_m\| \leqq D\left[\frac{m}{\sin((2l+1)\pi/2m)}\frac{\tau^r\|\bar{B}(\tau)\|}{\sqrt[m]{D/2}}+O\left(\left[\frac{\tau^r\bar{B}(\tau)}{\sqrt[m]{D/2}}\right]^2\right)\right]\|\varepsilon_0\| \quad \text{as } \tau\to 0. \tag{2.15}$$

Firstly, this estimate shows that for fixed damping factor D

$$\|y^{(m)}-y(t_{n+1})\| = \|\varepsilon_m\|+O(\tau^{p+\nu}) \leqq O(\tau^{p+\nu}+\tau^{q+\nu+r}) \quad \text{as } \tau\to 0,$$

where p and q are the orders of consistency of the generating multistep method and of the predictor formula used for $y^{(0)}$. Thus the order of consistency $\tilde{p}$ is given by

$$\tilde{p}=\min\{p, q+r\}. \tag{2.16}$$

Notice that $\tilde{p}=\min\{p, q\}$ if the consistency condition (2.13) is not satisfied.

Secondly, we observe that for given D the value of m should be minimized, that is in (2.8′) the value of $(b-a)/a$ should be made as small as possible. In view of (2.13′) this means that

$$\frac{b-a}{a}=\frac{2(b-\alpha_0)}{2\alpha_0+b(\cos((2l+1)\pi/2m)-1)}$$

should be minimized. This implies that $l=0$ is the best choice.

Given the operator L and the approximating operators $\tilde{L}_j$, the iteration parameters in the splitting method (2.3) can be explicitly derived from (2.6), (2.7) and (2.13′) with $b=\bar{b}$ and $l=0$. The only free parameter left is the number of iterations m which will be used to satisfy stability conditions (see § 4) and to monitor the damping of the low frequencies.

3. The approximate inverse. In order to define the approximate inverse $\tilde{L}_j^{-1}$ for the problem (1.1) we use the formalism developed in [4] and introduce the *splitting* function $F(t, u, v)$ which is such that

$$F(t, y, y)\equiv f(t, y). \tag{3.1}$$

This rather general splitting function includes a number of well-known splittings such as the ADI splittings [6] and the hopscotch splittings [1]. It is convenient to introduce the Jacobian matrices

$$Z_1=b_0\tau^\nu\frac{\partial F}{\partial u}, \quad Z_2=b_0\tau^\nu\frac{\partial F}{\partial v}, \quad Z=Z_1+Z_2, \tag{3.2}$$

which are both evaluated at (t_{n+1}, η, η). The eigenvalues of Z_i, Z will be denoted by z_i and z, respectively. We assume that Z has negative eigenvalues in the interval $[-S, 0)$ and that the algebraically large eigenvalues correspond to eigenvectors of low frequency. The spectral radius of $\partial f/\partial y$ is given by $S/b_0\tau^\nu$ and will be denoted by σ.

3.1. Successive corrections.

3.1.1. One-stage approximations. A relatively simple class of methods is based on the approximate inverse $\tilde{L}_j^{-1}$: $x\to y$ defined by the one-stage formula

$$\omega y+(1-\omega)y^{(j)}-b_0\tau^\nu F(t_{n+1}, y, y^{(j)})=x, \qquad \omega\neq 0. \tag{3.3}$$

Thus, in addition to the parameter m we also have the parameter ω optimizing the splitting method.

From (3.3) we derive that

$$\tilde{L}'^{-1}L' = [\omega - Z_1]^{-1}[I - Z_1 - Z_2]. \tag{3.4}$$

Examples of splitting functions which are suitable for use in (3.3) are the Jacobi and Gauss–Seidel splittings.

By writing (3.4) in the form (2.9), that is

$$\tilde{L}'^{-1}L' = \frac{1}{\omega} - b_0\tau^\nu\left[\omega - b_0\tau^\nu\frac{\partial F}{\partial u}\right]^{-1}\left[\frac{\omega-1}{\omega}\frac{\partial F}{\partial u} + \frac{\partial F}{\partial v}\right], \tag{3.4'}$$

we see that $\alpha_0 = 1/\omega$, $r = \nu$ and that $B(\tau)$ is uniformly bounded as $\tau \to 0$. Hence, the error equation (2.15) applies provided that (2.14) holds:

$$\frac{1}{b} < \omega < \frac{2}{b(1-\cos(\pi/2m))}. \tag{3.5}$$

Within this range of ω-values we try to make the factor D sufficiently small. In addition, however, we require that the interval $[a, b]$ contains sufficiently many eigenvalues of eigenvectors of low frequency.

Let us consider the important case where Z_1 is given by

$$Z_1 = -\theta SI. \tag{3.6}$$

Then

$$a := \frac{2+\omega b(\cos(\pi/2m)-1)}{\omega(\cos(\pi/2m)+1)}, \quad b := \bar{b} = \frac{1+S}{\omega+\theta S}, \quad \bar{a} = \frac{1}{\omega+\theta S}. \tag{3.7}$$

The eigenvalues corresponding to the lower frequencies are in the neighbourhood of $\bar{a}$. In Fig. 3.1 the corresponding polynomial $P_m(\alpha)$ is illustrated for $m = 2$.

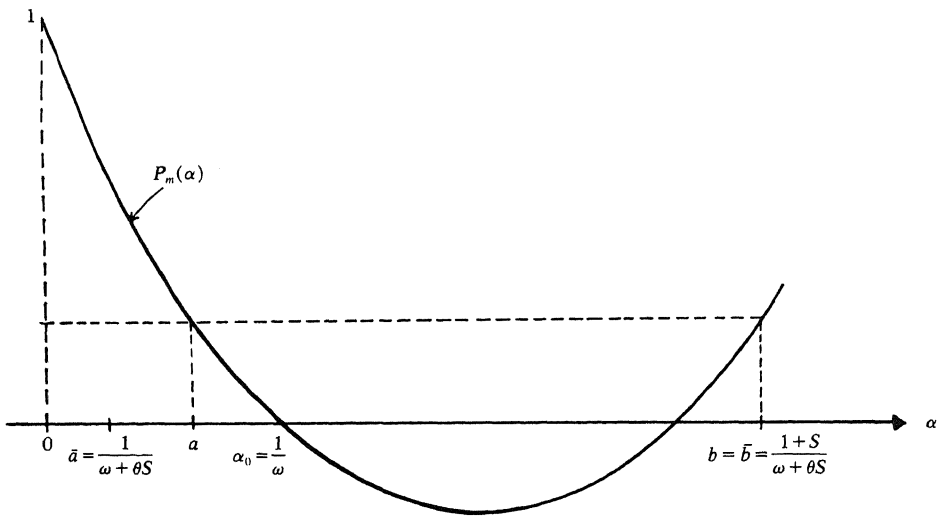

FIG. 3.1. *The polynomial $P_m(\alpha)$ for $m = 2$.*

If $\theta = 0$ the iteration process does not contain implicit relations and can be considered as an *explicit Runge–Kutta method* of a special form. Related methods were analysed in [5]. If $\theta \neq 0$, e.g. $\theta = \frac{1}{2}$, the iteration process only contains *scalarly implicit relations* which may be attractive from a computational point of view (notice

that this process is identical to nonlinear Jacobi iteration if $\operatorname{diag}[\partial f/\partial y] = -\theta SI$). However, from Fig. 3.1 we conclude that choosing $(\bar{b}-\bar{a})/\bar{b}$ small means that all eigenvectors of low frequency are not damped unless θS is so small that $\bar{a} \cong a$. In practice, $\theta S = \theta b_0 \tau^\nu \sigma$ is usually rather large because the integration step τ is much greater than $\sigma^{-1/\nu}$, σ being the spectral radius of $\partial f/\partial y$. Hence in order to damp low frequencies we should choose ω such that $\bar{a} \cong a$, that is

$$\omega = \frac{2\theta}{1-\cos(\pi/2m)}. \tag{3.8}$$

This value for ω satisfies the inequality (3.5) for all $\theta > 0$. Substitution in (3.7) yields $(b-a)/a = S$ so that for $S \gg 1$ (cf. (2.8′))

$$m \approx \frac{1}{2}\sqrt{S}\,\operatorname{arccosh}\left(\frac{1}{D}\right). \tag{3.9}$$

Unless $D \cong 1$ this value for m is extremely large because of the usually large values of S.

3.1.2. Two-stage approximations. In this section iteration matrices are considered in which the large eigenvalues correspond to eigenvectors of low frequency. This enables us to get a strong damping of the lower frequencies without an extremely large number of iterations.

Consider the operator $\tilde{L}_j^{-1}: x \to y$ defined by the two-stage formula

$$\omega y + (1-\omega)y^* - b_0\tau^\nu F(t_{n+1}, y, y^*) = x, \tag{3.10}$$

$$\omega y^* + (1-\omega)y^{(j)} - b_0\tau^\nu F(t_{n+1}, y^{(j)}, y^*) = x, \qquad \omega \neq 0, \tfrac{1}{2}. \tag{3.11}$$

The corresponding splitting method (2.3) again possesses the two free iteration parameters m and ω. An elementary calculation leads to the iteration matrix

$$\tilde{L}'^{-1}L' = (2\omega-1)[\omega - Z_1]^{-1}[\omega - Z_2]^{-1}[I - Z_1 - Z_2]. \tag{3.12}$$

By writing (3.12) in the form (2.9) we see that

$$\alpha_0 = \frac{2\omega-1}{\omega^2}, \qquad r = \nu \tag{3.13}$$

and that $B(\tau)$ is uniformly bounded as $\tau \to 0$. Thus, (2.15) holds provided that inequality (2.14) is satisfied. This inequality gives an interval of ω-values and within this interval one should try to minimize the factor $(b-a)/b$ occurring in (2.8′) and at the same time to include sufficiently many eigenvalues of low frequency eigenvectors in the interval $[a, b]$.

In the following we consider in more details the model problem where Z and $\omega - Z_i$ share the same eigensystem of which the eigenvectors of low frequency correspond to eigenvalues of small magnitude. Then from (2.13′) and (3.12) we find (with $l = 0$)

$$a = \frac{2(2\omega-1) + \omega^2 b(\cos(\pi/2m)-1)}{\omega^2(\cos(\pi/2m)+1)}, \tag{3.14a}$$

$$\bar{a} = (2\omega-1)\frac{S+1}{(S/2+\omega)^2}, \qquad \bar{b} = \frac{2\omega-1}{\omega}\frac{S+1}{S+\omega}, \tag{3.15}$$

where we have assumed that $\bar{a} \leqq \alpha_0 \leqq b$, that is

$$1 \leqq \omega \leqq \tfrac{1}{2}[1+\sqrt{2S+1}]. \tag{3.16}$$

Since S is usually rather large we choose instead of $b = \bar{b}$

$$b = \frac{2\omega-1}{\omega} \geqq \bar{b} = \frac{2\omega-1}{\omega}\left[1+O\left(\frac{1}{S}\right)\right] \quad \text{as } S \to \infty. \tag{3.14b}$$

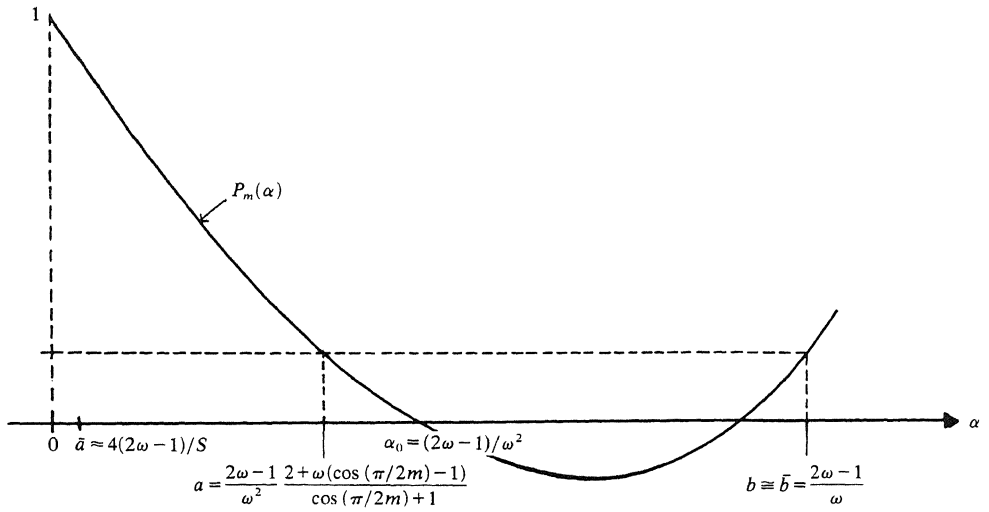

FIG. 3.2. *The polynomial $P_m(\alpha)$ for $S \gg 1$.*

In Fig. 3.2 the corresponding polynomial $P_m(\alpha)$ is illustrated for $S \gg 1$. Evidently, the low frequency eigenvectors have eigenvalues in the neighbourhood of α_0 which is different from the situation in the preceding section where these eigenvalues are in the neighbourhood of $\bar{a}$. In order to see what eigenvalues correspond to the damped eigenvectors we show in Fig. 3.3 in the (z_1, z_2)-plane the region corresponding to the interval $a \leqq \alpha \leqq b$. The magnitude of this damping region can be characterized by the quantity

$$S^* := \frac{\omega(1+\sqrt{1-a})-1}{1-\sqrt{1-a}}, \qquad a \leqq \frac{2\omega-1}{\omega^2} \tag{3.17}$$

(the inequality for a follows from $\alpha(0,0) \geqq a$).

In the following it is convenient to use the directly interpretable parameter S^* instead of ω. From (3.17) it follows that

$$a = \frac{(2\omega-1)(2S^*+1)}{(S^*+\omega)^2}, \qquad 1 \leqq \omega \leqq \tfrac{1}{2}[1+\sqrt{2S^*+1}] \tag{3.17'}$$

and from (3.14a) we find that ω and S^* are related by the equation

$$(2S^*+1)\left(\cos\left(\frac{\pi}{2m}\right)+1\right)\omega^2 = \left[2+\omega\left(\cos\left(\frac{\pi}{2m}\right)-1\right)\right](S^*+\omega)^2. \tag{3.18}$$

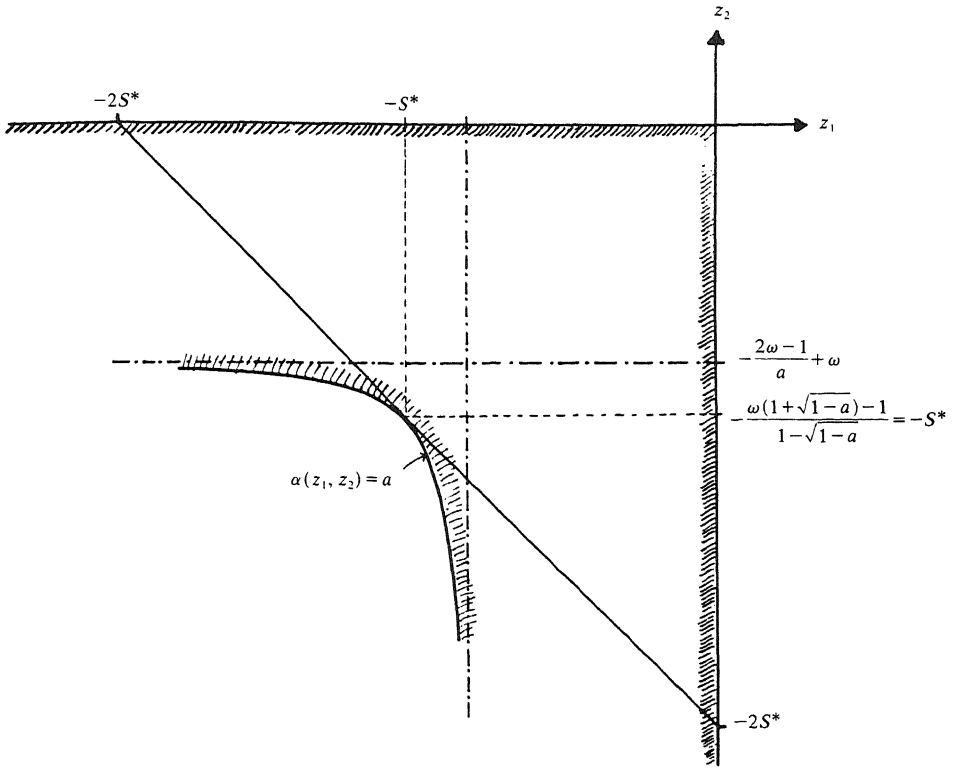

FIG. 3.3. *Region of damping in the* (z_1, z_2)*-plane.*

In order to compute the value of the damping factor D for given values of m and S^* we write

$$D = T_m^{-1}\left(\frac{\omega \cos(\pi/2m)+1}{\omega-1}\right). \tag{3.19}$$

In Table 3.1 the values (ω, D) are given satisfying (3.18) and (3.19) for various values of m and S^*. All ω-values turn out to be in the range (3.16). Choosing an appropriate value for S^*, this table gives the number of iterations required to obtain the desired damping factor D. However, as we will see in § 4, the parameters (m, S^*) have also to satisfy the stability conditions so that a definite choice has to be postponed.

TABLE 3.1
$(\omega; D)$-values for various values of m and S^*.

S^*	$m=1$	$m=2$	$m=3$	$m=4$	$m\to\infty$
1	(1.15; .15)	(1.29; .01)	(1.33; 10^{-3})	(1.34; $7_{10^{-5}}$)	(1.37; $2\exp(-2.5m)$)
2	(1.26; .26)	(1.50; .03)	(1.56; $4_{10^{-3}}$)	(1.58; $4_{10^{-4}}$)	(1.62; $2\exp(-2.1m)$)
4	(1.40; .40)	(1.80; .07)	(1.90; .01)	(1.94; $2_{10^{-3}}$)	(2.00; $2\exp(-1.7m)$)
6	(1.49; .49)	(2.02; .10)	(2.17; .02)	(2.23; $4_{10^{-3}}$)	(2.30; $2\exp(-1.6m)$)
8	(1.55; .55)	(2.20; .12)	(2.39; .03)	(2.46; $6_{10^{-3}}$)	(2.56; $2\exp(-1.5m)$)
10	(1.60; .60)	(2.36; .15)	(2.59; .04)	(2.67; $9_{10^{-3}}$)	(2.79; $2\exp(-1.4m)$)
50	(1.87; .87)	(3.84; .41)	(4.67; .16)	(5.02; .06)	(5.52; $2\exp(-0.9m)$)
100	(1.93; .93)	(4.58; .55)	(5.99; .26)	(6.63; .11)	(7.59; $2\exp(-0.7m)$)

For large values of m (and S^* not extremely large) the relations (3.18) and (3.19) can be approximated by the equations

$$\omega = \frac{1}{2}[1+\sqrt{2S^*+1}], \qquad D = 2\exp\left[-m \operatorname{arccosh}\left(\frac{\omega+1}{\omega-1}\right)\right].$$

Substitution of the expression for ω into (3.19) yields for m

$$\begin{aligned} m &\approx \left[\operatorname{arccosh}\left(1+2\frac{1+\sqrt{2S^*+1}}{S^*}\right)\right]^{-1} \operatorname{arccosh}\left(\frac{1}{D}\right) \\ &\approx .42[S^*]^{1/4} \ln\left(\frac{2}{D}\right) \quad \text{as } D \ll 1 \text{ and } S^* \gg 1. \end{aligned} \tag{3.20}$$

A comparison of (3.20) and (3.9) reveals that the number of iterations m_1 of the one-stage operator and the number of iterations m_2 of the two-stage operator, needed to produce the same damping D, are related by the formula

$$m_2 = .82\sqrt[4]{\frac{S^*}{S^2}}\, m_1 \quad \text{as } D \ll 1 \text{ and } S^* \gg 1.$$

Thus, even for $S^* = S$, the two-stage formula is usually much more efficient.

3.1.3. Multistage approximations. Next consider the operator $\tilde{L}_j^{-1}: x \to y$ defined by the $\tilde{m}$-stage formula (compare similar operators employed in linear elliptic equations e.g. in [12, p. 518])

$$\begin{aligned} &y_0^* = y^{(j)}, \\ &\omega_i y_i^* + (1-\omega_i)\bar{y}_{i-1}^* - b_0\tau^\nu F(t_{n+1}, y_i^*, \bar{y}_{i-1}^*) = x, \\ &\omega_i \bar{y}_{i-1}^* + (1-\omega_i) y_{i-1}^* - b_0\tau^\nu F(t_{n+1}, y_{i-1}^*, \bar{y}_{i-1}^*) = x, \qquad i = 1, 2, \cdots, \tilde{m}, \\ &y = y_{\tilde{m}}^*. \end{aligned} \tag{3.21}$$

The corresponding iteration matrix is given by

$$\tilde{L}'^{-1}L' = I - \prod_{i=\tilde{m}}^{1} [\omega_i - Z_1]^{-1}[\omega_i - Z_2]^{-1}[\omega_i - 1 + Z_1][\omega_i - 1 + Z_2] \tag{3.22}$$

which can be written in the form (2.9) with

$$\alpha_0 = 1 - \prod_{i=1}^{\tilde{m}} \left(\frac{\omega_i - 1}{\omega_i}\right)^2, \qquad r = \nu \tag{3.23}$$

and $B(\tau)$ uniformly bounded in τ. Assuming that the parameters ω_i satisfy the inequality (3.16) and restricting our considerations to the same class of model problems as in the preceding section, we find that

$$\bar{b} \leqq 1 + \frac{\omega_l - 1}{\omega_l}\frac{S - \omega_l + 1}{S + \omega_l} \prod_{i \neq l} \left[\frac{S - \omega_i + 1}{S + \omega_i}\right]^2 \cong \frac{2\omega_l - 1}{\omega_l} \quad \text{as } S \to \infty, \tag{3.24}$$

where l is such that $(\omega_l - 1)/\omega_l$ is maximal. We define a by (2.13) and put

$$b = \max_i \frac{2\omega_i - 1}{\omega_i}. \tag{3.25}$$

The damping region $a \leqq \alpha(z_1, z_2) \leqq b$ in the (z_1, z_2)-plane contains the region defined by

$$\prod_{i=1}^{\tilde{m}} \left| \frac{\omega_i - 1 + z_j}{\omega_i - z_j} \right| \leqq \sqrt{1-a}, \qquad j = 1, 2. \tag{3.26}$$

We now use the following lemma.

LEMMA 3.1. *In the interval $A \leqq x \leqq B$ the function*

$$\psi_m(x) = \prod_{i=1}^{m} \frac{x - \theta_i}{x + \theta_i}, \qquad \theta_i = B\left[\frac{A}{B}\right]^{(i-1)/(m-1)}, \qquad 0 < A < B, \quad m \geqq 2$$

is bounded by

$$\left[\frac{1 - C_m}{1 + C_m}\right]^2, \qquad C_m = \left[\frac{A}{B}\right]^{1/2(m-1)}.$$

Proof. See Young [12, p. 528]. □

The parameters θ_i were proposed by Wachspress [10]. We apply this lemma with

$$x = \tfrac{1}{2} - z_j, \quad A = \tfrac{1}{2}, \quad B = \tfrac{1}{2} + S^*, \quad m = \tilde{m}, \quad \theta_i = \omega_i - \tfrac{1}{2}.$$

Thus, if

$$\omega_i = \tfrac{1}{2} + \tfrac{1}{2}(2S^* + 1)^{(\tilde{m}-i)/(\tilde{m}-1)}, \qquad i = 1, 2, \cdots, \tilde{m} \geqq 2, \tag{3.27}$$

then

$$\prod_{i=1}^{\tilde{m}} \left| \frac{\omega_i - 1 + z_j}{\omega_i - z_j} \right| \leqq \left[\frac{1 - C_{\tilde{m}}}{1 + C_{\tilde{m}}}\right]^2, \qquad C_{\tilde{m}} = [2S^* + 1]^{-1/2(\tilde{m}-1)}$$

for $-S^* \leqq z_j \leqq 0$ (note that the left-hand side is bounded by one for all negative values of z_j). Hence, if a is chosen such that

$$\sqrt{1-a} = \left[\frac{1 - C_{\tilde{m}}}{1 + C_{\tilde{m}}}\right]^2, \tag{3.28}$$

i.e.

$$a = \frac{8C_{\tilde{m}}(1 + C_{\tilde{m}}^2)}{(1 + C_{\tilde{m}})^4},$$

then (3.26) is satisfied for all (z_1, z_2) in the square $-S^* \leqq z_1, z_2 \leqq 0$. However, a is also prescribed by (2.13'), so that the consistency equation

$$\frac{8C_{\tilde{m}}(1 + C_{\tilde{m}}^2)}{(1 + C_{\tilde{m}})^4} = \frac{2 + b(\cos(\pi/2m) - 1)}{\cos(\pi/2m) + 1}, \qquad b = \frac{2}{1 + C_{\tilde{m}}^{2(\tilde{m}-1)}}, \tag{3.29}$$

$$C_{\tilde{m}} = [2S^* + 1]^{-1/2(\tilde{m}-1)}$$

should be satisfied (notice that $\alpha_0 = 1$ because $\omega_{\tilde{m}} = 1$). Here, S^* cannot be chosen freely as in the preceding section. In Table 3.2 a few values of (S^*, D) are given, where S^* satisfies (3.29). The asymptotic error estimate (2.15) holds for the (m, D)-values occurring in this table.

In order to compare the efficiency of the two-stage operator and the multistage Wachspress operator we consider the number of iterations given by (3.20) and the quantity $\tilde{m}m$ giving the number of "iterations" of the present process. In terms of S^* and D we have

$$\tilde{m}m = \tilde{m}\frac{\operatorname{arccosh}(1/D)}{\operatorname{arccosh}((b+a)/(b-a))} \cong \frac{\tilde{m}}{4}[2S^*]^{1/4(\tilde{m}-1)} \ln\left(\frac{2}{D}\right) \tag{3.30}$$

as $D \ll 1$ and $S^* \gg 1$.

TABLE 3.2
$(S^*; D)$ values for various values of m and $\tilde{m}$ with S^* satisfying (3.29).

	$\tilde{m}=2$	$\tilde{m}=3$	$\tilde{m}=4$
$m=1$	$(\infty; 1)$	$(\infty; 1)$	$(\infty; 1)$
$m=2$	$(9; .080)$	$(223; .093)$	$(4805; .094)$
$m=3$	$(3.7; 8_{10^{-3}})$	$(47; .013)$	$(482; .014)$

Taking (3.20) and (3.30) as a measure for the computational effort of the two-stage and multistage methods, we may conclude that the two-stage approximation should be used if (3.20) yields a lower value than (3.30). In particular, we compare the D-values obtained for the two-stage operator for the same S^*-value and check if the number of iterations equals the value of $m\tilde{m}$ listed in Table 3.2. Writing $m = m_1 m_2$ the two-stage operator yields values given by Table 3.2′, showing that the two-stage operator has a considerably stronger damping in the same damping region unless S^* is extremely large.

TABLE 3.2′
$(S^*; D)$ values satisfying (3.18) and (3.19) for various values of $m = m_1 m_2$.

	$m_2=2$	$m_2=3$	$m_2=4$
$m_1=1$	$(\infty; 1)$	$(\infty; 1)$	$(\infty; 1)$
$m_1=2$	$(9; .008)$	$(223; .05)$	$(4805; .2)$
$m_1=3$	$(37; 4_{10^{-5}})$	$(47; 5_{10^{-4}})$	$(482; 4_{10^{-3}})$

3.1.4. Recommendation of a successive-correction scheme. In the preceding subsections it has been shown that the two-stage approximation (3.11) to the operator $\tilde{L}_j^{-1}$ is expected to be superior both to the one-stage approximation (3.3) and to the multistage approximation (3.21). Therefore, we conclude this section by writing down explicitly the complete scheme based on the recommended two-stage operator (3.11). Firstly, however, we simplify the scheme (2.3) by using (2.5) and (3.11). Let $x := \tilde{L}_j y^{(j)}$ then it follows from (3.11) that x and $y^{(j)}$ are related to each other by

$$\begin{aligned} x &= \omega y^{(j)} + (1-\omega) y^* - b_0 \tau^\nu F(t_{n+1}, y^{(j)}, y^*) \\ &= \omega y^* + (1-\omega) y^{(j)} - b_0 \tau^\nu F(t_{n+1}, y^{(j)}, y^*). \end{aligned}$$

Hence, $y^* = y^{(j)}$ so that

$$\tilde{L}_j y^{(j)} = x = y^{(j)} - b_0 \tau^\nu F(t_{n+1}, y^{(j)}, y^{(j)}) = y^{(j)} - b_0 \tau^\nu f(t_{n+1}, y^{(j)}) = L y^{(j)}.$$

Substitution of this and of (2.5) into (2.3) yields the scheme

$$y^{(j+1)} = (\mu_j - \lambda_j) y^{(j)} + (1-\mu_j) y^{(j-1)} + \lambda_j \tilde{L}_j^{-1} \Sigma, \qquad j = 0, 1, \cdots, m-1. \tag{3.31}$$

Again using (3.11) yields

$$y^{(j+1)} = (\mu_j - \lambda_j) y^{(j)} + (1-\mu_j) y^{(j-1)} + \lambda_j y^\bullet, \qquad j = 0, 1, \cdots, m-1, \tag{3.32a}$$

where $y^\bullet$ is to be computed by solving the system

$$\begin{aligned} &\omega y^\bullet + (1-\omega) y^* - b_0 \tau^\nu F(t_{n+1}, y^\bullet, y^*) = \Sigma, \\ &\omega y^* + (1-\omega) y^{(j)} - b_0 \tau^\nu F(t_{n+1}, y^{(j)}, y^*) = \Sigma. \end{aligned} \tag{3.32b}$$

As we already observed, the parameters μ_j and λ_j are obtained from the Chebyshev recursion formula, i.e.,

$$\mu_0=\frac{1}{2}(b+a)\lambda_0=1, \quad \mu_j=2w_0\frac{T_j(w_0)}{T_{j+1}(w_0)}, \quad \lambda_j=\frac{2\mu_j}{b+a}, \quad w_0=\frac{b+a}{b-a}, \tag{3.32c}$$
$$j=1,2,\cdots,m-1,$$

where a and b are defined by

$$a=\frac{(2\omega-1)(2S^*+1)}{(S^*+\omega)^2}, \qquad b=\frac{2\omega-1}{\omega}, \tag{3.32d}$$

and where ω and S^* are related by (3.18).

Given a corrector formula (1.3) and a splitting function $F(t, u, v)$ (see [4] for a survey), the successive-correction-scheme is now completely determined if we specify the predictor formula for $y^{(0)}$, the number of iterations m, and the frequency parameter S^*; this scheme will be denoted by SC $(y^{(0)}, m, S^*)$. It has already been observed that the choice of (m, S^*) also depends on the stability conditions. These conditions will be given in § 4.1, and on the basis of these conditions we come to a definite choice of (m, S^*) in § 4.2.

3.2. Fractional steps.

3.2.1. Two-stage approximations. In this section it will be assumed that the splitting function is of the special form (cf. [4])

$$F(t, u, v)=f_1(t, u)+f_2(t, v). \tag{3.33}$$

Examples of such splitting functions are the LOD splittings [11] and the hopscotch splittings [1].

We define the operator $\tilde{L}_j^{-1}: x \to y$ in two steps (cf. (3.11)):

$$\begin{aligned} &\omega y+(1-\omega)y^*-b_0\tau^\nu[f_1(t_{n+1}, y)+f_2(t_{n+1}, y^*)]=x, \\ &\omega y^*+(1-\omega)y^{(j)}-b_0\tau^\nu f_2(t_{n+1}, y^*)=x. \end{aligned} \tag{3.34}$$

Notice that the intermediate result y^* is obtained by using only a "fraction" of the right-hand side function $f(t, y)$.

A straightforward calculation reveals that the iteration matrix $\tilde{L}'^{-1}L'$ is identical to (3.12). Consequently, the analysis of the §§ 3.1.2 and 3.1.3 also applies to the approximation (3.34) if Z_1 and Z_2 are understood to be the Jacobian matrices of $b_0\tau^\nu f_1(t_{n+1}, y)$ and $b_0\tau^\nu f_2(t_{n+1}, y)$, respectively.

Verwer [8] studied the special case where $\omega=1$, $\Sigma=y_n$, $\nu=1$, $b_0=1$ (backward Euler) and where $F(t, u, v)$ corresponds to an LOD splitting [11]. However, in that case only eigenvectors of lowest frequency are damped, and just as in the case of multistep splitting methods considered in [3], the convergence turns out to be rather poor. Verwer therefore proposed the application of line Jacobi iteration after each LOD iteration in order to damp eigenvectors of higher frequencies which indeed improves the rate of convergence [9].

4. Stability. We recall that we want a relatively low number of iterations and consequently the stability properties of $y_{n+1}=y^{(m)}$ may considerably differ from those of the exact solution η of the linear k-step formula (1.3). Therefore, we investigate the sensitivity of y_{n+1} against perturbations Δy_n of previous y_n-values.

Our considerations will be confined to the SC $(y^{(0)}, m, S^*)$ method defined by (3.32). It is convenient to write $\tilde{L}_j^{-1}$ as the operator $K: (y^{(j)}, x) \to y$. Then (2.3) assumes

the form (cf. 3.20))

$$y^{(j+1)}=(\mu_j-\lambda_j)y^{(j)}+(1-\mu_j)y^{(j-1)}+\lambda_j K(y^{(j)},\Sigma), \tag{4.1}$$

where we used (2.5). Denoting the Jacobian matrices of K with respect to the successive arguments by K_1' and K_2' we obtain the variational equation

$$\Delta y^{(j+1)}=[\mu_j-\lambda_j+\lambda_j K_1']\Delta y^{(j)}+(1-\mu_j)\Delta y^{(j-1)}+\lambda_j K_2'\Delta\Sigma. \tag{4.2}$$

From (3.11) it follows that

$$K_1'=[\omega-Z_1]^{-1}[1-\omega-Z_2][\omega-Z_2]^{-1}[1-\omega-Z_1],$$
$$K_2'=[\omega-Z_1]^{-1}[I-[1-\omega-Z_2][\omega-Z_2]^{-1}]$$

so that

$$\Delta y^{(j+1)}=[\mu_j-\lambda_j\tilde{L}'^{-1}L']\Delta y^{(j)}+(1-\mu_j)\Delta y^{(j-1)}+\lambda_j K_2'\Delta\Sigma, \tag{4.2'}$$

where $\tilde{L}'^{-1}L'$ is given by (3.12).

We now use the following lemma (cf. [5]):

LEMMA 4.1. *For arbitrary vectors u_0 and v_0 the recurrence relation*

$$v_{j+1}=[\mu_j-\lambda_j\alpha]v_j+(1-\mu_j)v_{j-1}+\lambda_j u_0, \qquad j\geqq 0 \tag{4.3}$$

is satisfied by

$$v_j=P_j(\alpha)v_0+Q_j(\alpha)u_0, \tag{4.4}$$

where $P_j(\alpha)$ is defined by (2.6) *and $Q_j(\alpha)$ by*

$$Q_j(\alpha)=\frac{1-P_j(\alpha)}{\alpha}.$$

Proof. By substitution of (4.4') into (4.3). □

Applying this lemma to (4.2) leads to the variational equation

$$\begin{aligned}&\Delta y^{(j+1)}=P_{j+1}(A)\Delta y^{(0)}+Q_{j+1}(A)K_2'\Delta\Sigma,\\ &A=\tilde{L}'^{-1}L'=(2\omega-1)[\omega-Z_1]^{-1}[\omega-Z_2]^{-1}[I-Z_1-Z_2],\\ &K_2'=(2\omega-1)[\omega-Z_1]^{-1}[\omega-Z_2]^{-1}.\end{aligned} \tag{4.5}$$

4.1. Stability analysis for model problems. In this section we assume that Z and $\omega-Z_i$ share the same eigensystem with eigenvalues z_1 and z_2. Assuming that $y^{(0)}$ is computed by a formula of the form

$$y^{(0)}=\sum_{l=1}^{k}[\hat{a}_l y_{n+1-l}+\hat{b}_l\tau_r f(t_{n+1-l},y_{n+1-l})] \tag{4.6}$$

and substituting Σ into (4.5) according to (1.3), we arrive after m iterations at the characteristic equation

$$\zeta^k=\sum_{l=1}^{k}\left\{P_m(\alpha)\left[\hat{a}_l+\frac{\hat{b}_l}{b_0}(z_1+z_2)\right]+[1-P_m(\alpha)]\frac{a_l+b_l(z_1+z_2)/b_0}{1-(z_1+z_2)}\right\}\zeta^{k-l}, \tag{4.7}$$

where α is given by

$$\alpha=\frac{(2\omega-1)(1-z_1-z_2)}{(\omega-z_1)(\omega-z_2)}. \tag{4.8}$$

We define the *stability region* by the set of points (z_1, z_2) where (4.7) has its roots on the unit disk.

An important class of methods uses extrapolation formulas for $y^{(0)}$, i.e. $\hat{b}_l = 0$ for $l = 1(1)k$, and backward differentiation formulas for Σ, i.e. $b_l = 0$ for $l = 1(1)k$. Then (4.7) reduces to

$$\zeta^k = \sum_{l=1}^{k} \left\{ \hat{a}_l P_m(\alpha) + \frac{a_l}{1 - z_1 - z_2}[1 - P_m(\alpha)] \right\} \zeta^{k-l}. \tag{4.7'}$$

In order to illustrate this characteristic equation we derive the stability regions of two well-known iterated integration formulas for first order equations.

Example 4.1. Consider Euler's backward formula as the generating formula, i.e. $k = 1$ and $a_1 = 1$, and $y^{(0)} = y_n$ as predictor formula, i.e. $\hat{a}_1 = 1$. Evidently, the stability region consists of the set of points (z_1, z_2) where

$$|\zeta| = \left| P_m(\alpha) + \frac{1 - P_m(\alpha)}{1 - z_1 - z_2} \right| \leqq 1. \tag{4.9}$$

For $z_1, z_2 \leqq 0$ this yields the inequality

$$\frac{2 - z_1 - z_1}{z_1 + z_2} \leqq P_m(\alpha) \leqq 1 \tag{4.10}$$

which is satisfied if $0 \leqq \alpha \leqq b$ (see Fig. 3.2). Since $\bar{a} \leqq \alpha \leqq \bar{b}$ and $\bar{a} > 0$, $\bar{b} \leqq b$ provided $\omega \geqq 1$ we find that (4.9) is satisfied for all negative z_1 and z_2. Furthermore, by virtue of (2.16) the method is first order consistent. □

Example 4.2. Next we consider the two-step backward differentiation formula as the generating formula, i.e. $k = 2$, $a_1 = \frac{4}{3}$, $a_2 = -\frac{1}{3}$, and the predictor formula $y^{(0)} = 2y_n - y_{n-1}$, i.e. $\hat{a}_1 = 2$, $\hat{a}_2 = -1$, to obtain the characteristic equation

$$\zeta^2 - \left\{ 2P_m(\alpha) + \frac{4[1 - P_m(\alpha)]}{3(1 - z_1 - z_2)} \right\} \zeta + \left\{ P_m(\alpha) + \frac{1 - P_m(\alpha)}{3(1 - z_1 - z_2)} \right\} = 0. \tag{4.11}$$

This equation has its roots on the unit disk if

$$-\frac{3(z_1 + z_2) - 8}{9(z_1 + z_2) - 4} \leqq P_m(\alpha) \leqq 1, \tag{4.12}$$

which is certainly satisfied for all negative z_1 and z_2 if $-\frac{1}{3} \leqq P_m(\alpha) \leqq 1$. From Fig. 3.2 and the discussion in the preceding example it follows that this inequality holds if $\omega \geqq 1$ and $D \leqq \frac{1}{3}$. Using Table 3.1 we can determine stable values for (m, S^*). The order of consistency equals 2 according to (2.16). We remark that D is not restricted if the predictor $y^{(0)} = y_n$ would be used. □

Generally, the root condition for the characteristic equation (4.7) is satisfied if the polynomial $P_m(\alpha)$ satisfies the condition

$$-D_1 \leqq P_m(\alpha) \leqq D_2, \qquad 0 < D_1 \leqq D_2 \leqq 1 \tag{4.13}$$

for $\bar{a} \leqq \alpha \leqq \bar{b}$. From Fig. 3.2 it is clear that this condition is satisfied if

$$D \leqq D_1, \tag{4.14a}$$

$$\alpha \geqq \tilde{a} \quad \text{where } P_m(\tilde{a}) = D_2. \tag{4.14b}$$

Substitution of (3.19) into (4.14a) and of (3.15) into (4.14b) yields

$$\text{(4.14'a)}\quad \omega \leqq \frac{T_{1/m}(1/D_1)+1}{T_{1/m}(1/D_1)-\cos(\pi/2m)}, \qquad T_{1/m}(x) := \cosh\left(\frac{1}{m}\operatorname{arccosh} x\right),$$

$$\text{(4.14'b)}\quad \frac{S+1}{(S/2+\omega)^2} \geqq \frac{1+\omega\cos(\pi/2m)-(\omega-1)T_{1/m}(D_2 T_m((\omega\cos(\pi/2m)+1)/\omega-1))}{\omega^2(\cos(\pi/2m)+1)}.$$

This expresses the stability conditions in terms of ω and m, or using (3.18) in terms of S^* and m.

For smooth problems the stability condition (4.14a) seems to be the most important one, because violating this condition means that instabilities are developed in the *low frequency components* of the solution (recall that these components correspond to eigenvalues α in the damping interval $[a, b]$). If (4.14a) is satisfied but (4.14b) is not, then instabilities are developed only in the *high frequency components* of the solution. Since we assumed the solution to be smooth these instabilities will not directly ruin the numerical solution. Moreover, the characteristic roots do not increase polynomially with z_1 and z_2 as the region of instability is entered, a situation which occurs in explicit methods (all coefficients in (4.7) are bounded as $z_1, z_2 \to -\infty$). Therefore, the effect of instabilities due to too large a time step can be removed by now and then performing a smoothing operation on the numerical solution y_n[2].

4.2. Determination of the iteration parameters m and S^*. The free iteration parameters (m, S^*) in the SC $(y^{(0)}, m, S^*)$ method should satisfy the stability condition (4.14'). In the (m, ω)-plane this condition determines a stability region

$$\text{(4.14'')}\qquad \underline{\omega}(m, S) \leqq \omega \leqq \bar{\omega}(m),$$

where $\bar{\omega}(m)$ follows from (4.14'a) and where the function $\underline{\omega}(m, S)$ is implicitly determined by (4.14'b). Using (3.18) the region (4.14'') can be transformed into a region in the (S^*, m)-plane:

$$\text{(4.15)}\qquad \underline{S}^*(m, S) \leqq S^* \leqq \bar{S}^*(m).$$

In Figure 4.1 two typical situations are illustrated. It should be noticed that for $D_2 < 1$ the function $\underline{S}^*$ depends on the problem parameter $S := b_0\tau^\nu\sigma$ (recall that σ is the spectral radius of $\partial f/\partial y$). From this figure we conclude that the number of iterations m in a stable SC method is bounded below by $\underline{m}$ where $\underline{m} = 1$ if $D_2 = 1$ and $\underline{m}$ is the solution of

$$\text{(4.16)}\qquad \underline{S}^*(m, S) = \bar{S}^*(m)$$

if $D_2 < 1$. Evidently, $\underline{m}$ can be found by solving the equation $\underline{\omega}(m, S) = \bar{\omega}(m)$. From the definition of $\underline{\omega}$ and $\bar{\omega}$ it follows that $\underline{m}$ is the solution of the equations (cf. (4.14'))

$$\text{(4.17)}\quad \begin{aligned} &\frac{\omega^2(S+1)(\cos(\pi/2m)+1)}{(\omega-1)(S/2+\omega)^2} = T_{1/m}\left(\frac{1}{D_1}\right) - T_{1/m}\left(\frac{D_2}{D_1}\right), \\ &\omega = \frac{T_{1/m}(1/D_1)+1}{T_{1/m}(1/D_1)-\cos(\pi/2m)}, \qquad D_2 < 1. \end{aligned}$$

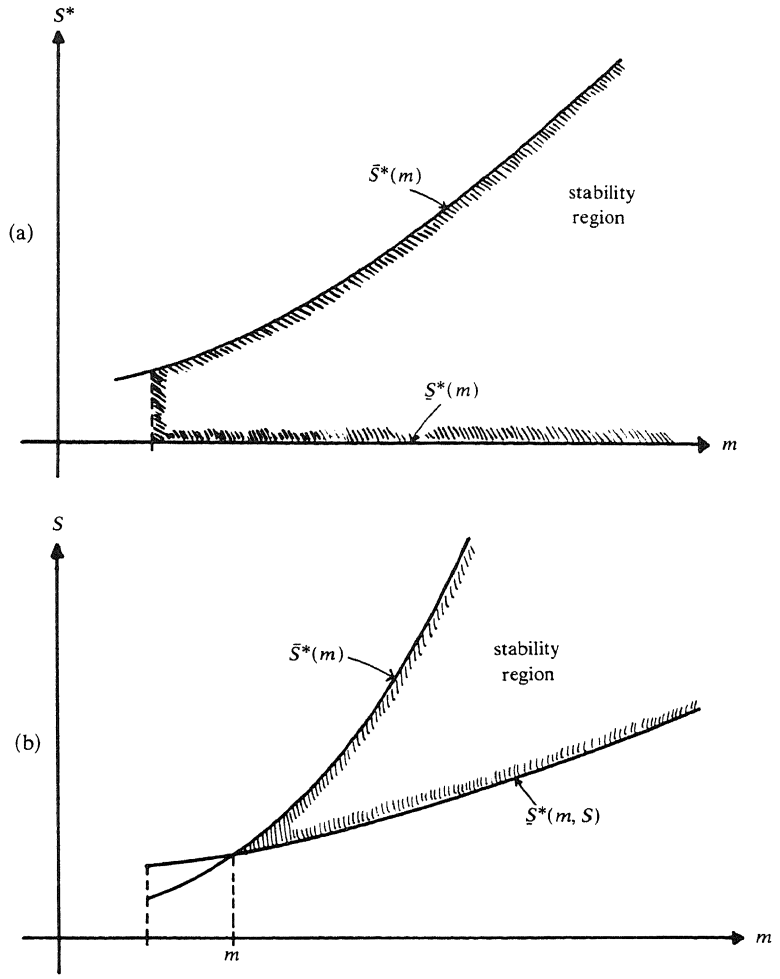

FIG. 4.1. a) *Stability region for* $D_2 = 1$, b) *Stability region for* $D_2 < 1$.

The solution $\underline{m}$ depends on the problem parameter S and the quantities (D_1, D_2) determined by the predictor-corrector pair $\{y^{(0)}, (1.3)\}$.

As an illustration, we give the values of $\underline{m}$ (when rounded to the first integer $\geqq$ the exact value of $\underline{m}$) as a function of S for the cases where (1.3) is the fourth-order backward differentiation formula (BDF_4) for first order ODEs and where $y^{(0)}$ corresponds to an extrapolation formula of order q (we shall write $y^{(0)} = y_q^{(0)}$). For $q = 0$ and $q = 1$ it turns out that $D_2 = 1$ (see also the Examples 4.1 and 4.2) so that $\underline{m} = 1$ for all S. For $q = 2$ and $q = 3$ we have respectively $(D_1, D_2) = (1/7, .495)$ and $(D_1, D_2) = (1/15, .199)$ which leads to the values listed in Table 4.1a.

The actual value of m should not be chosen larger than necessary for staying in the stability region (4.15). From Fig. 4.1 it follows that the pair

$$(m, S^*) = (\underline{m}, \bar{S}^*(\underline{m})) \tag{4.18}$$

is stable and at the same time optimal in the sense that the number of iterations is minimal. In Table 4.1b the values $\bar{S}^*(\underline{m})$ are listed for $y^{(0)} = y_q^{(0)}$, $q = 1, 2, 3$, together with the damping factor $D = D_1$ and the order $\tilde{p} = q + 1$ (cf. (2.16)).

TABLE 4.1a
Solution $\underline{m}$ of (4.17).

$y_2^{(0)}$		$y_3^{(0)}$	
S	$\underline{m}(S)$	S	$\underline{m}(S)$
(0, 6.6]	1	(0, 1.9]	1
(6.6, 47]	2	(1.9, 12.5]	2
(47, 198]	3	(12.5, 52]	3
(198, 587]	4	(52, 154]	4
(587, 1391]	5	(154, 360]	5
(1391, 2836]	6	(360, 732]	6
$S \gg 1$	$.82\sqrt[4]{S}$	$S \gg 1$	$1.17\sqrt[4]{S}$

TABLE 4.1b
Values of $\bar{S}^*(\underline{m})$, D and $\tilde{p}$.

	$m=1$	$m=2$	$m=3$	$m=4$	$m=5$	$m=6$	$m \gg 1$	D	$\tilde{p}$
$y_1^{(0)}$	2.96	33	157	486	1176	2425	$1.85m^4$	1/3	2
$y_2^{(0)}$	0.98	9.4	43	131	316	649	$0.49m^4$	1/7	3
$y_3^{(0)}$	0.48	4	18	54	129	264	$0.20m^4$	1/15	4

5. Numerical experiments. Before giving results obtained by the "optimal" SC method specified in § 4.2, we present in § 5.1 results obtained for various values of m and S in order to illustrate the effect of these iteration parameters. In § 5.2 we will compare the SC $(y_3^{(0)}, \underline{m}, \bar{S}^*(\underline{m}))$ method with the ADI method of Peaceman and Rachford [6] in nonlinear form:

$$(5.1) \qquad y^{(0)} = y_n + \tfrac{1}{2}\tau F(t_{n+1/2}, y^{(0)}, y_n), \qquad y_{n+1} = y^{(0)} + \tfrac{1}{2}\tau F(t_{n+1}, y^{(0)}, y_{n+1}).$$

It can be considered as a splitting method which "solves" the trapezoidal rule by one iteration.

The test problems are listed in Table 5.1. Initial conditions at $t=0$ and Dirichlet boundary conditions on the unit square $0 \leqq x_1, x_2 \leqq 1$ were taken from the exact solutions and the functions v were chosen such that the exact solutions are given by those listed in the table. These problems were semi-discretized using standard differences on a uniform grid with mesh spacing h.

TABLE 5.1
Test problems of the form $U_t = d\Delta(U^i) + (U_{x_1})^j + (U_{x_2})^j + v$.

Problem	Solution	d	i	j	σ
I	$1+e^{-t}(x_1^2+x_2^2)$	1	1	0	$8h^{-2}$
II	$1+e^{-t}(x_1^2+x_2^2)$	$(1+t)^{-1}$	1	2	Gerschgorin estimate
III	$\frac{1}{2}(x_1+x_2)\sin 2\pi t$	$\frac{1}{2}(x_1+x_2)(1+t)^{-1}$	3	0	$\approx 24\,h^{-2}(1+t)^{-1}\sin^2 2\pi t$

The starting values at $t=-3h, -2h, -h, 0$ were taken from the exact solution and the splitting function F in the SC method is identical to that used in the ADI method. The Jacobian matrices $\partial F/\partial u$ and $\partial F/\partial v$ were derived by hand and updated at the beginning of each integration step.

In the tables of results the accuracy delivered by the various methods is measured by the number of correct significant digits obtained in the end point $t=1$, i.e.

$$sd := -\log_{10}(\text{maximum absolute error at } t=1). \tag{5.2}$$

5.1. A linear problem. The matrix $Z=b_0\tau\partial f/\partial y$ corresponding to the semi-discrete form of problem I has eigenfunctions of the form

$$e_{k,l}=\sin(ik\pi h)\sin(jl\pi h), \tag{5.3}$$

where (ih, jh) with $i, j=1, 2, \cdots, h^{-1}-1$ refer to the grid points and k, l assume integer values. The corresponding eigenvalues are given by

$$z_{k,l}=[-4+2\cos(k\pi h)+2\cos(l\pi h)]b_0\tau h^{-2}. \tag{5.4}$$

The initial error ε_0 may be considered as an odd grid function (ε_0 vanishes on the boundary) and can therefore be expressed in terms of the eigenfunctions (5.3). Hence, the SC $(y_q^{(0)}, m, S^*)$ method damps all frequencies which satisfy $z_{k,l}\in[-2S^*, 0]$ by at least a factor D (see Fig. 3.3). If $h^2/\tau \ll 1$ it follows from (5.4) that the damped frequencies are those for which

$$k^2+l^2 \leqq \frac{2S^*}{b_0\tau\pi^2} \approx .42\frac{S^*}{\tau}. \tag{5.5}$$

Thus, for given values of τ and S^* those frequencies with $k, l \leqq l_{\max} := \sqrt{.21S^*/\tau}$ are damped (we observe that $l_{\max}$ does not depend on h).

In the following tables the values of sd/D together with the corresponding values of $l_{\max}$ are listed for a few values of m and S^*. Results obtained for (m, S^*)-values outside the stability region are indicated by an asterisk (see also Fig. 4.1). These results show that:

(i) There seems to be optimal S^*-value which for this problem corresponds to a damping factor $D\in(.15, .25)$ if $m=2$ and to $D\in(.01, .05)$ if $m=4$.

(ii) Unstable integration does not ruin the solution in a few steps.

TABLE 5.2a
Problem I; $q=1$, $\tau=h=\frac{1}{10}$.

m \ S^*	0	10	20	40
2	2.3/0	3.0/.15	2.8/.25	2.7*/.37
4	2.8/0	4.3/.01	3.7/.02	3.3/.05
$l_{\max}$	0	4	6	8

TABLE 5.2b
Problem I; $q=1$, $\tau=2h=\frac{1}{10}$.

m \ S^*	0	10	20	40
2	1.4/0	2.5/.15	2.8/.25	2.7*/.37
4	1.8/0	3.0/.01	3.5/.02	3.3/.05
$l_{\max}$	0	4	6	8

TABLE 5.3a
Problem I; $q=3$, $\tau=h=\frac{1}{10}$, $S=38.4$.

m \ S^*	0	4	10	20	40
2	3.9*/0	4.6*/.07	4.8*/.15	3.8*/.25	2.7*/.37
4	4.4*/0	5.7*/.002	6.1/.01	5.8/.02	5.3/.05
$l_{\max}$	0	2	4	6	8

TABLE 5.3b
Problem I; $q=3$, $\tau=2h=\frac{1}{10}$, $S=153.6$.

m \ S^*	0	4	10	40	50
2	3.2*/0	3.5*/.07	4.0*/.15	2.7*/.37	2.5*/.42
4	3.4*/0	4.2*/.002	4.5*/.01	5.3*/.05	5.2*/.07
$l_{\max}$	0	2	4	8	10

The first observation is strongly problem-dependent. The optimal value of S^* is determined by the right compromise of the number of dominant eigenfunctions to be damped and the damping factor D (note that the number of damped eigenfunctions corresponding to the highest accuracy varies from 4 until 8).

The second observation implies that an underestimation of σ, and consequently an unstable pair (m, S^*), is not too serious. However, a large number of steps with an unstable combination of the parameters m and S^* finally leads to an unstable result as is illustrated in Table 5.4, where the sd-values are listed obtained by the SC $(y_3^{(0)}, 4, S^*)$ method in $0 \leqq t \leqq 8$. Theoretically, $S^*=40$ and $S^*=80$ should be unstable and $S^*=50$ should be stable. The results confirm the stability theory, but they also show that the instability is of a rather mild character.

TABLE 5.4
Stability test. Problem I; $\tau=\frac{1}{10}$, $h=\frac{1}{20}$, $S=153.6$.

Method	$t=1$	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$	$t=7$	$t=8$
SC $(y_3^{(0)}, 4, 40)$	5.3	5.7	6.1	6.1	5.7	5.5	4.9	4.7
SC $(y_3^{(0)}, 4, 50)$	5.2	5.6	6.0	6.5	6.9	7.3	7.7	8.0
SC $(y_3^{(0)}, 4, 80)$	5.0	5.2	4.8	4.1	3.5	2.9	2.3	1.6

5.2. Comparison with the ADI method. In Tables 5.5, 5.6 and 5.7 the $sd/\tilde{m}$-values are listed obtained by the SC $(y_3^{(0)}, \underline{m}, \bar{S}^*(\underline{m}))$ method and by the ADI method (5.1) when applied to the problems I, II and III. Here, $\tilde{m}=1$ for the ADI method and $\tilde{m}$ is the average number of iterations per step for the SC method (recall that $\underline{m}$ depends on S and may vary during the integration process).

In order to compare the efficiency of the two methods one should take into account the computational effort per step. In addition to the iterations to be performed, both methods require the evaluation of the Jacobian matrix used in the Newton iteration

process and the LU-decompositions of the tridiagonal matrices. Thus, in the interpretation of the results listed in Tables 5.5, 5.6 and 5.7 the value of $\tilde{m}$ gives not more than an indication of the computational effort.

TABLE 5.5
sd/$\tilde{m}$ values obtained for problem I; $h = \frac{1}{24}$.

Method	$\tau = 1/10$	$\tau = 1/20$	$\tau = 1/40$	$\tau = 1/80$	$\tilde{p}_{\text{eff}}$
ADI	2.6/1	3.2/1	3.9/1	4.5/1	2
SC	5.1/5	6.3/4	7.4/4	8.6/3	4

TABLE 5.6
sd/$\tilde{m}$ values obtained for problem II; $h = \frac{1}{24}$.

Method	$\tau = 1/5$	$\tau = 1/10$	$\tau = 1/20$	$\tau = 1/40$	$\tau = 1/80$	$\tilde{p}_{\text{eff}}$
ADI	*	*	2.0/1	3.6/1	4.3/1	2
SC	3.8/5.2	4.9/4.4	6.1/4	7.3/3.1	8.5/3	3.9

TABLE 5.7
sd/$\tilde{m}$-values obtained for problem III; $h = \frac{1}{24}$.

Method	$\tau = 1/20$	$\tau = 1/40$	$\tau = 1/80$	$\tau = 1/160$	$\tilde{p}_{\text{eff}}$
ADI	*	*	2.1/1	2.7/1	2
SC	3.0/4.3	4.5/3.4	6.0/2.8	7.4/2.4	4.7

In all problems the superiority of the SC method, particularly in the high accuracy region, is evident. This is of course due to its fourth order behavior which is already demonstrated for relatively large steps (we have listed the effective order $\tilde{p}_{\text{eff}} := (sd(\tau) - sd(2\tau))/\log_{10}(2)$ in order to illustrate the order behavior). In this connection, we remark that the fourth order successive-correction scheme analysed in [3] does not show its fourth order unless $\tau\sigma$ is rather small ($\tau\sigma \leqq 200$). This scheme can be fitted into the framework of the SC methods by putting $y^{(0)} = y_n$ and $S^* = 0$ (cf. [2]).

In problems II and III the nonlinear relations (3.32b) were solved by performing just one Newton iteration. For too large an integration step both methods did not work (indicated by an asterisk). The SC method, however, is more robust for larger steps because the Jacobian matrix is evaluated at $(t_{n+1}, y_3^{(0)})$, whereas the ADI method has to use Jacobian matrices evaluated at (t_n, y_n).

Problem II illustrates that the SC method, although designed for problems possessing Jacobian matrices with a negative spectrum, can handle problems with "imaginary noise" (the derivatives U_{x_1} and U_{x_2} introduce imaginary parts into the eigenvalues of the matrix Z).

Problem III is rather nonlinear and has a rapidly changing spectral radius σ. The SC method adapts the iteration parameters m and S^* to the value of $S = 12\tau\sigma/25$, so that this problem tests the SC method when applied with rapidly changing values for m and S^*.

Acknowledgment. The author is grateful to Mr. H. B. de Vries for carefully reading the manuscript, and for carrying out the numerical experiments.

REFERENCES

[1] A. R. GOURLAY, *Hopscotch, a fast second order partial differential equation solver*, J. Inst. Math. Applics., 6 (1970), pp. 375–390.

[2] P. J. VAN DER HOUWEN AND H. B. DE VRIES, *A fourth order* ADI *method for semidiscrete parabolic equations*, J. Comp. Appl. Math., 9 (1983), pp. 41–63.

[3] P. J. VAN DER HOUWEN, *Multistep splitting methods of higher order for initial value problems*, this Journal, 17 (1980), pp. 410–427.

[4] P. J. VAN DER HOUWEN AND J. G. VERWER, *One-step splitting methods for semi-discrete parabolic equations*, Computing, 22 (1979), pp. 291–309.

[5] P. J. VAN DER HOUWEN AND B. P. SOMMEIJER, *A special class of multistep Runge-Kutta methods with extended real stability interval*, IMA J. Numer. Anal., 2 (1982), pp. 183–209.

[6] D. W. PEACEMAN AND H. H. RACHFORD JR., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math., 3 (1955), pp. 28–41.

[7] H. J. STETTER, *The defect correction principle and discretization methods*, Numer. Math., 29 (1978), pp. 425–443.

[8] J. G. VERWER, *The application of iterated defect correction to the* LOD *method for parabolic equations*, BIT, 19 (1979), pp. 384–394.

[9] ———, *On iterated defect correction and the* LOD *method for parabolic equations*, in Syllabus MCS 44, Mathematical Centrum, Amsterdam, 1980.

[10] E. L. WACHSPRESS, CURE, *A generalized two-space-dimension multigroup coding for the* IBM-704, Report KAPL-1724, Knolls Atomic Power Lab., Schenectady, NY, 1957.

[11] N. N. YANENKO, *The Method of Fractional Steps*, Springer, Berlin-Heidelberg-New York, 1971.

[12] D. M. YOUNG, *Iterative Solution of Large Linear Systems*, Academic Press, New York-London, 1971.