Numerische Mathematik manuscript No.
(will be inserted by the editor)

Fully implicit, linearly implicit and implicit–explicit backward difference formulae for quasi-linear parabolic equations

Georgios Akrivis · **Christian Lubich**

Received: 18 March 2014 / Revised: 6 November 2014, 16 December 2014

Abstract Quasi-linear parabolic equations are discretised in time by fully implicit backward difference formulae (BDF) as well as by implicit–explicit and linearly implicit BDF methods up to order 5. Under appropriate stability conditions for the various methods considered, we establish optimal order a priori error bounds by energy estimates, which become applicable via the Nevanlinna-Odeh multiplier technique.

Keywords quasi-linear parabolic equations · BDF methods · energy technique · multiplier technique · stability · error estimates

Mathematics Subject Classification (2000) 65M12 · 65M60 · 65L06

En l'honneur de Michel Crouzeix à l'occasion de son soixante-dixième anniversaire

1 Introduction

In this paper we study stability and convergence of time discretisations of quasi-linear parabolic equations. The time integration methods considered are variants of backward difference formulae (BDF) up to order 5, which include the standard fully

The work of G.A. was partially supported by GSRT-ESET "Excellence" grant 1456.

G. Akrivis
Department of Computer Science & Engineering, University of Ioannina,
451 10 Ioannina, Greece
E-mail: akrivis@cs.uoi.gr

Ch. Lubich (✉)
Mathematisches Institut, Universität Tübingen,
Auf der Morgenstelle, D-72076 Tübingen, Germany
E-mail: lubich@na.uni-tuebingen.de

implicit BDF method as well as computationally less expensive linearly implicit and implicit–explicit variants. Such methods have been studied previously for non-linear parabolic problems with temporally constant elliptic operator [1,3,5,8] and for linear parabolic problems with time-dependent operators [12], using spectral and Fourier techniques.

On the other hand, energy techniques for first- (implicit Euler) and second-order BDF methods have been used for parabolic problems in [14] and [13]. These energy arguments rely on the A-stability property of the methods in the equivalent form of Dahlquist's G-stability [6]. The restriction to A-stable multistep methods in the use of energy techniques has been overcome in [11], where the multiplier technique for A(θ)-stable methods is developed and applied to stiff ordinary differential equations. Apart from some preliminary remarks in [11], this powerful technique has not been used in the numerical analysis of parabolic problems until fairly recently, in [9], where a class of linear problems with time-dependent operators is considered.

For quasi-linear parabolic problems, implicit Runge–Kutta methods have been studied in [10] using both energy and Fourier techniques.

Here, we will use the Nevanlinna-Odeh multiplier technique of [11], in a way similar to [9], in studying BDF methods and their linearly implicit and implicit–explicit variants, up to order 5, when they are applied to quasi-linear parabolic problems in the setting of [10]. We give particular attention to the arising stability conditions.

In Section 2 we formulate the general problem setting and the numerical methods to be studied. We consider an abstract setting that encompasses quasi-linear parabolic partial differential equations as well as their finite element semi-discretisations in space. In Section 3 we discuss existence of the numerical solutions and the consistency errors of the various methods. Section 4 gives the stability and error analysis of the fully implicit BDF methods up to order 5, while Sections 5 and 6 deal with the implicit–explicit and linearly implicit BDF variants, respectively. In Section 7 we study the case of Hermitian elliptic operators in the quasi-linear parabolic problem, for which we require less stringent stability conditions that are independent of the boundedness-coercivity ratio of the operators for all BDF methods up to order 5. This substantial improvement in the stability conditions is obtained by using time- and state-dependent norms in the analysis.

2 Setting and Preliminaries

2.1 Abstract setting

Let $T>0, u^0 \in H$, and consider an abstract initial value problem for a possibly quasi-linear parabolic equation

$$\begin{cases} u'(t)+A(t,u(t))u(t)=B(t,u(t)), \quad 0<t<T,\\ u(0)=u^0, \end{cases} \tag{1}$$

in the following setting, cf. [10]: Let H and V be separable complex Hilbert spaces with norms $|\cdot|$ and $\|\cdot\|$, respectively, such that V is densely and continuously embedded in H. The norm of the dual space V' is denoted by $\|\cdot\|_\star$. We identify H and its dual H', so that $V \subset H = H' \subset V'$, and the duality pairing $(\cdot,\cdot)$ between V' and V coincides on $H \times V$ with the inner product of H. We assume that, uniformly for all $w \in V$, the sesquilinear form associated with the linear operators $A(t,w) : V \to V'$ satisfies the *coercivity* inequality

$$\mathrm{Re}(A(t,w)\upsilon,\upsilon) \geq \kappa(t)\|\upsilon\|^2 \quad \forall \upsilon \in V, \tag{2}$$

with a smooth positive function $\kappa : [0,T] \to \mathbb{R}$, and is *bounded* by

$$|(A(t,w)\upsilon,\tilde{\upsilon})| \leq \nu(t)\|\upsilon\|\,\|\tilde{\upsilon}\| \quad \forall \upsilon,\tilde{\upsilon} \in V, \tag{3}$$

with a smooth positive function $\nu : [0,T] \to \mathbb{R}$.

Furthermore, we assume that the operators $A(t,\cdot)$ satisfy the restricted Lipschitz condition along the exact solution $u(t)$,

$$\|\big(A(t,w) - A(t,\tilde{w})\big)u(t)\|_\star \leq \lambda(t)\|w-\tilde{w}\| + \mu|w-\tilde{w}| \quad \forall w,\tilde{w} \in V, \tag{4}$$

for all $t \in [0,T]$, with a smooth nonnegative function $\lambda : [0,T] \to \mathbb{R}$. Typically in the applications this is satisfied if the solution $u(t)$ has sufficient regularity such as gradients bounded in the $L^\infty-$norm; see, e.g., the example below.

We furthermore assume that $B(t,\cdot)$ satisfies the following local Lipschitz condition in a ball $\mathscr{B}_{u(t)} := \{\upsilon \in V : \|\upsilon - u(t)\| \leq 1\}$, centred at the value $u(t)$ of the solution u at time t, and, for simplicity, defined here in terms of the norm of V,

$$\|B(t,w) - B(t,\tilde{w})\|_\star \leq \tilde{\lambda}(t)\|w-\tilde{w}\| + \tilde{\mu}|w-\tilde{w}| \quad \forall w,\tilde{w} \in \mathscr{B}_{u(t)}, \tag{5}$$

for all $t \in [0,T]$, with a smooth nonnegative function $\tilde{\lambda} : [0,T] \to \mathbb{R}$ and an arbitrary constant $\tilde{\mu}$. We always assume

$$\lambda(t) + \tilde{\lambda}(t) < \kappa(t), \tag{6}$$

which yields stability of the implicit Euler method; see Theorem 1 below for $q = 1$. In many applications, one typically has that for every $\delta > 0$ one can choose $\lambda(t) \leq \delta$ and $\tilde{\lambda}(t) \leq \delta$ and appropriate μ and $\tilde{\mu}$ depending on δ. Nevertheless, it will be of interest to see how small $\lambda(t)$ and $\tilde{\lambda}(t)$ need to be to ensure stability of the various numerical schemes. Moreover, we are interested in understanding for which methods and under which assumptions on the operators we obtain stability estimates independently of the boundedness-coercivity ratio $\nu(t)/\kappa(t)$.

2.2 An example

On a smooth bounded domain $\Omega \subset \mathbb{R}^d$, consider the quasi-linear parabolic equation (with time-independent coefficient functions for notational simplicity)

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\Big(a_{ij}(u(x,t))\frac{\partial u}{\partial x_j}\Big) + \sum_{l=1}^{d} b_l(u(x,t))\frac{\partial u}{\partial x_l} + c(u(x,t)) \tag{7}$$

for $x \in \Omega,\ 0 < t \le T$, with homogeneous Neumann boundary conditions

$$\sum_{i,j=1}^{d} n_i \cdot a_{ij}(u(x,t)) \frac{\partial u}{\partial x_j} = 0, \quad x \in \partial\Omega,\ 0 < t \le T,$$

where $(n_1, \dots, n_d)(x)$ denotes the outward pointing unit normal vector to $\partial\Omega$. The coefficient functions $a_{ij}, b_l, c : \mathbb{R} \to \mathbb{C}$ are assumed to be bounded and Lipschitz bounded, and the matrices $\mathscr{A}(\mu) := \big(a_{ij}(\mu)\big), \mu \in \mathbb{R}$, satisfy the conditions

$$\operatorname{Re}\big(z^* \mathscr{A}(\mu) z\big) \ge \kappa z^* z \quad \forall z \in \mathbb{C}^d \quad \text{and} \quad \|\mathscr{A}(\mu)\|_2 \le \nu. \tag{8}$$

The variational formulation of this problem is of the form (1) on $H = L^2(\Omega)$ and $V = H^1(\Omega)$, with the operator $A(w) : V \to V'$ defined by

$$(A(w)\upsilon, \tilde{\upsilon}) = \int_\Omega \sum_{i,j=1}^{d} a_{ij}(w(x)) \frac{\partial \upsilon}{\partial x_j} \frac{\partial \tilde{\upsilon}}{\partial x_i}\, dx.$$

This satisfies (2) and (3) with $\kappa(t) = \kappa$ and $\nu(t) = \nu$ of (8). Condition (4) holds (even with $\lambda(t) = 0$) if $\nabla_x u(\cdot, t) \in L^\infty(\Omega)$ because for $w, \tilde{w}, \upsilon \in V$ we have

$$\Big| \int_\Omega \sum_{i,j=1}^{d} \big[a_{ij}(w(x)) - a_{ij}(\tilde{w}(x))\big] \frac{\partial u}{\partial x_j} \frac{\partial \upsilon}{\partial x_i}\, dx \Big| \le \ell \cdot \|w - \tilde{w}\|_{L^2} \cdot \|\nabla_x u\|_{L^\infty} \cdot \|\upsilon\|_{H^1},$$

where ℓ denotes a Lipschitz constant of the functions $a_{ij}(\cdot)$.

If the operator $B(u)$ is defined by the two last terms of (7), then we write, using Green's formula,

$$\begin{aligned}
\int_\Omega \sum_{l=1}^{d} \Big(b_l(w) \frac{\partial w}{\partial x_l} - b_l(\tilde{w}) \frac{\partial \tilde{w}}{\partial x_l} \Big)\, \upsilon\, dx &= \int_\Omega \sum_{l=1}^{d} \Big(b_l(w) - b_l(\tilde{w}) \Big) \frac{\partial w}{\partial x_l}\, \upsilon\, dx \\
&+ \int_{\partial\Omega} \sum_{l=1}^{d} b_l(\tilde{w})\, (w - \tilde{w})\, \upsilon\, n_l\, d\sigma \\
&- \int_\Omega \sum_{l=1}^{d} \Big(b_l'(\tilde{w}) \frac{\partial \tilde{w}}{\partial x_l}\, (w - \tilde{w})\, \upsilon\, dx + b_l(\tilde{w})\, (w - \tilde{w}) \frac{\partial \upsilon}{\partial x_l} \Big) dx.
\end{aligned}$$

For dimension $d \le 3$ we can use the Cauchy-Schwarz inequality in the form

$$\Big| \int_\Omega abc\, dx \Big| \le \|a\|_{L^4} \|b\|_{L^4} \|c\|_{L^2}$$

and the estimate, for arbitrary $\vartheta > 0$ and $\upsilon \in H^1(\Omega)$,

$$\|\upsilon\|_{L^4} \le \vartheta \|\upsilon\|_{H^1} + C(\vartheta) \|\upsilon\|_{L^2},$$

which is an easy consequence of the Gagliardo-Nirenberg and the Young inequalities,

$$\begin{aligned}
\|\upsilon\|_{L^4} &\le C \|\upsilon\|_{H^1}^{\frac{d}{4}} \|\upsilon\|_{L^2}^{1-\frac{d}{4}} = C\big(\tilde{\vartheta} \|\upsilon\|_{H^1}\big)^{\frac{d}{4}}\, \tilde{\vartheta}^{-\frac{d}{4}} \|\upsilon\|_{L^2}^{1-\frac{d}{4}} \\
&\le C \Big[\frac{d}{4} \tilde{\vartheta} \|\upsilon\|_{H^1} + \frac{4-d}{4} \tilde{\vartheta}^{-\frac{d}{4-d}} \|\upsilon\|_{L^2} \Big],
\end{aligned}$$

to conclude that condition (5) holds with $\tilde{\lambda}(t) \leq \delta$ for any given $\delta > 0$ (and with $\tilde{\mu}$ depending on δ).

Alternatively, we may regroup the first two terms on the right-hand side of (7) into $A(u)u$ and only the last term into $B(u)$. In this latter case, the conditions hold with $\lambda(t) = \tilde{\lambda}(t) = 0$.

Our abstract framework applies equally to finite element space discretisations of the initial-boundary value problem, uniformly in the spatial grid size h. In the spatially discrete case, condition (4) is required for a projection of the spatially continuous solution $u(\cdot,t)$ onto the finite element space.

2.3 The numerical methods

For $q = 1,\dots,5$, consider the implicit $q-$step BDF method (α,β) and the explicit $q-$step method (α,γ) described by the polynomials α,β and γ,

$$\begin{cases} \alpha(\zeta) = \sum_{j=1}^{q} \frac{1}{j}\zeta^{q-j}(\zeta-1)^j = \sum_{i=0}^{q}\alpha_i\zeta^i, \quad \beta(\zeta) = \zeta^q, \\ \gamma(\zeta) = \zeta^q - (\zeta-1)^q = \sum_{i=0}^{q-1}\gamma_i\zeta^i. \end{cases} \tag{9}$$

The BDF methods are $A-$stable for $q = 1$ and $q = 2$, i.e., $A(\vartheta_q)-$stable with $\vartheta_1 = \vartheta_2 = 90^\circ$, and $A(\vartheta_q)-$stable for $q = 3,\dots,5$ with $\vartheta_3 = 86.03^\circ$, $\vartheta_4 = 73.35^\circ$ and $\vartheta_5 = 51.84^\circ$; see [7, Section V.2]. Their order is q. For the α given in (9), the scheme (α,γ) is the unique explicit $q-$step scheme of order q; the order of all other explicit $q-$step schemes $(\alpha,\tilde{\gamma})$ is at most $q-1$.

Let the integer $N \geq q$, and consider a uniform partition $t^n := nk, n = 0,\dots,N$, of the interval $[0,T]$, with time step $k := T/N$. Assuming we are given starting approximations $U^0,\dots,U^{q-1}$, we discretise (1) in time by the fully implicit $(\alpha,\beta)-$scheme, i.e., we define approximations U^m to the nodal values $u^m := u(t^m)$ of the exact solution as follows:

$$\sum_{i=0}^{q}\alpha_i U^{n+i} + kA(t^{n+q},U^{n+q})U^{n+q} = kB(t^{n+q},U^{n+q}), \tag{10}$$

or by the $q-$step implicit–explicit $(\alpha,\beta,\gamma)-$scheme,

$$\sum_{i=0}^{q}\alpha_i U^{n+i} + kA(t^{n+q},U^{n+q})U^{n+q} = k\sum_{i=0}^{q-1}\gamma_i B(t^{n+i},U^{n+i}), \tag{11}$$

$n = 0,\dots,N-q$. The scheme (11) is referred to as the $q-$step implicit–explicit BDF method. The unknown U^{n+q} appears only on the left-hand side of (11).

Since equation (11) is in general nonlinear in the unknown U^{n+q}, we will also consider the following linearly implicit modification:

$$\sum_{i=0}^{q}\alpha_i U^{n+i} + kA\Big(t^{n+q},\sum_{i=0}^{q-1}\gamma_i U^{n+i}\Big)U^{n+q} = k\sum_{i=0}^{q-1}\gamma_i B(t^{n+i},U^{n+i}), \tag{12}$$

for $n = 0,\dots,N-q$. Notice that now the unknown U^{n+q} appears in (12) only linearly; therefore, to advance with (12) in time, we only need to solve, at each time level, just one linear equation, which reduces to a linear system if we discretise also in space.

To simplify the notation a little bit, for a given sequence $\upsilon^0,\dots,\upsilon^N$, we denote by $\hat{\upsilon}^{n+q}$ the following linear combination of $\upsilon^n,\dots,\upsilon^{n+q-1}$

$$\hat{\upsilon}^{n+q} := \sum_{i=0}^{q-1} \gamma_i \upsilon^{n+i}.$$

We let $B^m := B(t^m, U^m)$ and write (12) equivalently in the form

$$\sum_{i=0}^{q} \alpha_i U^{n+i} + kA(t^{n+q}, \widehat{U}^{n+q})U^{n+q} = k\widehat{B}^{n+q}, \tag{13}$$

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

2.4 Auxiliary results by Dahlquist and Nevanlinna & Odeh

We will use the following result from Dahlquist's G-stability theory.

Lemma 1 ([6]; see also [4] and [7, Section V.6]) *Let $\alpha(\zeta) = \alpha_q\zeta^q + \cdots + \alpha_0$ and $\mu(\zeta) = \mu_q\zeta^q + \cdots + \mu_0$ be polynomials of degree at most q (and at least one of them of degree q) that have no common divisor. Let $(\cdot,\cdot)$ be an inner product with associated norm $|\cdot|$. If*

$$\operatorname{Re}\frac{\alpha(\zeta)}{\mu(\zeta)} > 0 \quad \textit{for } |\zeta| > 1,$$

then there exists a symmetric positive definite matrix $G = (g_{ij}) \in \mathbb{R}^{q\times q}$ and real $\delta_0,\dots,\delta_q$ such that for $\upsilon^0,\dots,\upsilon^q$ in the inner product space,

$$\operatorname{Re}\Big(\sum_{i=0}^{q}\alpha_i\upsilon^i, \sum_{j=0}^{q}\mu_j\upsilon^j\Big) = \sum_{i,j=1}^{q} g_{ij}(\upsilon^i,\upsilon^j) - \sum_{i,j=1}^{q} g_{ij}(\upsilon^{i-1},\upsilon^{j-1}) + \Big|\sum_{i=0}^{q}\delta_i\upsilon^i\Big|^2.$$

In combination with the preceding result for $\mu(\zeta) = \zeta^q - \eta_q\zeta^{q-1}$, the following property of BDF methods up to order 5 will assume a key role in our stability analysis.

Lemma 2 ([11]) *For $q \le 5$, there exists $0 \le \eta_q < 1$ such that the generating polynomial of the qth order BDF method, $\alpha(\zeta) = \sum_{j=1}^{q}\frac{1}{j}\zeta^{q-j}(\zeta-1)^j$, satisfies*

$$\operatorname{Re}\frac{\alpha(\zeta)}{\zeta^q - \eta_q\zeta^{q-1}} > 0 \quad \textit{for } |\zeta| > 1.$$

The smallest possible values of η_q are

$$\eta_1 = \eta_2 = 0,\ \eta_3 = 0.0836,\ \eta_4 = 0.2878,\ \eta_5 = 0.8160.$$

3 Existence, uniqueness, consistency

3.1 Existence and uniqueness of the approximations

In the stability and error bounds given below, we will always tacitly assume that a numerical solution exists. It is, however, of interest to clarify situations where this assumption can be guaranteed to hold true.

First, in the case of the linearly implicit method (12), existence and uniqueness of the approximations $U^q,\dots,U^N$ can be easily established by the Lax-Milgram lemma, using (2) and (3).

In the case of the implicit–explicit method (11), we can prove existence of approximations $U^q,\dots,U^N$ by Brouwer's fixed-point theorem, assuming that our Hilbert spaces are finite dimensional; this applies, if we discretise also in space, for instance, by the finite element method. To this end, for $w\in V'$ and $t\in[0,T]$, we consider the continuous mapping $G:V\to V', G(\upsilon):=\alpha_q\upsilon+kA(t,\upsilon)\upsilon-w$, and will show that it vanishes at some point $\tilde\upsilon\in V$. With $\rho:=\big(\frac{1}{\alpha_q^2}|w|^2+1\big)^{1/2}$, let $\mathscr{B}_\rho:=\{\upsilon\in V:|\upsilon|\le\rho\}$ be the ball of radius ρ, centred at the origin. We shall show that G vanishes at some point $\tilde\upsilon\in\mathscr{B}_\rho$ by contradiction. First, using (2), for $\upsilon\in V$, we have

$$\begin{aligned}\operatorname{Re}(G(\upsilon),\upsilon)&=\alpha_q|\upsilon|^2+k\operatorname{Re}(A(t,\upsilon)\upsilon,\upsilon)-\operatorname{Re}(w,\upsilon)\\&\ge\alpha_q|\upsilon|^2+k\kappa(t)\|\upsilon\|^2-\frac{\alpha_q}{2}|\upsilon|^2-\frac{1}{2\alpha_q}|w|^2\\&\ge\frac{\alpha_q}{2}|\upsilon|^2-\frac{1}{2\alpha_q}|w|^2,\end{aligned}$$

and infer easily that

$$\operatorname{Re}(G(\upsilon),\upsilon)>0\quad\forall\upsilon\in V,|\upsilon|=\rho.\tag{14}$$

Now, if G does not vanish in $\mathscr{B}_\rho$, then the mapping

$$F:\mathscr{B}_\rho\to\mathscr{B}_\rho,\quad F(\upsilon):=-\rho\,\frac{G(\upsilon)}{|G(\upsilon)|},$$

is continuous, and, according to Brouwer's fixed-point theorem, it has a fixed point $\upsilon^\star\in\mathscr{B}_\rho, F(\upsilon^\star)=\upsilon^\star$. Since $|F(\upsilon^\star)|=\rho$, we also have $|\upsilon^\star|=\rho$, whence, in view of (14),

$$\rho^2=|\upsilon^\star|^2=\operatorname{Re}(F(\upsilon^\star),\upsilon^\star)=-\rho\,\frac{\operatorname{Re}(G(\upsilon^\star),\upsilon^\star)}{|G(\upsilon^\star)|}<0,$$

a contradiction.

The same argument applies also to the fully implicit scheme (10), for sufficiently small time step k, provided we replace the local Lipschitz condition (5) on B by its global counterpart, i.e., if we require the inequality to hold for all $w,\tilde w\in V$, instead of $w,\tilde w\in\mathscr{B}_{u(t)}=\{\upsilon\in V:\|\upsilon-u(t)\|\le1\}$. Indeed, if we first rewrite the mapping $G:V\to V', G(\upsilon):=\alpha_q\upsilon+kA(t,\upsilon)\upsilon-kB(t,\upsilon)-\tilde w$ in the form

$$G(\upsilon)=\alpha_q\upsilon+kA(t,\upsilon)\upsilon-k\big[B(t,\upsilon)-B(t,0)\big]-w,$$

with $w := kB(t,0) + \tilde{w}$, take the inner product with υ, and use (2) and the global version of the local Lipschitz condition (5), we obtain, for positive ε such that $\varepsilon + \tilde{\lambda}(t) \le \kappa(t)$ for all $t \in [0,T]$ (recall (6)),

$$\begin{aligned}
\mathrm{Re}(G(\upsilon),\upsilon) &= \alpha_q|\upsilon|^2 + k\,\mathrm{Re}(A(t,\upsilon)\upsilon,\upsilon) - k\,\mathrm{Re}\big(B(t,\upsilon)-B(t,0),\upsilon\big) - \mathrm{Re}(w,\upsilon)\\
&\ge \alpha_q|\upsilon|^2 + k\kappa(t)\|\upsilon\|^2 - k\tilde{\lambda}(t)\|\upsilon\|^2 - k\tilde{\mu}|\upsilon|\,\|\upsilon\| - \frac{\alpha_q}{2}|\upsilon|^2 - \frac{1}{2\alpha_q}|w|^2\\
&\ge \frac{\alpha_q}{2}|\upsilon|^2 + k\kappa(t)\|\upsilon\|^2 - k\tilde{\lambda}(t)\|\upsilon\|^2 - k\varepsilon\|\upsilon\|^2 - \frac{1}{4\varepsilon}k\tilde{\mu}^2|\upsilon|^2 - \frac{1}{2\alpha_q}|w|^2\\
&\ge \frac{1}{2}\Big[\Big(\alpha_q - \frac{1}{2\varepsilon}k\tilde{\mu}^2\Big)|\upsilon|^2 - \frac{1}{\alpha_q}|w|^2\Big],
\end{aligned}$$

and infer easily that $\mathrm{Re}(G(\upsilon),\upsilon)$ is positive for all $\upsilon \in V$ with $|\upsilon| = \rho$, with a suitable positive ρ, provided k is sufficiently small. As before, this shows that G vanishes at some point $\tilde{\upsilon} \in \mathscr{B}_\rho$, which in turn yields existence of $U^q,\dots,U^N$ satisfying the fully implicit scheme (10), again in the case of finite dimensional Hilbert spaces.

We next show local uniqueness of the approximations for the nonlinear schemes (10) and (11), assuming a stronger version of (4), namely

$$\|\big(A(t,w)-A(t,\tilde{w})\big)\upsilon\|_\star \le \lambda(t)\|w-\tilde{w}\| + \mu|w-\tilde{w}| \quad \forall w,\tilde{w} \in V, \tag{15}$$

for all $\upsilon \in \mathscr{B}_{u(t)}$, and that the time step k is sufficiently small. We shall only consider the scheme (11); the argument applies also to the scheme (10). For a given $w \in V'$, we assume that

$$\alpha_q\upsilon + kA(t,\upsilon)\upsilon = kB(t,\upsilon) + w \quad \text{and} \quad \alpha_q\tilde{\upsilon} + kA(t,\tilde{\upsilon})\tilde{\upsilon} = kB(t,\tilde{\upsilon}) + w,$$

with $\upsilon,\tilde{\upsilon} \in \mathscr{B}_{u(t)}$, and will show that $\upsilon = \tilde{\upsilon}$. Subtracting the second relation from the first, adding and subtracting the term $A(t,\upsilon)\tilde{\upsilon}$, and taking the inner product with $\upsilon - \tilde{\upsilon}$, we obtain

$$\begin{aligned}
&\alpha_q|\upsilon-\tilde{\upsilon}|^2 + k\,\mathrm{Re}\big(A(t,\upsilon)(\upsilon-\tilde{\upsilon}),\upsilon-\tilde{\upsilon}\big)\\
&\qquad = k\,\mathrm{Re}\big((A(t,\tilde{\upsilon})-A(t,\upsilon))\tilde{\upsilon},\upsilon-\tilde{\upsilon}\big) + k\,\mathrm{Re}\big(B(t,\upsilon)-B(t,\tilde{\upsilon}),\upsilon-\tilde{\upsilon}\big).
\end{aligned}$$

Now, estimating the second term on the left-hand side from below using (2), and the terms on the right-hand side using (15) and (5), respectively, we get

$$\begin{aligned}
\alpha_q|\upsilon-\tilde{\upsilon}|^2 + k\kappa(t)\|\upsilon-\tilde{\upsilon}\|^2 &\le k\lambda(t)\|\upsilon-\tilde{\upsilon}\|^2 + k\mu|\upsilon-\tilde{\upsilon}|\,\|\upsilon-\tilde{\upsilon}\|\\
&\quad + k\tilde{\lambda}(t)\|\upsilon-\tilde{\upsilon}\|^2 + k\tilde{\mu}|\upsilon-\tilde{\upsilon}|\,\|\upsilon-\tilde{\upsilon}\|.
\end{aligned}$$

Therefore, for positive ε small enough such that $\lambda(t) + \tilde{\lambda}(t) + \varepsilon \le \kappa(t)$ (recall (6)),

$$\begin{aligned}
\alpha_q|\upsilon-\tilde{\upsilon}|^2 + k\kappa(t)\|\upsilon-\tilde{\upsilon}\|^2 &\le k\lambda(t)\|\upsilon-\tilde{\upsilon}\|^2 + k\frac{\varepsilon}{2}\|\upsilon-\tilde{\upsilon}\|^2 + k\frac{\mu^2}{2\varepsilon}|\upsilon-\tilde{\upsilon}|^2\\
&\quad + k\tilde{\lambda}(t)\|\upsilon-\tilde{\upsilon}\|^2 + k\frac{\varepsilon}{2}\|\upsilon-\tilde{\upsilon}\|^2 + k\frac{\tilde{\mu}^2}{2\varepsilon}|\upsilon-\tilde{\upsilon}|^2,
\end{aligned}$$

whence

$$\left[2\varepsilon\alpha_q - k(\mu^2+\tilde{\mu}^2)\right]|\upsilon-\tilde{\upsilon}|^2 \le 0$$

and we infer that $\upsilon = \tilde{\upsilon}$, for sufficiently small k. This argument shows local uniqueness of the approximations U^m of the scheme (11) in $\mathscr{B}_{u(t^m)}$. Notice also that if (15) is satisfied for υ in a set $S_{u(t)}$, then the above argument shows local uniqueness of the approximations U^m in the intersection of $S_{u(t^m)}$ and $\mathscr{B}_{u(t^m)}$.

3.2 Consistency error

The order of the q-step methods (α,β) and (α,γ) is q, i.e.,

$$\sum_{i=0}^{q} i^\ell \alpha_i = \ell q^{\ell-1} = \ell \sum_{i=0}^{q-1} i^{\ell-1}\gamma_i, \quad \ell = 0,1,\dots,q. \tag{16}$$

The consistency errors d^n, $\tilde{d}^n$ and $\check{d}^n$ of the schemes (10), (11) and (13) for the solution u of (1), i.e., the amounts by which the exact solution misses satisfying (10), (11) and (13), respectively, are given by

$$d^n = \frac{1}{k}\Big(\sum_{i=0}^{q}\alpha_i u^{n+i} + kA(t^{n+q},u^{n+q})u^{n+q} - kB(t^{n+q},u^{n+q})\Big), \tag{17}$$

$$\tilde{d}^n = \frac{1}{k}\Big(\sum_{i=0}^{q}\alpha_i u^{n+i} + kA(t^{n+q},u^{n+q})u^{n+q} - k\sum_{i=0}^{q-1}\gamma_i B(t^{n+i},u^{n+i})\Big) \tag{18}$$

and

$$\check{d}^n = \frac{1}{k}\Big(\sum_{i=0}^{q}\alpha_i u^{n+i} + kA(t^{n+q},\hat{u}^{n+q})u^{n+q} - k\sum_{i=0}^{q-1}\gamma_i B(t^{n+i},u^{n+i})\Big), \tag{19}$$

$n = 0,\dots,N-q$, respectively. Here, $u^{n+i} := u(t^{n+i})$ denote the nodal values of the exact solution $u(t)$.

Lemma 3 *The consistency errors* (17)–(19) *are bounded by*

$$\max_{0\le n\le N-q}\|d^n\|_\star \le Ck^q, \quad \max_{0\le n\le N-q}\|\tilde{d}^n\|_\star \le \tilde{C}k^q, \quad \max_{0\le n\le N-q}\|\check{d}^n\|_\star \le \check{C}k^q, \tag{20}$$

provided that the solution u is sufficiently regular.

Proof We first focus on the implicit scheme (10). Using the differential equation in (1), we rewrite (17) in the form

$$d^n = \frac{1}{k}\sum_{i=0}^{q}\alpha_i u^{n+i} - u'(t^{n+q}). \tag{21}$$

Now, by Taylor expanding about t^n, we see that, due to the order conditions of the implicit $(\alpha,\beta)-$scheme, i.e., the first equality in (16), leading terms of order up to $q-1$ cancel, and we obtain

$$d^n = \frac{1}{q!}\Bigg[\frac{1}{k}\sum_{i=0}^{q}\alpha_i\int_{t^n}^{t^{n+i}}(t^{n+i}-s)^q u^{(q+1)}(s)\,ds \\ - q\int_{t^n}^{t^{n+q}}(t^{n+q}-s)^{q-1}u^{(q+1)}(s)\,ds\Bigg]. \tag{22}$$

Thus, under obvious regularity requirements, we obtain the desired optimal order consistency estimate (20) for the scheme (10).

Next, concerning the scheme (11), letting

$$\tilde{d}_2^n := B(t^{n+q},u^{n+q}) - \sum_{i=0}^{q-1}\gamma_i B(t^{n+i},u^{n+i}),$$

and using the differential equation in (1) and (21), we infer that

$$\tilde{d}^n = d^n + \tilde{d}_2^n. \tag{23}$$

Now, by Taylor expanding about t^n and using the second equality in (16), we obtain

$$\tilde{d}_2^n = \frac{1}{(q-1)!}\Bigg[\int_{t^n}^{t^{n+q}}(t^{n+q}-s)^{q-1}\widetilde{B}^{(q)}(s)ds - \sum_{i=0}^{q}\gamma_i\int_{t^n}^{t^{n+i}}(t^{n+i}-s)^{q-1}\widetilde{B}^{(q)}(s)ds\Bigg],$$

with $\widetilde{B}(t) := B(t,u(t)), t\in[0,T]$. Thus, taking the first bound of (20) also into account, we obtain the desired optimal order consistency estimate (20) for the scheme (11), under obvious regularity requirements.

Furthermore, from (18) and (19) we immediately obtain the following relation between $\check{d}^n$ and $\tilde{d}^n$

$$\check{d}^n = \tilde{d}^n + \big(A(t^{n+q},\hat{u}^{n+q}) - A(t^{n+q},u^{n+q})\big)u^{n+q} \tag{24}$$

and infer, in view of (4), that

$$\|\check{d}^n\|_\star \le \|\tilde{d}^n\|_\star + \lambda(t^{n+q})\|\hat{u}^{n+q}-u^{n+q}\|. \tag{25}$$

Now, by Taylor expanding about t^n and using the second equality in (16), exactly as with the term $\tilde{d}_2^n$ above, we obtain

$$\hat{u}^{n+q} - u^{n+q} = \frac{1}{(q-1)!}\Bigg[\int_{t^n}^{t^{n+q}}(t^{n+q}-s)^{q-1}u^{(q)}(s)\,ds \\ - \sum_{i=0}^{q}\gamma_i\int_{t^n}^{t^{n+i}}(t^{n+i}-s)^{q-1}u^{(q)}(s)\,ds\Bigg],$$

and immediately infer that, under obvious regularity requirements,

$$\|\hat{u}^{n+q} - u^{n+q}\| \le ck^q. \tag{26}$$

Combining (25), (26) and the bound for $\tilde{d}^n$, we obtain the desired optimal order consistency estimate (20) for the scheme (13). □

4 Stability and error bounds for the fully implicit BDF method

In this section we establish stability and optimal order error estimates for the fully implicit BDF methods (10).

Theorem 1 *Assume* (2)–(5) *and consider time discretisation by the fully implicit BDF method* (10) *of order* $q \le 5$. *Let* η_q *be as in Lemma* 2, $\eta_1 = \eta_2 = 0, \eta_3 = 0.0836, \eta_4 = 0.2878, \eta_5 = 0.8160$. *Under the stability condition*

$$\kappa(t) - \eta_q \nu(t) - (1+\eta_q)\big[\lambda(t) + \tilde{\lambda}(t)\big] \ge \rho > 0 \quad \forall t \in [0,T], \tag{27}$$

the method is convergent of order q: *for* $t^n \le T$,

$$c_q |U^n - u(t^n)|^2 + \tfrac{1}{3}\rho k \sum_{\ell=q}^{n} \|U^\ell - u(t^\ell)\|^2 \le C\rho^{-1} k^{2q}, \tag{28}$$

provided that the solution is sufficiently regular and the starting values are $O(k^q)$ *accurate in the* H*-norm* $|\cdot|$ *and* $O(k^{q-\frac{1}{2}})$ *in the* V*-norm* $\|\cdot\|$. *In* (28), $c_q > 0$ *depends only on* q *and the constant* C *is independent of* ρ, k *and* n *with* $nk \le T$.

Remark 1 Note that the stability condition (27) is independent of the step size k. As mentioned before, the Lipschitz constants $\lambda(t)$ and $\widetilde{\lambda}(t)$ of (4) and (5), respectively, are often zero or can be chosen arbitrarily small at the expense of increasing μ and $\widetilde{\mu}$, which just leads to a larger exponent in the exponential dependence of C on the final time T, but does not otherwise affect the finite-time stability of the method.

Under condition (27) the boundedness-coercivity ratio $\nu(t)/\kappa(t)$ must be bounded such that $\eta_q \nu(t)/\kappa(t) < 1$ and hence cannot be arbitrarily large when $\eta_q > 0$, i.e., for the BDF methods having order $q \ge 3$, which are not A-stable. This restriction cannot be avoided in the general setting of Section 2.1. In fact, consider the example of a constant operator $A = e^{i\vartheta} H$ $(0 < \vartheta < \pi/2)$ with a Hermitian positive definite operator H. Here $\kappa = \nu \cos\vartheta$ and the method behaves in a stable way only if it is $A(\vartheta)$-stable, so that $\nu/\kappa = 1/\cos\vartheta$ cannot be arbitrarily large for a method that is not A-stable; see [2] for a detailed discussion and a comparison of η_q and the cosine of the angle ϑ_q of $A(\vartheta_q)$-stability for $q = 3,4,5$. In Section 7 we will show, however, that in the case of Hermitian operators $A(t,w)$ we can obtain stability of the BDF methods up to order 5 independently of the boundedness-coercivity ratio $\nu(t)/\kappa(t)$.

Proof Let $e^n := u^n - U^n$ denote the discretisation error of the scheme (10) and $b^n := B(t^n, u^n) - B(t^n, U^n), n = 0, \dots, N$. Subtracting (10) from (17), we obtain

$$\sum_{i=0}^{q} \alpha_i e^{n+i} + k\big[A(t^{n+q}, u^{n+q})u^{n+q} - A(t^{n+q}, U^{n+q})U^{n+q}\big] = kb^{n+q} + kd^n, \tag{29}$$

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

Following the approach of [11] and [9], we take in (29) the inner product with $e^{n+q} - \eta_q e^{n+q-1}$, and take real parts to obtain

$$\begin{aligned} \mathrm{Re}\Big(\sum_{i=0}^{q} \alpha_i e^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big) + kI^{n+q} = kJ^{n+q} & \\ + k\,\mathrm{Re}(d^n, e^{n+q} - \eta_q e^{n+q-1}) & \end{aligned} \tag{30}$$

with

$$I^{n+q} := \operatorname{Re}\big(A(t^{n+q},u^{n+q})u^{n+q} - A(t^{n+q},U^{n+q})U^{n+q}, e^{n+q} - \eta_q e^{n+q-1}\big) \tag{31}$$

and

$$J^{n+q} := \operatorname{Re}\big(b^{n+q}, e^{n+q} - \eta_q e^{n+q-1}\big). \tag{32}$$

The first term on the left-hand side of (30) can be taken care of exactly as in [11] and [9]: From Lemmas 1 and 2, with the notation $E^{n+q} = (e^{n+1},\dots,e^{n+q})^T$ and

$$|E^{n+q}|_G^2 = \sum_{i,j=1}^{q} g_{ij}(e^{n+i}, e^{n+j}),$$

we have

$$\operatorname{Re}\Big(\sum_{i=0}^{q} \alpha_i e^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big) \geq |E^{n+q}|_G^2 - |E^{n+q-1}|_G^2. \tag{33}$$

Since the last term on the right-hand side of (30) can be easily estimated from above, see the estimate following (44) in the sequel, essentially all that remains to be done is to estimate I^{n+q} from below and J^{n+q} from above in an appropriate way.

We start with I^{n+q}. First, we rewrite the expression

$$A(t^{n+q},u^{n+q})u^{n+q} - A(t^{n+q},U^{n+q})U^{n+q}$$

as $A(t^{n+q},U^{n+q})e^{n+q} + \big[A(t^{n+q},u^{n+q}) - A(t^{n+q},U^{n+q})\big]u^{n+q}$; then, (31) reads

$$I^{n+q} = I_1^{n+q} + I_2^{n+q} \tag{34}$$

with

$$I_1^{n+q} := \operatorname{Re}\big(A(t^{n+q},U^{n+q})e^{n+q}, e^{n+q} - \eta_q e^{n+q-1}\big) \tag{35}$$

and

$$I_2^{n+q} := \operatorname{Re}\big(\big[A(t^{n+q},u^{n+q}) - A(t^{n+q},U^{n+q})\big]u^{n+q}, e^{n+q} - \eta_q e^{n+q-1}\big). \tag{36}$$

One part of I_1^{n+q}, namely $\operatorname{Re}\big(A(t^{n+q},U^{n+q})e^{n+q}, e^{n+q}\big)$, can be easily estimated from below using the coercivity condition (2): with $\kappa^{n+q} = \kappa(t^{n+q})$ we get

$$I_1^{n+q} \geq \kappa^{n+q}\|e^{n+q}\|^2 - \eta_q \operatorname{Re}\big(A(t^{n+q},U^{n+q})e^{n+q}, e^{n+q-1}\big); \tag{37}$$

now, the second term on the right-hand side of (37) can be estimated using the bound (3). We then get, with $\nu^{n+q} = \nu(t^{n+q})$,

$$I_1^{n+q} \geq \kappa^{n+q}\|e^{n+q}\|^2 - \eta_q \nu^{n+q}\|e^{n+q}\|\,\|e^{n+q-1}\|,$$

whence

$$I_1^{n+q} \geq \big[\kappa^{n+q} - \tfrac{1}{2}\eta_q \nu^{n+q}\big]\|e^{n+q}\|^2 - \tfrac{1}{2}\eta_q \nu^{n+q}\|e^{n+q-1}\|^2. \tag{38}$$

For simplicity of presentation, we assume $\mu = 0$ in the following, since the case of general μ is treated similarly without any substantial further difficulty. (A larger μ just leads to a larger exponential growth with T of C in (28) via a straightforward use of a discrete Gronwall inequality at the end of the proof.) Then, I_2^{n+q} can be estimated from below using the Lipschitz condition (4); indeed, with $\lambda^{n+q} = \lambda(t^{n+q})$ we have

$$\begin{aligned} I_2^{n+q} &\geq -\|[A(t^{n+q},u^{n+q}) - A(t^{n+q},U^{n+q})]u^{n+q}\|_\star \|e^{n+q} - \eta_q e^{n+q-1}\| \\ &\geq -\lambda^{n+q}\|u^{n+q} - U^{n+q}\|\,\|e^{n+q} - \eta_q e^{n+q-1}\|, \end{aligned}$$

i.e.,

$$I_2^{n+q} \geq -\lambda^{n+q}\|e^{n+q}\|\,\|e^{n+q} - \eta_q e^{n+q-1}\|,$$

whence

$$I_2^{n+q} \geq -\lambda^{n+q}\big(1 + \tfrac{1}{2}\eta_q\big)\|e^{n+q}\|^2 - \tfrac{1}{2}\lambda^{n+q}\eta_q\|e^{n+q-1}\|^2. \tag{39}$$

From (34), (38) and (39), we obtain the desired estimate of I^{n+q} from below, namely

$$\begin{aligned} I^{n+q} \geq \Big[\kappa^{n+q} - \lambda^{n+q} - \tfrac{1}{2}\eta_q\big[\nu^{n+q} + \lambda^{n+q}\big]\Big]\|e^{n+q}\|^2 \\ - \tfrac{1}{2}\eta_q\big[\nu^{n+q} + \lambda^{n+q}\big]\|e^{n+q-1}\|^2. \end{aligned} \tag{40}$$

As far as J^{n+q} is concerned, in view of the Lipschitz condition (5), we have, assuming also $\tilde{\mu} = 0$ for ease of presentation,

$$\begin{aligned} J^{n+q} &\leq \|b^{n+q}\|_\star \|e^{n+q} - \eta_q e^{n+q-1}\| \\ &\leq \tilde{\lambda}^{n+q}\|e^{n+q}\|\,\|e^{n+q} - \eta_q e^{n+q-1}\|, \end{aligned}$$

whence

$$J^{n+q} \leq \tilde{\lambda}^{n+q}\big(1 + \tfrac{1}{2}\eta_q\big)\|e^{n+q}\|^2 + \tfrac{1}{2}\tilde{\lambda}^{n+q}\eta_q\|e^{n+q-1}\|^2. \tag{41}$$

Summarizing our estimates, from (30), (33), (40) and (41), we get

$$\begin{aligned} &|E^{n+q}|_G^2 - |E^{n+q-1}|_G^2 + k\Big[\kappa^{n+q} - \lambda^{n+q} - \tfrac{1}{2}\eta_q\big[\nu^{n+q} + \lambda^{n+q}\big]\Big]\|e^{n+q}\|^2 \\ &- k\tfrac{1}{2}\eta_q\big[\nu^{n+q} + \lambda^{n+q}\big]\|e^{n+q-1}\|^2 \leq k\tilde{\lambda}^{n+q}\big(1 + \tfrac{1}{2}\eta_q\big)\|e^{n+q}\|^2 \\ &+ k\tfrac{1}{2}\tilde{\lambda}^{n+q}\eta_q\|e^{n+q-1}\|^2 + k\,\mathrm{Re}(d^n, e^{n+q} - \eta_q e^{n+q-1}). \end{aligned} \tag{42}$$

Now, with $\sigma(t) := \frac{1}{2}\big[\nu(t) + \lambda(t) + \tilde{\lambda}(t)\big]$, condition (27) reads

$$\kappa(t) - \tfrac{1}{2}\eta_q\nu(t) - (1 + \tfrac{1}{2}\eta_q)\big[\lambda(t) + \tilde{\lambda}(t)\big] \geq \rho + \eta_q\sigma(t),$$

and (42) can be equivalently written in the form

$$\begin{aligned} |E^{n+q}|_G^2 - |E^{n+q-1}|_G^2 + k\rho\|e^{n+q}\|^2 + k\eta_q\sigma^{n+q}\big[\|e^{n+q}\|^2 - \|e^{n+q-1}\|^2\big] \\ \leq k\,\mathrm{Re}(d^n, e^{n+q} - \eta_q e^{n+q-1}). \end{aligned} \tag{43}$$

Therefore, for σ Lipschitz continuous, we have

$$\begin{aligned} |E^{n+q}|_G^2 - |E^{n+q-1}|_G^2 + k\rho\|e^{n+q}\|^2 + k\eta_q\big[\sigma^{n+q}\|e^{n+q}\|^2 - \sigma^{n+q-1}\|e^{n+q-1}\|^2\big] \\ \leq ck^2\|e^{n+q-1}\|^2 + k\,\mathrm{Re}(d^n, e^{n+q} - \eta_q e^{n+q-1}). \end{aligned} \tag{44}$$

Thus, bounding

$$\mathrm{Re}(d^n, e^{n+q} - \eta_q e^{n+q-1}) \le \frac{\rho}{2(1+\eta_q)} \|e^{n+q}\|^2 + \frac{\eta_q \rho}{2(1+\eta_q)} \|e^{n+q-1}\|^2 + \frac{1+\eta_q}{\rho} \|d^n\|_\star^2$$

and summing the inequalities (44) from $n = 0$ to m yields the stability estimate

$$|E^{m+q}|_G^2 + \tfrac{1}{3}\rho k \sum_{n=0}^{m} \|e^{n+q}\|^2 \le |E^{q-1}|_G^2 + k\eta_q c \|e^{q-1}\|^2 + k \frac{1+\eta_q}{\rho} \sum_{n=0}^{m} \|d^n\|_\star^2. \quad (45)$$

We note the lower bound $|E^{m+q}|_G^2 \ge c_q |e^{m+q}|^2$ with $c_q > 0$ equal to the smallest eigenvalue of the symmetric positive definite matrix G. Together with the consistency error bound (20) and the fact that, under our accuracy assumptions on the starting values, $|E^{q-1}|_G^2 + k\eta_q c\|e^{q-1}\|^2$ is of order k^{2q}, this implies (28). □

Remark 2 Theorem 1 bounds the error of a BDF semi-discretisation in time of a quasilinear parabolic problem with a temporally smooth solution and also the error between a temporally smooth solution of a finite element semi-discretisation of the parabolic problem and the full discretisation. Alternatively, to bound the error of the full discretisation one can study the defect d^n in the fully discrete BDF method of a (Ritz or interpolation) projection of the exact solution of the parabolic problem to the finite element space and then use the stability estimate (45) to bound the error between the projected exact solution and the fully discrete numerical solution; see [9] for this procedure in a situation of a full discretisation by finite elements in space and BDF in time for a class of linear parabolic equations with time-dependent operators. Since the general procedure is conceptually clear but technically cumbersome, we will not work out the details of error bounds for a full discretisation in this paper.

5 Stability and error bounds for the implicit–explicit BDF method

In this section we establish optimal order error estimates for the implicit–explicit BDF methods (11).

Theorem 2 *Assume* (2)–(5) *and consider time discretisation by the implicit–explicit BDF method* (11) *of order* $q \le 5$. *Let* η_q *be as in Lemma* 2, $\eta_1 = \eta_2 = 0, \eta_3 = 0.0836, \eta_4 = 0.2878, \eta_5 = 0.8160$. *Under the stability condition*

$$\kappa(t) - \eta_q \nu(t) - (1+\eta_q)\big[\lambda(t) + (2^q - 1)\tilde{\lambda}(t)\big] \ge \rho \ge c_0 k \quad \forall t \in [0,T] \quad (46)$$

with a sufficiently large constant $c_0 > 0$, *the method is convergent of order* q: *for* $t^n \le T$,

$$c_q |U^n - u(t^n)|^2 + \tfrac{1}{3}\rho k \sum_{\ell=q}^{n} \|U^\ell - u(t^\ell)\|^2 \le C\rho^{-1} k^{2q}, \quad (47)$$

provided that the solution is sufficiently regular and the starting values are $O(k^q)$ *accurate in the* H*-norm* $|\cdot|$ *and* $O(k^{q-\frac{1}{2}})$ *in the* V*-norm* $\|\cdot\|$. *In* (47), $c_q > 0$ *depends only on* q *and the constant* C *is independent of* ρ, k *and* n *with* $nk \le T$.

We note that Remarks 1 and 2 also apply to the implicit–explicit BDF method.

Proof Let $e^n := u^n - U^n$ denote the discretisation error of the scheme (11) and $b^n := B(t^n,u^n) - B(t^n,U^n), n = 0,\dots,N$. Subtracting (11) from (18), we obtain

$$\sum_{i=0}^{q} \alpha_i e^{n+i} + k\big[A(t^{n+q},u^{n+q})u^{n+q} - A(t^{n+q},U^{n+q})U^{n+q}\big] = k\sum_{i=0}^{q-1} \gamma_i b^{n+i} + k\tilde{d}^n, \quad (48)$$

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

We now take in (48) the inner product with $e^{n+q} - \eta_q e^{n+q-1}$ and then real parts to obtain

$$\begin{aligned} \mathrm{Re}\Big(\sum_{i=0}^{q} \alpha_i e^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big) + kI^{n+q} &= kJ^{n+q} \\ &+ k\,\mathrm{Re}(\tilde{d}^n, e^{n+q} - \eta_q e^{n+q-1}) \end{aligned} \quad (49)$$

with I^{n+q} as in (31) and

$$J^{n+q} := \mathrm{Re}\Big(\sum_{i=0}^{q-1} \gamma_i b^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big). \quad (50)$$

We have already estimated the terms on the left-hand side of (49) in the previous proof. Since the last term on the right-hand side can be easily estimated from above, cf. the analogous estimate following (44), all that remains to be done is to estimate J^{n+q} from above in an appropriate way. In view of the Lipschitz condition (5), we have

$$\begin{aligned} J^{n+q} &= \mathrm{Re}\Big(\sum_{i=0}^{q-1} \gamma_i b^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big) \\ &\le \sum_{i=0}^{q-1} |\gamma_i|\, \|b^{n+i}\|_\star \big(\|e^{n+q}\| + \eta_q \|e^{n+q-1}\|\big) \\ &\le \sum_{i=0}^{q-1} |\gamma_i|\, \tilde{\lambda}^{n+i} \|e^{n+i}\| \big(\|e^{n+q}\| + \eta_q \|e^{n+q-1}\|\big) \\ &\le \sum_{i=0}^{q-1} |\gamma_i|\, \tilde{\lambda}^{n+i} \big(\|e^{n+i}\|^2 + \|e^{n+q}\|^2 + \eta_q \|e^{n+i}\|^2 + \eta_q \|e^{n+q-1}\|^2\big). \end{aligned}$$

Here we bound $\tilde{\lambda}^{n+i} \le \tilde{\lambda}^{n+q} + qk\max_{0\le t\le T} |\tilde{\lambda}'(t)| =: \tilde{\Lambda}^{n+q}$. Since

$$\sum_{i=0}^{q-1} |\gamma_i| = |\gamma(-1)| = 2^q - 1,$$

we thus obtain (compare with (41))

$$\begin{aligned} J^{n+q} &\le \tfrac{1}{2}\tilde{\Lambda}^{n+q}(1+\eta_q)\sum_{i=0}^{q-1} |\gamma_i|\, \|e^{n+i}\|^2 \\ &+ \tfrac{1}{2}\tilde{\Lambda}^{n+q}(2^q-1)\big(\|e^{n+q}\|^2 + \eta_q \|e^{n+q-1}\|^2\big). \end{aligned} \quad (51)$$

Proceeding further as in the proof of Theorem 1 yields the stated result. □

6 Stability and error bounds for the linearly implicit BDF method

In this section we establish optimal order error estimates for the linearly implicit BDF methods (13).

Theorem 3 *Assume* (2)–(5) *and consider time discretisation by the linearly implicit BDF method* (12) *of order* $q \le 5$*. Let* η_q *be as in Lemma* 2*,* $\eta_1 = \eta_2 = 0, \eta_3 = 0.0836, \eta_4 = 0.2878, \eta_5 = 0.8160$*. Under the stability condition*

$$\kappa(t) - \eta_q \nu(t) - (1+\eta_q)(2^q-1)\big[\lambda(t) + \tilde{\lambda}(t)\big] \ge \rho \ge c_0 k \quad \forall t \in [0,T] \qquad (52)$$

with a sufficiently large constant $c_0 > 0$*, the method is convergent of order* q: *for* $t^n \le T$*,*

$$c_q |U^n - u(t^n)|^2 + \tfrac{1}{3}\rho k \sum_{\ell=q}^{n} \|U^\ell - u(t^\ell)\|^2 \le C\rho^{-1}k^{2q}, \qquad (53)$$

provided that the solution is sufficiently regular and the starting values are $O(k^q)$ *accurate in the* H*-norm* $|\cdot|$ *and* $O(k^{q-\frac{1}{2}})$ *in the* V*-norm* $\|\cdot\|$*. In* (53)*,* $c_q > 0$ *depends only on* q *and the constant* C *is independent of* ρ, k *and* n *with* $nk \le T$*.*

We note that Remarks 1 and 2 also apply to the linearly implicit BDF method.

Proof Let $e^n := u^n - U^n$ denote the discretisation error of the scheme (13) and $b^n := B(t^n, u^n) - B(t^n, U^n), n = 0, \dots, N$. Subtracting (13) from (19), we obtain

$$\sum_{i=0}^{q} \alpha_i e^{n+i} + k\big[A(t^{n+q}, \hat{u}^{n+q})u^{n+q} - A(t^{n+q}, \widehat{U}^{n+q})U^{n+q}\big] = k\sum_{i=0}^{q-1}\gamma_i b^{n+i} + k\check{d}^n, \quad (54)$$

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

We now take in (54) the inner product with $e^{n+q} - \eta_q e^{n+q-1}$ and then real parts to obtain

$$\begin{aligned} \mathrm{Re}\Big(\sum_{i=0}^{q}\alpha_i e^{n+i}, e^{n+q} - \eta_q e^{n+q-1}\Big) + kI^{n+q} &= kJ^{n+q} \\ &+ k\,\mathrm{Re}(\check{d}^n, e^{n+q} - \eta_q e^{n+q-1}) \end{aligned} \qquad (55)$$

with J^{n+q} as in (50) and

$$I^{n+q} := \mathrm{Re}\big(A(t^{n+q}, \hat{u}^{n+q})u^{n+q} - A(t^{n+q}, \widehat{U}^{n+q})U^{n+q}, e^{n+q} - \eta_q e^{n+q-1}\big). \qquad (56)$$

We have already estimated the first term on the left-hand side of (55) in (33). Since the last term on the right-hand side can be easily estimated from above, cf. the analogous estimate following (44), essentially all that remains to be done is to estimate I^{n+q} from below and J^{n+q} from above in an appropriate way. Furthermore, we already estimated J^{n+q} in (51).

To estimate I^{n+q} from below, we proceed as in Section 4. First, we rewrite the expression

$$A(t^{n+q}, \hat{u}^{n+q})u^{n+q} - A(t^{n+q}, \widehat{U}^{n+q})U^{n+q}$$

as $A(t^{n+q},\widehat{U}^{n+q})e^{n+q}+\big[A(t^{n+q},\hat{u}^{n+q})-A(t^{n+q},\widehat{U}^{n+q})\big]u^{n+q}$; then, (56) reads

$$I^{n+q}=I_1^{n+q}+I_2^{n+q} \tag{57}$$

with

$$I_1^{n+q}:=\operatorname{Re}\big(A(t^{n+q},\widehat{U}^{n+q})e^{n+q},e^{n+q}-\eta_q e^{n+q-1}\big) \tag{58}$$

and

$$I_2^{n+q}:=\operatorname{Re}\big(\big[A(t^{n+q},\hat{u}^{n+q})-A(t^{n+q},\widehat{U}^{n+q})\big]u^{n+q},e^{n+q}-\eta_q e^{n+q-1}\big). \tag{59}$$

I_1^{n+q} is estimated as in (38):

$$I_1^{n+q}\geq\big[\kappa^{n+q}-\tfrac{1}{2}\eta_q\nu^{n+q}\big]\|e^{n+q}\|^2-\tfrac{1}{2}\eta_q\nu^{n+q}\|e^{n+q-1}\|^2. \tag{60}$$

Furthermore, I_2^{n+q} can be estimated from below using the Lipschitz condition (4); indeed, we have

$$\begin{aligned}I_2^{n+q}&\geq-\|[A(t^{n+q},\hat{u}^{n+q})-A(t^{n+q},\widehat{U}^{n+q})\big]u^{n+q}\|_\star\|e^{n+q}-\eta_q e^{n+q-1}\|\\&\geq-\lambda^{n+q}\|\hat{u}^{n+q}-\widehat{U}^{n+q}\|\,\|e^{n+q}-\eta_q e^{n+q-1}\|,\end{aligned}$$

i.e.,

$$I_2^{n+q}\geq-\lambda^{n+q}\|\hat{e}^{n+q}\|\,\|e^{n+q}-\eta_q e^{n+q-1}\|.$$

Using here the definition of $\hat{e}^{n+q}$ and proceeding as in the derivation of estimate (51), we arrive at the estimate

$$\begin{aligned}I_2^{n+q}\geq&-\tfrac{1}{2}\lambda^{n+q}(1+\eta_q)\sum_{i=0}^{q-1}|\gamma_i|\,\|e^{n+i}\|^2\\&-\tfrac{1}{2}\lambda^{n+q}(2^q-1)\big(\|e^{n+q}\|^2+\eta_q\|e^{n+q-1}\|^2\big).\end{aligned} \tag{61}$$

From (57), (60) and (61), we obtain the desired estimate of I^{n+q} from below, namely

$$\begin{aligned}I_2^{n+q}\geq&\,\kappa^{n+q}\|e^{n+q}\|^2-\tfrac{1}{2}\eta_q\big[\nu^{n+q}+(2^q-1)\lambda^{n+q}\big]\big(\|e^{n+q}\|^2+\|e^{n+q-1}\|^2\big)\\&-\tfrac{1}{2}\lambda^{n+q}(1+\eta_q)\sum_{i=0}^{q-1}|\gamma_i|\,\|e^{n+i}\|^2.\end{aligned} \tag{62}$$

Proceeding further as in the proof of Theorem 1 yields the stated result. □

7 Stability and error estimates in the Hermitian positive definite case

The above stability conditions impose a restriction for the methods of orders 3 to 5 (which have $\eta_q > 0$) even if $\lambda(t)$ and $\tilde{\lambda}(t)$ are arbitrarily small: $\eta_q \nu(t) < \kappa(t)$. We will now show that no such restriction appears when the operators $A(t,w)$ are Hermitian and positive definite. We require the following assumptions in addition to (2)–(5):

The sesquilinear forms on V defined by $A(t,w)$ are Hermitian for all $t \in [0,T]$ and $w \in V$. (63)

There is a subspace S of V such that we have the modified Lipschitz condition

$$\|\big(A(t,w) - A(\tilde{t},\tilde{w})\big)\upsilon\|_\star \le (\hat{\lambda}\|w-\tilde{w}\|_S + \hat{\sigma}|t-\tilde{t}|)\|\upsilon\| \tag{64}$$

for all $w,\tilde{w} \in S,\ \upsilon \in V$ and all $t,\tilde{t} \in [0,T]$. Moreover, there exists a (possibly small) $\varepsilon > 0$ such that we have an inverse estimate

$$\|\upsilon\|_S \le \frac{1}{\varepsilon}\|\upsilon\| \quad \forall \upsilon \in V. \tag{65}$$

Example 1 In the example of Section 2.2 we can take $S = V = H^1(\Omega)$ in the case of space dimension $d = 1$. In dimension $d = 2$, we can take $S = V = H^{1+s}(\Omega)$ and $H = H^s(\Omega)$ with $0 < s < 1/2$, see [10, Section 4]. In dimension $d = 3$, the choice $S = V$ is not possible. With $H = L^2(\Omega)$ and $V = H^1(\Omega)$, condition (64) is satisfied for $S = L^\infty(\Omega)$. In this case, for $d > 1$ conditions (64)–(65) are not satisfied in the spatially continuous problem, but they hold for finite element discretisations with ε depending on the spatial grid size h: on quasi-uniform meshes we have $\varepsilon \sim 1/|\log h|$ for $d = 2$ and $\varepsilon \sim h$ for $d = 3$. We recall that our abstract framework applies equally to the spatially discretised problem.

We assume that the exact solution values $u(t)$ lie in S and

$$\|u(t+k) - u(t)\|_S \le C_S k \quad \forall t \in [0, T-k]. \tag{66}$$

Theorem 4 *Assume* (2)–(5) *and* (63)–(66)*, and consider time discretisation by the BDF method* (10) *of order q with $3 \le q \le 5$ under the time step restriction $k^{q-3/2} \le C_0\varepsilon$. Let η_q be as in Lemma* 2, $\eta_3 = 0.0836, \eta_4 = 0.2878, \eta_5 = 0.8160$. *Under the stability condition*

$$\kappa(t) - \frac{1+\eta_q}{1-\eta_q}\big[\lambda(t) + \tilde{\lambda}(t)\big] \ge \rho \ge c_0 k \quad \forall t \in [0,T] \tag{67}$$

with a sufficiently large constant $c_0 > 0$, the method is convergent of order q: for $t^n \le T$,

$$|U^n - u(t^n)|^2 + c\rho k \sum_{\ell=q}^{n} \|U^\ell - u(t^\ell)\|^2 \le C\rho^{-1}k^{2q}, \tag{68}$$

with constants $c > 0$ and C independent of ρ and ε, provided that the solution is sufficiently regular and the starting values are $O(k^q)$ accurate in the H-norm $|\cdot|$ and $O(k^{q-\frac{1}{2}})$ in the V-norm $\|\cdot\|$.

Analogous results hold also for the implicit–explicit and linearly implicit BDF methods (11) and (12), under the stability conditions

$$\kappa(t) - \frac{1+\eta_q}{1-\eta_q}\big[\lambda(t) + (2^q-1)\tilde{\lambda}(t)\big] \geq \rho \geq c_0 k \qquad \forall t \in [0,T] \tag{69}$$

and

$$\kappa(t) - \frac{1+\eta_q}{1-\eta_q}(2^q-1)\big[\lambda(t) + \tilde{\lambda}(t)\big] \geq \rho \geq c_0 k \qquad \forall t \in [0,T], \tag{70}$$

respectively. Note that in contrast to conditions (27), (46) and (52), the corresponding stability conditions (67), (69) and (70), respectively, do not depend on the operator bound $\nu(t)$ of (3). Hence, in the Hermitian case we have stability and error bounds uniformly in the boundedness-coercivity ratio $\nu(t)/\kappa(t)$ for all BDF methods up to order 5.

Proof In the Hermitian case,

$$\|\upsilon\|_{t,w} = (A(t,w)\upsilon,\upsilon)^{1/2}, \qquad \upsilon \in V, \tag{71}$$

defines a time- and state-dependent norm that is uniformly equivalent to the norm $\|\cdot\|$ on V: in view of conditions (2)–(3),

$$\kappa(t)\|\upsilon\|^2 \leq \|\upsilon\|_{t,w}^2 \leq \nu(t)\|\upsilon\|^2 \qquad \forall \upsilon \in V \tag{72}$$

and for all $t \in [0,T]$ and all $w \in V$. We denote by $\|\cdot\|_{\star,t,w}$ the dual norm on V' defined by

$$\|\varphi\|_{\star,t,w} = \sup_{\|\upsilon\|_{t,w}=1} |(\varphi,\upsilon)|.$$

In the time- and state-dependent norm, the restricted Lipschitz condition (4) then becomes

$$\|\big(A(t,\upsilon) - A(\tilde{t},\tilde{\upsilon})\big)u(t)\|_{\star,t,w} \leq \lambda_1(t)\|\upsilon-\tilde{\upsilon}\|_{t,w} + \mu_1(t)|\upsilon-\tilde{\upsilon}| \tag{73}$$

for all $\upsilon, \tilde{\upsilon} \in V$ and all $t \in [0,T]$, with

$$\lambda_1(t) = \frac{\lambda(t)}{\kappa(t)}, \quad \mu_1(t) = \frac{\mu}{\sqrt{\kappa(t)}}.$$

Similarly, we have

$$\|B(t,\upsilon) - B(t,\tilde{\upsilon})\|_{\star,t,w} \leq \tilde{\lambda}_1(t)\|\upsilon-\tilde{\upsilon}\|_{t,w} + \tilde{\mu}_1(t)|\upsilon-\tilde{\upsilon}| \quad \forall \upsilon,\tilde{\upsilon} \in \mathscr{B}_{u(t)}, \tag{74}$$

with

$$\tilde{\lambda}_1(t) = \frac{\tilde{\lambda}(t)}{\kappa(t)}, \quad \tilde{\mu}_1(t) = \frac{\tilde{\mu}}{\sqrt{\kappa(t)}}.$$

We then aim to carry out the stability and error analysis as in the previous sections, but now working with the time- and state-dependent norms. See also [9] for a related stability analysis of BDF methods in time-dependent norms. We estimate the critical term, which previously introduced $\nu(t)$ into the error bounds, as

$$(A(t^{n+q},U^{n+q})e^{n+q},e^{n+q-1}) \le \|e^{n+q}\|_{t^{n+q},U^{n+q}} \cdot \|e^{n+q-1}\|_{t^{n+q},U^{n+q}},$$

and we need to relate the last factor back to $\|e^{n+q-1}\|_{t^{n+q-1},U^{n+q-1}}$. Since

$$\begin{aligned}\|e^{n+q-1}\|^2_{t^{n+q},U^{n+q}} &= \|e^{n+q-1}\|^2_{t^{n+q-1},U^{n+q-1}} \\ &\quad + ((A(t^{n+q},U^{n+q}) - A(t^{n+q-1},U^{n+q-1}))e^{n+q-1},e^{n+q-1})\end{aligned}$$

we obtain the desired estimate

$$\|e^{n+q-1}\|^2_{t^{n+q},U^{n+q}} \le (1+ck)\|e^{n+q-1}\|^2_{t^{n+q-1},U^{n+q-1}} + ck\|e^{n+q}\|^2_{t^{n+q},U^{n+q}} \tag{75}$$

as follows: we can estimate (for simplicity we take here $\hat{\sigma} = 0$)

$$\begin{aligned}&((A(t^{n+q},U^{n+q}) - A(t^{n+q-1},U^{n+q-1}))e^{n+q-1},e^{n+q-1}) \\ &= ((A(t^{n+q},u^{n+q}) - A(t^{n+q-1},u^{n+q-1}))e^{n+q-1},e^{n+q-1}) \\ &\quad + ((A(t^{n+q},U^{n+q}) - A(t^{n+q},u^{n+q}))e^{n+q-1},e^{n+q-1}) \\ &\quad - ((A(t^{n+q-1},U^{n+q-1}) - A(t^{n+q-1},u^{n+q-1}))e^{n+q-1},e^{n+q-1}) \\ &\le \hat{\lambda} C_S k\|e^{n+q-1}\|^2 + \frac{\hat{\lambda}}{\varepsilon}\|e^{n+q}\| \cdot \|e^{n+q-1}\|^2 + \frac{\hat{\lambda}}{\varepsilon}\|e^{n+q-1}\| \cdot \|e^{n+q-1}\|^2.\end{aligned}$$

This is bounded by $ck\|e^{n+q-1}\|^2 + ck\|e^{n+q}\|^2$ as long as $\|e^{n+q-1}\| \le Ck^{q-1/2}$ for some constant C, which is ensured recursively for $t^n \le T$. We then obtain (75). The proof now proceeds as in Section 4, but working with the time- and state-dependent norms, for which the corresponding coercivity and boundedness constants are trivially $\kappa_1(t) = \nu_1(t) = 1$. We therefore obtain stability under the condition that corresponds to (27) for $\kappa(t) = \nu(t) = 1$, viz.,

$$1 - \eta_q - (1+\eta_q)[\lambda_1(t) + \tilde{\lambda}_1(t)] \ge \rho_1 \qquad \forall t \in [0,T],$$

which is equivalent to (67). By the same arguments as in Section 4 we then obtain a stability estimate in the time- and state-dependent norms that is analogous to (45),

$$\begin{aligned}|E^{m+q}|^2_G + \tfrac{1}{2}\rho k \sum_{n=0}^{m} \|e^{n+q}\|^2_{t^{n+q},U^{n+q}} &\le |E^{q-1}|^2_G + ck\|e^{q-1}\|^2_{t^{q-1},U^{q-1}} \\ &\quad + \frac{ck}{\rho}\sum_{n=0}^{m} \|d^n\|^2_{\star,t^{n+q},U^{n+q}}.\end{aligned} \tag{76}$$

Using the uniform equivalence of the norms $\|\cdot\|_{t,w}$ and $\|\cdot\|$, the error estimate in the time- and state-dependent norms is finally transfered back to the standard V-norm $\|\cdot\|$. □

References

1. G. Akrivis, *Implicit–explicit multistep methods for nonlinear parabolic equations*, Math. Comp. **82**, 45–68 (2013).
2. G. Akrivis, *Stability of implicit–explicit backward difference formulas for nonlinear parabolic equations*, SIAM J. Numer. Anal. (to appear).
3. G. Akrivis, M. Crouzeix and Ch. Makridakis, *Implicit–explicit multistep methods for quasilinear parabolic equations*, Numer. Math. **82**, 521–541 (1999).
4. C. Baiocchi and M. Crouzeix, *On the equivalence of A-stability and G-stability*, Appl. Numer. Math. **5**, 19–22 (1989).
5. M. Crouzeix, *Une méthode multipas implicite–explicite pour l'approximation des équations d'évolution paraboliques*, Numer. Math. **35**, 257–276 (1980).
6. G. Dahlquist, *G-stability is equivalent to A-stability*, BIT **18**, 384–401 (1978).
7. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II*: *Stiff and Differential–Algebraic Problems*, 2^{nd} revised ed., Springer–Verlag, Berlin Heidelberg, Springer Series in Computational Mathematics v. 14 (2002).
8. C. Lubich, *On the convergence of multistep methods for nonlinear stiff differential equations*, Numer. Math. **58**, 839–853 (1991). Erratum: Numer. Math. **61**, 277–279 (1992).
9. C. Lubich, D. Mansour and C. Venkataraman, *Backward difference time discretisation of parabolic differential equations on evolving surfaces*, IMA J. Numer. Anal. **33**, 1365–1385 (2013).
10. C. Lubich and A. Ostermann, *Runge–Kutta approximations of quasi-linear parabolic equations*, Math. Comp. **64**, 601–627 (1995) .
11. O. Nevanlinna and F. Odeh, *Multiplier techniques for linear multistep methods*, Numer. Funct. Anal. Optim. **3**, 377–423 (1981).
12. G. Savaré, $A(\Theta)-$*stable approximations of abstract Cauchy problems*, Numer. Math. **65**, 319–335 (1993).
13. V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*. 2^{nd} ed., Springer–Verlag, Berlin (2006).
14. M. Zlámal, *Finite element methods for nonlinear parabolic equations*, RAIRO **11**, 93–107 (1977).