

## IMPLICIT–EXPLICIT MULTISTEP METHODS FOR NONLINEAR PARABOLIC EQUATIONS

GEORGIOS AKRIVIS

ABSTRACT. Implicit–explicit multistep methods for nonlinear parabolic equations were recently analyzed in [2, 3, 1]. In these papers the linear operator of the equation is assumed time-independent, self-adjoint and positive definite; then, the linear part is discretized implicitly and the remaining part explicitly. Here we slightly relax the hypotheses on the linear operator by allowing part of it to be time-dependent or nonself-adjoint. We establish optimal order a priori error estimates.

### 1. INTRODUCTION

Let  $T > 0$ ,  $u^0 \in H$ , and consider the initial value problem of seeking  $u : [0, T] \rightarrow \mathcal{D}(A)$  satisfying

$$(1.1) \quad \begin{cases} u'(t) + Au(t) = B(t, u(t)), & 0 < t < T, \\ u(0) = u^0, \end{cases}$$

with  $A$  a positive definite, self-adjoint, linear operator on a Hilbert space  $(H, (\cdot, \cdot))$  with domain  $\mathcal{D}(A)$  dense in  $H$ , and  $B(t, \cdot) : \mathcal{D}(A) \rightarrow H$ ,  $t \in [0, T]$ , a (possibly) nonlinear operator. We assume that (1.1) possesses a smooth solution.

Let  $(\alpha, \beta)$  be a strongly  $A(0)$ –stable  $q$ –step scheme and  $(\alpha, \gamma)$  be an explicit  $q$ –step scheme, characterized by three polynomials  $\alpha, \beta$  and  $\gamma$ ,

$$\alpha(\zeta) = \sum_{i=0}^q \alpha_i \zeta^i, \quad \beta(\zeta) = \sum_{i=0}^q \beta_i \zeta^i, \quad \gamma(\zeta) = \sum_{i=0}^{q-1} \gamma_i \zeta^i$$

We consider three multistep schemes for (1.1): A first possibility is the discretization of (1.1) by the implicit scheme  $(\alpha, \beta)$ ; an advantage is that these schemes have good stability properties, and a drawback that they require solving a *nonlinear* problem at every time level to advance in time (in the case of nonlinear operator  $B(t, \cdot)$ ). A second possibility is the discretization of (1.1) by the implicit–explicit  $(\alpha, \beta, \gamma)$ –scheme; such schemes were constructed and analyzed for linear equations in [6] and later extended to nonlinear equations in [2, 3, 1]. In this case the linear part of the equation is discretized implicitly and the nonlinear part explicitly. The computational advantage is that, to advance in time, at every time level we need to solve one linear problem, with the same operator for all levels. The disadvantage

---

Received by the editor April 15, 2011 and, in revised form, September 12, 2011.

2010 *Mathematics Subject Classification*. Primary 65M12, 65M60; Secondary 65L06.

*Key words and phrases*. Nonlinear parabolic equations, implicit–explicit multistep methods, BDF methods, strong  $A(0)$ –stability,  $A(\vartheta)$ –stability,  $G$ –stability.

Work supported in part by University of Cyprus grant #8037P-3/311-21028.

is that these schemes are locally stable under restrictive conditions, namely that  $B(t, \cdot)$  is dominated, in a certain sense, by the operator  $A$ ; see the discussion in subsection 1.2. Our purpose here is to propose and analyze an intermediate alternative, namely to discretize part of  $B(t, \cdot)$  implicitly, essentially in an attempt to relax the hypotheses on the linear part of the equation. If  $B(t, \cdot)$  consists of linear and nonlinear parts, this alternative allows us to advance in time by solving one linear problem and it has, in general, better stability properties than the standard  $(\alpha, \beta, \gamma)$ -scheme.

**1.1. Three multistep schemes.** Let  $N \in \mathbb{N}$ ,  $k := T/N$  be the constant time step, and  $t^n := nk, n = 0, \dots, N$ , be a uniform partition of the interval  $[0, T]$ . Since we will consider  $q$ -step schemes, we assume that starting approximations  $U^0, \dots, U^{q-1}$  are given. The first choice is to discretize (1.1) by the implicit scheme  $(\alpha, \beta)$ . Thus, we recursively define a sequence of approximations  $U^m$  to the nodal values  $u^m := u(t^m)$  by

$$(1.2) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) U^{n+i} = k \sum_{i=0}^q \beta_i B(t^{n+i}, U^{n+i});$$

this is the standard  $(\alpha, \beta)$ -scheme, used for both the linear and the nonlinear part. In the case of nonlinear  $B$ , to advance in time, i.e., to compute  $U^{n+q}$ , we need to solve one *nonlinear* problem at every time level.

The second choice is to use the implicit scheme  $(\alpha, \beta)$  for the discretization of the linear part and the explicit scheme  $(\alpha, \gamma)$  for the discretization of the nonlinear part of the equation; see [2], [3]. We recursively define a sequence of approximations  $U^m$  to  $u^m$  by the corresponding implicit-explicit  $(\alpha, \beta, \gamma)$ -scheme for (1.1),

$$(1.3) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) U^{n+i} = k \sum_{i=0}^{q-1} \gamma_i B(t^{n+i}, U^{n+i}).$$

The main computational advantage of this scheme is that it is easy to implement: to advance in time, we need to solve one linear problem with the *same operator* at every time level, namely  $\alpha_q I + k\beta_q A$ ; the linear problems reduce to linear systems with the same coefficient matrix in the fully discrete case, i.e., when the time discretization is combined with discretization in space. These schemes are analyzed in [2] and [3]; see also [1] for a wider class of linearly implicit schemes. Let us emphasize that in the analysis in [2], [3] and [1] the assumption that the linear, time-independent operator  $A$  is self-adjoint and positive definite plays a key role.

Our purpose here is to propose an intermediate alternative to the two extreme cases (1.3) and (1.2). These schemes have in general better stability properties than (1.3) (but worse than (1.2)) and are easier to implement than (1.2) (but more expensive than (1.3)). To construct the schemes, we decompose  $B$  into two parts,  $B_1$  and  $B_2$ ,  $B(t, v) = B_1(t, v) + B_2(t, v)$ . In the third choice, the alternative we will consider here, both the linear part as well as  $B_1$  are discretized by the implicit scheme  $(\alpha, \beta)$  and the remaining part  $B_2$  of  $B$  is discretized by the explicit scheme  $(\alpha, \gamma)$ . Thus, we define approximations  $U^m$  to the values  $u^m$  as follows

$$(1.4) \quad \sum_{i=0}^q [\alpha_i U^{n+i} + k\beta_i (AU^{n+i} - B_1(t^{n+i}, U^{n+i}))] = k \sum_{i=0}^{q-1} \gamma_i B_2(t^{n+i}, U^{n+i}).$$

Assume now that  $B_1$  is linear; interesting cases are when  $B_1$  is not self-adjoint and/or has time-dependent coefficients. Then, to advance with (1.4) in time, we need to solve one linear problem, however, in general with an operator depending on the time level. This may be viewed as an attempt to relax the hypotheses imposed in [2, 3] on the linear part of the equation in (1.1); indeed, the operator  $B_1(t, \cdot)$  may be viewed as a perturbation of the self-adjoint and positive definite operator  $A$  of [3, 2, 1]. We will see that in some cases (1.4) has better stability properties than (1.3). If the original equation is given in the form  $u'(t) + A(t)u(t) = B_2(t, u(t))$ , then a natural decomposition of the linear operator  $A(t)$  in the desired form  $A - B_1(t, \cdot)$  would be  $A := [A(0) + A(0)^*]/2$  and  $B_1(t, \cdot) := A - A(t)$ , with  $A(0)^*$  the adjoint of  $A(0)$ . The analysis may be easily modified to allow a finite number of different decompositions of  $A(t)$  in a partition of  $[0, T]$ .

**1.2. Stability assumptions.** Here we will formulate stability conditions for the schemes (1.3), (1.2) and (1.4). To this end we first introduce notation.

Let  $|\cdot|$  denote the norm of  $H$ , and introduce in  $V$ ,  $V := \mathcal{D}(A^{1/2})$ , the norm  $\|\cdot\|$  by  $\|v\| := |A^{1/2}v|$ . We identify  $H$  with its dual, and denote by  $V'$  the dual of  $V$ , and by  $\|\cdot\|_*$  the dual norm on  $V'$ ,  $\|v\|_* := |A^{-1/2}v|$ . We use the notation  $(\cdot, \cdot)$  also for the duality pairing between  $V'$  and  $V$ ; then  $\|v\| = (Av, v)^{1/2}$  and  $\|v\|_* = (v, A^{-1}v)^{1/2}$ . For stability purposes, we assume that  $B(t, \cdot)$  can be extended to an operator from  $V$  into  $V'$ , and satisfies a local Lipschitz condition of the form

$$(1.5) \quad \|B(t, v) - B(t, w)\|_* \leq \lambda \|v - w\| + \mu |v - w| \quad \forall v, w \in T_u$$

in a tube  $T_u, T_u := \{v \in V : \min_t \|v - u(t)\| \leq 1\}$ , around the solution  $u$ , uniformly in  $t$ , with the *stability constant*  $\lambda < 1$  and a constant  $\mu$ .

We also assume that  $B_1(t, \cdot)$  and  $B_2(t, \cdot)$  can be extended to operators from  $V$  into  $V'$ , and satisfy the global Lipschitz condition (see, however, Remark 6.1 below)

$$(1.6) \quad \|B_1(t, v) - B_1(t, w)\|_* \leq \lambda_1 \|v - w\| + \mu_1 |v - w| \quad \forall v, w \in V$$

and the local Lipschitz condition

$$(1.7) \quad \|B_2(t, v) - B_2(t, w)\|_* \leq \lambda_2 \|v - w\| + \mu_2 |v - w| \quad \forall v, w \in T_u,$$

respectively, uniformly in  $t$ , with the nonnegative *stability constants*  $\lambda_1$  and  $\lambda_2$ , and constants  $\mu_1$  and  $\mu_2$ . The decomposition  $B = B_1 + B_2$  is particularly useful if

$$(1.8) \quad \lambda = \lambda_1 + \lambda_2.$$

For simplicity, we assume in the sequel that (1.8) is satisfied.

Depending on the particular scheme we will use for discretizing (1.1) in time, it will be essential for our analysis that  $\lambda$ ,  $\lambda_1$  and  $\lambda_2$  be appropriately small. To make this precise, we introduce the constants  $K_{(\alpha, \beta)}$  and  $K_{(\alpha, \beta, \gamma)}$  by

$$(1.9) \quad K_{(\alpha, \beta)} := \sup_{x>0} \max_{\zeta \in \mathcal{K}} \left| \frac{x\beta(\zeta)}{(\alpha + x\beta)(\zeta)} \right|, \quad K_{(\alpha, \beta, \gamma)} := \sup_{x>0} \max_{\zeta \in \mathcal{K}} \left| \frac{x\gamma(\zeta)}{(\alpha + x\beta)(\zeta)} \right|,$$

with  $\mathcal{K}$  denoting the unit circle in the complex plane,  $\mathcal{K} := \{z \in \mathbb{C} : |z| = 1\}$ . Under our hypotheses, the constants  $K_{(\alpha, \beta, \gamma)}$  and  $K_{(\alpha, \beta)}$  are finite; cf. [3].

It is shown in [3] that the implicit-explicit  $(\alpha, \beta, \gamma)$ -scheme (1.3) is locally stable for (1.1), provided (1.5) is satisfied with a constant  $\lambda$  small enough such that

$$(1.10) \quad K_{(\alpha, \beta, \gamma)} \lambda < 1.$$

Let us also note that (1.10) is sharp in the following sense: for any constant  $\lambda$  such that  $K_{(\alpha,\beta,\gamma)}\lambda > 1$ , there exists a (linear) operator  $B$  satisfying (1.5) such that the implicit–explicit  $(\alpha, \beta, \gamma)$ –scheme is unstable for (1.1).

Also, it follows from the analysis of [3] that the implicit  $(\alpha, \beta)$ –scheme (1.2) is locally stable for (1.1), provided (1.5) is satisfied and

$$(1.11) \quad K_{(\alpha,\beta)}\lambda < 1;$$

this stability condition is sharp in the same sense as well.

Concerning the intermediate scheme (1.4) for (1.1), we will see that it is locally stable, provided the constants  $\lambda_1$  and  $\lambda_2$  in the Lipschitz conditions (1.6) and (1.7) are small enough such that

$$(1.12) \quad K_{(\alpha,\beta)}\lambda_1 + K_{(\alpha,\beta,\gamma)}\lambda_2 < 1.$$

It seems natural that in this stability condition the constants  $\lambda_1$  and  $\lambda_2$  are multiplied by  $K_{(\alpha,\beta)}$  and  $K_{(\alpha,\beta,\gamma)}$ , respectively, since  $B_1$  and  $B_2$  are discretized implicitly and explicitly, respectively. Furthermore, we give a necessary stability condition and will see that (1.12) is sharp, provided the scheme  $(\alpha, \beta)$  is  $A$ –stable.

Finally, concerning the tube  $T_u$ , we emphasize that it is defined in terms of the norm of  $V$  for concreteness. The analysis may be modified to yield convergence under conditions analogous to (1.5), (1.6) and (1.7) for  $v$  and  $w$  belonging to tubes defined in terms of other norms, not necessarily the same for both arguments; see [3]. Actually, in the error estimation we only need to apply the Lipschitz conditions with one of the arguments being  $u(t)$ , for  $t \in [0, T]$ .

**1.3. Consistency assumptions.** For simplicity, we assume that the order of both  $q$ –step schemes, the implicit  $(\alpha, \beta)$  and the explicit  $(\alpha, \gamma)$ , is  $p$ , i.e.,

$$(1.13) \quad \sum_{i=0}^q i^\ell \alpha_i = \ell \sum_{i=0}^q i^{\ell-1} \beta_i = \ell \sum_{i=0}^{q-1} i^{\ell-1} \gamma_i, \quad \ell = 0, 1, \dots, p.$$

The implicit–explicit BDF methods, see (2.15) below, satisfy (1.13) with  $p = q$ . For further examples of schemes satisfying our stability and consistency assumptions we refer to [2] and [3], and the references therein. Let us also note that, if  $p \geq 1$ , then  $\gamma(1) = \beta(1)$ , whence  $K_{(\alpha,\beta,\gamma)} \geq 1$ .

An outline of the paper is as follows: In Section 2 we present some properties of the constants  $K_{(\alpha,\beta,\gamma)}$  and  $K_{(\alpha,\beta)}$ . In Section 3 we establish existence and uniqueness of the approximations  $U^q, \dots, U^N$ , and in Section 4 we prove consistency. In Section 5 we establish our main result, namely local stability of the scheme (1.4) under the stability condition (1.12). In Section 6 we combine the results of Sections 4 and 5 and derive optimal order error estimates for the implicit–explicit scheme (1.4) in the norm  $|\cdot|$ , assuming (1.6), (1.7) and (1.12), that the order of our scheme is  $p$ , and that appropriate starting approximations  $U^0, \dots, U^{q-1}$  are given. Finally, in Section 7 we focus on first- and second-order schemes and relax the Lipschitz condition (1.6) on  $B_1$  to a one-sided one.

The analysis in Sections 4, 5 and 6 is similar to the one in [3]; see also [1].

Monotonicity and boundedness properties of implicit–explicit multistep methods are discussed in [11]. For various time-stepping schemes for parabolic equations and their properties we refer to the classical monograph in this field, namely [17].

2. SOME PROPERTIES OF  $K_{(\alpha,\beta,\gamma)}$  AND  $K_{(\alpha,\beta)}$ 

After summarizing well-known properties of  $A(\vartheta)$ -stable multistep schemes, in this section we discuss some properties of the constants  $K_{(\alpha,\beta,\gamma)}$  and  $K_{(\alpha,\beta)}$ . In particular, we give some useful representations for them. The implicit-explicit BDF schemes are discussed in detail. We also comment on the sharpness of the stability condition (1.12).

**2.1.  $A(\vartheta)$ -stability.** For the convenience of the reader, we summarize some well-known facts about  $A(\vartheta)$ -stability and strong  $A(0)$ -stability for an implicit multistep scheme  $(\alpha, \beta)$ ; see [9, 15, 13, 12, 8, 7]. First, for  $z \in \mathbb{C}$ , consider the polynomial  $\chi(z, \cdot) := \alpha(\cdot) - z\beta(\cdot)$ . The *stability region*  $S$  of the scheme  $(\alpha, \beta)$  consists of the complex numbers  $z$  such that the roots  $\zeta_j(z)$  of  $\chi(z, \cdot)$  do not exceed 1 in modulus and unimodular roots are simple, i.e., the modulus of every multiple root is (strictly) less than 1. Now let  $0 < \vartheta < \pi/2$ . The scheme  $(\alpha, \beta)$  is called  *$A(\vartheta)$ -stable*, if the sector  $S_\vartheta := \{z \in \mathbb{C} : z = -\rho e^{i\varphi}, \rho \geq 0, |\varphi| \leq \vartheta\}$  is contained in its stability region  $S$ . If the scheme  $(\alpha, \beta)$  is  $A(\vartheta)$ -stable, then, for  $z$  in the interior of the sector  $S_\vartheta$ , i.e.,  $z = -\rho e^{i\varphi}$  with  $\rho > 0$  and  $|\varphi| < \vartheta$ , the polynomial  $\chi(z, \cdot)$  does not have unimodular roots. The scheme is called  *$A$ -stable*, if it is  $A(\vartheta)$ -stable for all  $0 < \vartheta < \pi/2$ , and it is called  *$A(0)$ -stable*, if it is  $A(\vartheta)$ -stable for some  $\vartheta > 0$ . The highest attainable order of an  $A(\vartheta)$ -stable  $q$ -step scheme is  $p = q$ , with only one exception, namely the trapezoidal method, for which  $q = 1$  and  $p = q + 1 = 2$ . Such schemes are implicit; indeed the product  $\alpha_q \beta_q$  is positive. See, e.g., [9, §V.1 and §V.2].

An  $A(0)$ -stable scheme  $(\alpha, \beta)$  is called *strongly  $A(0)$ -stable*, if the roots of  $\beta$  are strictly less than 1 in modulus. This can be equivalently formulated as follows: We order the roots  $\zeta_j(x)$ , respectively  $\zeta_j(-\infty)$ ,  $1 \leq j \leq q$ , of the polynomial  $\chi(x, \cdot)$ , respectively  $\beta(\cdot)$ , in such a way that the functions  $\zeta_j(\cdot)$  are continuous on  $[-\infty, 0]$  and that the roots  $\xi_j := \zeta_j(0)$ ,  $j = 1, \dots, s$ , satisfy  $|\xi_j| = 1$ ; these unimodular roots are called the *principal* roots of  $\alpha$  and the complex numbers  $\lambda_j := \frac{\beta(\xi_j)}{\xi_j \alpha'(\xi_j)}$  are called the *growth factors* of  $\xi_j$ . Then, the method  $(\alpha, \beta)$  is strongly  $A(0)$ -stable, if and only if

$$(i) \quad \text{for all } -\infty \leq x < 0 \text{ and for all } j = 1, \dots, q, \text{ there holds } |\zeta_j(x)| < 1,$$

and

$$(ii) \quad \text{the principal roots of } \alpha \text{ are simple and satisfy: } \operatorname{Re} \lambda_j > 0, j = 1, \dots, s.$$

The motivation for the term *growth factor* is the fact that  $\zeta_j'(0) = \lambda_j \xi_j$ . Therefore,  $\zeta_j(x) = \xi_j(1 + \lambda_j x) + O(x^2)$ , whence  $|\zeta_j(x)|^2 = 1 + x \operatorname{Re} \lambda_j + O(x^2)$ , and we infer that, for  $\operatorname{Re} \lambda_j > 0$ , the root  $\zeta_j(x)$  moves inside the unit circle as  $x$  moves to the negative real axis. Concerning condition (ii), let us mention that for a *zero-stable* scheme  $(\alpha, \beta)$ , i.e., such that 0 is contained in its stability region  $S$ , and *consistent* scheme, i.e.,  $\alpha(1) = 0$  and  $\alpha'(1) = \beta(1)$ ,  $\xi_1 = 1$  is a principal root of  $\alpha$  and its growth factor is  $\lambda_1 = 1$ . All other principal roots are called *parasitic*; practical methods with stability regions  $S$  with nonempty interior, do not have parasitic roots, in which case condition (ii) becomes void. Also, practical methods are irreducible, i.e., the polynomials  $\alpha$  and  $\beta$  have no roots in common.

For any  $\vartheta \in (0, \pi/2)$  and  $q \in \mathbb{N}$ , there are known  $A(\vartheta)$ -stable and strongly  $A(0)$ -stable  $q$ -step schemes of order  $p = q$ . A drawback of these schemes is that their consistency constants deteriorate as  $\vartheta$  approaches  $\pi/2$ . See, e.g., [9, §V.2].

Let  $d(\zeta) := \alpha(\zeta)/\beta(\zeta) = -\rho e^{i\varphi(\zeta)}$ , for  $\zeta$  in the unit circle  $\mathcal{K}$  such that  $\beta(\zeta) \neq 0$ , represent the points of the *root locus curve* of the scheme  $(\alpha, \beta)$ , with  $\rho \geq 0$  and  $-\pi \leq \varphi(\zeta) < \pi$ . Since the coefficients of  $\alpha$  and  $\beta$  are real, we have  $d(\bar{\zeta}) = \overline{d(\zeta)}$ , i.e., the root locus curve is symmetric with respect to the real axis. The complement of the root locus curve is open and consists of a finite number of connected components. If a point of such a component belongs to the (interior of the) stability region, then all its points belong to the interior of the stability region. The boundary of the stability region consists of parts of the root locus curve; see, e.g., [9, §V.1]. If the scheme  $(\alpha, \beta)$  is zero-stable and consistent, then  $d(1) = 0$  and the root locus curve cuts the real axis at the origin. For the trapezoidal method, we have  $d(\zeta) = 2(\zeta - 1)/(\zeta + 1)$ , whence  $d(\zeta) + d(\bar{\zeta}) = 0$ , and the root locus curve is the imaginary axis; in particular, the root locus curve cuts the real axis only at the origin in this case. If  $\alpha(-1)\beta(-1) \neq 0$ , the root locus curve cuts the real axis at the additional point  $d(-1)$ .

In analogy to the growth factors  $\lambda_j$ , let the growth factors  $\mu_j$  of unimodular roots  $\zeta_j$  of  $\beta$  be given by  $\mu_j := \frac{\alpha(\zeta_j)}{\zeta_j \beta'(\zeta_j)}$ . Then we have the following criterion for the  $A(\vartheta)$ -stability. With  $\vartheta \in (0, \pi/2)$ , the scheme  $(\alpha, \beta)$  is  $A(\vartheta)$ -stable, if and only if the following conditions are satisfied, cf. [8], [7, p. 225]:

- a) The roots of  $\beta$  do not exceed 1 in modulus and its unimodular roots are simple.
- b) All growth factors  $\mu_j$  of the unimodular roots of  $\beta$  have positive real parts,  $\operatorname{Re} \mu_j > 0$ .
- c) For all points  $d(\zeta)$  of the root locus curve there holds

$$(2.1) \quad |\operatorname{Im} d(\zeta)| + \tan \vartheta \operatorname{Re} d(\zeta) \geq 0.$$

Obviously, the geometric interpretation of (2.1) is that the root locus curve is located outside the sector  $S_\vartheta$ . According to the above criterion, the location of the root locus curve determines the  $A(\vartheta)$ -stability of the scheme, if and only if conditions a) and b) are satisfied.

If  $\vartheta_{\max} < \pi/2$  is the maximum value of  $\vartheta$ , for which the scheme  $(\alpha, \beta)$  is  $A(\vartheta)$ -stable, then we obtain from (2.1) the relation

$$(2.2) \quad \sin \vartheta_{\max} = \inf \left\{ \frac{|\operatorname{Im} d(\zeta)|}{|d(\zeta)|} : \zeta \in \mathcal{K}, \operatorname{Re} d(\zeta) < 0 \right\}.$$

This relation can be rewritten in the form

$$(2.3) \quad \sin \vartheta_{\max} = \inf \left\{ \frac{|\operatorname{Im} \alpha(\zeta)\beta(\bar{\zeta})|}{|\alpha(\zeta)\beta(\bar{\zeta})|} : \zeta \in \mathcal{K}, \operatorname{Re}[\alpha(\zeta)\beta(\bar{\zeta})] < 0 \right\}.$$

Since  $d(\bar{\zeta}) = \overline{d(\zeta)}$ , we can replace the unit circle  $\mathcal{K}$  in (2.3) by the half-circle  $\{\zeta = e^{i\varphi} : 0 \leq \varphi \leq \pi\}$ . Now,

$$(2.4) \quad \alpha(e^{i\varphi})\beta(e^{-i\varphi}) = \sum_{j=0}^q d_j \cos(j\varphi) + i \sum_{j=0}^q \tilde{d}_j \sin(j\varphi),$$

with real coefficients  $d_j, \tilde{d}_j$ . Changing variables,  $x = \cos \varphi$ , and using the Chebyshev polynomials  $T_j(x) = \cos(j\varphi)$  and  $U_j(x) = \sin(j\varphi)/\sin \varphi$  of the first and the second

kind, respectively, and setting

$$R_q(x) = \sum_{j=0}^q d_j T_j(x), \quad I_q(x) = \sum_{j=0}^q \tilde{d}_j U_j(x),$$

we can rewrite (2.3) in the form

$$(2.5) \quad \sin \vartheta_{\max} = \inf \left\{ \frac{\sqrt{1-x^2} |I_q(x)|}{\sqrt{[I_q(x)]^2 + [R_q(x)]^2}} : x \in [-1, 1], R_q(x) < 0 \right\};$$

see [15].

**2.2. The constants  $K_{(\alpha, \beta, \gamma)}$  and  $K_{(\alpha, \beta)}$ .** For a fixed scheme  $(\alpha, \beta)$ , we introduce the parts  $\mathcal{H}^+$  and  $\mathcal{H}^-$  of the unit circle  $\mathcal{H}$  according to the sign of  $\operatorname{Re} d(\zeta)$ ,

$$\mathcal{H}^+ := \{\zeta \in \mathcal{H} : \operatorname{Re} d(\zeta) \geq 0\}, \quad \mathcal{H}^- := \{\zeta \in \mathcal{H} : \operatorname{Re} d(\zeta) < 0\},$$

with  $d(\zeta)$  the points of the root locus curve as in subsection 2.1. Now, obviously, for  $\zeta \in \mathcal{H}^+$ , we have  $\sup_{x>0} \frac{1}{|1+x^{-1}d(\zeta)|} = 1$ , and, for  $\zeta \in \mathcal{H}^-$ ,

$$\sup_{x>0} \frac{1}{|1+x^{-1}d(\zeta)|} = \left( \sup_{y>0} \frac{1}{1+2y \operatorname{Re} d(\zeta) + y^2 |d(\zeta)|^2} \right)^{1/2} = \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|} = \frac{1}{|\sin \varphi(\zeta)|},$$

i.e.,

$$(2.6) \quad \sup_{x>0} \frac{1}{|1+x^{-1}d(\zeta)|} = \begin{cases} 1 & \forall \zeta \in \mathcal{H}^+, \\ \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|} = \frac{1}{|\sin \varphi(\zeta)|} & \forall \zeta \in \mathcal{H}^-. \end{cases}$$

Therefore, with

$$k(x, \zeta) := \frac{x\gamma(\zeta)}{(\alpha + x\beta)(\zeta)} = \frac{1}{1+x^{-1}d(\zeta)} \frac{\gamma(\zeta)}{\beta(\zeta)}$$

we have

$$(2.7) \quad \forall \zeta \in \mathcal{H}^+ \quad \sup_{x>0} |k(x, \zeta)| = \frac{|\gamma(\zeta)|}{|\beta(\zeta)|}$$

and

$$(2.8) \quad \forall \zeta \in \mathcal{H}^- \quad \sup_{x>0} |k(x, \zeta)| = \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|}.$$

Now, from (2.7), (2.8) and the definition of  $K_{(\alpha, \beta, \gamma)}$  we infer that

$$(2.9) \quad K_{(\alpha, \beta, \gamma)} = \max \left\{ \max_{\zeta \in \mathcal{H}^+} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|}, \sup_{\zeta \in \mathcal{H}^-} \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|} \right\}$$

or, equivalently,

$$(2.10) \quad K_{(\alpha, \beta, \gamma)} = \max \left\{ \max_{\zeta \in \mathcal{H}^+} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|}, \sup_{\zeta \in \mathcal{H}^-} \frac{1}{|\sin \varphi(\zeta)|} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|} \right\}.$$

Notice that in (2.10) the function  $|\gamma(\zeta)/\beta(\zeta)|$  is multiplied by 1 or by  $1/|\sin \varphi(\zeta)|$ , respectively, if  $\zeta \in \mathcal{H}$  is mapped to a point  $d(\zeta)$  of the root locus curve on the right or the left complex half-plane, respectively. Here,  $\varphi(\zeta)$  is the angle between the negative real half-axis and the half-line passing through the origin and the point

$d(\zeta)$  of the root locus curve. In particular, if the implicit scheme  $(\alpha, \beta)$  is  $A$ -stable, then we have  $\operatorname{Re} d(\zeta) \geq 0$  for all  $\zeta \in \mathcal{H}$ , i.e.,  $\mathcal{H}^+ = \mathcal{H}$ , and (2.9) reduces to

$$(2.11) \quad K_{(\alpha, \beta, \gamma)} = \max_{\zeta \in \mathcal{H}} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|}.$$

The boundedness of  $K_{(\alpha, \beta, \gamma)}$  was established in [3]; it is also a trivial consequence of (2.10) and the definition of the strong  $A(0)$ -stability of the scheme  $(\alpha, \beta)$ .

Furthermore, concerning  $K_{(\alpha, \beta)}$ , we immediately infer from (2.11) that

$$(2.12) \quad K_{(\alpha, \beta)} = 1$$

in case the scheme  $(\alpha, \beta)$  is  $A$ -stable, and from (2.9) that

$$(2.13) \quad K_{(\alpha, \beta)} = \sup_{\zeta \in \mathcal{H}^-} \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|}$$

otherwise. Combining (2.13) with (2.2), we conclude that

$$(2.14) \quad K_{(\alpha, \beta)} = \frac{1}{\sin \vartheta}$$

for an  $A(\vartheta)$ -stable scheme with  $\vartheta$  as large as possible. Relation (2.5) can be used to compute  $K_{(\alpha, \beta)}$  in case  $\vartheta$  is not known a priori.

*Remark 2.1* (The role of the strong  $A(0)$ -stability). The strong  $A(0)$ -stability of the implicit scheme  $(\alpha, \beta)$  is essential for the boundedness of  $K_{(\alpha, \beta, \gamma)}$  in the sense that it cannot be relaxed to  $A(0)$ -stability or be replaced by  $A$ -stability. We illustrate this with an example, namely the trapezoidal method. In this case we have  $\alpha(\zeta) = \zeta - 1$ ,  $\beta(\zeta) = (\zeta + 1)/2$ , and  $\gamma$  is constant; we choose  $\gamma(\zeta) = 1$ , for consistency. Since the trapezoidal method is  $A$ -stable, we have, for  $\zeta \neq -1$ ,

$$\sup_{x > 0} \frac{x}{|\alpha(\zeta) + x\beta(\zeta)|} = \frac{1}{|\beta(\zeta)|} = \frac{2}{|\zeta + 1|};$$

as  $\zeta$  tends to  $-1$  this quantity tends to infinity. Notice, however, that  $A(\vartheta)$ -stability suffices for the boundedness of  $K_{(\alpha, \beta)}$ ; see (2.14).  $\square$

**2.3. The BDF schemes.** An interesting example of multistep methods satisfying our assumptions are the BDF methods,

$$(2.15) \quad \alpha(\zeta) = \sum_{j=1}^q \frac{1}{j} \zeta^{q-j} (\zeta - 1)^j, \quad \beta(\zeta) = \zeta^q, \quad \gamma(\zeta) = \zeta^q - (\zeta - 1)^q.$$

The corresponding implicit  $(\alpha, \beta)$ -schemes are the well-known BDF methods, which are strongly  $A(0)$ -stable for  $q = 1, \dots, 6$ ; their order is  $p = q$ . For a given  $\alpha$ , the scheme  $(\alpha, \gamma)$  is the unique explicit  $q$ -step scheme of order  $p = q$ . For these methods, the constants  $K_{(\alpha, \beta, \gamma)}$  in (1.9) are known, namely

$$(2.16) \quad K_{(\alpha, \beta, \gamma)} = |\gamma(-1)| = 2^q - 1;$$

see [3]. Let us mention that  $d(-1) = \sum_{j=1}^q 2^j/j > 0$ , whence, in the notation of (2.9) and (2.10), the maximum is attained at a point of the unit circle mapped at a point of the root locus curve in the right complex half-plane. Indeed, the second term in (2.9) is bounded by  $2^q - 5$ ; see [3]. Furthermore, in view of (2.14),

$$(2.17) \quad K_{(\alpha, \beta)} = \frac{1}{\sin \vartheta_q}$$

with  $\vartheta_1 = \vartheta_2 = 90^\circ$ ,  $\vartheta_3 = 86.03^\circ$ ,  $\vartheta_4 = 73.35^\circ$ ,  $\vartheta_5 = 51.84^\circ$  and  $\vartheta_6 = 17.84^\circ$ ; see [9]. We refer to [9, Fig. 1.6] for the root locus curves and the stability regions of the BDF schemes. The quantities  $K_{(\alpha,\beta,\gamma)}$ ,  $K_{(\alpha,\beta)}$  and their ratios are given in Table 2.1.

$q$	$K_{(\alpha,\beta,\gamma)}$	$K_{(\alpha,\beta)}$	$K_{(\alpha,\beta,\gamma)}/K_{(\alpha,\beta)}$	$\lambda_{(\alpha,\beta)}$
2	3	1	3	1
3	7	1.002402461	6.983223079	0.9976032970
4	15	1.043752810	14.371218795	0.9580812530
5	31	1.271802188	24.374859780	0.7862857993
6	63	3.264173630	19.300443892	0.3063562523

TABLE 2.1. The quantities  $K_{(\alpha,\beta,\gamma)}$ ,  $K_{(\alpha,\beta)}$  and their ratios for the  $q$ -step BDF schemes,  $q = 2, \dots, 6$ , as well as the stability constants  $\lambda_{(\alpha,\beta)} = 1/K_{(\alpha,\beta)}$  for the implicit BDF schemes  $(\alpha, \beta)$ .

Obviously, for the BDF schemes with  $q = 2, \dots, 6$ , the stability condition (1.12) for the scheme (1.4) is less restrictive than the stability condition (1.10) for the implicit–explicit  $(\alpha, \beta, \gamma)$ -scheme (1.3), provided that (1.8) is satisfied and  $\lambda_1$  is positive. For every  $q$ , the improvement is quantified by the ratio  $\lambda_1/\lambda_2$ : if  $\lambda_1/\lambda_2 = 0$ , the stability conditions (1.12) and (1.10) coincide; however, as the value of this ratio becomes larger, condition (1.12) becomes weaker and therefore  $\lambda$  is allowed to attain larger values in scheme (1.4) than in scheme (1.3). In the extreme case  $\lambda_2 = 0$ , the stability constant in (1.12) may be  $K_{(\alpha,\beta,\gamma)}/K_{(\alpha,\beta)}$  times larger than the one in (1.10); cf. Table 2.1. In this latter case the stability condition for the scheme (1.4) reduces to the one for the implicit  $(\alpha, \beta)$ -scheme (1.2).

Since in general the scheme (1.4) is more expensive than (1.3) while their stability conditions coincide for  $\lambda_1 = 0$ , from the stability point of view, the scheme (1.4) is of interest only for positive  $\lambda_1$ . In the applications, this means that  $B_1(t, \cdot)$  and  $A$  are differential operators of the same order; cf. [3, 2].

*Remark 2.2* (A modified two-step BDF scheme). The modified implicit–explicit two-step BDF scheme  $(\alpha, \tilde{\beta}, \gamma)$ ,

$$\alpha(\zeta) = \frac{3}{2}\zeta^2 - 2\zeta + \frac{1}{2}, \quad \tilde{\beta}(\zeta) = \frac{3}{2}\zeta^2 - \zeta + \frac{1}{2}, \quad \gamma(\zeta) = 2\zeta - 1,$$

was proposed in [4]; notice that the polynomials  $\alpha$  and  $\gamma$  coincide with the ones of the implicit–explicit two-step BDF scheme, while  $\beta$  has been modified to  $\tilde{\beta}$ . Since the roots of  $\tilde{\beta}$  are strictly less than 1 in modulus and  $\operatorname{Re}[\alpha(\zeta)\tilde{\beta}(\bar{\zeta})] = 3(\operatorname{Re} \zeta - 1)^2 \geq 0$ , for all  $\zeta \in \mathcal{K}$ , the second order implicit scheme  $(\alpha, \tilde{\beta})$  is  $A$ -stable and strongly  $A(0)$ -stable; cf. (2.1). Thus,  $K_{(\alpha,\tilde{\beta})} = 1$ . Furthermore, for the  $(\alpha, \tilde{\beta}, \gamma)$ -scheme there holds

$$K_{(\alpha,\tilde{\beta},\gamma)} = \max_{\zeta \in \mathcal{K}} \frac{|\gamma(\zeta)|}{|\tilde{\beta}(\zeta)|} = \frac{|\gamma(e^{i\pi/3})|}{|\tilde{\beta}(e^{i\pi/3})|} = 2,$$

while  $K_{(\alpha,\beta,\gamma)} = 3$  for the implicit–explicit two-step BDF scheme  $(\alpha, \beta, \gamma)$ . For further implicit–explicit schemes with good stability properties we refer to [4].  $\square$

**2.4. On the sharpness of the stability condition (1.12).** Here, we will give a necessary stability condition for (1.4) and discuss its discrepancy to the sufficient condition (1.12). We shall also see that (1.12) is sharp, if the implicit scheme  $(\alpha, \beta)$  is  $A$ -stable.

For nonnegative  $\lambda_1, \lambda_2$ , we shall see that a necessary stability condition for (1.4) is

$$(2.18) \quad \sup_{x>0} \max_{\zeta \in \mathcal{K}} \frac{\lambda_1 |\beta(\zeta)| + \lambda_2 |\gamma(\zeta)|}{|x^{-1}\alpha(\zeta) + \beta(\zeta)|} \leq 1.$$

As before, (2.18) is necessary if we want (1.4) to be locally stable for all equations satisfying our assumptions with the given stability constants  $\lambda_1$  and  $\lambda_2$ . Unfortunately, it is not clear whether (2.18) is also sufficient for stability. The difference in the left-hand sides of the sufficient and necessary, respectively, stability conditions (1.12) and (2.18) is that the sum of the suprema in the former is replaced by the supremum of the sum in the latter. Also, in view of (2.6), condition (2.18) can be equivalently written in the form

$$(2.19) \quad \lambda_1 + \lambda_2 \max_{\zeta \in \mathcal{K}^+} \frac{|\gamma(\zeta)|}{|\beta(\zeta)|} \leq 1, \quad \sup_{\zeta \in \mathcal{K}^-} \left[ \frac{|d(\zeta)|}{|\operatorname{Im} d(\zeta)|} \left( \lambda_1 + \lambda_2 \frac{|\gamma(\zeta)|}{|\beta(\zeta)|} \right) \right] \leq 1.$$

A necessary, in general weaker, stability condition for (1.4), namely,

$$(2.20) \quad \lambda_1 + K_{(\alpha, \beta, \gamma)} \lambda_2 \leq 1, \quad K_{(\alpha, \beta)} (\lambda_1 + c_{(\beta, \gamma)}^* \lambda_2) \leq 1,$$

with  $c_{(\beta, \gamma)}^* := |\gamma(\zeta^*)|/|\beta(\zeta^*)|$ , if  $K_{(\alpha, \beta)}$  is attained at  $\zeta^* \in \mathcal{K}^-$ , follows immediately from (2.19). Condition (2.20) is simpler than (2.19) in that the constants involved are independent of  $\lambda_1$  and  $\lambda_2$ ; this is the case also for the sufficient condition (1.12).

In particular, in the case of an  $A$ -stable implicit scheme  $(\alpha, \beta)$ , the second condition in (2.19) is void, and the left-hand side of the first condition coincides with the one of the sufficient condition (1.12), whence, in this case, (1.12) is sharp.

Assume now that (2.18) is not satisfied. Then, obviously, with an appropriate  $\Theta \in [0, 2\pi)$ , for the function  $k$ ,

$$(2.21) \quad k(x, \zeta) := \frac{\lambda_1 x e^{i\Theta} \beta(\zeta) + \lambda_2 x \gamma(\zeta)}{\alpha(\zeta) + x \beta(\zeta)}, \quad x > 0, \quad |\zeta| \geq 1,$$

we have

$$(2.22) \quad \exists z \in \mathcal{K}, \quad x > 0 \quad |k(x, z)| > 1.$$

Since

$$\lim_{|\zeta| \rightarrow \infty} |k(x, \zeta)| = \lambda_1 \frac{x \beta_q}{\alpha_q + x \beta_q} < 1,$$

we infer that there exists a  $\zeta^* \in \mathbb{C}$  with  $|\zeta^*| > 1$  such that  $|k(x, \zeta^*)| = 1$ , i.e.,

$$\frac{\lambda_1 x e^{i\Theta} \beta(\zeta^*) + \lambda_2 x \gamma(\zeta^*)}{\alpha(\zeta^*) + x \beta(\zeta^*)} = e^{-i\varphi},$$

for a  $\varphi \in [0, 2\pi)$ . Therefore,

$$(2.23) \quad \alpha(\zeta^*) + x \beta(\zeta^*) - \lambda_1 x e^{i(\Theta+\varphi)} \beta(\zeta^*) - \lambda_2 x e^{i\varphi} \gamma(\zeta^*) = 0.$$

Then choosing the linear operators  $B_1(t, \cdot) := \lambda_1 e^{i(\Theta+\varphi)} A$  and  $B_2(t, \cdot) := \lambda_2 e^{i\varphi} A$ , we easily see that the Lipschitz conditions (1.6) and (1.7) are satisfied. According

to the von Neumann criterion, a necessary stability condition is that, if  $\nu$  is an eigenvalue of  $A$ , the solutions of

$$(2.24) \quad \sum_{i=0}^q [\alpha_i + k\nu(\beta_i - \lambda_1 e^{i(\Theta+\varphi)}\beta_i - \lambda_2 e^{i\varphi}\gamma_i)] v^{n+i} = 0$$

are bounded; for  $k\nu = x$  this is not the case, since in view of (2.23) the root condition is not satisfied. Therefore, the scheme is not unconditionally stable. (In (2.24) we used the notation  $\gamma_q = 0$ .)

Let us now focus on the stability properties of the numerical scheme (1.4) under the assumptions (1.6) and (1.7) with constants  $\lambda_1$  and  $\lambda_2$  such that

$$(2.25) \quad K_{(\alpha,\beta)}\lambda_1 + K_{(\alpha,\beta,\gamma)}\lambda_2 > 1;$$

cf. (1.12). First, clearly, if one of the terms on the left-hand side of (2.25) exceeds 1, then the necessary stability condition (2.18) is not satisfied, i.e., the scheme (1.4) is unstable for appropriate choices of operators satisfying our assumptions. It thus remains to investigate the case that  $(\lambda_1, \lambda_2)$  belongs to the quadrilateral  $R$ ,

$$R := \{(x, y) \in \mathbb{R}^2 : 0 < x \leq \frac{1}{K_{(\alpha,\beta)}}, 0 < y \leq \frac{1}{K_{(\alpha,\beta,\gamma)}}, \\ K_{(\alpha,\beta)}x + K_{(\alpha,\beta,\gamma)}y \geq 1, x + y \leq 1\}.$$

We did not succeed to settle the problem in this form and confine ourselves with a partial answer, a consequence of (2.20). More precisely, our scheme is in general unstable for  $(\lambda_1, \lambda_2) \in R \setminus T$ , with the triangle  $T$ ,

$$T := \{(x, y) \in \mathbb{R}^2 : K_{(\alpha,\beta)}x + K_{(\alpha,\beta,\gamma)}y \geq 1, x + K_{(\alpha,\beta,\gamma)}y \leq 1, \\ K_{(\alpha,\beta)}(x + c_{(\beta,\gamma)}^*y) \leq 1\}.$$

Obviously, the triangle  $T$  is the discrepancy between the necessary and sufficient stability conditions (2.20) and (1.12), respectively. The stability triangle  $S$ , the quadrilateral  $R$ , and the triangle  $T$  are illustrated in Figure 2.1. In Table 2.2 we give the areas of  $T, S$  and  $R$ , as well as the ratio of the area of  $T$  to the areas of both  $S$  and  $R$  for the case of the implicit–explicit BDF methods. Our answer is very satisfactory for  $q = 3$ , in which case the discrepancy between the necessary and sufficient stability conditions is less than 0.28%, quite satisfactory for  $q = 4$ , satisfactory to some extent for  $q = 5$ , and unsatisfactory for  $q = 6$ . The discrepancy between the conditions (2.18) and (1.12) is, of course, a more interesting quantity; here, we only presented upper bounds of it.

### 3. EXISTENCE AND UNIQUENESS OF THE APPROXIMATIONS

For simplicity, we assume in this section that  $\mu_1 = 0$  in (1.6); the general case can be treated analogously, provided the time step  $k$  is sufficiently small.

We shall show that the approximations  $U^q, \dots, U^N$  are well defined by the implicit–explicit scheme (1.4). As already mentioned,  $\alpha_q\beta_q$  is positive. Assume, without loss of generality, that  $\alpha_q$  is positive. For a given  $w \in V'$ , we will show that equation

$$(3.1) \quad \alpha_q v + k\beta_q[Av - B_1(t, v)] = w$$

possesses a unique solution  $v \in V$ . We consider the operator  $G : V \rightarrow V'$ ,

$$G(v) := \alpha_q v + k\beta_q[Av - B_1(t, v)] - w,$$

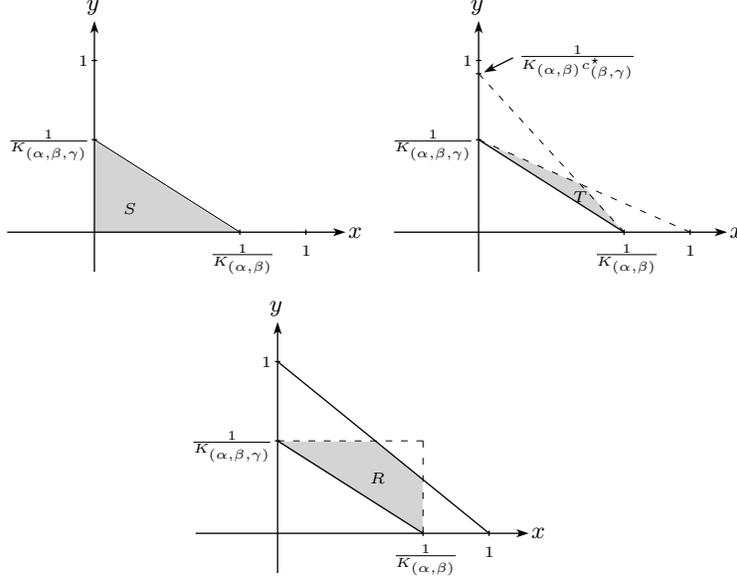


FIGURE 2.1. The stability region  $S$ , the triangle  $T$  —the discrepancy between the sufficient and the necessary stability conditions (1.12) and (2.20)— and the quadrilateral under investigation  $R$ .

$q$	$ T $	$ T / S $	$ T / R $	$c_{(\beta,\gamma)}^*$
3	$1.706476 \times 10^{-4}$	$2.394807 \times 10^{-3}$	$2.779603 \times 10^{-3}$	1.734178
4	$1.320922 \times 10^{-3}$	$4.136150 \times 10^{-2}$	$4.176194 \times 10^{-2}$	3.495466
5	$2.693041 \times 10^{-3}$	$2.135099 \times 10^{-1}$	$2.135099 \times 10^{-1}$	0.710828
6	$1.624825 \times 10^{-3}$	$6.682675 \times 10^{-1}$	$6.682675 \times 10^{-1}$	1.001753

TABLE 2.2. The area  $|T|$  of the triangle  $T$  as well as its ratio to the areas of the stability triangle  $S$  and of the quadrilateral under investigation  $R$ , for the implicit–explicit  $q$ –step BDF schemes,  $q = 3, \dots, 6$ ; cf. Figure 2.1. For  $q = 5, 6$ , we have  $|R| = |S|$ .

and write (3.1) in the form  $G(v) = 0$ . Thus, to show existence and uniqueness of the solution of (3.1) it suffices to show that the equation  $G(v) = 0$  possesses a unique solution. To this end, we will use the Zarantonello fixed-point theorem. First, for  $v, \tilde{v} \in V$ ,

$$\begin{aligned} (G(v) - G(\tilde{v}), v - \tilde{v}) &= \alpha_q |v - \tilde{v}|^2 + k\beta_q \|v - \tilde{v}\|^2 - k\beta_q (B_1(t, v) - B_1(t, \tilde{v}), v - \tilde{v}) \\ &\geq \alpha_q |v - \tilde{v}|^2 + k\beta_q \|v - \tilde{v}\|^2 - k\beta_q \|B_1(t, v) - B_1(t, \tilde{v})\|_* \|v - \tilde{v}\|, \end{aligned}$$

whence, in view of (1.6),

$$(G(v) - G(\tilde{v}), v - \tilde{v}) \geq \alpha_q |v - \tilde{v}|^2 + k\beta_q (1 - \lambda_1) \|v - \tilde{v}\|^2;$$

therefore,

$$(3.2) \quad (G(v) - G(\tilde{v}), v - \tilde{v}) \geq k\beta_q (1 - \lambda_1) \|v - \tilde{v}\|^2 \quad \forall v, \tilde{v} \in V.$$

Similarly, we have

$$(3.3) \quad \|G(v) - G(\tilde{v})\|_* \leq \alpha_q \|v - \tilde{v}\|_* + k\beta_q(1 + \lambda_1) \|v - \tilde{v}\| \quad \forall v, \tilde{v} \in V.$$

We easily infer from (3.2) and (3.3), respectively, that the operator  $\tilde{G} : H \rightarrow H$ ,  $\tilde{G}(v) := A^{-1/2}G(A^{-1/2}v)$ , is strongly monotone and Lipschitz continuous, i.e.,

$$(3.4) \quad (\tilde{G}(v) - \tilde{G}(\tilde{v}), v - \tilde{v}) \geq c|v - \tilde{v}|^2 \quad \forall v, \tilde{v} \in V,$$

for a positive constant  $c$ , and

$$(3.5) \quad |\tilde{G}(v) - \tilde{G}(\tilde{v})| \leq L|v - \tilde{v}| \quad \forall v, \tilde{v} \in V.$$

It readily follows from (3.4) and (3.5) that  $F : H \rightarrow H$ ,  $F(v) := v - \frac{c}{L^2}\tilde{G}(v)$ , is a contraction in  $(H, |\cdot|)$ ; this fact is referred to as Zarantonello's fixed-point theorem. If  $v \in H$  is the unique fixed point of  $F$ , then  $A^{-1/2}v \in V$  is the unique solution of  $G(v) = 0$ .

*Remark 3.1* (Linear  $B_1(t, \cdot)$ ). Since the scheme (1.4) is implementable only for linear operators  $B_1(t, \cdot)$ , and existence and uniqueness of the approximate solutions are then easier to establish, we briefly consider this case separately here. According to the Lax–Milgram lemma, it obviously suffices to show that the bilinear form  $a : V \times V \rightarrow \mathbb{R}$ ,

$$a(v, \tilde{v}) := \alpha_q(v, \tilde{v}) + k\beta_q(Av, \tilde{v}) - k\beta_q(B_1(t, v), \tilde{v}),$$

is coercive and continuous. This is indeed the case since, for  $v \in V$ ,

$$\begin{aligned} a(v, v) &= \alpha_q|v|^2 + k\beta_q\|v\|^2 - k\beta_q(B_1(t, v), v) \\ &\geq \alpha_q|v|^2 + k\beta_q\|v\|^2 - k\beta_q\|B_1(t, v)\|_* \|v\|, \end{aligned}$$

whence, in view of (1.6),

$$a(v, v) \geq \alpha_q|v|^2 + k\beta_q(1 - \lambda_1)\|v\|^2,$$

and, similarly, for  $v, \tilde{v} \in V$ ,

$$|a(v, \tilde{v})| \leq \alpha_q|v| |\tilde{v}| + k\beta_q(1 + \lambda_1)\|v\| \|\tilde{v}\|. \quad \square$$

#### 4. CONSISTENCY

In this section we will derive an optimal order estimate for the consistency error of the implicit–explicit scheme (1.4); for the analogous result for (1.3) we refer to [3, 1]. We assume that the order of both schemes  $(\alpha, \beta)$  and  $(\alpha, \gamma)$  is  $p$ ; cf. (1.13).

The consistency error  $E^n$  of the scheme (1.4) for the solution  $u$  of (1.1), i.e., the amount by which the exact solution misses satisfying (1.4), is given by

$$(4.1) \quad kE^n = \sum_{i=0}^q [\alpha_i u^{n+i} + k\beta_i(Au^{n+i} - B_1(t^{n+i}, u^{n+i}))] - k \sum_{i=0}^{q-1} \gamma_i B_2(t^{n+i}, u^{n+i}),$$

$n = 0, \dots, N - q$ . First, letting

$$E_1^n := \sum_{i=0}^q [\alpha_i u^{n+i} - k\beta_i u'(t^{n+i})], \quad E_2^n := k \sum_{i=0}^q (\beta_i - \gamma_i) B_2(t^{n+i}, u^{n+i}),$$

with  $\gamma_q := 0$ , and using the differential equation in (1.1), we infer that

$$(4.2) \quad kE^n = E_1^n + E_2^n.$$

Furthermore, by Taylor expanding around  $t^n$  and using the order conditions of the implicit  $(\alpha, \beta)$ -scheme, i.e., the first equality in (1.13), and the second equality in (1.13), respectively, we obtain

$$(4.3) \quad \begin{cases} E_1^n = \frac{1}{p!} \sum_{i=0}^q \int_{t^n}^{t^{n+i}} (t^{n+i} - s)^{p-1} [\alpha_i(t^{n+i} - s) - pk\beta_i] u^{(p+1)}(s) ds, \\ E_2^n = \frac{k}{(p-1)!} \sum_{i=0}^q (\beta_i - \gamma_i) \int_{t^n}^{t^{n+i}} (t^{n+i} - s)^{p-1} \frac{d^p}{dt^p} B_2(s, u(s)) ds. \end{cases}$$

Thus, under obvious regularity requirements, we obtain the desired consistency estimate

$$(4.4) \quad \max_{0 \leq n \leq N-q} \|E^n\|_* \leq Ck^p.$$

## 5. LOCAL STABILITY

In this section we present the main result of the paper, namely local stability of the scheme (1.4) under the Lipschitz conditions (1.6) and (1.7) as well as (1.12).

Let  $U^m, V^m \in T_u, m = 0, \dots, N$ , satisfy (1.4) and

$$(5.1) \quad \sum_{i=0}^q [\alpha_i V^{n+i} + k\beta_i (AV^{n+i} - B_1(t^{n+i}, V^{n+i}))] = k \sum_{i=0}^{q-1} \gamma_i B_2(t^{n+i}, V^{n+i}),$$

$n = 0, \dots, N - q$ , respectively. Let  $\vartheta^m := U^m - V^m$  and  $b_i^m := B_i(t^m, U^m) - B_i(t^m, V^m)$ ,  $i = 1, 2, m = 0, \dots, N$ . Subtracting (5.1) from (1.4) we obtain

$$(5.2) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) \vartheta^{n+i} = k \sum_{i=0}^q \beta_i b_1^{n+i} + k \sum_{i=0}^{q-1} \gamma_i b_2^{n+i}, \quad n = 0, \dots, N - q.$$

The rational functions  $e(\ell, \cdot)$ ,  $f(\ell, \cdot)$  and  $\tilde{f}(\ell, \cdot)$  defined through the expansions

$$(5.3) \quad \begin{cases} \frac{1}{\alpha(\zeta) + x\beta(\zeta)} = \sum_{\ell=q}^{\infty} e(\ell, x) \zeta^{-\ell}, \\ \frac{\gamma(\zeta)}{\alpha(\zeta) + x\beta(\zeta)} = \sum_{\ell=1}^{\infty} f(\ell, x) \zeta^{-\ell}, \\ \frac{\beta(\zeta)}{\alpha(\zeta) + x\beta(\zeta)} = \sum_{\ell=0}^{\infty} \tilde{f}(\ell, x) \zeta^{-\ell}, \end{cases}$$

will play a crucial role in the stability analysis. Due to the strong  $A(0)$ -stability, for all  $x \in (0, \infty]$ , the modulus of all roots of  $\alpha(\cdot) + x\beta(\cdot)$  is less than 1. Therefore, the expansions are valid for all  $|\zeta| \geq 1$ . We also note that the only pole of these rational functions is  $-\alpha_q/\beta_q < 0$  and that they vanish at  $\infty$ . Thus, we can define  $e(\ell, kA)$ ,  $f(\ell, kA)$  and  $\tilde{f}(\ell, kA)$ . Now let

$$(5.4) \quad \vartheta_1^n := k \sum_{\ell=0}^n \tilde{f}(n - \ell, kA) b_1^\ell, \quad n = 0, \dots, N.$$

Then, in view of the last relation in (5.3), we have

$$(5.5) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) \vartheta_1^{n+i} = k \sum_{i=0}^q \beta_i b_1^{n+i}, \quad n = 0, \dots, N - q;$$

cf., e.g., [10, pp. 242–244]. Similarly, for

$$(5.6) \quad \vartheta_2^0 := 0, \quad \vartheta_2^n := k \sum_{\ell=0}^{n-1} f(n-\ell, kA) b_2^\ell, \quad n = 1, \dots, N,$$

we have

$$(5.7) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) \vartheta_2^{n+i} = k \sum_{i=0}^{q-1} \gamma_i b_2^{n+i}, \quad n = 0, \dots, N-q.$$

Therefore, according to (5.2), (5.5) and (5.7), the sequence  $\vartheta_3^n$ ,  $\vartheta_3^n := \vartheta^n - \vartheta_1^n - \vartheta_2^n$ , satisfies the relation

$$(5.8) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A) \vartheta_3^{n+i} = 0, \quad n = 0, \dots, N-q,$$

and, consequently, with  $g_j(n, x) := \sum_{\ell=j+1}^q e(n+\ell-j, x)(\alpha_\ell + x\beta_\ell)$ ,

$$(5.9) \quad \vartheta_3^n = \sum_{j=0}^{q-1} g_j(n, kA) \vartheta_3^j, \quad n = q, \dots, N.$$

Notice that, in this way, we decompose the quantities  $\vartheta^n$  of (5.2) into three parts, namely  $\vartheta_1^n$ ,  $\vartheta_2^n$  and  $\vartheta_3^n$ , respectively, which account for the nonlinearity  $b_1$ , the nonlinearity  $b_2$ , and the initial entries  $U^0, \dots, U^{q-1}, V^0, \dots, V^{q-1}$ , respectively.

We quote the following estimates for  $\vartheta_2^n$  and  $\vartheta_3^n$  from [3, Lemma 2.1]:

**Lemma 5.1** (Estimation of  $\vartheta_2^n$  and  $\vartheta_3^n$ ). *For  $n = 0, \dots, N$ , the following estimates are valid*

$$(5.10a) \quad k \sum_{\ell=0}^n \|\vartheta_2^\ell\|^2 \leq K_{(\alpha, \beta, \gamma)}^2 k \sum_{\ell=0}^{n-1} \|b_2^\ell\|_\star^2,$$

$$(5.10b) \quad |\vartheta_2^n|^2 \leq K_2 k \sum_{\ell=0}^{n-1} \|b_2^\ell\|_\star^2,$$

and

$$(5.11a) \quad k \sum_{\ell=0}^n \|\vartheta_3^\ell\|^2 \leq qN_1 \sum_{j=0}^{q-1} (|\vartheta_3^j|^2 + k\|\vartheta_3^j\|^2),$$

$$(5.11b) \quad |\vartheta_3^n| \leq N_2 \sum_{j=0}^{q-1} |\vartheta_3^j|,$$

with the constants  $K_2, N_1$  and  $N_2$  from [3].  $\square$

The proof of [3, Lemma 2.1] is based on a spectral expansion and Parseval's identity. Similar techniques are used in [14] and [16]; see also [1].

In a completely analogous manner we can estimate also  $\vartheta_1^n$ :

**Lemma 5.2** (Estimation of  $\vartheta_1^n$ ). *For  $n = 0, \dots, N$ , the following estimates are valid*

$$(5.12a) \quad k \sum_{\ell=0}^n \|\vartheta_1^\ell\|^2 \leq K_{(\alpha, \beta)}^2 k \sum_{\ell=0}^n \|b_1^\ell\|_\star^2,$$

$$(5.12b) \quad |\vartheta_1^n|^2 \leq \tilde{K}_2 k \sum_{\ell=0}^n \|b_1^\ell\|_\star^2,$$

with

$$\tilde{K}_2 := \sup_{x>0} \int_0^1 \left| \frac{\sqrt{x}\beta(e^{-2i\pi t})}{(\alpha+x\beta)(e^{-2i\pi t})} \right|^2 dt. \quad \square$$

In the local stability result, Theorem 5.1, we will estimate  $\vartheta^n$  in terms of  $\vartheta^0, \dots, \vartheta^{q-1}$ . Part of  $\vartheta^n$ , namely  $\vartheta_3^n$ , will be estimated in terms of  $\vartheta^0, \dots, \vartheta^{q-1}$  in the following Lemma.

**Lemma 5.3** (Estimation of  $\vartheta_3^n$ ). *There exists a constant  $C$  such that*

$$(5.13) \quad |\vartheta_3^n|^2 + k \sum_{\ell=0}^n \|\vartheta_3^\ell\|^2 \leq C \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k\|\vartheta^j\|^2), \quad n = 0, \dots, N.$$

*Proof.* Obviously, since  $\vartheta_3^n = \vartheta^n - \vartheta_1^n - \vartheta_2^n$ , we have

$$\vartheta_3^j = \vartheta^j - k \sum_{\ell=0}^j \tilde{f}(j-\ell, kA)b_1^\ell - k \sum_{\ell=0}^{j-1} f(j-\ell, kA)b_2^\ell, \quad j = 0, \dots, q-1.$$

Therefore

$$\begin{cases} |\vartheta_3^j| \leq |\vartheta^j| + \sqrt{k} \sum_{\ell=0}^j \tilde{m}_{j-\ell} \|b_1^\ell\|_\star + \sqrt{k} \sum_{\ell=0}^{j-1} m_{j-\ell} \|b_2^\ell\|_\star, \\ \|\vartheta_3^j\| \leq \|\vartheta^j\| + \sum_{\ell=0}^j \tilde{n}_{j-\ell} \|b_1^\ell\|_\star + \sum_{\ell=0}^{j-1} n_{j-\ell} \|b_2^\ell\|_\star, \end{cases}$$

with

$$\begin{cases} \tilde{m}_\ell := \sup_{x>0} |\sqrt{x}\tilde{f}(\ell, x)|, & \tilde{n}_\ell := \sup_{x>0} |x\tilde{f}(\ell, x)|, \\ m_\ell := \sup_{x>0} |\sqrt{x}f(\ell, x)|, & n_\ell := \sup_{x>0} |xf(\ell, x)|. \end{cases}$$

Then, (5.13) follows from the estimates (5.11a) and (5.11b) and the Lipschitz conditions (1.6) and (1.7).  $\square$

The main result in this section, the local stability of the scheme (1.4), is given in the following theorem:

**Theorem 5.1** (Local stability). *Let  $U^m, V^m \in T_u$  satisfy (1.4) and (5.1), respectively, and assume that the stability condition (1.12) is satisfied. Then, with  $\vartheta^m = U^m - V^m$ , we have the local stability estimate*

$$(5.14) \quad |\vartheta^n|^2 + k \sum_{\ell=0}^n \|\vartheta^\ell\|^2 \leq C \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k\|\vartheta^j\|^2), \quad n = q-1, \dots, N,$$

with a constant  $C$  independent of  $U^m, V^m$  and  $k$ .

*Proof.* Using the Lipschitz condition (1.6) we have

$$\left( k \sum_{\ell=0}^n \|b_1^\ell\|_\star^2 \right)^{1/2} \leq \left( k \sum_{\ell=0}^n (\lambda_1 \|\vartheta^\ell\| + \mu_1 |\vartheta^\ell|)^2 \right)^{1/2},$$

whence, in view of Minkowski's inequality,

$$(5.15) \quad \left( k \sum_{\ell=0}^n \|b_1^\ell\|_*^2 \right)^{1/2} \leq \lambda_1 a_n + \mu_1 d_n + e_n$$

with

$$a_n := \left( k \sum_{\ell=0}^n \|\vartheta_1^\ell + \vartheta_2^\ell\|^2 \right)^{1/2}, \quad d_n := \left( k \sum_{\ell=0}^n |\vartheta_1^\ell + \vartheta_2^\ell|^2 \right)^{1/2},$$

and

$$e_n := \left( k \sum_{\ell=0}^n (\lambda_1 \|\vartheta_3^\ell\| + \mu_1 |\vartheta_3^\ell|)^2 \right)^{1/2}.$$

Similarly, using (1.7), we obtain

$$(5.16) \quad \left( k \sum_{\ell=0}^{n-1} \|b_2^\ell\|_*^2 \right)^{1/2} \leq \lambda_2 a_{n-1} + \mu_2 d_{n-1} + \tilde{e}_{n-1}$$

with

$$\tilde{e}_n := \left( k \sum_{\ell=0}^n (\lambda_2 \|\vartheta_3^\ell\| + \mu_2 |\vartheta_3^\ell|)^2 \right)^{1/2}.$$

Thus, (5.10a) and (5.12a) yield

$$\begin{aligned} a_n &\leq \left( k \sum_{\ell=0}^n \|\vartheta_1^\ell\|^2 \right)^{1/2} + \left( k \sum_{\ell=0}^n \|\vartheta_2^\ell\|^2 \right)^{1/2} \\ &\leq K_{(\alpha,\beta)} (\lambda_1 a_n + \mu_1 d_n + e_n) + K_{(\alpha,\beta,\gamma)} (\lambda_2 a_{n-1} + \mu_2 d_{n-1} + \tilde{e}_{n-1}) \\ &\leq (K_{(\alpha,\beta)} \lambda_1 + K_{(\alpha,\beta,\gamma)} \lambda_2) a_n + (K_{(\alpha,\beta)} \mu_1 + K_{(\alpha,\beta,\gamma)} \mu_2) d_n \\ &\quad + (K_{(\alpha,\beta)} e_n + K_{(\alpha,\beta,\gamma)} \tilde{e}_{n-1}), \end{aligned}$$

whence

$$(5.17) \quad a_n \leq c_1 (c_2 d_n + K_{(\alpha,\beta)} e_n + K_{(\alpha,\beta,\gamma)} \tilde{e}_{n-1})$$

with

$$c_1 := (1 - K_{(\alpha,\beta)} \lambda_1 - K_{(\alpha,\beta,\gamma)} \lambda_2)^{-1} \quad \text{and} \quad c_2 := K_{(\alpha,\beta)} \mu_1 + K_{(\alpha,\beta,\gamma)} \mu_2.$$

Similarly, in view of (5.15) and (5.16), estimates (5.10b) and (5.12b) yield

$$\begin{aligned} \frac{1}{k} (d_n^2 - d_{n-1}^2) &= |\vartheta_1^n + \vartheta_2^n|^2 \leq 2(|\vartheta_1^n|^2 + |\vartheta_2^n|^2) \\ &\leq 2\tilde{K}_2 (\lambda_1 a_n + \mu_1 d_n + e_n)^2 + 2K_2 (\lambda_2 a_n + \mu_2 d_n + \tilde{e}_n)^2; \end{aligned}$$

therefore, using (5.17), we have

$$(5.18) \quad \frac{1}{k} (d_n^2 - d_{n-1}^2) \leq C (d_n^2 + e_n^2 + \tilde{e}_n^2), \quad n = 1, \dots, N.$$

Summing here, we get

$$d_n^2 \leq Ck \sum_{i=0}^n d_i^2 + CT (e_n^2 + \tilde{e}_n^2) + d_0^2,$$

$n = 1, \dots, N$ , and using the Gronwall inequality, we infer, for sufficiently small  $k$ ,

$$(5.19) \quad d_n \leq C (d_0 + e_n + \tilde{e}_n), \quad n = 0, \dots, N;$$

thus, in view of (5.17),

$$(5.20) \quad a_n \leq \tilde{C}(d_0 + e_n + \tilde{e}_n), \quad n = 0, \dots, N.$$

Furthermore, from (5.18) we obtain

$$|\vartheta_1^n + \vartheta_2^n| \leq \sqrt{C}(d_n + e_n + \tilde{e}_n),$$

whence, in view of (5.19),

$$(5.21) \quad |\vartheta_1^n + \vartheta_2^n| \leq c(d_0 + e_n + \tilde{e}_n), \quad n = 0, \dots, N.$$

To estimate  $d_0$  we first notice that  $\tilde{f}(0, x) = \beta_q/(\alpha_q + \beta_q x)$  and  $(\alpha_q I + \beta_q k A)\vartheta_1^0 = \beta_q k b_1^0$ ; see (5.3) and (5.4). We easily infer that  $|\vartheta_1^0|^2 \leq \beta_q k \|b_1^0\|_*^2 / (4\alpha_q)$ , whence, combining  $d_0^2 = k|\vartheta_1^0|^2$  with the Lipschitz condition (1.6),

$$(5.22) \quad d_0 \leq ck(\|\vartheta^0\| + |\vartheta^0|).$$

Now, (5.20), (5.21), (5.22) and (5.13) yield

$$(5.23) \quad |\vartheta_1^n + \vartheta_2^n|^2 + k \sum_{\ell=0}^n \|\vartheta_1^\ell + \vartheta_2^\ell\|^2 \leq C \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k\|\vartheta^j\|^2).$$

From (5.23) and (5.13) we easily obtain (5.14) and the proof is complete.  $\square$

## 6. ERROR ESTIMATES

In this section we assume that the stability condition (1.12) is satisfied and that the order of both schemes  $(\alpha, \beta)$  and  $(\alpha, \gamma)$  is  $p$ , and shall establish optimal order error estimates.

We note that we will use similar notation as in Section 5; however, several quantities here, like  $\vartheta^m$  and  $b^m$ , do not coincide with those in Section 5.

Let  $\vartheta^m := u^m - U^m$ ,  $b_i^m := B_i(t^m, u^m) - B_i(t^m, U^m)$ ,  $m = 0, \dots, N$ ,  $i = 1, 2$ . Subtracting (1.4) from (4.1) we obtain

$$(6.1) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A)\vartheta^{n+i} = k \sum_{i=0}^q \beta_i b_1^{n+i} + k \sum_{i=0}^{q-1} \gamma_i b_2^{n+i} + kE^n,$$

$n = 0, \dots, N - q$ . Let

$$\begin{aligned} \vartheta_1^n &:= k \sum_{\ell=0}^n \tilde{f}(n-\ell, kA)b_1^\ell, \\ \vartheta_2^0 &:= 0, \quad \vartheta_2^n := k \sum_{\ell=0}^{n-1} f(n-\ell, kA)b_2^\ell, \\ \vartheta_4^n &:= k \sum_{\ell=0}^{n-q} e(n-\ell, kA)E^\ell, \end{aligned}$$

with  $e(0, x) = \dots = e(q-1, x) = 0$ . It is then easily seen that  $\vartheta_3^n := \vartheta^n - \vartheta_1^n - \vartheta_2^n - \vartheta_4^n$  satisfies

$$(6.2) \quad \sum_{i=0}^q (\alpha_i I + k\beta_i A)\vartheta_3^{n+i} = 0, \quad n = 0, \dots, N - q;$$

cf. (5.8). Now  $\vartheta_4^j = 0$  for  $j \leq q-1$ ; therefore  $\vartheta_3^0, \dots, \vartheta_3^{q-1}$ , and thus all  $\vartheta_3^n$ , depend only on the initial entries  $u^0, \dots, u^{q-1}, U^0, \dots, U^{q-1}$ .

The quantities  $\vartheta_1^n, \vartheta_2^n$  and  $\vartheta_3^n$  can be estimated as in Lemmata 5.1 and 5.2. Similarly, for  $\vartheta_4^n$  we have

$$(6.3a) \quad k \sum_{\ell=0}^n \|\vartheta_4^\ell\|^2 \leq M_1^2 k \sum_{\ell=0}^{n-q} \|E^\ell\|_\star^2,$$

$$(6.3b) \quad |\vartheta_4^n|^2 \leq M_2 k \sum_{\ell=0}^{n-q} \|E^\ell\|_\star^2,$$

with

$$M_1 := \sup_{x>0} \max_{\zeta \in S_1} \left| \frac{x}{(\alpha + x\beta)(\zeta)} \right| \quad \text{and} \quad M_2 := \sup_{x>0} \int_0^1 \frac{x}{|(\alpha + x\beta)(e^{-2i\pi t})|^2} dt.$$

Under our assumptions,  $M_1$  and  $M_2$  are finite; see Section 2 and [3].

In our main result, Theorem 6.1, we will need to estimate  $\vartheta^n$ . Part of it, namely  $\vartheta_3^n + \vartheta_4^n$ , can be estimated in terms of  $\vartheta^0, \dots, \vartheta^{q-1}$  and the consistency errors  $E^0, \dots, E^{N-q}$ . This result follows immediately by combining Lemma 5.3 with the estimates (6.3a) and (6.3b).

**Lemma 6.1** (Estimate for  $\vartheta_3^n + \vartheta_4^n$ ). *There exists a constant  $C$  such that, for  $n = 0, \dots, N$ ,*

$$(6.4) \quad |\vartheta_3^n + \vartheta_4^n|^2 + k \sum_{\ell=0}^n \|\vartheta_3^\ell + \vartheta_4^\ell\|^2 \leq C \left\{ \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k \|\vartheta^j\|^2) + k \sum_{\ell=0}^{n-q} \|E^\ell\|_\star^2 \right\}. \quad \square$$

In the following theorem we establish optimal order error estimates.

**Theorem 6.1** (Error estimate). *Let the order of both schemes  $(\alpha, \beta)$  and  $(\alpha, \gamma)$  be  $p$  and let the stability condition (1.12) be satisfied. Assume we are given starting approximations  $U^0, U^1, \dots, U^{q-1} \in V$  to  $u^0, \dots, u^{q-1}$  such that*

$$(6.5) \quad \max_{0 \leq j \leq q-1} \left( |u^j - U^j| + k^{1/2} \|u^j - U^j\| \right) \leq Ck^p.$$

*Let  $U^n \in V$ ,  $n = q, \dots, N$ , be recursively defined by (1.4). Let  $\vartheta^n = u^n - U^n$ ,  $n = 0, \dots, N$ . Then, there exists a constant  $C$ , independent of  $k$  and  $n$ , such that, for  $k$  sufficiently small,*

$$(6.6) \quad |\vartheta^n|^2 + k \sum_{\ell=0}^n \|\vartheta^\ell\|^2 \leq C \left\{ \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k \|\vartheta^j\|^2) + k \sum_{\ell=0}^{n-q} \|E^\ell\|_\star^2 \right\},$$

$n = q - 1, \dots, N$ , and

$$(6.7) \quad \max_{0 \leq n \leq N} |u(t^n) - U^n| \leq Ck^p.$$

*Proof.* According to (6.5) and (4.4), there exists a constant  $C_\star$  such that the right-hand side of (6.6) can be estimated by  $C_\star^2 k^{2p}$ ,

$$(6.8) \quad C \left\{ \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k \|\vartheta^j\|^2) + k \sum_{\ell=0}^{N-q} \|E^\ell\|_\star^2 \right\} \leq C_\star^2 k^{2p}.$$

Now, obviously, (6.7) is a consequence of (6.6) and (6.8). Thus, it remains to prove (6.6). The estimate (6.6) is obviously valid for  $n = q - 1$ . Assume that it holds for

$q - 1, \dots, n - 1, q \leq n \leq N$ . Then, according to (6.8) and the induction hypothesis, we have, for  $k$  small enough,

$$(6.9) \quad \max_{0 \leq j \leq n-1} \|\vartheta^j\| \leq C_* k^{p-1/2} \leq 1,$$

and thus  $U^j \in T_u, j = 0, \dots, n - 1$ . It is then easily seen, cf. the derivation of (5.22), that

$$(6.10) \quad |\vartheta_1^n + \vartheta_2^n|^2 + k \sum_{\ell=0}^n \|\vartheta_1^\ell + \vartheta_2^\ell\|^2 \leq C \sum_{j=0}^{q-1} (|\vartheta^j|^2 + k \|\vartheta^j\|^2).$$

From (6.4) and (6.10) it easily follows that (6.6) holds for  $n$  as well, and the proof is complete.  $\square$

*Remark 6.1* (Local versus global Lipschitz condition on  $B_1$ ). The *global* Lipschitz condition (1.6) is too restrictive in the case of nonlinear operator  $B_1$  but it can be relaxed to a *local* one,

$$(6.11) \quad \|B_1(t, v) - B_1(t, w)\|_* \leq \lambda_1 \|v - w\| + \mu_1 |v - w| \quad \forall v, w \in T_u,$$

in a straightforward way: Assume that  $B_1(t, \cdot)$  can be modified to yield an operator  $\tilde{B}_1(t, \cdot) : V \rightarrow V'$  coinciding with  $B_1(t, \cdot)$  in the tube  $T_u$ ,  $\tilde{B}_1(t, v) = B_1(t, v)$  for all  $t \in [0, T]$  and all  $v \in T_u$ , and satisfying the global Lipschitz condition (1.6) (with  $B_1$  replaced by  $\tilde{B}_1$ ). Consider then the *modified* scheme (1.4) with  $B_1$  replaced by  $\tilde{B}_1$ . According to our results, for sufficiently small  $k$ , we then have existence and uniqueness of the approximate solutions  $U^m$  as well as the error estimate (6.7). Now, in view of (6.9),  $U^m \in T_u$ . Therefore the solutions of the modified scheme satisfy also the original scheme (1.4). Hence, for sufficiently small  $k$ , we have existence of solutions of (1.4), *local* uniqueness in the tube  $T_u$  as well as the error estimate (6.7).  $\square$

For several examples of multistep schemes as well as for partial differential equations satisfying the conditions of this paper we refer the reader to [3] and [2].

## 7. IMPROVEMENTS FOR LOW ORDER SCHEMES

In this section we focus on the implicit–explicit Euler method and the implicit–explicit two-step BDF scheme. We will use the fact that the corresponding implicit methods are  $G$ –stable to relax the condition on  $B_1$ ; more precisely, we will relax the Lipschitz condition (1.6) to a one-sided one, namely

$$(7.1) \quad (B_1(t, v) - B_1(t, w), v - w) \leq \lambda_1 \|v - w\|^2 + \mu_1 \|v - w\| |v - w| \quad \forall v, w \in V,$$

in which  $\lambda_1$  is allowed to take on also negative values.

In this Section we generalize the results of [5]; in [5] the operator  $B_1$  was assumed time-independent, linear and dispersive, i.e., such that  $(B_1 v, v) = 0$ .

**7.1. Implicit–explicit Euler method.** It is well known that the implicit Euler scheme is  $B$ –stable, or equivalently  $G$ –stable, when viewed as a Runge–Kutta or one-step scheme, respectively. We will use this property here to relax the condition on  $B_1$ .

Our assumption in this subsection is that  $\lambda_1 + \lambda_2 < 1$ , since for the implicit–explicit Euler method  $K_{(\alpha, \beta)} = K_{(\alpha, \beta, \gamma)} = 1$ ; see subsection 2.3 with  $q = 1$ .

7.1.1. *Local stability.* With the notation of Section 5, we have, for the implicit–explicit Euler method,

$$(7.2) \quad (I + kA)\vartheta^{n+1} = \vartheta^n + kb_1^{n+1} + kb_2^n, \quad n = 0, \dots, N-1.$$

Taking here the inner product with  $\vartheta^{n+1}$  and using the one-sided Lipschitz condition (7.1) and the local Lipschitz condition (1.7), we easily obtain

$$\begin{aligned} |\vartheta^{n+1}|^2 + k\|\vartheta^{n+1}\|^2 &\leq \frac{1}{2}|\vartheta^n|^2 + \frac{1}{2}|\vartheta^{n+1}|^2 + \lambda_1 k\|\vartheta^{n+1}\|^2 + \mu_1 k\|\vartheta^{n+1}\| |\vartheta^{n+1}| \\ &\quad + \lambda_2 k\|\vartheta^n\| \|\vartheta^{n+1}\| + \mu_2 k|\vartheta^n| \|\vartheta^{n+1}\|. \end{aligned}$$

Therefore, we have

$$\begin{aligned} |\vartheta^{n+1}|^2 + k(2 - 2\lambda_1 - \lambda_2 - \mu\varepsilon)\|\vartheta^{n+1}\|^2 &\leq |\vartheta^n|^2 + \frac{\mu_1}{\varepsilon}k|\vartheta^{n+1}|^2 \\ &\quad + \lambda_2 k\|\vartheta^n\|^2 + \frac{\mu_2}{\varepsilon}k|\vartheta^n|^2, \end{aligned}$$

with  $\mu := \mu_1 + \mu_2$ , for all positive  $\varepsilon$ ; thus, with  $\varepsilon := (1 - \lambda_1 - \lambda_2)/(2\mu)$  and  $\tilde{\mu}_i := \mu_i/\varepsilon$ ,

$$\begin{aligned} |\vartheta^{n+1}|^2 + \frac{k}{2}(5 - 5\lambda_1 - 3\lambda_2)\|\vartheta^{n+1}\|^2 &\leq |\vartheta^n|^2 + \tilde{\mu}_1 k|\vartheta^{n+1}|^2 \\ &\quad + \lambda_2 k\|\vartheta^n\|^2 + \tilde{\mu}_2 k|\vartheta^n|^2. \end{aligned}$$

Hence, with  $c := (5 - 5\lambda_1 - 3\lambda_2)/2$ , we have  $\lambda_2 < c$ , and, consequently,

$$(7.3) \quad |\vartheta^{n+1}|^2 + ck\|\vartheta^{n+1}\|^2 \leq |\vartheta^n|^2 + ck\|\vartheta^n\|^2 + \tilde{\mu}_1 k|\vartheta^{n+1}|^2 + \tilde{\mu}_2 k|\vartheta^n|^2.$$

With the norm  $\|\cdot\|_c$ ,  $\|v\|_c := (|v|^2 + ck\|v\|^2)^{1/2}$ , we write (7.3) in the form

$$\|\vartheta^{n+1}\|_c^2 \leq \|\vartheta^n\|_c^2 + \tilde{\mu}_1 k|\vartheta^{n+1}|^2 + \tilde{\mu}_2 k|\vartheta^n|^2$$

and infer easily, for sufficiently small  $k$ , with an appropriate constant  $C$ , that

$$(7.4) \quad \|\vartheta^{n+1}\|_c^2 \leq (1 + Ck)\|\vartheta^n\|_c^2, \quad n = 0, \dots, N-1.$$

This relation leads to the desired stability estimate,

$$(7.5) \quad \max_{0 \leq n \leq N} (|\vartheta^n|^2 + ck\|\vartheta^n\|^2) \leq e^{CT} (|\vartheta^0|^2 + ck\|\vartheta^0\|^2).$$

7.1.2. *Error estimates.* Combining the stability estimate (7.5) with the consistency estimate (4.4), with  $p = 1$  for the implicit–explicit Euler scheme, we easily establish the desired error estimate

$$(7.6) \quad \max_{0 \leq n \leq N} |u(t^n) - U^n| \leq Ck.$$

**7.2. Implicit–explicit two–step BDF method.** It is well known that the (implicit) two–step BDF scheme is  $G$ –stable.

Our assumption here is that  $\lambda_1 + 3\lambda_2 < 1$ , since for the implicit–explicit two–step BDF method  $K_{(\alpha,\beta)} = 1$  and  $K_{(\alpha,\beta,\gamma)} = 3$ ; see subsection 2.3 with  $q = 2$ .

7.2.1. *Local stability.* With the notation of Section 5, we have, for the implicit–explicit two–step BDF method

$$(7.7) \quad \frac{3}{2}\vartheta^{n+2} - 2\vartheta^{n+1} + \frac{1}{2}\vartheta^n + kA\vartheta^{n+2} = kb_1^{n+2} + 2kb_2^{n+1} - b_2^n,$$

$n = 0, \dots, N - 2$ . As is well known from the  $G$ –stability of the two–step BDF method,

$$\begin{aligned} \left(\frac{3}{2}\vartheta^{n+2} - 2\vartheta^{n+1} + \frac{1}{2}\vartheta^n, \vartheta^{n+2}\right) &= \frac{5}{4}|\vartheta^{n+2}|^2 - |\vartheta^{n+1}|^2 - \frac{1}{4}|\vartheta^n|^2 \\ &\quad - [(\vartheta^{n+2}, \vartheta^{n+1}) - (\vartheta^{n+1}, \vartheta^n)] + \frac{1}{4}|\vartheta^{n+2} - 2\vartheta^{n+1} + \vartheta^n|^2. \end{aligned}$$

Therefore, taking in (7.7) the inner product with  $\vartheta^{n+2}$  and using (7.1), we obtain

$$\begin{aligned} &\frac{5}{4}|\vartheta^{n+2}|^2 - |\vartheta^{n+1}|^2 - \frac{1}{4}|\vartheta^n|^2 - [(\vartheta^{n+2}, \vartheta^{n+1}) - (\vartheta^{n+1}, \vartheta^n)] + k\|\vartheta^{n+2}\|^2 \\ &\leq \lambda_1 k\|\vartheta^{n+2}\|^2 + \mu_1 k\|\vartheta^{n+2}\||\vartheta^{n+2}| + 2k\|b_2^{n+1}\|_*\|\vartheta^{n+2}\| + k\|b_2^n\|_*\|\vartheta^{n+2}\|. \end{aligned}$$

Estimating here  $\|b_2^{n+1}\|_*$  and  $\|b_2^n\|_*$  using the Lipschitz condition (1.7), and subsequently applying elementary inequalities, we easily arrive at

$$\begin{aligned} &\frac{5}{4}(|\vartheta^{n+2}|^2 - |\vartheta^{n+1}|^2) + \frac{1}{4}(|\vartheta^{n+1}|^2 - |\vartheta^n|^2) - [(\vartheta^{n+2}, \vartheta^{n+1}) - (\vartheta^{n+1}, \vartheta^n)] \\ &\quad + k(1 - \lambda_1 - \frac{3}{2}\lambda_2 - 3\varepsilon)\|\vartheta^{n+2}\|^2 \leq \lambda_2 k\|\vartheta^{n+1}\|^2 + \frac{1}{2}\lambda_2 k\|\vartheta^n\|^2 \\ &\quad + \frac{\mu_1^2}{4\varepsilon}k|\vartheta^{n+2}|^2 + \frac{\mu_2^2}{\varepsilon}k|\vartheta^{n+1}|^2 + \frac{\mu_2^2}{4\varepsilon}k|\vartheta^n|^2. \end{aligned}$$

Choosing here  $\varepsilon := (1 - \lambda_1 - 3\lambda_2)/6$ , we obtain

$$\begin{aligned} &\frac{5}{4}(|\vartheta^{n+2}|^2 - |\vartheta^{n+1}|^2) + \frac{1}{4}(|\vartheta^{n+1}|^2 - |\vartheta^n|^2) - [(\vartheta^{n+2}, \vartheta^{n+1}) - (\vartheta^{n+1}, \vartheta^n)] \\ &\quad + \frac{1}{2}(1 - \lambda_1)k\|\vartheta^{n+2}\|^2 \leq \lambda_2 k\|\vartheta^{n+1}\|^2 + \frac{1}{2}\lambda_2 k\|\vartheta^n\|^2 \\ &\quad + Ck(|\vartheta^{n+2}|^2 + |\vartheta^{n+1}|^2 + |\vartheta^n|^2), \end{aligned}$$

with  $C := C_\varepsilon := \max(\mu_1^2, 4\mu_2^2)/(4\varepsilon)$ . Now let  $\varepsilon := 4(1 - \lambda_1)$ . Then,  $\lambda_2 < c/6$  and we have

$$\begin{aligned} &\frac{5}{4}(|\vartheta^{n+2}|^2 - |\vartheta^{n+1}|^2) + \frac{1}{4}(|\vartheta^{n+1}|^2 - |\vartheta^n|^2) - [(\vartheta^{n+2}, \vartheta^{n+1}) - (\vartheta^{n+1}, \vartheta^n)] \\ &\quad + \frac{c}{2}k(\|\vartheta^{n+2}\|^2 - \|\vartheta^{n+1}\|^2) + \frac{c}{12}k(\|\vartheta^{n+1}\|^2 - \|\vartheta^n\|^2) \\ &\leq Ck(|\vartheta^{n+2}|^2 + |\vartheta^{n+1}|^2 + |\vartheta^n|^2). \end{aligned}$$

Summing here from  $n = 0$  to  $n = \ell$ , we obtain

$$\begin{aligned} &\frac{5}{4}(|\vartheta^{\ell+2}|^2 - |\vartheta^1|^2) + \frac{1}{4}(|\vartheta^{\ell+1}|^2 - |\vartheta^0|^2) - (\vartheta^{\ell+2}, \vartheta^{\ell+1}) + \frac{c}{4}k\|\vartheta^{n+2}\|^2 \\ &\leq 3Ck \sum_{n=0}^{\ell+2} |\vartheta^n|^2 + \frac{c}{12}k(2\|\vartheta^1\|^2 + \|\vartheta^0\|^2) - (\vartheta^1, \vartheta^0), \end{aligned}$$

whence, easily,

$$\|\vartheta^{\ell+2}\|_c^2 \leq 12Ck \sum_{n=0}^{\ell+2} |\vartheta^n|^2 + \tilde{c}(\|\vartheta^0\|_c^2 + \|\vartheta^1\|_c^2).$$

Now, a straightforward application of the discrete Gronwall inequality, for sufficiently small  $k$ , leads to the desired stability estimate

$$(7.8) \quad \|\vartheta^n\|_c^2 \leq \tilde{C}(\|\vartheta^0\|_c^2 + \|\vartheta^1\|_c^2),$$

for  $n = 1, \dots, N$ .

7.2.2. *Error estimates.* Let  $U^0 := u^0$ ,  $U^1$  be defined by the implicit–explicit Euler scheme,

$$(I + kA)U^1 - kB_1(t^1, U^1) = U^0 + kB_2(t^0, U^0),$$

and  $U^2, \dots, U^N$  be defined by the implicit–explicit two–step BDF method,

$$\begin{aligned} \frac{3}{2}U^{n+2} - 2U^{n+1} + \frac{1}{2}U^n + kAU^{n+2} - kB_1(t^{n+2}, U^{n+2}) \\ = 2kB_2(t^{n+1}, U^{n+1}) - kB_2(t^n, U^n), \end{aligned}$$

$n = 0, \dots, N-2$ . Assuming that the solution  $u$  is sufficiently smooth and combining the stability estimate (7.8) with the consistency estimate (4.4), with  $p = 2$  for the implicit–explicit two–step BDF scheme, and the corresponding results for the first step by the implicit–explicit Euler method, we easily establish the desired error estimate

$$(7.9) \quad \max_{0 \leq n \leq N} |u(t^n) - U^n| \leq Ck^2.$$

**Acknowledgment.** The author is grateful to Professor Michel Crouzeix for useful discussions.

#### REFERENCES

1. G. Akrivis and M. Crouzeix, *Linearly implicit methods for nonlinear parabolic equations*, Math. Comp. **73** (2004) 613–635. MR **2005a**:65097
2. G. Akrivis, M. Crouzeix and Ch. Makridakis, *Implicit–explicit multistep finite element methods for nonlinear parabolic problems*, Math. Comp. **67** (1998) 457–477. MR **98g**:65088
3. G. Akrivis, M. Crouzeix and Ch. Makridakis, *Implicit–explicit multistep methods for quasi-linear parabolic equations*, Numer. Math. **82** (1999) 521–541. MR **2000e**:65075
4. G. Akrivis and F. Karakatsani, *Modified implicit–explicit BDF methods for nonlinear parabolic equations*, BIT Numer. Math. **43** (2003) 467–483. MR **2004m**:65139
5. G. Akrivis and Y.–S. Smyrlis, *Linearly implicit schemes for a class of dispersive–dissipative systems*, Calcolo **48** (2011) 145–172.
6. M. Crouzeix, *Une méthode multipas implicite–explicite pour l’approximation des équations d’évolution paraboliques*, Numer. Math. **35** (1980) 257–276. MR **82b**:65084
7. R. D. Grigorieff, *Numerik gewöhnlicher Differentialgleichungen, Bd. 2, Mehrschrittverfahren*, Teubner Studienbücher, Stuttgart, 1977. MR **58** #31842
8. R. D. Grigorieff and J. Schroll, *Über  $A(\alpha)$ –stabile Verfahren hoher Konsistenzordnung*, Computing **20** (1978) 343–350. MR **83b**:65086
9. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential–Algebraic Problems*, 2<sup>nd</sup> revised ed., Springer–Verlag, Berlin Heidelberg, Springer Series in Computational Mathematics v. 14, 2010. MR **2011f**:65001
10. P. Henrici, *Discrete Variable Methods in Ordinary Differential Equations*, J. Wiley & Sons, New York, London, 1962. MR **24** #B1772
11. W. Hundsdorfer and S. J. Ruuth, *IMEX extensions of linear multistep methods with general monotonicity and boundedness properties*, J. Comp. Phys. **225** (2007) 2016–2042. MR **2009f**:65185
12. R. Jeltsch, *Stiff stability and its relation to  $A_0$ – and  $A(0)$ –stability*, SIAM J. Numer. Anal. **13** (1976) 8–17. MR **53** #14913
13. W. Liniger, *A criterion for  $A$ –stability of linear multistep integration formulae*, Computing **3** (1968) 280–285. MR **39** #1120

14. C. Lubich, *On the convergence of multistep methods for nonlinear stiff differential equations*, Numer. Math. **58** (1991) 839–853. MR **92d**:65127
15. S. Nørsett, *A criterion for  $A(\alpha)$ –stability of linear multistep methods*, BIT **9** (1969) 259–263. MR **41** #1227
16. G. Savaré,  *$A(\Theta)$ –stable approximations of abstract Cauchy problems*, Numer. Math. **65** (1993) 319–335. MR **94h**:65062
17. V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*. 2<sup>nd</sup> ed., Springer–Verlag, Berlin, 2006. MR **2007b**:65003

COMPUTER SCIENCE DEPARTMENT, UNIVERSITY OF IOANNINA, 451 10 IOANNINA, GREECE  
*E-mail address*: akrivis@cs.uoi.gr