stichting
mathematisch
centrum

MC

AFDELING NUMERIEKE WISKUNDE
(DEPARTMENT OF NUMERICAL MATHEMATICS)

NW 128/82 MEI

P.J. VAN DER HOUWEN & H.B. DE VRIES

# A FOURTH ORDER ADI METHOD FOR SEMIDISCRETE PARABOLIC EQUATIONS

Preprint

kruislaan 413 1098 SJ amsterdam

*Printed at the Mathematical Centre, 413 Kruislaan, Amsterdam.*

*The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).*

---

1980 Mathematics subject classification: 65M10, 65M20

---

# A fourth order ADI method for semidiscrete parabolic equations*)

by

P.J. van der Houwen & H.B. de Vries

ABSTRACT

A fourth order fourstep ADI method is described for solving the systems of ordinary differential equations which are obtained when a (nonlinear) parabolic initial-boundary value problem in two dimensions is semi-discretized. The local time-discretization error and the stability conditions are derived. By numerical experiments it is demonstrated that the (asymptotic) fourth order behaviour does not degenerate if the time step increases to relatively large values. Also a comparison is made with the classical ADI method of Peaceman and Rachford showing the superiority of the fourth order method in the higher accuracy region, particularly in nonlinear problems.

KEY WORDS & PHRASES: *Numerical analysis, parabolic equations, method of lines, ADI methods*

---

*) This report will be submitted for publication elsewhere.

## 1. INTRODUCTION

In a few recent papers [1,4,5] *multistep splitting methods* were analysed for solving two-dimensional parabolic initial-boundary value problems. By using the method of lines the problem is first reduced to a (usually very large) system of ODE's of the explicit form

$$(1.1) \qquad \frac{dy}{dt} = f(t,y), \quad y(t_0) = y_0,$$

and then an implicit linear multistep method (LMM) is applied to obtain at each integration step an *implicit* equation for the numerical solution $y_{n+1}$ at $t_{n+1}$:

$$(1.2) \qquad y_{n+1} - b_0\tau f(t_{n+1},y_{n+1}) = \sum_{\ell=1}^{k} [a_\ell y_{n+1-\ell} + b_\ell \tau f(t_{n+1-\ell},y_{n+1-\ell})].$$

Here, $\tau$ is the integration step and $\{a_\ell,b_\ell\}$ are coefficients defining the LMM. The papers mentioned above describe methods for approximating the solution of (1.2) by using a splitting of the right-hand side function $f(t,y)$, e.g. $f(t,y) = f_1(t,y) + f_2(t,y)$ where $f_1$ and $f_2$ have "simply structured" Jacobian matrices. More generally, one may use splitting functions $F(t,u,v)$ such that $F(t,y,y) = f(t,y)$ and $\partial F/\partial u$, $\partial F/\partial v$ are again "simply structured".

The method analysed and tested in this paper is a special case of a class of methods described in [5]. These methods explicitly use the information that (1.1) originates from a parabolic problem so that the eigenvalues of $\partial f/\partial y$ will be located in a long narrow strip along the *negative axis*. At the same time, this is also a restriction in the applicability of these methods.

An outline of the construction of the method is as follows. The system of equations (1.2) is solved by a (nonlinear) splitting method (e.g. ADI) and this iteration process is accelerated by using *Chebyshev polynomials* [10,p.344]. The relaxation parameter in the splitting method is chosen such that the approximation obtained for the solution of (1.2) has a maximal order of accuracy as $\tau \to 0$ for a given LMM and a given initial approximation used for starting the iteration process. The iteration parameters in the

Chebyshev polynomials are chosen such that the *lower frequencies* in the initial error are strongly *damped*. As a consequence we obtain a rather fast convergence to the solution of (1.2) in problems where the low frequencies are *dominant* in the solution of the initial-boundary value problem. For details of the construction we will refer to [5].

In section 2 the specification of the method will be given, its local error will be derived and the characteristic equation for a class of model problems will be analysed. In section 3 the results of section 2 will be applied to the case where the LMM (1.2) is identified with the fourth order backward differentiation formula (BDF4) and where the initial approximation used in the iteration process is a "smoothed" extrapolation formula. It will be shown that the resulting fourth order, fourstep splitting method has a real stability boundary bounded below by $cm^4$, m being the number of iterations and c some constant (numerical verification reveals that $c \cong 4$). Finally, in section 4 the method proposed in this paper is compared with the classical ADI method of PEACEMAN and RACHFORD [8] showing the superiority of the present method (particularly in nonlinear problems) if higher accuracies are desired. This favourable behaviour is due to the property that the fourth order method really behaves as a fourth order method even for relatively large integration steps (this behaviour is not shared by the high order splitting methods analysed in [4], the order of which degenerates for larger values of the time step).

## 2. MULTISTEP SPLITTING METHODS FOR NONLINEAR EQUATIONS

### 2.1. Specification of the method

The method constructed in [5] and more fully analysed in this paper is defined by

$$y^{(0)} = \text{some predictor formula for } y_{n+1}, \tag{2.1a}$$

$$y^{(j+1)} = (\mu_j-\lambda_j)y^{(j)} + (1-\mu_j)y^{(j-1)} + \lambda_j y^{**}, \quad j = 0,1,\ldots,m-1, \tag{2.1b}$$

$$y_{n+1} = y^{(m)},$$

where $y^{**}$ is determined by the two equations

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

(2.2) ,

$$\omega y^{*} + (1-\omega)y^{(j)} - b_0\tau F(t_{n+1},y^{(j)},y^{*}) = \Sigma$$

with

$$\Sigma = \sum_{\ell=1}^{k} [a_\ell y_{n+1-\ell} + b_\ell \tau f(t_{n+1-\ell},y_{n+1-\ell})] \tag{2.3}$$

and $F(t,u,v)$ a splitting function such that $F(t,y,y) \equiv f(t,y)$.

The coefficients $\mu_j$ and $\lambda_j$ are defined by

$$\mu_0 = 1, \quad \mu_j = 2\omega_0 \frac{T_j(\omega_0)}{T_{j+1}(\omega_0)}, \quad \omega_0 = \frac{b+a}{b-a},$$

$$\lambda_0 = \frac{2}{b+a}, \quad \lambda_j = \frac{2\mu_j}{b+a}, \quad j = 1,2,\ldots,m-1, \tag{2.4}$$

$$a = \frac{(2\omega-1)(2S^{*}+1)}{(S^{*}+\omega)^2}, \quad b = \frac{2\omega-1}{\omega},$$

where $\omega$ is the largest real solution of the equation

$$(2S^{*}+1)(\cos\tfrac{\pi}{2m}+1)\omega^2 = [2 + \omega(\cos\tfrac{\pi}{2m}-1)](S^{*}+\omega)^2 \tag{2.5}$$

and $S^{*}$ is a free parameter to be used for maximizing the stability interval. The method is completely defined if we specify $y^{(0)}$ by some predictor formula, $\Sigma$ by choosing an appropriate LMM$\{a_\ell,b_\ell\}$, and $(m,S^{*})$ on basis of stability considerations.

We remark that for a class of model problems (cf. section 2.2) the interval $[a,b]$ corresponds to the eigenvalue interval of those eigenvectors which are strongly damped by the Chebyshev iteration. Furthermore, as we will see in the next section, the relaxation parameter $\omega$ defined by (2.5) decreases the magnitude of the local error as $\tau \to 0$.

## 2.2. The local error

Let $\eta$ denote the solution of (1.2) and define the iteration error $\varepsilon_j = y^{(j)} - \eta$. Furthermore, we write $\varepsilon^{*} = y^{*} - \eta$ and $\varepsilon^{**} = y^{**} - \eta$. The

local error is given by

$$y^{(m)} - y(t_{n+1}) = \varepsilon_m + \eta - y(t_{n+1}), \tag{2.6}$$

where we assume that $y_j = y(t_j)$ for $j \le n$ (localizing assumption). Thus, when we are given the local error of the generating multistep method (1.2) and if we can find an estimate for the iteration error, then we have found an estimate for the local error of the splitting method (2.1)-(2.5).

In order to derive a recurrence relation for the iteration error as $\tau \to 0$ we first deduce from (2.2) the relations

$$\begin{aligned} [\omega - Z_2]\varepsilon^* &= [\omega - 1 + Z_1]\varepsilon_j + \tau O(\|\varepsilon_j\|^2 + \|\varepsilon_j\|\|\varepsilon^*\| + \|\varepsilon^*\|^2) \\ [\omega - Z_1]\varepsilon^{**} &= [\omega - 1 + Z_2]\varepsilon^* + \tau O(\|\varepsilon^*\|^2 + \|\varepsilon^*\|\|\varepsilon^{**}\| + \|\varepsilon^{**}\|^2) \end{aligned} \tag{2.7}$$

as $\tau \to 0$. Here $Z_1$ and $Z_2$ are defined by $b_0\tau\partial F/\partial u$ and $b_0\tau\partial F/\partial v$ where the derivatives are evaluated at $(t_{n+1}, \eta, \eta)$.

Let us assume that all iteration errors $\varepsilon_j$ are $O(\tau^s)$ as $\tau \to 0$ for $s \ge 0$. Then it follows from (2.7) that

$$\varepsilon^{**} = [\omega - Z_1]^{-1}[\omega - 1 + Z_2][\omega - Z_2]^{-1}[\omega - 1 + Z_1]\varepsilon_j + O(\tau^{2s+1})$$

as $\tau \to 0$. From (2.1) we derive the recurrence relation

$$\varepsilon_{j+1} = [\mu_j - \lambda_j A]\varepsilon_j + (1 - \mu_j)\varepsilon_{j-1} + O(\tau^{2s+1}) \text{ as } \tau \to 0, \tag{2.1'}$$

where the matrix A is defined by

$$A = I - [\omega - Z_1]^{-1}[\omega - 1 + Z_2][\omega - Z_2]^{-1}[\omega - 1 + Z_1]. \tag{2.8}$$

Let us write

$$\varepsilon_j = P_j(A)\varepsilon_0 + c_j\tau^{2s+1} \quad \text{as } \tau \to 0, \tag{2.9}$$

where $P_j$ is the polynomial which satisfies the recurrence relation

$$(2.10) \qquad P_0 = 1,\ P_{j+1} = (\mu_j - \lambda_j A)P_j + (1-\mu_j)P_{j-1}, \qquad j = 0,1,\ldots$$

and $c_j$ yet to be determined. Substitution of (2.9) into (2.1') reveals that the representation (2.9) is correct provided that $c_j$ satisfies the recurrence relation

$$c_0 = 0, \quad c_{j+1} = (\mu_j - \lambda_j A)c_j + (1-\mu_j)c_{j-1} + O(1) \qquad \text{as } \tau \to 0.$$

Evidently, for $j \le m$ the coefficients $c_j$ are bounded as $\tau \to 0$, hence

$$(2.9') \qquad \varepsilon_j = P_j(A)\varepsilon_0 + O(\tau^{2s+1}) \qquad \text{as } \tau \to 0.$$

Finally, we have to determine s, that is the order in $\tau$ of $\varepsilon_j$ as $\tau \to 0$. Suppose that the linear multistep method (1.2) and the predictor formula for $y^{(0)}$ have orders of accuracy p and q, respectively. Then

$$\varepsilon_0 = y^{(0)} - \eta = y^{(0)} - y(t_{n+1}) + y(t_{n+1}) - \eta = O(\tau^{q+1} + \tau^{p+1}).$$

From (2.9') it follows that apparently $\varepsilon_j = O(\tau^{q+1}+\tau^{p+1})$ provided that $P_j(A)$ is bounded as $\tau \to 0$ for $j = 1,2,\ldots,m$. Thus $s = \min\{p+1,q+1\}$. Since it follows from (2.4) that $P_j(A)$ can be identified with the *shifted Chebyshev polynomial*

$$P_j(A) = \frac{T_j(w_0+w_1A)}{T_j(w_0)}, \qquad w_1 = \frac{-2}{b-a},$$

so that $P_j(A)$ is bounded as $\tau \to 0$ for all finite j and $S^*$, we may summarize the results in the form of the following theorem.

THEOREM 2.1. *The local error of the method* (2.1) - (2.5) *is given by*

$$(2.11) \qquad y^{(m)} - y(t_{n+1}) = [I-P_m(A)](\eta - y(t_{n+1})) + P_m(A)(y^{(0)} - y(t_{n+1}))$$
$$+ O(\tau^{2p+3}+\tau^{2q+3}). \qquad \square$$

In practice, we often have p > q so that the term originating from the predictor formula will largely determine the magnitude of the local error. It is therefore of interest to estimate the norm of the amplification matrix $P_m(A)$ as $\tau \to 0$. In [5] an estimate is given which assumes the form

$$(2.12) \qquad \|P_m(A)\| \le D[T_m(w_0+w_1\alpha_0) + O(\frac{\tau}{D^{1/m}})] \quad \text{as } \tau \to 0,$$

provided that $S^* \neq 0$ and where

$$(2.13) \qquad D = T_m^{-1}(w_0), \quad \alpha_0 = \frac{2\omega-1}{\omega^2} .$$

From (2.4) and (2.5) it follows that $T_m(w_0+w_1\alpha_0) = 0$, i.e. $P_m(\alpha_0) = 0$, hence by assuming that $D = O(\tau^r)$ as $\tau \to 0$ it follows from (2.12) and theorem 2.1 that the order of consistency of the method (2.1) - (2.5) is given by

$$(2.14) \qquad \tilde{p} = \text{Min}\{p, q+r+1-\frac{r}{m}\}, \quad S^* \neq 0.$$

In our experiments we used fixed $S^*$ values for given m and therefore by virtue of (2.5) fixed values for $\omega$. Since D is related to $\omega$ by the formula

$$(2.13') \qquad D = T_m^{-1}\left(\frac{1+\omega \cos\frac{\pi}{2m}}{\omega-1}\right),$$

we conclude that D does not depend on $\tau$, i.e. r = 0, so that

$$(2.14') \qquad \tilde{p} = \min(p, q+1), \quad S^* \neq 0.$$

Thus, even for zero order predictor formulas (q=0) the method (2.1) - (2.5) is still a consistent integration method provided (of course) that the generating LMM is consistent ($p \ge 1$).

Apart from the estimate (2.12) it is of interest how the operator $P_m(A)$ damps the lower and higher frequencies in the predictor error $y^{(0)} - y(t_{n+1})$. In [5] a result is given for the following class of *model problems*

(i) The matrices $Z_1 = b_0\tau \frac{\partial F}{\partial u}(t_{n+1},\eta,\eta)$ and $Z_2 = b_0\tau \frac{\partial F}{\partial v}(t_{n+1},\eta,\eta)$

have a *common eigensystem* $\{e_i\}$

(ii) The eigenvalues $z_j^{(i)}$ of $Z_j$, $j = 1,2$, are *negative*.

THEOREM 2.2. *Let the local errors of* $y^{(0)}$ *and* $y^{(m)}$ *have the eigenvector expansions*

$$y^{(0)} - y(t_{n+1}) = \sum_i c_0^{(i)} e_i, \quad y^{(m)} - y(t_{n+1}) = \sum_i c^{(i)} e_i$$

*then*

$$D|c_0^{(i)}| \quad \textit{for} \quad -S^* \le z_1^{(i)}, z_2^{(i)} < 0$$

$$|c^{(i)}| \le$$

$$|c_0^{(i)}| \quad \textit{otherwise} \qquad \square$$

We will choose the splitting function $F(t,u,v)$ such that the eigenvectors $e_i$ of low frequency correspond to eigenvalues $z_j^{(i)}$, $j = 1,2$, on the right end of the eigenvalue interval (e.g. if F corresponds to ADI splitting). Then, theorem 2.2 implies that the *low frequencies* in the local error of the predictor formula are *damped* by a *factor* D. Thus, if the problem is smooth so that no high frequencies are involved and if the LMM and the predictor formula themselves do not introduce high frequencies, we may expect a fast convergence to the solution of the LMM (1.2).

## 2.3. The characteristic equation

Here, we confine our considerations to the class of model problems specified in the preceding section. For such problems it was derived in [5] that the variational equation of the method (2.1) - (2.5) is given by

$$\Delta y_{n+1} = P_m(A)\Delta y^{(0)} + [1-P_m(A)][1-Z_1-Z_2]^{-1}\,\Delta\Sigma, \tag{2.15}$$

where $\Delta y^{(0)}$ denotes a perturbation of $y^{(0)}$, and $\Delta y_{n+1}$ and $\Delta\Sigma$ denote perturbations caused by perturbations of $y_n, y_{n-1}, \ldots, y_{n+1-k}$.

Let us assume that $\Delta y^{(0)}$ can also be expressed in terms of the perturbations $\Delta y_n, \ldots, \Delta y_{n+1-k}$, say

$$\Delta y^{(0)} = \sum_{\ell=1}^{k} c_\ell(Z_1,Z_2)\Delta y_{n+1-\ell}. \tag{2.16}$$

Then using the definition of $\Sigma$ in (2.3) we find the characteristic equation

$$(2.17)\qquad \zeta^k = \sum_{\ell=1}^{k} \{P_m(\alpha)c_\ell(z_1,z_2) + \frac{1-P_m(\alpha)}{1-z_1-z_2}[a_\ell + \frac{b_\ell}{b_0}(z_1+z_2)]\}\zeta^{k-\ell},$$

where $\alpha$ denotes an eigenvalue of A, i.e.

$$(2.18)\qquad \alpha = (2\omega-1)\frac{1-z_1-z_2}{(\omega-z_1)(\omega-z_2)},$$

with $z_1$ and $z_2$ assuming values in the eigenvalue intervals of $Z_1$ and $Z_2$.

The method (2.1) - (2.5) will be called *stable* if (2.17) has its roots on the unit disk for all $z_1$ and $z_2$ in the eigenvalue intervals of $Z_1$ and $Z_2$, respectively. This condition will be called the *root condition*. If the functions $c_\ell$ are constant and $b_\ell = 0, \ell > 0$, this root condition usually leads to a condition on $P_m(\alpha)$ of the type

$$(2.19)\qquad -1 \le -D_1 \le P_m(\alpha(z_1,z_2)) \le D_2 \le 1,$$

where $D_1$ and $D_2$ are also constant (see example 2.1).
In the following stability theorem $T_{1/m}(\cdot)$ means $\cosh(\frac{\operatorname{arccosh}(\cdot)}{m})$.

THEOREM 2.3. *Let $\sigma$ be the spectral radius of $\partial f/\partial y$ at $(t_{n+1},\eta)$ and let the root condition be satisfied if $P_m$ satisfies (2.19) for all $z_1$ and $z_2$ in the eigenvalue intervals of $Z_1$ and $Z_2$. Then the method (2.1) - (2.5) is stable if*

$$(2.20a)\qquad \omega \le \tilde{\omega},\ \tilde{\omega} = \frac{T_{1/m}(\frac{1}{\tilde{D}})+1}{T_{1/m}(\frac{1}{\tilde{D}})-\cos\frac{\pi}{2m}},\ \tilde{D} = \min\{D_1,D_2\},$$

$$(2.20b)\qquad \tau \le \frac{\beta}{\sigma},\ \beta = \frac{\omega(1+\sqrt{1-\tilde{a}})-1}{b_0(1-\sqrt{1-\tilde{a}})},$$

*where $\tilde{a}$ is the point where $P_m(\alpha)$ assumes the value $D_2$, i.e.*

$$(2.21)\qquad \tilde{a} = \begin{cases} 0 & \text{if } D_2 = 1 \\ \dfrac{2\omega-1}{\omega^2(\cos\frac{\pi}{2m}+1)}\left[1+\omega\cos\frac{\pi}{2m} - (\omega-1)\,T_{1/m}(\frac{D_2}{D})\right] & \text{if } D_2 < 1 \end{cases}$$

*with* D *given by* (2.13').

PROOF. It may be helpful to consider the behaviour of $P_m(\alpha)$ as a function of $\alpha$ (see figure 2.1). $P_m(\alpha)$ assumes the value 1 at $\alpha = 0$ and it has a zero at $\alpha = \alpha_0$ where $\alpha_0$ is defined in (2.13). Furthermore, $P_m(\alpha)$ assumes a minimal maximum norm of magnitude D defined in (2.13) over the interval [a,b]. We observe that the eigenvalue interval of the matrix A is given by (cf. (2.4) and (2.18))

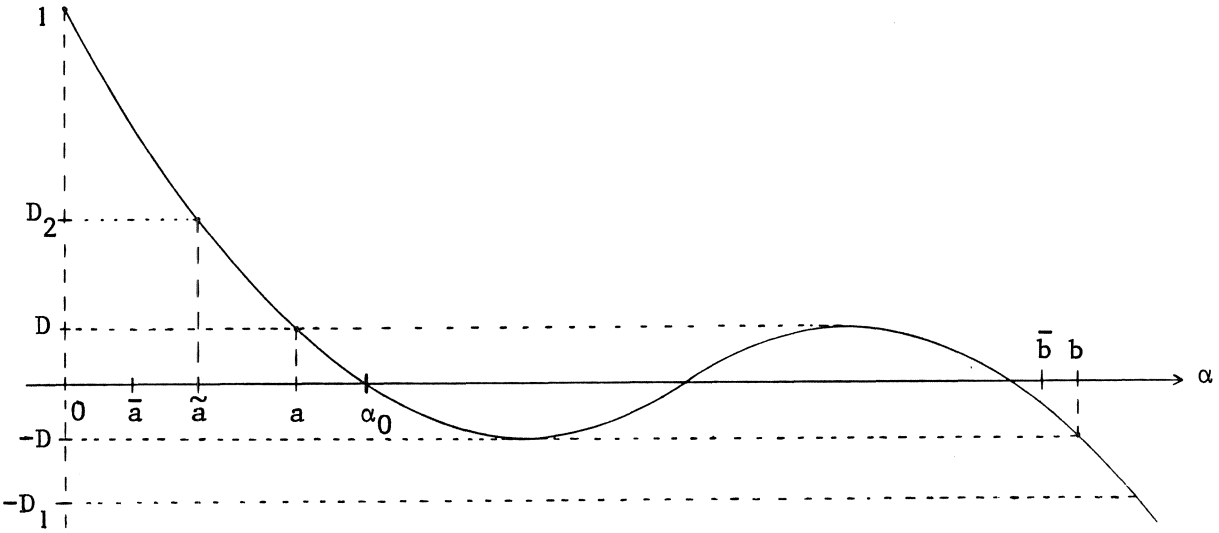


Figure 2.1. The polynomial $P_m(\alpha)$ for m = 3

$$(2.22) \qquad [\bar{a},\bar{b}],\ \bar{a} = (2\omega-1)\frac{2S+1}{(S+\omega)^2},\quad \bar{b} = \frac{2\omega-1}{\omega}\frac{S+1}{S+\omega},$$

where $S = b_0\tau\sigma$.

It is now evident from (2.19) that $D \le \tilde{D}$, hence from (2.13') it easily follows that $\omega$ should satisfy (2.20a). Thus, (2.20a) implies that $P_m(\alpha) \ge -D_1$ for all eigenvalues $\alpha \in [\bar{a},\bar{b}]$.

The condition $P_m(\alpha) \le D_2$ for all $\alpha \in [\bar{a},\bar{b}]$ is satisfied if $\bar{a} \ge \tilde{a}$ where $\tilde{a}$ is the (first) point where $P_m(\alpha) = D_2$ (see figure 2.1), i.e. the point defined by (2.21). From (2.22) it follows that $\bar{a} \ge \tilde{a}$ if S satisfies the inequality

$$(2\omega-1)\frac{2S+1}{(S+\omega)^2} \ge \tilde{a}.$$

Replacing S by $b_0\tau\sigma$ leads to condition (2.20b). □

From this theorem it follows that (2.20a) presents an upperbound for $\omega$ and by (2.5) an upperbound for $S^*$. This means that only a limited number of low frequencies in the local error of $y^{(0)}$ can be strongly damped (cf. theorem 2.2). Of course, by decreasing $\tau$ the number of strongly damped eigenvectors can be increased.

The condition (2.20b) on the integration step $\tau$ is illustrated in the following subsections.

2.3.1. Stability boundaries for the method of successive corrections

Suppose we choose $S^* = 0$, then $\omega = \lambda_j = \mu_j = 1$ and the method reduces to the *method of successive corrections* analysed in [4]. For this method we have $D = 0$ so that the equation for $\tilde{a}$ reduces to

$$P_m(\tilde{a}) = (1-\tilde{a})^m = D_2.$$

Solving this equation and substitution into (2.20b) yields

$$(2.20'b) \qquad \tau \leq \frac{D_2^{1/2m}}{b_0\sigma(1-D_2^{1/2m})} = \frac{2m}{b_0\sigma \ln(1/D_2)}\left(1+O\left(\frac{1}{m^2}\right)\right) \quad \text{as } m \to \infty.$$

Let us in particular consider the BDF4 formula defined by (cf. [7, p.242])

$$(2.23) \qquad b_0 = \frac{12}{25}, \quad \Sigma = \frac{1}{25}[48y_n - 36y_{n-1} + 16y_{n-2} - 3y_{n-3}],$$

and let the predictor $y^{(0)}$ be defined by extrapolation. In order to satisfy the root condition we obtain the bounds for $P_m(\alpha)$ listed in table 2.1 (for a proof we refer to [6, Appendix]).

From (2.20'b) the *stability boundaries* $\beta = \beta(m)$ can now be derived. Evidently, they are infinitely large if $D_2 = 1$, that is in the case of zero-order and first order extrapolation. For higher order extrapolation predictors we have conditional stability. In table 2.1 the values of $\beta$ are listed for $m = 2$ and for $m >> 1$. From these values we may conclude that the method of successive corrections based on BDF4 is of less practical value if higher order predictors are used.

Table 2.1. Stability boundaries for the method of successive corrections based on BDF4 and extrapolation

| Predictor formula | q | $-D_1 \leq P_m(\alpha) \leq D_2$ | $\beta(m=2)$ | $\beta(m>>1)$ |
|---|---|---|---|---|
| $y^{(0)} = y_n$ | 0 | $-0.7493 \leq P_m(\alpha) \leq 1$ | $\infty$ | $\infty$ |
| $y^{(0)} = 2y_n - y_{n-1}$ | 1 | $-\frac{1}{3} \leq P_m(\alpha) \leq 1$ | $\infty$ | $\infty$ |
| $y^{(0)} = 3y_n - 3y_{n-1} + y_{n-2}$ | 2 | $-\frac{1}{7} \leq P_m(\alpha) \leq 0.4951$ | 11.0 | 6.0 m |
| $y^{(0)} = 4y_n - 6y_{n-1} + 4y_{n-2} - y_{n-3}$ | 3 | $-\frac{1}{15} \leq P_m(\alpha) \leq 0.1999$ | 4.2 | 2.6 m |

2.3.2. Optimal stability boundaries for the iterated BDF4

Again we consider the BDF4 with the second and third order extrapolation predictors listed in table 2.1, but now with the maximal value for $\omega$ allowed by condition (2.20a). In table 2.2 the corresponding $S^*$ and $\beta$ values are listed for various values of m.

Table 2.2. Stability boundaries for BDF4 combined with second and third order extrapolation as predictor formula

| | | m = 1 | m = 2 | m = 3 | m = 4 | m = 5 | m = 6 |
|---|---|---|---|---|---|---|---|
| q = 2 | $S^* \cong$ | .98 | 9.4 | 43 | 131 | 316 | 649 |
| | $\beta \cong$ | 6.9 | 49 | 206 | 612 | 1449 | 2954 |
| q = 3 | $S^* \cong$ | .44 | 4 | 18 | 54 | 129 | 264 |
| | $\beta \cong$ | 2 | 13 | 54 | 159 | 376 | 763 |

We notice that the stability boundary obtained for m = 2 and for optimal (maximal)$S^*$ is already considerably larger than that for m = 2 and $S^* = 0$.

2.3.3. Behaviour of the stability boundary for large values of m

In order to get an impression of the stability boundary for large values of m we prove the following corollary of theorem 2.3.

COROLLARY 2.1. *If the conditions of theorem 2.3 are satisfied then*

$$\beta = \frac{32\omega^2 m^2}{b_0[16m^2-\pi^2\omega-4\omega d_2^2]}\,[1+0(\tfrac{1}{m^2})] \qquad m \to \infty \tag{2.24a}$$

*as*

$$\beta \le \frac{128}{b_0[\pi^2+4d_1^2][d_1^2-d_2^2]}\, m^4[1+0(\tfrac{1}{m^2})] \qquad \omega = 0(m^2) \tag{2.24b}$$

*where*

$$d_1 = \operatorname{arccosh}\frac{1}{\tilde{D}}\,, \quad d_2 = \operatorname{arccosh}\frac{D_2}{D}\,,$$

*with* $\tilde{D} = \min\{D_1, D_2\}$ *and* $D$ *given by* (2.13').

PROOF. From (2.21) it follows that for $m \gg 1$

$$\tilde{a} = \frac{2\omega-1}{\omega^2(c+1)}\,[1+\omega c - (\omega-1)\cosh(\tfrac{d_2}{m})]$$

$$= \frac{2\omega-1}{\omega^2(c+1)}\,[2+\omega c-\omega-\tfrac{1}{2}(\omega-1)\tfrac{d_2^2}{m^2}\,(1+\tfrac{1}{12}\tfrac{d_2^2}{m^2}+0(\tfrac{1}{m^4}))]$$

where we have written $c = \cos \pi/2m$. Expansion of $c$ yields for $\tilde{a}$

$$\tilde{a} = \frac{1}{\omega}\,[2-\frac{\omega(\pi^2+4d_2^2)}{8m^2}+0(\tfrac{1}{\omega})+0(\tfrac{1}{m^2})+0(\tfrac{\omega}{m^4})].$$

From (2.20b) it follows that

$$\beta = \frac{4\omega}{b_0\tilde{a}}\,[1+0(\tfrac{1}{\omega})+0(\tfrac{1}{\tilde{a}})].$$

Substitution of $\tilde{a}$ and using $\omega = 0(m^2)$ yields (2.24a).

The upperbound (2.24b) is obtained by substituting

$$\omega = \tilde{\omega} = \frac{16m^2}{\pi^2+4d_1^2}\,(1+0(\tfrac{1}{m^2}))\,.$$

□

Thus, under the conditions of theorem 2.3 the stability boundary $\beta = \beta(m)$ has at least an $O(m^4)$ behaviour as $m \to \infty$. In this connection, we remark that explicit Runge-Kutta methods with maximal real stability interval have a stability boundary $\beta(m)$ of $O(m^2)$ as $m \to \infty$ (see [9]). The partial implicitness of the splitting method (2.1)-(2.5) is apparently compensated by a considerably larger stability interval (we recall that $\beta = \infty$ if $D_2 = 1$). However, if we use higher order predictors we still need a relatively large number of iterations if $\tau\sigma$ is large.

As an example we consider BDF4 with third order extrapolation for $y^{(0)}$. Choosing $S^*$ maximal (i.e. $\omega=\tilde{\omega}$) we find from (2.24) and table 2.1 ($D_1=1/15$, $D_2= 0.1999$)

$$\beta(m) \cong .56\, m^4 \quad \text{as} \quad m >> 1.$$

Hence, for given $\tau$ and $\sigma$ we need at least

$$m = \sqrt[4]{\frac{\tau\sigma}{.56}}$$

iterations in order to have stability. For instance, if $\tau\sigma = 1000$ we need 7 iterations which is rather expensive.

## 3. A FOURTH ORDER SPLITTING METHOD WITH LARGE STABILITY BOUNDARIES

The stability boundaries derived in the preceding section for the fourth order method based on BDF4 and third order extrapolation as predictor formula, are relatively low. This is caused by a too fast increase of $P_m(\alpha)$ as $\alpha \to 0$, that is the *high frequencies* are not damped sufficiently. We can not correct this by choosing $S^*$ larger because $S^*$ is limited to the values given in table 2.2. In this section we investigate the effect of performing an *adjusted Jacobi iteration* on an extrapolation predictor in order to remove the high frequencies. This iteration leads to the following smoothed predictor

$$\tilde{y}^{(0)} = \Sigma + b_0\tau\tilde{F}(t_{n+1},\tilde{y}^{(0)},y^{(0)}) \,, \tag{3.1a}$$

$$(3.1b)\qquad y^{(0)} = \sum_{\ell=1}^{k} \hat{a}_\ell \, y_{n+1-\ell} \ ,$$

where $\tilde{F}(t,u,v)$ denotes a Jacobi type splitting function which is required to satisfy the conditions

$$(3.2)\qquad \tilde{F}(t,y,y) \equiv f(t,y), \ \frac{\partial \tilde{F}}{\partial u} = - \theta\tilde{\sigma} I$$

with I the identity matrix, $\tilde{\sigma}$ an estimate of the spectral radius of $\partial f/\partial y$, and $\theta$ a positive parameter to be determined below.

Let us define $\tilde{F}$ by specifying its i-th component $\tilde{F}_i$ according to

$$(3.3)\qquad \tilde{F}_i(t,u,v) = f_i(t,v_1,\ldots,v_{i-1},\gamma_i u_i+(1-\gamma_i)v_i,v_{i+1},\ldots,v_N),$$

$$(3.4)\qquad \gamma_i = - \theta\tilde{\sigma}\left[\frac{\partial f_i}{\partial y_i}\right]^{-1}, \quad i = 1,2,\ldots,N,$$

where $v_1,v_2,\ldots$; $u_1,u_2,\ldots$ and $y_1,y_2,\ldots$ denote the components of the vectors v, u and y, respectively, and N is the number of vector components.

Then using the definition of $\Sigma$ in (2.3) we prove the following comparison theorem.

THEOREM 3.1. *Let the method* (2.1)-(2.5) *be stable if* $y^{(0)}$ *is defined by* (3.1b) *and if* $P_m(\alpha)$ *satisfies* (2.19). *Then this method is also stable if the predictor is given by* $\tilde{y}^{(0)}$ *(defined by* (3.1)) *and* $P_m(\alpha)$ *satisfies the condition*

$$(3.5)\qquad - D_1 \le \frac{z+\theta b_0\tau\tilde{\sigma}}{1+\theta b_0\tau\tilde{\sigma}} P_m(\alpha(z_1,z_2)) \le D_2, \quad z=z_1+z_2.$$

PROOF. The variational equation for the predictor $\tilde{y}^{(0)}$ is given by

$$\Delta\tilde{y}^{(0)} = \sum_{\ell=1}^{k} \frac{a_\ell \, I + \frac{b_\ell}{b_0}(Z_1+Z_2)+\hat{a}_\ell(Z_1+Z_2+\theta b_0\tau\tilde{\sigma}I)}{1+\theta b_0\tau\tilde{\sigma}} \Delta y_{n+1-\ell} \ .$$

Substitution into (2.17) reveals that the characteristic equation assumes the form

$$(3.6)\qquad \zeta^k = \sum_{\ell=1}^{k} \{P_m(\alpha) \frac{a_\ell+ \frac{b_\ell}{b_0} z+\hat{a}_\ell(z+\theta b_0\tau\tilde{\sigma})}{1+\theta b_0\tau\tilde{\sigma}} + (a_\ell+ \frac{b_\ell}{b_0} z) \frac{1-P_m(\alpha)}{1-z}\} \zeta^{k-\ell}.$$

It is easily verified that this equation can be written in the form

$$(3.6') \qquad \zeta^k = \sum_{\ell=1}^{k} \{\hat{a}_\ell \tilde{P}_m(z_1,z_2) + (a_\ell + \frac{b_\ell}{b_0} z) \frac{1-\tilde{P}_m(z_1,z_2)}{1-z}\} \zeta^{k-\ell}$$

where

$$(3.7) \qquad \tilde{P}_m(z_1,z_2) = \frac{z+\theta b_0 \tau\tilde{\sigma}}{1+\theta b_0 \tau\tilde{\sigma}} P_m(\alpha(z_1,z_2)), \quad z = z_1 + z_2.$$

We observe that replacing the predictor $\tilde{y}^{(0)}$ (defined by (3.1)) by the predictor $y^{(0)}$ (defined by (3.1b)) implies that in the characteristic equation one should replace $\tilde{P}_m$ by $P_m$. Since this latter equation satisfies the root condition if $P_m$ satisfies (2.19), equation (3.6') will also satisfy the root condition if $-D_1 \le \tilde{P}_m \le D_2$, that is if (3.5) is satisfied. □

We apply this theorem to the case where $b_0$ and $\Sigma$ are defined by the BDF4 given by (2.23), and where $y^{(0)}$ is given by third order extrapolation. Using $\tilde{y}^{(0)}$ as defined by (3.1) as the predictor formula, theorem 3.1 implies that the resulting method is stable if (see table 2.1)

$$(3.5') \qquad -\frac{1}{15} \le \tilde{P}_m(z_1,z_2) \le 0.1999 \ .$$

Let us assume that $P_m(\alpha) \ge -1/15$ for $z_i \ge -S^*$, that is $S^*$ is bounded by the values listed in table 2.2 for $q = 3$. Then $\tilde{P}_m(z_1,z_2)$ is certainly bounded below by $-1/15$ if $z + \theta b_0 \tau\tilde{\sigma} \ge 0$. If $z + \theta b_0 \tau\tilde{\sigma} < 0$ (this happens when $\theta < \sigma/\tilde{\sigma}$) then $\tilde{P}_m(z_1,z_2)$ is bounded below by $-1/15$ if

$$\frac{-b_0\tau\sigma+\theta b_0\tau\tilde{\sigma}}{1+\theta b_0\tau\tilde{\sigma}} \ge -\frac{1}{15} \ .$$

By choosing

$$(3.8) \qquad \theta = \frac{15}{16}, \quad \tilde{\sigma} \ge \sigma - \frac{1}{15 b_0 \tau}$$

this inequality is satisfied.

In order to satisfy the inequality $\tilde{P}_m(z_1,z_2) \le 0.1999$ it is sufficient to consider the case $z_i \le -S^*$. Let us consider the function $\tilde{P}_m(z_1,z_2)$ along

the line $z_1 + z_2 = \text{const}$. Since in this region $P_m(\alpha(z_1,z_2))$ is larger as $\alpha(z_1,z_2)$ is smaller, we find a maximal value of $\tilde{P}_m(z_1,z_2)$ at the point where $\alpha(z_1,z_2)$ is minimal. From the definition of $\alpha$ it follows that along the line $z_1 + z_2 = \text{const}$. the minimum is reached at $z_1 = z_2$.

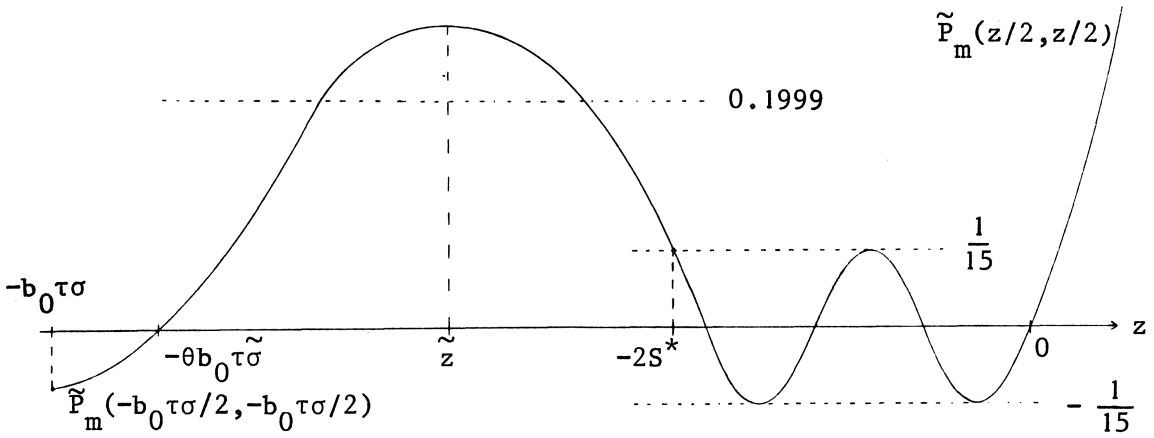


Figure 3.1. The function $\tilde{P}_m(z/2,z/2)$ for $-\sigma < -\theta\tilde{\sigma}$

Thus the function $\tilde{P}_m(z_1,z_2)$ is maximal in magnitude along the line $z_1 = z_2 = z/2$. In figure 3.1 the behaviour is illustrated. If $\tilde{z}$ is the point where $\tilde{P}_m$ assumes a maximum value in the interval $[-b_0\tau\sigma,-2S^*]$, then we should choose $\tau\tilde{\sigma}$ such that $\tilde{P}_m(-\tilde{z}/2,-\tilde{z}/2) \leq 0.1999$. This yields an upper bound for $\tau\tilde{\sigma}$ which is just the stability boundary $\beta$ of the method. In table 3.1 these values are listed for m = 1 until 6 and $S^*$ as large as allowed

Table 3.1. Maximal stability boundaries for BDF4 combined with the fourth order predictor (3.1)

| | m = 1 | m = 2 | m = 3 | m = 4 | m = 5 | m = 6 |
|---|---|---|---|---|---|---|
| $S^*$ | .44 | 4 | 18 | 54 | 129 | 264 |
| $\beta(m)$ | 20 | 101 | 385 | 1095 | 2549 | 5150 |
| $\beta/m^4$ | 20 | 6.3 | 4.8 | 4.3 | 4.1 | 4.0 |

by (2.20a). A comparison with the maximal boundaries attainable for the third order extrapolation as listed in table 2.2 reveals that we have

gained a factor of about 7 by performing the Jacobi iteration (3.1). For large values of m the stability boundary tends to behave as $4\ m^4$.

Summarizing, we conclude that the iterated BDF4 together with the predictor (3.1) where $y^{(0)}$ is defined by third order extrapolation, generate a fourth order accurate splitting method which is stable for the $S^*$-values listed in table 3.1 and for integration steps satisfying the condition

$$\tau \leq \frac{\beta(m)}{\tilde{\sigma}} \ . \tag{3.9}$$

This method will be denoted by *SC method* since it can be considered as a variant of the method of successive corrections introduced in [4] (see also section 2.3.1).

In actual application of the method we will choose for m the smallest integer such that (3.9) is satisfied when $\tau$ and $\tilde{\sigma}$ are prescribed.

## 4. NUMERICAL EXPERIMENTS

### 4.1. Methods used

We tested the SC method by comparing it with the ADI method of PEACEMAN and RACHFORD [8]. This method is denoted by PR($\nu$) in the tables of results, where $\nu$ indicates the number of Newton iterations used for solving each implicit relation. In the SC method we performed only one Newton iteration in solving both (2.2) and (3.1).

The splitting function F(t,u,v) used in the SC method was defined by the same ADI splitting function as used in the PR method. The Jacobian matrices $\partial F/\partial u$ and $\partial F/\partial v$ needed in the Newton iterations were updated at the beginning of each integration step. The examples are such that an analytical expression for the Jacobian matrices was available.

The estimate $\tilde{\sigma}$ needed for determining a safe number of iterations was either given in analytical form or computed by applying Gerschgorin's theorem to the matrix $\partial f/\partial y = \partial F/\partial u + \partial F/\partial v$ (this hardly requires additional effort).

The starting values needed by the SC method were obtained by computing

them from the exact values prescribed at $t_{-3}$, $t_{-2}$, $t_{-1}$ and $t_0$.

## 4.2. Numerical examples

The problems we chose were all of the form

$$(4.1) \qquad \frac{\partial U}{\partial t} = d\left[\frac{\partial^2 U^r}{\partial x_1^2} + \frac{\partial^2 U^r}{\partial x_2^2}\right] + \left[\frac{\partial U}{\partial x_1}\right]^s + \left[\frac{\partial U}{\partial x_2}\right]^s + v, \quad 0 \le t \le 1$$

where the diffusion coefficient d and the term v are functions of $U, t, x_1$ and $x_2$ to be specified below, and the integer r and s are nonlinearity parameters. The domain in the $(x_1,x_2)$-plane is given by the square $0 \le x_1, x_2 \le 1$; the Dirichlet boundary and the initial conditions at $t_0 = 0$ either follow from the exact solution given in table 4.1.

The initial-boundary value problems were semi-discretized by using standard differences on a uniform grid with grid points $(jh,\ell h)$. The components of the right-hand side function of the resulting system of ODE's (1.1) are evidently coupled according to a five-point molecule which allows the use of ADI splitting functions.

In the examples I and III we used for $\tilde{\sigma}$ the expression listed in table 4.1. In example II, a Gerschgorin estimate was used.

Table 4.1. Specification of the testproblems

| Example | Solution | d | r | s | v | $\tilde{\sigma}$ |
|---|---|---|---|---|---|---|
| I | $1+e^{-t}(x_1^2+x_2^2)$ | 1 | 1 | 0 | $-e^{-t}(x_1^2+x_2^2+4)-2$ | $8h^{-2}$ |
| II | $1+e^{-t}(x_1^2+x_2^2)$ | $\frac{1}{1+t}$ | 1 | 2 | $-e^{-t}[4d+(1+4e^{-t})(x_1^2+x_2^2)]$ | |
| III | $\frac{1}{2}(x_1+x_2)\sin 2\pi t$ | $\frac{x_1+x_2}{2(1+t)}$ | 3 | 0 | $-[\frac{3}{4}\frac{(x_1+x_2)^2}{1+t}\sin^3 2\pi t + 2 - \pi(x_1+x_2)\cos 2\pi t]$ | $\frac{24\sin^2 2\pi t}{(1+t)h^2}$ |

## 4.3. Numerical results

In the tables of results we listed the accuracy measured by the number *sd* of correct significant digits defined by

$$(4.2) \qquad sd = -\log_{10} \left| \text{maximum absolute error at } t = 1 \right|$$

and the computational effort *ce* measured by the total number of right-hand side evaluations needed in the integration process.

Problem I belongs to the class of model problems specified in section 2.2, so that the stability theory can be rigorously applied. The results in table 4.2 show that the SC method does behave in a stable way.

Table 4.2. sd/ce-values obtained for problem I with h = 1/24

| Method | $\tau = 1/2$ | $\tau = 1/5$ | $\tau = 1/10$ | $\tau = 1/20$ | $\tau = 1/40$ | $\tau = 1/80$ |
|---|---|---|---|---|---|---|
| PR(1) | 1.1/4 | 2.0/10 | 2.6/20 | 3.2/40 | 3.9/80 | 4.5/160 |
| SC | 2.0/22 | 4.0/45 | 5.1/90 | 6.3/140 | 7.4/280 | 8.7/400 |

The second order and fourth order behaviour of the PR and SC method is also reflected in the sd-values (on halving the integration step the sd-value should increase with $\log_{10} 2^{\tilde{p}}$, $\tilde{p}$ being the order of the method). Due to the fourth order behaviour of the SC method this method is much more accurate than the PR method for the same step size. This makes the fourth order method much more efficient than the PR method if high accuracy is desired. For example, to get four correct digits

Table 4.3. sd/ce-values obtained for problem II with h = 1/24

| Method | $\tau = 1/5$ | $\tau = 1/10$ | $\tau = 1/20$ | $\tau = 1/40$ | $\tau = 1/80$ |
|---|---|---|---|---|---|
| PR(1) | * | * | 2.0/40 | 3.6/80 | 4.3/160 |
| PR(2) | 1.6/20 | 2.4/40 | 3.1/80 | 3.7/160 | 4.3/320 |
| SC | * | * | 6.1/140 | 7.5/212 | 8.7/400 |

the PR method needs 80 right-hand side evaluations, whereas the SC method requires only 45 evaluations.

Problem II is mildly nonlinear; as a consequence one Newton iteration for solving the implicit relations is not sufficient if larger integration steps are used (indicated by * in table 4.3). For smaller steps we see a similar behaviour as exhibited by problem I.

Problem III is rather nonlinear with a rapidly changing spectral radius. It turned out to be a more

Table 4.4. sd/ce-values obtained for problem III with h = 1/24

| Method | $\tau \geq 1/40$ | $\tau = 1/80$ | $\tau = 1/160$ |
|---|---|---|---|
| PR(1) | * | 2.1/160 | 2.7/320 |
| PR(2) | * | 3.0/320 | 4.1/640 |
| SC | * | 5.9/390 | 6.9/676 |

difficult problem for both methods than problem II. For $\tau \geq 1/40$ both methods failed because the Newton process did not converge (indicated by *). Again the SC method is superior to the PR method if high accuracies are desired.

## 4.4. Concluding remarks

In the high accuracy region, the experiments reported in the preceding section show the superiority of the SC method over the classical PR method because of the order four behaviour which is maintained for realistic integration steps. In this connection we remark that the method of successive corrections with $S^* = 0$, $\omega = \lambda_j = \mu_j = 1$, analysed in [4], shows its fourth order behaviour only for relatively small integration steps (relative to the spectral radius). For realistic integration steps the order of this method degenerates so that it is hardly more efficient than e.g. the PR method. By virtue of the effective fourth order behaviour of the SC method, a variable order splitting method, e.g. composed of the SC and the PR method,

may turn out to be an efficient method for solving parabolic equations with an arbitrary degree of accuracy.

In the SC method described in this paper there are several choices which are not necessarily the best possible. For instance, the Jacobi iteration (3.1) may be replaced by Gauss-Seidel iteration (although this would complicate the theoretical analysis considerably). Furthermore, the relaxation parameter $\omega$ (or equivalently the parameter $S^*$) and the number of iterations m were chosen such that the effective stability boundary $\beta(m)/m$ is as large as possible. An alternative may be the use of another predictor formula (e.g. linear extrapolation) which yields an infinite stability boundary (cf. table 2.1). This leaves $\omega$ and m free for minimizing the local error. Finally, the choice of the splitting function F(t,u,v): Alternatives are odd-even hopscotch splitting which reduces the computational effort of solving the implicit relations, and line hopscotch which allows the integration of equations with mixed derivatives (for a treatment of hopscotch splittings we refer to [2,3]).

In [6, Appendix] a lot of additional experiments are reported which give some insight into these questions.

REFERENCES

[1] BEAM, R.M. & R.F. WARMING, *Alternating direction implicit methods for parabolic equations with a mixed derivative*, SIAM J. Sci. Stat. Comput. 1, 131-159 (1980).

[2] GOURLAY, A.R., *Hopscotch, a fast second order partial differential equation solver*, J. Inst. Math. Appl. 6, 375-390 (1970).

[3] _____ , *Splitting methods for time-dependent partial differential equations*, In: Proceedings of the 1976 Conference: The State of the Art in Numerical Analysis (Jacobs, D.A.H. ed.), Academic Press 1977.

[4] HOUWEN, P.J. VAN DER, *Multistep splitting methods of high order for initial value problems*, SIAM J. Numer. Anal. 17, 291-309 (1979).

[5] _____ , *Defect correction iteration and splitting methods for time-dependent partial differential equations*, Report NW 116/81, Mathematisch Centrum, Amsterdam (preprint 1981).

[6] HOUWEN, P.J. VAN DER & H.B. DE VRIES, *A fourth order ADI method for semi-discrete parabolic equations*, Report NW /82, Mathematisch Centrum, Amsterdam (1982).

[7] LAMBERT, J.D., *Computational methods in ordinary differential equations*, John Wiley & Sons, London 1973.

[8] PEACEMAN, D.W. & H.H. RACHFORD jr., *The numerical solution of parabolic and elliptic differential equations*, J. Soc. Ind. Appl. Math. 3, 28-41 (1955).

[9] RIHA, W., *Optimal stability polynomials*, Computing 9, 37-43 (1972).

[10] YOUNG, D.M., *Iterative solution of large linear systems*, Academic Press, New York 1971.

APPENDIX

For the BDF4 defined by (2.23) with $y^{(0)}$ defined by extrapolation we show how the bounds $D_1$ and $D_2$ (cf. (2.19)) for $P_m(\alpha)$ listed in table 2.1 are obtained by applying Hurwitz's criterion to the characteristic equation

$$\zeta^4 + \left[-P_m(\alpha)\hat{a}_1 - \frac{48}{25}\frac{1-P_m(\alpha)}{1-z}\right]\zeta^3 + \left[-P_m(\alpha)\hat{a}_2 + \frac{36}{25}\frac{1-P_m(\alpha)}{1-z}\right]\zeta^2$$

(A1)

$$+ \left[-P_m(\alpha)\hat{a}_3 - \frac{16}{25}\frac{1-P_m(\alpha)}{1-z}\right]\zeta + \left[-P_m(\alpha)\hat{a}_4 + \frac{3}{25}\frac{1-P_m(\alpha)}{1-z}\right] = 0.$$

Here, $z = z_1 + z_2$ and $y^{(0)} = \sum_{\ell=1}^{4} \hat{a}_\ell\, y_{n+1-\ell}$, where the coefficients $\hat{a}_\ell$ are specified in table 2.1. Note that (A1) is a special case of (2.17).

We also describe a number of experiments for the linear problem I with the splitting method (2.1)-(2.5) based on BDF4. These results show the effect of the parameter $S^*$ and the number of iterations m, the dependence of the accuracy on the mesh spacing h, the order of accuracy, the use of other predictor formulas and the effect of violating the stability conditions.

In order to test the theory developed for the splitting method (2.1)-(2.5) by performing a large number of experiments, it is convenient to denote this splitting method in a slightly different manner than in section 3. The method (2.1)-(2.5) with a given value of m and $S^*$ will be denoted by $SC(q,m,S^*)$ where q indicates the predictor formula, i.e. q = 1,2,3 and 4 corresponding to the first order extrapolation, second order extrapolation, third order extrapolation and the smoothed predictor (3.1) where $y^{(0)}$ is defined by third order extrapolation, respectively. We recall that the generating multistep formula defining $b_0$ and $\Sigma$ is the BDF4 defined by (2.23).

The results in the following sections have led us to the choice of the SC method defined in section 3, i.e.:

$$SC \equiv SC(4,m,S^*_{max}), \qquad m = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{cases} \quad \text{if } \tau\tilde{\sigma} \in \begin{cases} (0,20) \\ [20,101) \\ [101,385) \\ [385,1095) \\ [1095,2549) \\ [2549,5150) \end{cases}.$$

Here, $\tilde{\sigma}$ is an estimate of the spectral radius of $\partial f/\partial y$ as used in the predictor formula (3.1) and $S^*_{max}$ is given in table 3.1 ($S^*_{max}$ is the maximal value of $S^*$ allowed by condition (2.20a) for the third order extrapolation (see also table 2.2)).

In our experiments for the linear example I (see table 4.1) presented in this appendix, the starting values needed in the SC(q,m,$S^*$) method were prescribed by the exact solution.

In the tables of results we also listed the maximal step $\tau_{max}$ allowed by the stability condition (2.20b) for $S^* > 0$ and (2.20'b) for $S^* = 0$. For more details concerning the implementation and notation we refer to section 4.

A1. The stability interval

Writing (A1) as

$$\zeta^4 + c_1\zeta^3 + c_2\zeta^2 + c_3\zeta + c_4 = 0 \tag{A1'}$$

we find by applying Hurwitz's criterion [7,p.80] that (A1') has its roots on the unit disk if

$$\begin{aligned} &\gamma_i \geq 0, \quad i = 0(1)4 \\ &\gamma_1\gamma_2\gamma_3 - \gamma_1^2\gamma_4 - \gamma_0\gamma_3^2 \geq 0, \end{aligned} \tag{A2}$$

where

$$\begin{aligned}
\gamma_0 &= 1 - c_1 + c_2 - c_3 + c_4 = 1 + P_m(\alpha)(\hat{a}_1 - \hat{a}_2 + \hat{a}_3 - \hat{a}_4) + \frac{103}{25}\frac{1-P_m(\alpha)}{1-z}, \\
\gamma_1 &= 4 - 2c_1 + 2c_3 - 4c_4 = 4 + 2P_m(\alpha)(\hat{a}_1 - \hat{a}_3 + 2\hat{a}_4) + \frac{52}{25}\frac{1-P_m(\alpha)}{1-z}, \\
\gamma_2 &= 6 - 2c_2 + 6c_4 = 6 + 2P_m(\alpha)(\hat{a}_2 - 3\hat{a}_4) - \frac{54}{25}\frac{1-P_m(\alpha)}{1-z}, \\
\gamma_3 &= 4 + 2c_1 - 2c_3 - 4c_4 = 4 + 2P_m(\alpha)(\hat{a}_3 - \hat{a}_1 + 2\hat{a}_4) - \frac{76}{25}\frac{1-P_m(\alpha)}{1-z}, \\
\gamma_4 &= 1 + c_1 + c_2 + c_3 + c_4 = 1 - P_m(\alpha)\sum_{\ell=1}^{4}\hat{a}_\ell - \frac{1-P_m(\alpha)}{1-z}.
\end{aligned} \tag{A3}$$

Note that for all predictor formulas listed in table 2.1 $\sum_{\ell=1}^{4}\hat{a}_\ell = 1$, i.e.

$$\gamma_4 = - z \frac{1-P_m(\alpha)}{1-z} \quad .$$

Evidently the conditions (A2) are satisfied for all negative z if we put $P_m(\alpha) \equiv 0$, because (A1) then corresponds to the BDF4. For nonzero $P_m(\alpha)$ we have the following theorem.

THEOREM A1. *The characteristic equation* (A1) *has its roots on the unit disk for all negative* $z_1$ *and* $z_2$ *if* $P_m(\alpha)$ *satisfies the condition* (2.19), *i.e.,*

$$-1 \le -D_1 \le P_m(\alpha) \le D_2 \le 1,$$

*where for each extrapolation formula the bounds* $D_1$ *and* $D_2$ *are already given in table* 2.1.

PROOF. 1. Consider the third order extrapolation formula (i.e., q=3 in table 2.1). The relations for $\gamma_i$ with i = 0,1,2,3 (see (A3)) can be simplified as follows:

$$\gamma_0 = 1 + 15\, P_m(\alpha) + \frac{103}{25} \frac{1-P_m(\alpha)}{1-z},$$

$$\gamma_1 = \frac{1-P_m(\alpha)}{1-z} \left(6 \frac{2}{25} - 4z\right),$$

$$\gamma_2 = \frac{1-P_m(\alpha)}{1-z} \left(3 \frac{21}{25} - 6z\right),$$

$$\gamma_3 = \frac{1-P_m(\alpha)}{1-z} \left(\frac{24}{25} - 4z\right).$$

It can be easily shown that if $z \le 0$ and $-1/15 \le P_m(\alpha) \le 1$ then $\gamma_i \ge 0$ for all i. Substituting $\gamma_i$ in the nonlinear condition in (A2) gives after a tedious calculation the inequality

$$\text{(A4)} \qquad \frac{(1-P_m(\alpha))^2}{(1-z)^3} R(z) \ge 0,$$

where

$$R(z) = r_3 z^3 + r_2 z^2 + r_1 z + r_0,$$

$$r_3 = 320\, P_m(\alpha) - 64, \quad r_2 = 92.16 - 471.04\, P_m(\alpha),$$

$$r_1 = 203.5712\, P_m(\alpha) - 65.9456, \quad r_0 = 17.69472 - 32.44032\, P_m(\alpha).$$

It is easily numerically verified that $R(z) \geq 0$ if $z \leq 0$ and $P_m(\alpha) \leq 0.1999$. Thus, the inequality (A4) holds for all $z \leq 0$ and $-1 \leq P_m(\alpha) \leq 0.1999$. From Hurwitz's criterion (see (A2)) it follows that the characteristic equation (A1) with the third order extrapolation formula has its roots on the unit disk if $z \leq 0$ and

$$-\frac{1}{15} \leq P_m(\alpha) \leq 0.1999.$$

2. Consider the first order extrapolation formula (i.e., q=1 in table 2.1). The relations for $\gamma_i$ with i = 0,1,2,3 (see (A3)) may now be written as follows:

$$\gamma_0 = 1 + 3\,P_m(\alpha) + \frac{103}{25}\,\frac{1-P_m(\alpha)}{1-z},$$

$$\gamma_1 = 4 + 4\,P_m(\alpha) + \frac{52}{25}\,\frac{1-P_m(\alpha)}{1-z},$$

$$\gamma_2 = 4 + \frac{1-P_m(\alpha)}{1-z}\,(-\frac{4}{25} - 2z),$$

$$\gamma_3 = \frac{1-P_m(\alpha)}{1-z}\,(\frac{24}{25} - 4z).$$

It can be easily shown that if $z \leq 0$ and $-1/3 \leq P_m(\alpha) \leq 1$ then $\gamma_i \geq 0$ for all i. Substitution of $\gamma_i$ in the nonlinear condition in (A2) gives

$$\text{(A5)} \qquad \frac{1-P_m(\alpha)}{(1-z)^3} \qquad R(z) \geq 0,$$

where

$$R(z) = r_3 z^3 + r_2 z^2 + r_1 z + r_0,$$

$$r_3 = -64,$$

$$r_2 = 92.16 + 97.28\,P_m(\alpha) - 30.72\,P_m^2(\alpha),$$

$$r_1 = -62.9456 - 70.8608\,P_m(\alpha) + 11.3664\,P_m^2(\alpha),$$

$$r_0 = 17.69472 + 13.76256\,P_m(\alpha) - 0.73728\,P_m^2(\alpha).$$

It is easily established that for $P_m(\alpha)$ in the interval $[-1/3,1]$ $r_2 \geq 0$,

$r_1 \le 0$ and $r_0 \ge 0$. Thus, the inequality (A5) holds for all $z \le 0$ and $P_m(\alpha)$ in the interval $[-1/3,1]$. Again, application of Hurwitz's criterion reveals that the characteristic equation (A1) with the first order extrapolation formula has its roots on the unit disk if $z \le 0$ and $-1/3 \le P_m(\alpha) \le 1$.

In a completely analogous manner we can analyse the BDF4 with the other extrapolation formulas and derive for $P_m(\alpha)$ the bounds already given in table 2.1. □

REMARK 1. When in (A1) the variables $z_1$ and $z_2$ are largely negative, then the characteristic equation can be approximated by

$$\text{(A6)} \qquad \zeta^4 = P_m(\alpha) \sum_{\ell=1}^{4} \hat{a}_\ell \zeta^{4-\ell}.$$

Applying Hurwitz's criterion to this equation we obtain a condition on $P_m(\alpha)$ for each extrapolation formula. For the zero order, first order, second order and third order extrapolation the bounds $(D_1,D_2)$ for $P_m(\alpha)$ are in this case (1,1), (1/3,1), (1/7,1/2) and (1/15,1/5), respectively.

## A2. Stability tests

In this section the stability of $SC(q,m,S^*)$ is tested. Therefore the $SC(q,m,S^*)$ method is applied to the model problem I with a large number of integration steps and a relatively large step, say $\tau = 1/10$.

The second order method $SC(1,m,S^*)$ should not give difficulties, because theoretically it is unconditionally stable provided $S^*$ is not too large (see table A0). This is confirmed by the results in table A1, where the model problem I is integrated from $t = 0$ until $t = t_e = 10$. Here the accuracy is measured in the points $t_j = \{j \mid j = 1,2,\ldots,10\}$ by $sd = -\log_{10}$ | maximum absolute error in $t_j$ |. Note that the maximal stable step $\tau_{max}$ follows from theorem 2.3.

Table A0. Maximal $S^*$ values for the first order predictor (q=1).

| m | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $S^* \simeq$ | 2.96 | 33.2 | 157 | 486 | 1176 | 2425 |

Table A1. Results obtained by $SC(1,m,S^*)$ with $\tau = 1/10$ when applied to problem I with $t_e = 10$.

| Method | h | $\tau_{max}$ | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | t=7 | t=8 | t=9 | t=10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SC(1,2,10) | 1/10 | ∞ | 3.0 | 3.4 | 3.9 | 4.3 | 4.7 | 5.2 | 5.6 | 6.0 | 6.5 | 6.9 |
| SC(1,4,10) | 1/20 | ∞ | 3.0 | 3.4 | 3.9 | 4.3 | 4.8 | 5.2 | 5.6 | 6.0 | 6.5 | 6.9 |

The fourth order method $SC(3,m,S^*)$ is only stable for relatively small values of $\tau$. Taking again problem I we found the results listed in table A2.

Table A2. Results obtained by $SC(3,m,S^*)$ with $\tau = 1/10$ when applied to problem I with $t_e = 10$.

| Method | h | $\tau_{max}$ | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | t=7 | t=8 | t=9 | t=10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SC(3,2,10) | 1/10 | * | 4.8 | 4.5 | 3.7 | 2.7 | 1.6 | .5 | * | * | * | * |
| SC(3,4,10) | 1/10 | .06 | 6.1 | 6.5 | 6.9 | 7.4 | 7.8 | 8.3 | 8.7 | 9.1 | 9.6 | 10.0 |
| SC(3,2,10) | 1/20 | * | 4.0 | 2.8 | 1.4 | .1 | * | * | * | * | * | * |
| SC(3,4,10) | 1/20 | .015 | 4.5 | 3.7 | 2.7 | 1.5 | .2 | * | * | * | * | * |
| SC(3,4,40) | 1/20 | .04 | 5.3 | 5.7 | 6.1 | 6.1 | 5.7 | 5.5 | 4.9 | 4.7 | 4.2 | 3.7 |
| SC(3,4,52) | 1/20 | .05 | 5.2 | 5.6 | 6.0 | 6.5 | 6.9 | 7.3 | 7.7 | 8.0 | 8.2 | 8.2 |
| SC(3,4,80) | 1/20 | * | 5.0 | 5.2 | 4.8 | 4.1 | 3.5 | 2.9 | 2.3 | 1.6 | 1.0 | .4 |

Here, an asterisk means instability (for $\tau_{max}$ it indicates that $S^*$ is larger than allowed by table 2.2). According to the theory all experiments are unstable. Yet if $\tau$ is sufficiently close to $\tau_{max}$ (say $\tau \leq 2\tau_{max}$) the instabilities seem to vanish. By decreasing the step size $\tau$ such that $\tau \cong \tau_{max}$ we should get completely stable results. In table A3 these results are listed confirming the theory.

Table A3. Results obtained by SC(3,4,10) with $\tau = 1/70 = .0145....$ when applied to problem I with $t_e = 10$.

| Method | h | $\tau_{max}$ | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | t=7 | t=8 | t=9 | t=10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SC(3,4,10) | 1/20 | .015 | 9.1 | 9.5 | 10.0 | 10.4 | 10.8 | 11.3 | 11.8 | 12.4 | 12.4 | 12.3 |

A.2.1. Amplification factors

In order to see by what factor perturbations are amplified we have plotted the magnitude of the largest characteristic root of equation (A1) with the third order extrapolation (q=3) as a function of $\alpha$ (see (2.18)) with $z_1 = z_2 = z/2$. In the figures A1 and A2 the value of $|\zeta|_{max}$ is shown for SC(3,2,10) and SC(3,4,10), respectively. These plots show that amplifications by a factor as large as 1.4 occur if $\alpha$ is less than .1 and .05, respectively.

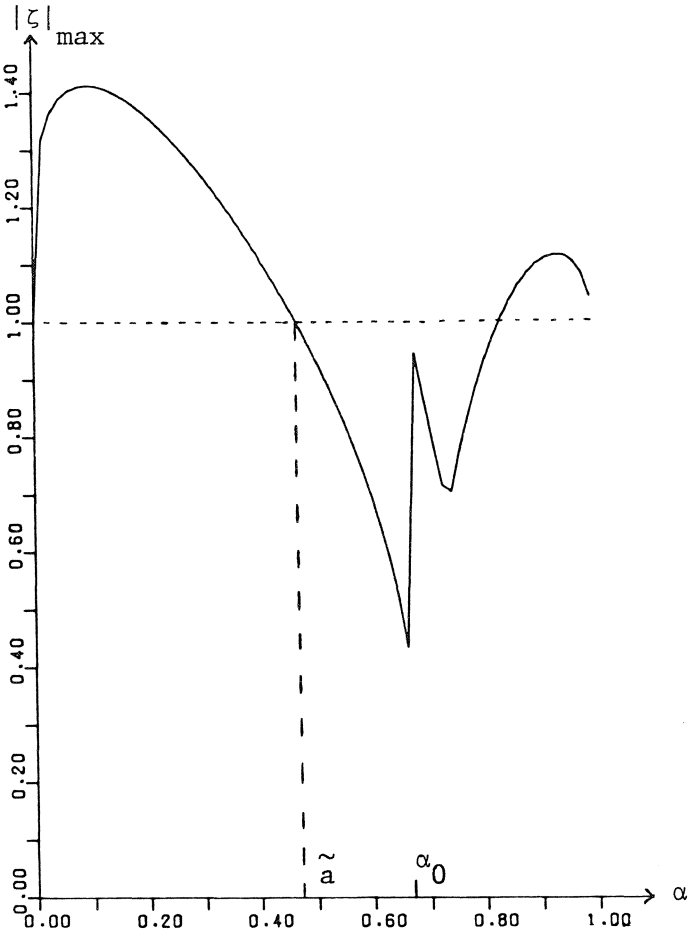


Fig. A1. The value of $|\zeta|_{max}$ of (A1) for SC(3,2,10) as a function of $\alpha$ with $z_1 = z_2 = z/2$.

This happens if $\tau\sigma$ is larger than 150 in SC(3,2,10) and larger than 350 in SC(3,4,10). For convenience we list for both methods in table A4 the values of $\omega$ defined by (2.5), b defined in (2.4), $\alpha_0$ defined in (2.13), D defined by (2.13') and $\tilde{a}$ given by (2.21). The SC(3,2,10) method is unstable unless the integration step $\tau$ is so small that $\alpha$ lies in the interval $[\tilde{a},.83]$, i.e. $\tau\sigma < 8.54$. It can be easily shown that if $z_1 = z_2 = z/2$ and $z \leq 0$ then $\alpha$ (given by (2.18)) lies in the interval [0,1]. Notice that the curves in the figures $A_1$ and $A_2$ are obtained by calculating the values of $|\zeta|_{max}$ for a large number of values of $\alpha$ in the interval [0,1].

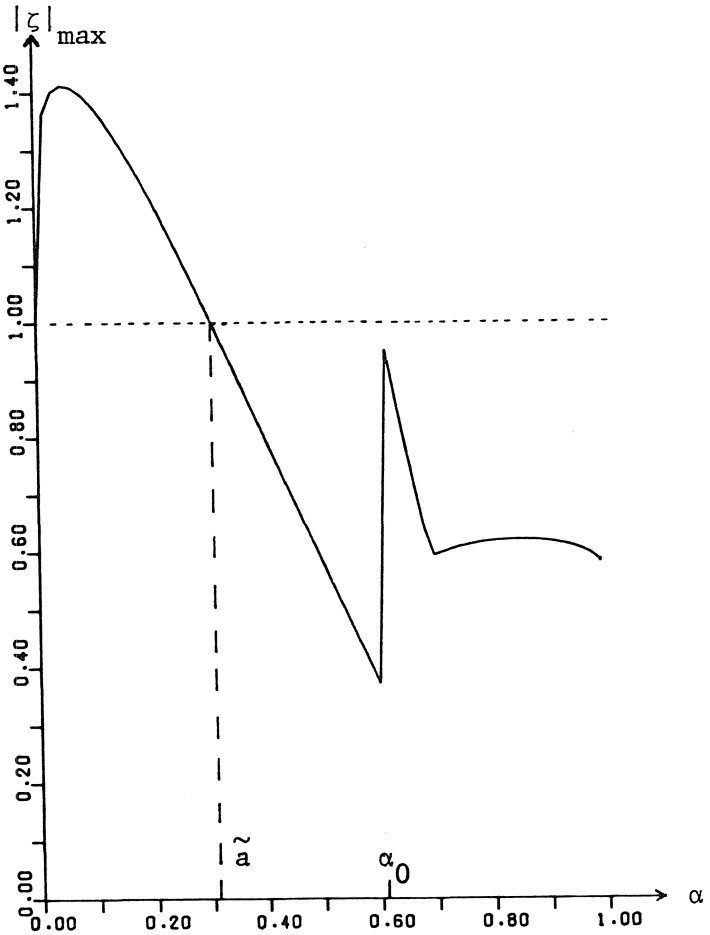


Fig. A2. The value of $|\zeta|_{max}$ of (A1) for SC(3,4,10) as a function of $\alpha$ with $z_1 = z_2 = z/2$.

Table A4. The values of $\omega$,b,$\alpha_0$, D and $\tilde{a}$ in the SC(3,2,10) and SC(3,4,10) method.

| Method | $\omega$ | b | $\alpha_0$ | D | $\tilde{a}$ |
|---|---|---|---|---|---|
| SC(3,2,10) | 2.36 | 1.5763 | 0.6679 | 0.1492 | 0.4685 |
| SC(3,4,10) | 2.67 | 1.6255 | 0.6088 | 0.0087 | 0.3074 |

Figure A2 shows that SC(3,4,10) has a stability interval $[\tilde{a},b]$, i.e. $\tau\sigma < 47.3$ (see table A4 and theorem 2.3); serious instabilities are to be expected if $\alpha \to 0.05$, i.e. $\tau\sigma$ becomes as large as 350. This explains the apparently stable behaviour of the experiment $(h,m,S^*) = (1/10,4,10)$ in table A2, and the unstable behaviour of the experiment$(h,m,S^*) = (1/20,4,10)$ in table A2.

A.2.2. Smoothed third order extrapolation

Finally, we consider the method SC(4,4,$S^*$) with its extended stability intervals (see section 3). Theoretically, the experiments in table A5 should be stable which is confirmed by the sd-values obtained.

Table A5. Results obtained by SC(4,4,$S^*$) with $\tau = 1/10$ when applied to problem I with $t_e = 10$.

| Method | h | $\tau_{max}$ | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | t=7 | t=8 | t=9 | t=10 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| SC(4,4,10) | 1/20 | .11 | 5.5 | 5.7 | 6.6 | 6.5 | 7.1 | 7.1 | 7.6 | 7.8 | 8.3 | 8.8 |
| SC(4,4,40) | 1/20 | .28 | 5.3 | 5.7 | 6.2 | 6.6 | 7.0 | 7.5 | 7.9 | 8.3 | 8.8 | 9.2 |
| SC(4,4,52) | 1/20 | .33 | 5.2 | 5.6 | 6.0 | 6.5 | 6.9 | 7.3 | 7.8 | 8.2 | 8.6 | 9.1 |

In order to suppress instabilities in the high frequency region we performed an adjusted Jacobi iteration on the third order extrapolation predictor which leads to the SC(4,m,$S^*$) method (see section 3). It is

well known that also Point Gauss-Seidel (PGS) iteration removes the high frequencies from the iteration error. Replacing the Jacobi iteration (3.1) by Point Gauss-Seidel iteration complicates the theoretical analysis considerably. Therefore, we give only some numerical results for the linear modelproblem I with a third order extrapolation predictor smoothed by PGS iteration.

The PGS iteration leads to the following smoothed predictor

$$\text{(A7a)} \qquad \hat{y}^{(0)} = \Sigma + b_0 \tau F^*(t_{n+1}, \hat{y}^{(0)}, y^{(0)}),$$

$$\text{(A7b)} \qquad y^{(0)} = 4y_n - 6y_{n-1} + 4y_{n-2} - y_{n-3},$$

where $F^*(t,u,v)$ denotes a Gauss-Seidel type splitting function.

Let us define $F^*$ by specifying its i-th component $F_i^*$ according to

$$\text{(A8)} \qquad F_i^*(t,u,v) = f_i(t,u_1,\ldots,u_{i-1},u_i,v_{i+1},\ldots,v_N),$$

where $v_1,v_2,\ldots$ and $u_1,u_2,\ldots.$ denote the components of the vectors v and u, respectively, and N is the number of vector components. Then the components $\hat{y}_i^{(0)}$, i = 1,2,...,N of the smoothed predictor $\hat{y}^{(0)}$ (A7a) are determined by solving

$$\text{(A9)} \qquad y - b_0 \tau f_i(t_{n+1}, \hat{y}_1^{(0)}, \ldots, \hat{y}_{i-1}^{(0)}, y, y_{i+1}^{(0)}, \ldots, y_N^{(0)}) = \Sigma_i, \; i = 1(1)N,$$

for y and setting $\hat{y}_i^{(0)} = y$, i = 1(1)N. One (nonlinear) Point Gauss-Seidel iteration is now defined by (approximately) solving (A9) performing just one Newton iteration for each scalar equation. For the linear problem I only linear scalar equations have to be solved. The iteration matrix in the PGS iteration, i.e., $\partial F^*/\partial u$ and $\partial F^*/\partial v$, does not have the same eigensystem for the class of model problems specified in section 2.2.

The iterated BDF4 together with the predictor (A7) will be denoted by PGSSC(4,m,S*). In table A6 some results obtained by this method are listed for the linear example I.

Table A6. Results obtained by PGSSC(4,4,$S^*$) with $\tau$ = 1/10 when applied to problem I with $t_e$ = 10.

| Method | h | t=1 | t=2 | t=3 | t=4 | t=5 | t=6 | t=7 | t=8 | t=9 | t=10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| PGSSC(4,4,10) | 1/20 | 4.9 | 2.7 | .4 | * | * | * | * | * | * | * |
| PGSSC(4,4,40) | 1/20 | 5.3 | 5.7 | 5.7 | 5.3 | 4.6 | 4.1 | 4.2 | 4.9 | 6.2 | 7.6 |
| PGSSC(4,4,52) | 1/20 | 5.2 | 5.6 | 6.0 | 6.5 | 6.9 | 7.4 | 7.8 | 8.2 | 8.7 | 9.1 |

Only the PGSSC(4,4,52) method gives stable results and is competitive with the SC(4,4,52) method (see table A5). Comparing the results listed in table A5 and A6 it seems that it is better to smooth the third order extrapolation predictor with Jacobi iteration (see section 3) than with Point Gauss-Seidel iteration ((A7)-(A9)).

## A3. The order of accuracy

From (4.2) it follows that a p-th order method satisfies the relation

(A10) $$\Delta sd := sd(\tau) - sd(a\tau) = \log_{10} a^p \quad \text{as } \tau \to 0.$$

Thus, on halving the step length we have $\Delta sd \cong .3\ p$; for the SC(q,m,$S^*$) method we expect (cf.(2.14'))

(A11) $$\Delta sd = .3 \min\{4, q+1\},\ S^* \text{ fixed and } \neq 0.$$

If $S^* = 0$ the SC method reduces to the method discussed in [4] where it was shown that

(A12) $$\Delta sd \cong .3 \min\{4, q+2m\},\quad S^* = 0.$$

In the following tables of results the sd-values (cf.(4.2)) are listed for the modelproblem I obtained by the various methods. We also listed the maximal step $\tau_{max}$ allowed by the stability conditions (see theorem 2.3 and section 2.3.1).

Table A7. Results obtained by SC(q,2,$S^*$) when applied to problem I with h = 1/10.

| τ | SC(q,2,0) | | SC(q,2,4) | | | SC(1,2,20) | SC(1,2,30) |
|---|---|---|---|---|---|---|---|
| | q=1 | q=3 | q=1 | q=3 | q=4 | | |
| 1/5 | 1.5 | 2.6 | 1.9 | 3.1 | 3.5 | 2.3 | 2.2 |
| 1/10 | 2.3 | 3.9 | 3.2 | 4.6 | 5.3 | 2.8 | 2.7 |
| 1/20 | 3.3 | 4.6 | 3.9 | 6.4 | 6.5 | 3.4 | 3.3 |
| 1/40 | 4.4 | 5.5 | 4.4 | 7.6 | 7.6 | 3.9 | 3.9 |
| 1/80 | 5.5 | 9.2 | 4.9 | 8.7 | 8.8 | 4.5 | 4.4 |
| $\tau_{max}$ | ∞ | .005 | ∞ | .017 | .13 | ∞ | ∞ |

Table A8. Results obtained by SC(q,4,$S^*$) when applied to problem I with h = 1/10.

| τ | SC(q,4,0) | | SC(q,4,10) | | SC(q,4,20) | | SC(q,4,52) | | |
|---|---|---|---|---|---|---|---|---|---|
| | q=1 | q=3 | q=1 | q=3 | q=1 | q=3 | q=1 | q=3 | q=4 |
| 1/5 | 1.8 | 2.9 | 2.9 | 4.1 | 3.2 | 4.4 | 2.8 | 4.0 | 4.1 |
| 1/10 | 2.8 | 4.4 | 4.3 | 6.1 | 3.7 | 5.8 | 3.2 | 5.2 | 5.2 |
| 1/20 | 3.9 | 5.9 | 4.5 | 7.3 | 4.1 | 6.7 | 3.7 | 6.3 | 6.3 |
| 1/40 | 5.2 | 8.3 | 5.0 | 8.3 | 4.6 | 7.8 | 4.2 | 7.4 | 7.5 |
| 1/80 | 6.6 | 10.0 | 5.4 | 9.3 | 5.1 | 8.9 | 4.7 | 8.6 | 8.7 |
| $\tau_{max}$ | ∞ | .012 | ∞ | .06 | ∞ | .096 | ∞ | .2 | 1.3 |

The results obtained for $S^* = 0$ should show a fourth order behaviour both for q = 1,3 and m = 2,4. Hence by virtue of (A12) we expect $\Delta sd \cong 1.2$ as $\tau \to 0$. For q=1 this behaviour is roughly confirmed by the tables A7 and A8, but for q=3 a much higher order of accuracy is shown inspite of the integration steps exceeding the maximal stable step $\tau_{max}$. If $S^* > 0$ relation (A11) indicates that $\Delta sd \cong .6$ for q=1 and $\Delta sd \cong 1.2$ for q=3,4, as $\tau \to 0$. This behaviour is more or less reflected in the tables

of results. The most interesting method seems to be SC(4,m,$S^*$) because of its rather high accuracy, particularly for larger integration steps.

In order to increase the "stiffness" of the problem we choose h = 1/20. In table A9 results are listed obtained by SC(4,m,$S^*$). The results again show the correct order behaviour. We also observe that comparing sd-values obtained for equal m/τ-values, that is requiring roughly the same computational effort, reveals that SC(4,2,4) is more efficient than SC(4,4,52), although SC(4,2,4) is stable only for τ = 1/40 and 1/80. A similar conclusion can be drawn for SC(q,2,4) and SC(q,4,52) from the tables A7 and A8, where q=1,3 and 4.

Table A9. Results obtained by SC(4,m,$S^*$) when applied to problem I with h = 1/20.

| τ | SC(4,2,4) | SC(4,4,52) |
|---|---|---|
| 1/5 | 2.9 | 4.0 |
| 1/10 | 4.1 | 5.2 |
| 1/20 | 6.1 | 6.3 |
| 1/40 | 7.6 | 7.4 |
| 1/80 | 8.7 | 8.6 |
| $\tau_{max}$ | .03 | .33 |

Comparing the results listed in the tables A7, A8 and A9 obtained by SC(4,2,4) and SC(4,4,52) we observe that for large τ-values the SC(4,2,4) method is more sensitive to grid refinement than the SC(4,4,52) method.

From the results presented in this section we may draw the following conclusions:

(i) The asymptotic order of accuracy of the SC methods is roughly as predicted by the theory.

(ii) The instability is rather mild, which could be expected from the results presented in section A2.

(iii) SC(3,m,$S^*$) is considerably more accurate than SC(1,m,$S^*$). SC(4,m,$S^*$) is more accurate than SC(3,m,$S^*$) for larger τ.

(iv) The accuracy increases as $S^*$ decreases for $\tau \to 0$.

(v) $SC(q,m_1,S^*_{max})$ is more efficient than $SC(q,m_2,S^*_{max})$ if $m_1 < m_2$.

A4. The effect of grid refinement

It is well known that splitting methods loose accuracy if the mesh width h is decreased, particularly for large time steps. Therefore, we tested $SC(q,2,S^*)$ by performing calculations for a sequence of h values with the test problem I.

Table A10. sd-values (4.2) obtained by $SC(q,2,S^*)$ with $\tau = 1/10$ when applied to problem I.

| h | $\tau\sigma$ | SC(q,2,0) | | SC(q,2,10) | | SC(q,2,20) | | SC(q,2,40) | |
|---|---|---|---|---|---|---|---|---|---|
| | | q=1 | q=3 | q=1 | q=3 | q=1 | q=3 | q=1 | q=3 |
| 1/5 | 20 | 3.2 | 4.8 | 3.0 | 4.9 | 2.8 | 4.0 | 2.7 | 3.1 |
| 1/10 | 80 | 2.3 | 3.9 | 3.0 | 4.8 | 2.8 | 3.8 | 2.7 | 2.7 |
| 1/20 | 320 | 1.4 | 3.2 | 2.5 | 4.0 | 2.8 | 3.8 | 2.7 | 2.7 |
| 1/40 | 1280 | 0.9 | 2.3 | 1.5 | 3.2 | 1.7 | 3.4 | 2.0 | 2.7 |
| $\beta$ | | $\infty$ | 4.2 | $\infty$ | * | $\infty$ | * | * | * |

All experiments listed in table A10 with q=3 are theoretically unstable (an asterisk for the stability boundary $\beta$ given in (2.20b) means that $S^*$ exceeds its maximal value listed in table 2.2) and with q=1 the method is only unstable for $S^* = 40$ (see table A0). Because of the small number of integration steps (only 7) the instabilities have not yet developed to an extent which seriously affects the numerical solution.

Therefore, these experiments suggest the following conclusions:

(vi) The accuracy of the $SC(q,2,S^*)$ methods is strongly sensitive to grid refinement.

(vii) The sensitivity decreases if $S^*$ increases.

(viii) The accuracy increases if $S^*$ increases as $h \to 0$.

## A5. The effect of the parameters $S^*$ and m

The method $SC(1,m,S^*)$ is completely defined if the parameters m and $S^*$ are specified. In particular, we are interested in the effect of $S^*$. In table A10 this is investigated for the splitting method $SC(q,m,S^*)$ for various values of h with m fixed (m=2) and q=1 and q=3. In table A11 sd-values (4.2) are given for h fixed (h=1/20) and various values of m and $S^*$ (An asterisk indicates that $S^*$ exceeds its maximal value listed in table A0).

Table A11. Results obtained by $SC(1,m,S^*)$ with $\tau = 1/10$ when applied to problem I with $h = 1/20$.

| m | $S^*=0$ | $S^*=10$ | $S^*=20$ | $S^*=40$ | $S^*=80$ |
|---|---|---|---|---|---|
| 1 | 1.1 | 1.6* | 1.7* | 1.8* | 1.9* |
| 2 | 1.4 | 2.5 | 2.8 | 2.7* | 2.3* |
| 3 | 1.7 | 2.8 | 3.2 | 2.9 | 2.7 |
| 4 | 1.8 | 3.0 | 3.5 | 3.3 | 3.0 |
| 5 | 2.0 | 3.2 | 3.7 | 4.0 | 3.4 |
| 6 | 2.2 | 3.4 | 3.9 | 4.2 | 4.0 |

From the results listed in table A10 and A11 we may draw the following conclusions:

(ix) For fixed h and m there is an optimal value for $S^*$.

(x) This optimal value increases if h decreases and is less sensitive to changes in m.

(xi) The rate of convergence slows down for m > 2.

The conclusions (ix) - (xi) roughly apply to $SC(3,m,S^*)$ too.

B. SMOOTHED SECOND ORDER EXTRAPOLATION

Consider the following smoothed predictor

$$\text{(B1a)} \qquad \tilde{y}^{(0)} = \Sigma + b_0\tau\tilde{F}(t_{n+1},\tilde{y}^{(0)},y^{(0)}),$$

$$\text{(B1b)} \qquad y^{(0)} = 3y_n - 3y_{n-1} + y_{n-2},$$

where $\Sigma$ and $b_0$ are defined by (2.23) and $\tilde{F}(t,u,v)$ denotes a Jacobi type splitting function as defined in section 3. The order of the predictor $\tilde{y}^{(0)}$ (B1) is 3 (cf.[4]). Then the order of the method (2.1) - (2.5) with the predictor (B1) is 4 (cf.(2.14')).

The parameter $\theta$ occurring in $\tilde{F}(t,u,v)$ ((3.2) - (3.4)) and the stability boundary $\beta$ for this method can be derived in a similar way as for the SC method (see section 3). Theorem 3.1 implies that the resulting method is stable if (see table 2.1)

$$\text{(B2)} \qquad -1/7 \le \tilde{P}_m(z_1,z_2) \le 0.4951,$$

where $\tilde{P}_m(z_1,z_2)$ is defined by (3.7). The inequality $\tilde{P}_m(z_1,z_2) \ge -1/7$ is satisfied by choosing $P_m(\alpha) \ge -1/7$, $\theta = \frac{7}{8}$, $\tilde{\sigma} \ge \sigma - 1/7b_0\tau$, where $\tilde{\sigma}$ is an estimate of the spectral radius of $\partial f/\partial y$ (see section 3). The inequality $\tilde{P}_m(z_1,z_2) \le 0.4951$ leads to the stability boundary $\beta$ of the method. In table B1 these values for $\beta$ are listed for m = 1 until 4 and $S^*$ as large as allowed by (2.20a).

Table B1. Maximal stability boundaries for BDF4 combined with the third order predictor (B1)

| | m=1 | m=2 | m=3 | m=4 |
|---|---|---|---|---|
| $S^* \simeq$ | .98 | 9.4 | 43 | 131 |
| $\beta(m) \simeq$ | 120 | 755 | 3090 | 9040 |
| $\beta/m^4 \simeq$ | 120 | 47.1 | 38.1 | 35.3 |

The method (2.1) - (2.5) with the predictor (B1b) and the predictor (B1) are denoted by $SC(2,m,S^*)$ and $SC_2(3,m,S^*)$, respectively. In table B2 the sd-values are listed for the modelproblem I obtained by $SC(2,2,4)$ and $SC_2(3,2,4)$.

Table B2. Results obtained by $SC(2,2,4)$ and $SC_2(3,2,4)$ when applied to problem I with $h = 1/10$.

| $\tau$ | $SC(2,2,4)$ | $SC_2(3,2,4)$ |
|---|---|---|
| 1/5 | 2.5 | 3.0 |
| 1/10 | 4.1 | 4.3 |
| 1/20 | 5.2 | 5.2 |
| 1/40 | 6.0 | 6.0 |
| 1/80 | 6.8 | 6.9 |

If $S^* > 0$ relation (A11) indicates that $\Delta sd \cong .9$ for $q=2$ and $\Delta sd \cong 1.2$ for $q=3$, as $\tau \to 0$. For $SC(2,2,4)$ the third order behaviour is more or less reflected in table B2. The fourth order $SC_2(3,2,4)$ method shows only a third order behaviour. It appears that the smoothed predictor (B1) shows its third order behaviour only for relatively small integration steps just as the method of successive corrections [4].

The iterated BDF4 together with the predictor (B1) generate a fourth order accurate splitting method which is stable for the $S^*$-values listed in table B1 and $\tau \leq \beta(m)/\tilde{\sigma}$, where $\beta(m)$ is also listed in table B1 (see also section 3). This method will be denoted by $SC_2$ method.

In table B3 the sd/ce-values are listed for problem I with $h = 1/24$ obtained by $SC_2$.

Table B3. sd/ce-values for problem I with $h = 1/24$ obtained by $SC_2$

| Method | $\tau = 1/5$ | $\tau = 1/10$ | $\tau = 1/20$ | $\tau = 1/40$ |
|---|---|---|---|---|
| $SC_2$ | 3.0/35 | 4.0/50 | 4.8/100 | 5.3/120 |

The order behaviour of the $SC_2$ method is not reflected in table B3. A comparison of the sd/ce-values for problem I with h = 1/24 obtained by SC and $SC_2$ as listed in table 4.2 and B3 shows that the SC method is to be preferred. However, for a given value of m the maximal boundaries for the smoothed second order extrapolation (B1) as listed in table B1 are much larger than the maximal boundaries for the smoothed third order extrapolation as listed in table 3.1.

Acknowledgement. The authors are grateful to Mr. A.J.M. Kaaij for his programming assistance.