ASC Report No. 29/2009

On the Stability and Error Structure of BDF Schemes Applied to Linear Parabolic Evolution Equations

Winfried Auzinger, Felix Kramer

Institute for Analysis and Scientific Computing
Vienna University of Technology — TU Wien
www.asc.tuwien.ac.at ISBN 978-3-902627-02-5

Most recent ASC Reports

28/2009 *Ansgar Jüngel*
New Velocity Formulation of Euler Equations with Third-order Derivatives: Application to Viscous Korteweg-type and Quantum Navier-Stokes Models

27/2009 *Markus Aurada, Samuel Ferraz-Leite, Dirk Praetorius*
Estimator Reduction and Convergence of Adaptive FEM and BEM

26/2009 *Anton Arnold, Laurent Desvillettes, Céline Prévost*
Existence of Nontrivial Steady States for Populations Structured with Respect to Space and a Continuous Trait

25/2009 *Bertram Düring, Daniel Matthes, Josipa Pina Milišić*
A Gradient Flow Scheme for Nonlinear Fourth Order Equations

24/2009 *Winfried Auzinger*
Normal Forms for Companion Matrices and Contractivity in Inner Product Norms

23/2009 *Harald Woracek*
Existence of Zerofree Functions N-associated to a de Branges Pontryagin Space

22/2009 *Michael Kaltenbäck, Harald Woracek*
Pontryagin Spaces of Entire Functions VI

21/2009 *Michael Kaltenbäck, Harald Woracek*
Pontryagin Spaces of Entire Functions V

20/2009 *Christoph Erath, Stefan Funken, Petra Goldenits, Dirk Praetorius*
Simple Error Estimations for the Galerkin Bem for some Hypersingular Integral Equation in 2D

19/2009 *Daniel Matthes, Ansgar Jüngel, Giuseppe Toscani*
Convex Sobolev Inequalities Derived from Entropy Dissipation

Institute for Analysis and Scientific Computing
Vienna University of Technology
Wiedner Hauptstraße 8–10
1040 Wien, Austria

E-Mail: admin@asc.tuwien.ac.at
WWW: http://www.asc.tuwien.ac.at
FAX: +43-1-58801-10196

ISBN 978-3-902627-02-5

On the stability and error structure of BDF schemes applied to linear parabolic evolution equations

Winfried Auzinger
Institute for Analysis and Scientific Computing
Vienna University of Technology
Wiedner Hauptstrasse 8–10/E101
A-1040 Wien, Austria, EU
w.auzinger@tuwien.ac.at

Felix Kramer
Department of Mathematics
University of Innsbruck
Technikerstrasse 13
A-6020 Innsbruck, Austria, EU
Felix.Kramer@uibk.ac.at

Dedicated to Ernst Hairer on the occasion of his 60th birthday.

Abstract

We continue the work of various authors on the stability and the structure of the global error of linear multistep schemes applied to linear evolution equations. Here, BDF schemes are considered, and, as far as reasonable, explicit expressions for all occurring bounds are specified, exploiting prior work on the location of characteristic roots. The 2-step BDF scheme is considered in particular detail, and for problems of sectorial type, an asymptotic error expansion is derived based on damping properties of the scheme.

2000 Mathematics Subject Classification: primary 65L06, 65M12, 65M20; secondary 35K90.

Key words and phrases: BDF schemes; sectorial operator; stability; asymptotic error expansion.

1 Introduction

We consider linear multistep methods of backward differentiation type (BDF) applied to evolution equations

$$u'(t) \;=\; A\,u(t) + f(t), \qquad u(t_0) = u_0, \tag{1.1}$$

where A is a linear sectorial operator, densely defined in a Hilbert space $\mathcal{H}$ with inner product $\langle\cdot,\cdot\rangle$ and associated norm $\|\cdot\|$; see Section 2. For more specific assumptions see Section 4.

The k-step BDF schemes are denoted as 'BDFk'. Let $h > 0$ denote the stepsize which is assumed to be constant, and $t_\nu := \nu h$. The BDF2 scheme is A-stable, and up to $k = 6$ these schemes are well known to be $A(\alpha)$-stable. Several authors have studied the stability of linear multistep methods for various types of evolution equations on the basis of their stability domains, cf. for instance [5, 8, 9, 12, 15]; see also [7, Chapter 5] and [16, Section 32].

Since, in current practice, BDF methods play a prominent role, we considered it worthwile to investigate their stability and approximation properties in more detail. In the present paper we show how to obtain explicit, quantitative stability bounds for the practically relevant BDF schemes, exploiting quantitative information about the location of their characteristic roots. Moreover, we study the error structure of the BDF2 scheme in more detail. We derive an asymptotic error expansion, including precise damping estimates for the initially irregular behavior of the remainder term. These results can be seen as an extension of earlier results, in particular concerning the backward Euler scheme (see [2]), and the error estimates for multistep schemes from [9]. All results are specified in a way showing their dependence on characteristic problem and method parameters.

The paper is organized as follows: Details about the problem class and about the BDF schemes are specified in Section 2. In particular, we give the relevant estimates for the location of the characteristic roots. In Section 3 we provide explicit discrete resolvent estimates for problem class (1.1); the underlying analysis is related to the techniques used in [12]. In Section 4 we derive an asymptotic error expansion for the BDF2 scheme, including careful estimates for the behavior of the remainder. Section 5 gives a numerical illustration via the behavior of Richardson extrapolation. Some auxiliary results are collected in Appendix A.

k	a_k	
2	$\frac{1}{2}$	$=\ 0.5$
3	$\frac{1}{\sqrt{3}}$	$\approx\ 0.5774$
4	$\frac{1}{\sqrt{2}}$	$\approx\ 0.7071$
5	$\frac{\sqrt{5+\sqrt{5}}}{\sqrt{10}}$	$\approx\ 0.8507$

Table 1: Inner radii a_k of annuli $\mathcal{A}_k$, for $k = 2 \ldots 5$ (Proposition 2.1)

2 Fundamentals

2.1 Characteristic polynomials and discrete resolvent

The BDF k scheme applied to $u' = \lambda\, u$ takes the form

$$\sum_{\ell=0}^{k} \alpha_\ell\, \tilde{u}_{\nu-k+\ell} \,=\, \mu\, \tilde{u}_\nu \quad \nu \geq k, \quad \text{with} \quad \mu := h\lambda \tag{2.1}$$

The coefficients α_ℓ can be expressed via the associated characteristic polynomials

$$\rho(\zeta) \,=\, \sum_{j=1}^{k} \tfrac{1}{j}\, \zeta^{k-j}\, (\zeta-1)^j, \qquad \sigma(\zeta) \,=\, \zeta^k, \tag{2.2}$$

see [6]. Collecting terms in $\rho(\zeta)$ gives[1]

$$\rho(\zeta) \,=\, \sum_{\ell=0}^{k} \alpha_\ell\, \zeta^\ell, \quad \text{with} \quad \alpha_\ell \,=\, (-1)^{k-\ell} \sum_{j=1}^{k} \frac{1}{j} \binom{j}{k-\ell}. \tag{2.3}$$

The leading coefficient of

$$p(\zeta) \,:=\, \rho(\zeta) - \mu\, \sigma(\zeta) \tag{2.4}$$

is $(\alpha_k - \mu)$, with $\alpha_k = \sum_{j=1}^{k} \frac{1}{j} > 0$. By $\hat{p}(\zeta)$ we denote the monic version of $p(\zeta)$ (with leading coefficient 1),

$$\hat{p}(\zeta) \,=\, \hat{p}(\zeta;\mu) \,:=\, (\alpha_k - \mu)^{-1} p(\zeta). \tag{2.5}$$

The rational function of degree k,

$$r(\zeta) \,=\, r(\zeta;\mu) \,:=\, p^{-1}(\zeta) \tag{2.6}$$

is called the discrete resolvent of the scheme. We shall make use of established techniques based on the so-called discrete resolvent calculus (cf. Section 2.3 and [4, 7, 12]). This favorable technique enables us to analyze multistep equations in a similar fashion as one-step recursions, on the basis of identity (2.17).

2.2 Location of characteristic roots

In the sequel, $\mathcal{S}_k$ denotes the stability domain of the BDF k scheme.

For the scalar case and for $k = 2 \ldots 5$, the following result was proved in [3]: [2]

[1] As usual, we suppress the dependence of ρ, σ, α_ℓ, etc. on k in our notation.

[2] For $k = 6$, the corresponding value a_6 would amount to $a_6 = 1$, which provides no further information beyond $A(\alpha)$-stability. The case $k = 6$ was not further analyzed in [3] because it is of little practical relevance due to its small stability angle α.

Proposition 2.1 *For $k = 2 \dots 5$ and arbitrary $\mu \in \mathcal{S}_k$, any root of $p(\zeta) = p(\zeta;\mu)$ which is contained in the annulus*

$$\mathcal{A}_k := \{ \zeta \in \mathbb{C} : \; a_k < |\zeta| \le 1 \} \tag{2.7}$$

is simple and solitary within $\mathcal{A}_k$, i.e., $\mathcal{A}_k$ contains no other root. Here the inner radii a_k are given by (cf. Table 1)

$$a_k = \frac{1}{|1-\omega_k|}, \quad \textit{with} \quad \omega_k = \mathrm{e}^{2\pi \mathrm{i}/k}. \tag{2.8}$$

Proposition 2.1 implies that for arbitrary $\mu = h\lambda \in \mathcal{S}_k$ at least $k-1$ of the k roots ζ_j satisfy[3] $|\zeta_j| \le a_k < 1$. Only a single root,

$$\text{the 'principal' root } \; \zeta_1 \; \text{ of maximal modulus} \tag{2.9}$$

may be contained in $\mathcal{A}_k$, which is the case for moderately sized μ. In [3], this quantitative information about the location of the characteristic roots was used to derive quantitative stability bounds for the scalar model problem.[4] With a_k from (2.8), we denote

$$b_k := \frac{1}{\alpha_k (1-a_k)^{k-1}}, \qquad c_k := \sum_{\ell=0}^{k-1} |\alpha_\ell|. \tag{2.10}$$

2.3 Application to $u' = Au + f$. Discrete variation of constants

Application of the BDF k scheme to (1.1) takes the form

$$\sum_{\ell=0}^{k} \alpha_\ell \, \tilde{u}_{\nu-k+\ell} = M\tilde{u}_\nu + hf_\nu, \quad \nu \ge k, \quad \text{with} \quad M := hA, \tag{2.11}$$

where $f_\nu := f(t_\nu)$. The required initial values $\tilde{u}_0 = s_0, \dots, \tilde{u}_{k-1} = s_{k-1}$ are usually computed by a sufficiently accurate strongly stable one-step scheme. We assume that A is densely defined and sectorial, i.e.,

$$\langle Av, v \rangle \in \mathcal{T}_\theta \quad \text{for all} \quad v \in \mathcal{H}, \tag{2.12}$$

where $\mathcal{T}_\theta$ is a sector

$$\mathcal{T}_\theta = \{ z \in \mathbb{C} : \; |\arg(-z)| \le \theta \} \in \mathbb{C}_-, \qquad 0 \le \theta \le \tfrac{\pi}{2}. \tag{2.13}$$

A sectorial operator satisfies a resolvent condition of the form

$$\|(zI-A)^{-1}\| \le \frac{1}{\operatorname{dist}(z, \mathcal{T}_\theta)} \quad \text{for all} \quad z \in \mathbb{C} \backslash \mathcal{T}_\theta, \tag{2.14}$$

cf. e.g. [1, 4, 16]. We shall require that $\mathcal{T}_\theta$ is contained in the stability domain $\mathcal{S}_k$ of the scheme considered.

For the general problem class (1.1), the question now is in what way Proposition 2.1 can be exploited to obtain quantitative bounds. The crucial quantity is the discrete resolvent (cf. (2.4)–(2.6))

$$R(\zeta) = R(\zeta; M) := P^{-1}(\zeta), \tag{2.15}$$

where

$$P(\zeta) = P(\zeta; M) := \rho(\zeta) I - \sigma(\zeta) M. \tag{2.16}$$

In terms of the resolvent, the solution of (2.11) can be expressed via a discrete variation of constants formula (similarly as in [4, 7, 12]),

$$\tilde{u}_\nu = - \sum_{j=k}^{\min\{2k-1,\nu\}} R_{\nu+k-j}(M) \Big(\sum_{\ell=0}^{2k-1-j} \alpha_\ell \, s_{j-k+\ell} \Big) + h \sum_{j=k}^{\nu} R_{\nu+k-j}(M) f_j, \quad \nu \ge k\,. \tag{2.17}$$

---

[3] The proof of Proposition 2.1 is technically involved. The easy part is to show that multiple roots can only be located on the circle $\{\zeta \in \mathbb{C} : \; |\zeta| = a_k\}$, see [3].

[4] For the case $k = 2$ we shall use more precise estimates for the characteristic roots, which are specified in Appendix A.

Here, $(R_\nu) = (R_\nu(M))$ is the discrete fundamental solution satisfying the homogeneous difference equation: With $R_0 := -\alpha_0^{-1} I$ and $R_1 := R_2 := \ldots := R_{k-1} := 0$, we have

$$\sum_{\ell=0}^{k} \alpha_\ell \, R_{\nu-k+\ell} \;=\; M R_\nu, \quad \nu \geq k\,, \tag{2.18}$$

in particular, $R_k = (\alpha_k I - M)^{-1}$. On the other hand, the R_ν are the coefficients in the series expansion of $R(\zeta)$ for $|\zeta| > 1$, i.e.

$$R(\zeta) \;=\; R(\zeta; M) \;=\; \sum_{\nu=k}^{\infty} \zeta^{-\nu} R_\nu, \tag{2.19}$$

obeying the Cauchy representation

$$R_\nu \;=\; R_\nu(M) \;=\; \frac{1}{2\pi\mathrm{i}} \oint_\Gamma \zeta^{\nu-1} R(\zeta)\,\mathrm{d}\zeta, \quad \nu = k, k{+}1, \ldots\,. \tag{2.20}$$

Here, Γ can be chosen as a contour containing all characteristic roots in its interior, e.g.

$$\Gamma \;=\; \Gamma_{1+\gamma} \;:=\; \{\, \zeta \in \mathbb{C}:\; |\zeta| = 1{+}\gamma \,\}, \quad \gamma > 0. \tag{2.21}$$

The monic version of $P(\zeta)$ is (cf. (2.5))

$$\hat{P}(\zeta) \;=\; \hat{P}(\zeta; M) \;:=\; (\alpha_k I - M)^{-1} P(\zeta), \tag{2.22}$$

and we shall also use the denotation

$$\hat{R}(\zeta) \;=\; \hat{R}(\zeta; M) \;:=\; (\alpha_k I - M) R(\zeta), \qquad \hat{R}_\nu = \hat{R}_\nu(M) \;:=\; (\alpha_k I - M) R_\nu(M). \tag{2.23}$$

3 Quantitative estimates for discrete resolvent and fundamental solution

The approach of analyzing stability via discrete resolvent estimates is essentially due to [12]. We do not intend to repeat these arguments in detail; our purpose here is simply to show how the information about location of characteristic roots enters the stability bounds. In particular, we provide such bounds for the discrete resolvent $R(\zeta)$ and the fundamental solution R_ν.

The BDF 2 scheme is A-stable: Its stability angle is $\alpha = \frac{\pi}{2}$, i.e., the nonpositive complex half plane $\mathbb{C}_-$ is contained in the stability domain $\mathcal{S}_2$. For $k = 3 \ldots 5$ the BDF k scheme is $A(\alpha)$-stable, i.e., for some stability angle $\alpha < \frac{\pi}{2}$, the sector $\mathcal{T}_\alpha$ is contained in the stability domain $\mathcal{S}_k$ (cf. e.g. [7] for the values α). Concerning A the natural assumption now is that it satisfies a resolvent condition (2.14) with respect to $\mathcal{T}_\theta$ with $\theta \leq \alpha$, and then the same is true for $M = hA$ ($h > 0$ arbitrary), i.e.,

$$\|(zI - M)^{-1}\| \;\leq\; \frac{1}{\operatorname{dist}(z, \mathcal{T}_\theta)} \quad \text{for all} \;\; z \in \mathbb{C}\backslash\mathcal{T}_\theta. \tag{3.1}$$

3.1 Discrete resolvent estimates

The following result is an extended, quantitative version of the estimate given in the proof of Theorem 3.8 in [12]. The b_k and c_k are method parameters defined by (2.10). For the definition of $P(\zeta), R(\zeta)$, etc., cf. Section 2.3.

Proposition 3.1 *Let $k \in \{2 \ldots 5\}$.*
– *If A satisfies* (2.14) *with $\theta \leq \alpha$ (α = stability angle), then*

$$\|R(\zeta)\| \;\leq\; \frac{b_k}{|\zeta|^{k-1}\,(|\zeta| - 1)} \quad \textit{for all} \;\; |\zeta| > 1\,. \tag{3.2}$$

– *A similar estimate is valid for $\hat{R}(\zeta) = (a_k I - M) R(\zeta)$:*

$$\|\hat{R}(\zeta)\| \;\leq\; \frac{b_k\, c_k}{|\zeta|^{k-1}\,(|\zeta| - 1)} \quad \textit{for all} \;\; |\zeta| > 1. \tag{3.3}$$

Proof:

– The scalar polynomial $p(\zeta) = p(\zeta;\mu)$ is given by

$$p(\zeta) \;=\; (a_k - \mu)\zeta^k + \sum_{\ell=0}^{k-1} \alpha_\ell\, \zeta^\ell, \tag{3.4}$$

cf. (2.3). Proposition 2.1 implies that, for arbitrary $\mu \in \mathcal{T}_\theta \subseteq \mathcal{S}_k$, $p(\zeta)$ factors into

$$p(\zeta) \;=\; (a_k - \mu) \cdot \prod_{j=1}^{k} (\zeta - \zeta_j), \tag{3.5}$$

or equivalently (due to $\sigma(\zeta) = \zeta^k$)

$$\frac{\rho}{\sigma}(\zeta) - \mu \;=\; (a_k - \mu) \cdot \prod_{j=1}^{k} \big(1 - \frac{\zeta_j}{\zeta}\big), \tag{3.6}$$

where the roots $\zeta_j = \zeta_j(\mu)$ satisfy

$$|\zeta_1| \le 1, \qquad |\zeta_j| \le a_k, \quad j = 2 \ldots k, \qquad (\zeta_1 \;=\; \text{principal root})\,. \tag{3.7}$$

Now, let $\mu \in \mathcal{T}_\theta$ and $\zeta \in \mathbb{C}$ with $|\zeta| > 1$ be given. (3.7) implies

$$\Big|\,\frac{\zeta_1}{\zeta}\,\Big| \le \frac{1}{|\zeta|} \,<\, 1, \qquad \Big|\,\frac{\zeta_j}{\zeta}\,\Big| \le \frac{a_k}{|\zeta|} \,<\, a_k, \quad j = 2 \ldots k, \tag{3.8}$$

hence

$$\Big|1 - \frac{\zeta_1}{\zeta}\,\Big| \ge 1 - \Big|\,\frac{\zeta_1}{\zeta}\,\Big| \ge \frac{|\zeta| - 1}{|\zeta|} \,>\, 0, \qquad \Big|1 - \frac{\zeta_j}{\zeta}\,\Big| \ge 1 - \Big|\,\frac{\zeta_j}{\zeta}\,\Big| \,>\, 1 - a_k, \quad j = 2 \ldots k. \tag{3.9}$$

For (3.6) this gives the estimate

$$\Big|\,\frac{\rho}{\sigma}(\zeta) - \mu\,\Big| \;>\; |a_k - \mu|\,(1 - a_k)^{k-1}\, \frac{|\zeta| - 1}{|\zeta|}\,. \tag{3.10}$$

Since (3.10) is valid for all $\mu \in \mathcal{T}_\theta$, we obtain

$$\operatorname{dist}\big(\,\frac{\rho}{\sigma}(\zeta), \mathcal{T}_\theta\,\big) \;>\; \alpha_k (1 - a_k)^{k-1}\, \frac{|\zeta| - 1}{|\zeta|} \;>\; 0 \quad \text{for all} \;\; |\zeta| > 1. \tag{3.11}$$

Together with (3.1), with $z = \frac{\rho}{\sigma}(\zeta)$, this implies

$$\|\,(\sigma P^{-1})(\zeta)\,\| \;=\; \big\|\big(\,\frac{\rho}{\sigma}(\zeta) I - M\,\big)^{-1}\big\| \;\le\; \frac{1}{\operatorname{dist}(\frac{\rho}{\sigma}(\zeta), \mathcal{T}_\theta)} \;\le\; \frac{1}{\alpha_k (1 - a_k)^{k-1}}\, \frac{|\zeta|}{|\zeta| - 1} \tag{3.12}$$

for all $|\zeta| > 1$, which is equivalent to (3.2) due to $\sigma(\zeta) = \zeta^k$.

– In order to prove (3.3) we note that

$$P(\zeta) \;=\; \zeta^k (a_k I - M) + \sum_{\ell=0}^{k-1} \alpha_\ell\, \zeta^\ell I \tag{3.13}$$

implies

$$\hat{R}(\zeta) \;=\; \hat{P}^{-1}(\zeta) \;=\; \zeta^{-k} \big(I - \sum_{\ell=0}^{k-1} \alpha_\ell\, \zeta^\ell\, P^{-1}(\zeta)\big). \tag{3.14}$$

Together with (3.2) this leads to the estimate

$$
\begin{aligned}
\|\hat{R}(\zeta)\| \;&\le\; |\zeta^{-k}| \Big(1+\sum_{\ell=0}^{k-1} |\alpha_\ell|\,|\zeta|^\ell \,\frac{b_k}{|\zeta|^{k-1}(|\zeta|-1)}\Big) &(3.15)\\
&=\; \frac{1}{|\zeta|^k\,(|\zeta|-1)} \Big(|\zeta|-1+b_k\sum_{\ell=0}^{k-1} |\alpha_\ell|\,|\zeta|^{\ell+1-k}\Big)\\
&=\; \frac{1}{|\zeta|^{k-1}\,(|\zeta|-1)} \Big(1-|\zeta|^{-1}+b_k\sum_{\ell=0}^{k-1} |\alpha_\ell|\,|\zeta|^{\ell-k}\Big)\\
&=\; \frac{1}{|\zeta|^{k-1}\,(|\zeta|-1)} \Big(1+(b_k\,|\alpha_{k-1}|-1)\,|\zeta|^{-1}+b_k\sum_{\ell=0}^{k-2} |\alpha_\ell|\,|\zeta|^{\ell-k}\Big)\\
&=\; \frac{1}{|\zeta|^{k-1}\,(|\zeta|-1)} \Big(1+(b_k\,|\alpha_{k-1}|-1)+b_k\sum_{\ell=0}^{k-2} |\alpha_\ell|\Big)
\end{aligned}
$$

for all $|\zeta| > 1$, which is equivalent to (3.3). □

3.2 Estimates for the discrete fundamental solution

We shall make use of a well-known Lemma due to Spijker (see [14]), which in its original version goes back to Leveque and Trefethen [10]; see also [16, Section 18].

'Spijker Lemma' from [9]: Let $q(\zeta)$ be a rational function of degree N with no poles on a circle $\Gamma \subset \mathbb{C}$. Then,

$$\frac{1}{2\pi}\oint_\Gamma |q'(\zeta)|\,|\mathrm{d}\zeta| \;\le\; N \sup_{\zeta\in\Gamma} |q(\zeta)|\,. \tag{3.16}$$

The following result is a quantitative extension of Theorem 3.8 in [12].

Proposition 3.2 *Let $k \in \{2\ldots 5\}$.*
– *If A satisfies* (2.14) *with $\theta \le \alpha$, then the following estimate is valid for $\nu \ge k$:*

$$\|R_\nu(M)\| \;\le\; \mathrm{e}\, b_k\,\nu\,, \qquad \|\hat{R}_\nu(M)\| \;\le\; \mathrm{e}\, b_k\, c_k\,\nu\,. \tag{3.17}$$

– *If the underlying space is finite dimensional, $\mathcal{H} = \mathbb{C}^n$, then*

$$\|R_\nu(M)\| \;\le\; \mathrm{e}\, b_k\, k\, n\,, \qquad \|\hat{R}_\nu(M)\| \;\le\; \mathrm{e}\, b_k\, c_k\, k\, n\,. \tag{3.18}$$

For the special case $k = 2$, a sharper, uniform estimate for the $R_\nu(M)$ and $\hat{R}_\nu(M)$ is easy to prove, see Appendix A.2.

Proof of Proposition 3.2 :

– From (2.20) and Proposition 3.1, (3.2) we obtain (integration along contour $\Gamma_{1+\gamma}$, see (2.21))

$$\|R_\nu(M)\| \;\le\; \frac{1}{2\pi}\oint_{\Gamma_{1+\gamma}} |\zeta|^{\nu-1}\,\|R(\zeta)\|\,|\mathrm{d}\zeta| \;\le\; \frac{b_k}{2\pi}\oint_{\Gamma_{1+\gamma}} \frac{|\zeta|^{\nu-k}}{|\zeta|-1}\,|\mathrm{d}\zeta| \;=\; (1{+}\gamma)\,b_k\,\frac{(1{+}\gamma)^{\nu-k}}{\gamma}\,. \tag{3.19}$$

For $\nu = k$ we consider $\gamma \to \infty$; for $\nu > k$ we choose $\gamma = 1/(\nu - k)$. In this way we obtain

$$\frac{(1{+}\gamma)^{\nu-k+1}}{\gamma} \;\le\; (\nu{-}k{+}1)\mathrm{e} \tag{3.20}$$

for all $\nu \ge k$, which proves the first bound in (3.17).

– For $v, w \in \mathbb{C}^n$ with $\|v\| = \|w\| = 1$,

$$q(\zeta) \;:=\; v^* R(\zeta)\, w \tag{3.21}$$

is a rational function of degree $N = kn$. Due to Proposition 3.1, (3.2) we have

$$|q(\zeta)| \;\le\; \frac{b_k}{|\zeta|^{k-1}\,(|\zeta|-1)} \quad \text{for all} \quad |\zeta| > 1. \tag{3.22}$$

With (2.20) and integrating by parts we obtain

$$v^* R_\nu(M) w \;=\; \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma_{1+\gamma}} \zeta^{\nu-1}\, q(\zeta)\, \mathrm{d}\zeta \;=\; \frac{-1}{2\pi \mathrm{i}\, \nu} \oint_{\Gamma_{1+\gamma}} \zeta^{\nu}\, q'(\zeta)\, \mathrm{d}\zeta. \tag{3.23}$$

With the same choice for γ as in the first part of the proof, (3.16) yields

$$|v^* R_\nu(M) w| \;\le\; \frac{(1{+}\gamma)^{\nu-k}}{\nu}\, k\, n \sup_{\zeta\in\Gamma} |q(\zeta)| \;=\; \frac{k\, n}{\nu}\, b_k\, \frac{(1{+}\gamma)^{\nu-k+1}}{\gamma} \;\le\; \frac{\nu-k+1}{\nu}\, \mathrm{e}\, b_k k\, n \tag{3.24}$$

for all $\nu \ge k$, which proves the first bound in (3.18).

The corresponding bounds for the $\hat{R}_\nu(M)$ in (3.17) and (3.18) are obtained in a completely analogous way by replacing $R(\zeta)$ by $\hat{R}(\zeta)$ and using (3.3) instead of (3.2). ☐

For the scalar case, where

$$r_\nu(\mu) \;=\; \frac{1}{2\pi \mathrm{i}} \oint_{\Gamma} \zeta^{\nu-1} r(\zeta)\, \mathrm{d}\zeta \tag{3.25}$$

are the coefficients in the series representation (2.19) for the scalar resolvent $r(\zeta)$, the (qualitative) estimates from [9] and [11] concerning damping properties of the r_ν can also be rewritten in a more quantitative sense. We do not go into detail here. These damping properties are, e.g., essential for estimating the behavior of the remainder ε_ν in the expansion (4.8) below; however, we shall use a more explicit representation for this purpose.

4 Asymptotic error expansion for the BDF2 scheme

Like the analysis given in [1] for the backward Euler scheme, the following results are valid for the sectorial case, with A invertible. To be more specific, additionally to (2.12) we assume that A is strictly coercive,

$$\mathrm{Re}\,\langle Av, v\rangle \le -\kappa \langle v, v\rangle \quad \text{for all} \;\; v \in \mathcal{H}, \quad \text{with} \;\; \kappa > 0, \tag{4.1}$$

such that (1.1) represents a non-degenerate parabolic problem, with

$$\|A^{-1}\| \;\le\; \frac{1}{\kappa}, \quad \text{and} \quad \|\mathrm{e}^{tA}\| \;\le\; e^{-\kappa t} \quad \text{for all} \;\; t > 0. \tag{4.2}$$

We proceed along similar lines as in [1], where one-step methods were considered. In the following, the analysis of the error expansion is done step by step, and the outcome will be summed up in Proposition 4.3 at the end of this section. The relevant stability and damping estimates for the BDF2 scheme which are used in our analysis are listed in Appendix A.

Our analysis applies under appropriate smoothness assumptions. In particular, the right-hand side $f(t)$ of (1.1) is assumed to have a sufficient degree of differentiability. From the theory of analytic semigroups it is well-known that for sufficiently smooth right-hand side $f(t)$ the solution $u(t)$ also enjoys a certain degree of differentiability for $t > 0$; cf., e.g. [13]. More specifically, we assume that $f \in C^5([0,T],\mathcal{H})$, and we assume that f is consistent with the initial data u_0 in such a way that $u \in C^6([0,T],\mathcal{H})$. Essentially, this means that that transient effects are masked out, and we study the error structure of BDF2 under somewhat idealized assumptions. Even for this best case scenario it turns out that the remainder of the error expansion shows a transient behavior (similar as in the backward Euler case), see Proposition 4.3. On the other hand, our results are formulated in such a way such that bounds on the $\|f^{(j)}\|$ are not explicitly used but only weaker bounds on $\|A^{-1} f^{(j)}\|$. (Typical examples where $\|f\|$ is large in contrast to $\|A^{-1} f\|$ stem from PDE discretizations where boundary conditions are a priori eliminated.)

Consider the solution of an initial value problem of the type (1.1), $v' = Av + g(t)$, $v(0) = v_0$ with sufficiently smooth $g(t)$. With the particular solution operator S,

$$(Sg)(t) := \int_0^t \mathrm{e}^{(t-s)A} g(s)\,\mathrm{d}s, \tag{4.3}$$

we have

$$v^{(j)}(t) = \mathrm{e}^{tA}\, v^{(j)}(0) + (Sg^{(j)})(t), \quad j = 0, 1, \ldots, \tag{4.4}$$

with initial values

$$v^{(j)}(0) = A^j v_0 + \sum_{\ell=0}^{j} A^{j-\ell} g^{(\ell)}(0). \tag{4.5}$$

Due to assumption (4.1), $v(t)$ can be estimated by

$$\|v(t)\| \le \|v(0)\| + \mathcal{C} \sup_{0\le s\le t} \|g(s)\| \quad \text{for all } \ t > 0. \tag{4.6}$$

As before we denote $hA =: M$. Consider the BDF scheme (2.11) for $k = 2$,

$$\alpha_0 \tilde{u}_{\nu-2} + \alpha_1 \tilde{u}_{\nu-1} + (\alpha_2 I - M)\tilde{u}_\nu = h f_\nu, \quad \nu \ge 2\,, \tag{4.7}$$

with[5] $\alpha_0 = \frac{1}{2}$, $\alpha_1 = -2$, $\alpha_2 = \frac{3}{2}$. For our analysis we make the assumption that the integration starts on the smooth solution $u(t)$ of (1.1), $\tilde{u}_0 = y(0)$, $\tilde{u}_1 = u(h)$, i.e., effects of an initial perturbation are also masked out. (As in [9], the analysis below can be extended including the effects of an initial perturbation, typically involving a backward Euler step at the start, but we refrain from working out the details.)

Here and in the sequel, $\mathcal{C}$ denotes a generic, or not explicitly evaluated, moderate-sized constant which is not influenced by 'critical', large quantities of the type $A^q v$, but only by natural, by derivatives of $f(t)$ or the solution $u(t)$, which are assumed to be moderate-sized. The symbol $\mathcal{O}(h^q)$ is used in the same spirit, i.e., $\mathcal{O}(h^q)$ means 'bounded by $\mathcal{C}h^q$'. Furthermore we use the denotation $\mathcal{C}(\theta)$ in a generic sense, with $\mathcal{C}(\theta)$ bounded for fixed θ but unbounded as $\theta \to \frac{\pi}{2}$, with a pole-like behavior. Note that all bounds derived in the following are valid on a fixed integration interval $t \in [0, T]$. The long-term behavior for $t \to \infty$ is not discussed.

4.1 The formal expansion

For the global error of the BDF2 scheme we make the ansatz

$$\tilde{u}_\nu - u(t_\nu) = h^2 e_2(t_\nu) + h^3 e_3(t_\nu) + \varepsilon_\nu\,. \tag{4.8}$$

As usual, the first step is a local error expansion: Inserting (4.8) into (4.7) gives

$$\begin{aligned} 0 = {} & \underbrace{\alpha_0\, u(t_{\nu-2}) + \alpha_1\, u(t_{\nu-1}) + (\alpha_2 I - M)u(t_\nu) - h f_\nu}_{\text{local truncation error}} + \\ & + h^2 \big(\alpha_0\, e_2(t_{\nu-2}) + \alpha_1\, e_2(t_{\nu-1}) + (\alpha_2 I - M)e_2(t_\nu)\big) + \\ & + h^3 \big(\alpha_0\, e_3(t_{\nu-2}) + \alpha_1\, e_3(t_{\nu-1}) + (\alpha_2 I - M)e_3(t_\nu)\big) + \\ & + \alpha_0\, \varepsilon_{\nu-2} + \alpha_1\, \varepsilon_{\nu-1} + (\alpha_2 I - M)\varepsilon_\nu. \end{aligned} \tag{4.9}$$

For the functions $y = u$, e_2 and e_3, we express the truncation error of the difference quotients in terms of Peano kernels (cf. e.g. [6, Section III.2]). With

$$(\ell - \xi)_+^q := \begin{cases} (\ell - \xi)^q, & \xi < \ell, \\ 0, & \xi \ge \ell, \end{cases} \tag{4.10}$$

and the Peano kernels

$$K_q(\xi) := \frac{1}{q!}\big(\alpha_0 (0-\xi)_+^q + \alpha_1 (1-\xi)_+^q + \alpha_2 (2-\xi)_+^q\big) - \frac{1}{(q-1)!}\,(2-\xi)_+^{q-1} \tag{4.11}$$

[5] For clearness of notation, we do not insert these particular values for the α_ℓ at the moment.

we have for $q = 1, 2$:

$$\alpha_0\, y(t-2h) + \alpha_1\, y(t-h) + \alpha_2\, y(t) - hy'(t) \;=\; h^{q+1}\int_0^2 K_q(\xi)\, y^{(q+1)}(t+(\xi-2)h)\,\mathrm{d}\xi. \tag{4.12}$$

For $y = e_2$ and $y = e_3$ we use $q = 2, 1$ at $t = t_\nu$,

$$\alpha_0\, e_2(t_{\nu-2}) + \alpha_1\, e_2(t_{\nu-1}) + \alpha_2\, e_2(t_\nu) - h\, e_2'(t_\nu) \;=\; h^3 \underbrace{\int_0^2 K_2(\xi)\, e_2'''(t_{\nu-2}+\xi h)\,\mathrm{d}\xi}_{=:\, I_{\nu,2}}, \tag{4.13}$$

$$\alpha_0\, e_3(t_{\nu-2}) + \alpha_1\, e_3(t_{\nu-1}) + \alpha_2\, e_3(t_\nu) - h\, e_3'(t_\nu) \;=\; h^2 \underbrace{\int_0^2 K_1(\xi)\, e_3''(t_{\nu-2}+\xi h)\,\mathrm{d}\xi}_{=:\, I_{\nu,3}}. \tag{4.14}$$

For $y = u$, we expand representation (4.12) with $q = 2$ at $t = t_\nu$ further, via double integration by parts. Together with $\alpha_0 = \frac{1}{2}$, $\alpha_1 = -2$, $\alpha_2 = \frac{3}{2}$ this results in

$$\begin{aligned} &\alpha_0\, u(t_{\nu-2}) + \alpha_1\, u(t_{\nu-1}) + \alpha_2\, u(t_\nu) - hu'(t_\nu) \\ &= \; -\tfrac{1}{3}\, h^3 u'''(t_\nu) + \tfrac{1}{4}\, h^4 u^{IV}(t_\nu) + h^5 \underbrace{\int_0^2 (K_4(\xi) - \tfrac{\xi}{3} + \tfrac{5}{12})\, u^V(t_{\nu-2}+\xi h)\,\mathrm{d}\xi}_{=:\, I_{\nu,0}}. \end{aligned} \tag{4.15}$$

Now, using (1.1) and collecting terms in powers of h, (4.9) evaluates to

$$\begin{aligned} 0 \;=\;\; & h^3\big(e_2'(t_\nu) - A\, e_2(t_\nu) - \tfrac{1}{3}u'''(t_\nu)\big) \;+ \\ & + h^4\big(e_3'(t_\nu) - A\, e_3(t_\nu) + \tfrac{1}{4}u^{IV}(t_\nu)\big) \;+ \\ & + \alpha_0\, \varepsilon_{\nu-2} + \alpha_1\, \varepsilon_{\nu-1} + (\alpha_2 I - M)\varepsilon_\nu + h^5 I_{\nu,0} + h^5 I_{\nu,2} + h^5 I_{\nu,3}. \end{aligned} \tag{4.16}$$

Equating coefficients of h^3 and h^4 in (4.16) gives rise to the variational equations

$$e_2' \;=\; A\, e_2 + \tfrac{1}{3}\, u''', \tag{4.17}$$

$$e_3' \;=\; A\, e_3 - \tfrac{1}{4}\, u^{IV}. \tag{4.18}$$

For the discrete remainder ε_ν we obtain the difference equation

$$\alpha_0\, \varepsilon_{\nu-2} + \alpha_1\, \varepsilon_{\nu-1} + (\alpha_2 I - M)\varepsilon_\nu \;=\; h\, \tau_\nu, \tag{4.19}$$

where

$$\tau_\nu \;:=\; h^4 I_{\nu,0} + h^4 I_{\nu,2} + h^4 I_{\nu,3} \tag{4.20}$$

depends on derivatives of $u(t)$, $e_2(t)$ and $e_3(t)$, respectively.

4.2 The principal error functions $e_2(t)$ and $e_3(t)$

With $e_2(0) = 0$ and $e_3(0) = 0$, the solutions of (4.17) and (4.18) are given by (see (4.4))

$$e_2(t) \;=\; \tfrac{1}{3}\,(Su''')(t), \tag{4.21}$$

$$e_3(t) \;=\; -\tfrac{1}{4}\,(Su^{IV})(t). \tag{4.22}$$

From (4.4) we obtain the relevant derivatives of e_2 and e_3 which influence τ_ν,

$$e_2'''(t) \;=\; \mathrm{e}^{tA} e_2'''(0) + \tfrac{1}{3}(Su^{VI})(t), \tag{4.23}$$

$$e_3''(t) \;=\; \mathrm{e}^{tA} e_3''(0) - \tfrac{1}{4}\,(Su^{VI})(t), \tag{4.24}$$

where the initial values can be expressed by means of (4.5). Setting $t_\nu = 0$ in (4.16) and substituting the lower derivatives of $e_2(0)$ and $e_3(0)$ gives

$$e_2'''(0) \;=\; u^V(0) - \tfrac{2}{3}\, f^{IV}(0) - \tfrac{1}{3}\, Af'''(0), \tag{4.25}$$

$$e_3''(0) \;=\; -\tfrac{1}{2}\, u^V(0) + \tfrac{1}{4}\, f^{IV}(0). \tag{4.26}$$

4.3 The remainder ε_ν

With $\varepsilon_0 = \varepsilon_1 = 0$, the solution of the remainder equation (4.19) reads

$$\varepsilon_\nu = h \sum_{j=2}^{\nu} R_{\nu+2-j}(M)\, \tau_j, \quad \nu \geq 2\,, \tag{4.27}$$

with τ_ν from (4.20). To describe the behavior of ε_ν, we now study the various contributions of $u(t)$, $e_2(t)$ and $e_3(t)$ on the inhomogeneity τ_ν.

In the rest of this section we suppress the dependence of R_ν on $M = hA$ in our notation. We also introduce the shortcut denotation

$$E(t) := \mathrm{e}^{tA}, \quad E := E(h) = \mathrm{e}^{M}, \quad \text{with} \quad \|E\| \leq \mathrm{e}^{-h\kappa} < 1. \tag{4.28}$$

We now proceed in several steps.

- **Representation of the inhomogeneity τ_ν (cf. (4.20))**

We list the contributions from $u(t)$, $e_2(t)$ and $e_3(t)$ to τ_ν and split/recombine them for further analysis.

– **Contribution of $u(t)$ to τ_ν** (first term in (4.20)): From (4.15) we immediately have

$$h^4 I_{\nu,0} = \mathcal{O}(h^4). \tag{4.29}$$

– **Contribution of $e_2(t)$ to τ_ν** (second term in (4.20)): Inserting representation (4.23) for $e_2'''(t)$ into $I_{\nu,2}$ from (4.13) gives

$$h^4 I_{\nu,2} = h^4 E^{\nu-2} \hat{T}_2\, e_2'''(0) + \mathcal{O}(h^4), \tag{4.30}$$

where $\hat{T}_2$ is a Peano integral satisfying an identity analogous to (4.13),

$$\begin{aligned} \hat{T}_2 &= \int_0^2 K_2(\xi) E(\xi h)\,\mathrm{d}\xi = A^{-3} \int_0^2 K_2(\xi) E'''(\xi h)\,\mathrm{d}\xi \\ &= h^{-3} A^{-3} \big(\alpha_0 E(0) + \alpha_1 E(h) + \alpha_2 E(2h) - hE'(2h)\big) = M^{-3} \big(\alpha_0 I + \alpha_1 E + (\alpha_2 I - M) E^2\big). \end{aligned} \tag{4.31}$$

– **Contribution of $e_3(t)$ to τ_ν** (third term in (4.20)): Inserting representation (4.24) for $e_3''(t)$ into $I_{\nu,3}$ from (4.14) gives

$$h^4 I_{\nu,3} = h^4 E^{\nu-2} \hat{T}_3\, e_3''(0) + \mathcal{O}(h^4), \tag{4.32}$$

where analogously as for (4.31),

$$\begin{aligned} \hat{T}_3 &= \int_0^2 K_1(\xi) E(\xi h)\,\mathrm{d}\xi = A^{-2} \int_0^2 K_1(\xi) E''(\xi h)\,\mathrm{d}\xi \\ &= h^{-2} A^{-2} \big(\alpha_0 E(0) + \alpha_1 E(h) + \alpha_2 E(2h) - hE'(2h)\big) = M^{-2} \big(\alpha_0 I + \alpha_1 E + (\alpha_2 I - M) E^2\big). \end{aligned} \tag{4.33}$$

Here we have introduced the quantities

$$\hat{T}_2 = M^{-3}\, T(E), \tag{4.34}$$
$$\hat{T}_3 = M^{-2}\, T(E), \tag{4.35}$$

where

$$T(E) = T(E; M) := \alpha_0 I + \alpha_1 E + (\alpha_2 I - M) E^2 \tag{4.36}$$

is a truncation error, namely the characteristic polynomial (2.16) evaluated at $\zeta = E$. Note that $\hat{T}_2$ and $\hat{T}_3$ are uniformly bounded,

$$\|\hat{T}_2\| = \mathcal{O}(1), \quad \|\hat{T}_3\| = \mathcal{O}(1), \tag{4.37}$$

which is obvious from the Peano kernel representations (4.31) and (4.33).

- **Representation of the remainder ε_ν**

Proceeding from (4.27) and with $\tau_\nu = (4.29) + (4.30) + (4.32)$ we write ε_ν as

$$\varepsilon_\nu \;=\; h\sum_{j=2}^{\nu} R_{\nu+2-j}\,\tau_j \;=\; h\sum_{j=2}^{\nu} R_{\nu+2-j}\,\big(h^4 E^{j-2}(\hat{T}_2\, e_2'''(0) + \hat{T}_3\, e_3''(0)) + \mathcal{O}(h^4)\big), \quad \nu \geq 2\,. \tag{4.38}$$

The 'critical' terms in this representation, resulting from $e_2'''(0)$ and $e_3''(0)$, are denoted by $\hat{\varepsilon}_{\nu,2}$ and $\hat{\varepsilon}_{\nu,3}$. These can be expressed as

$$\hat{\varepsilon}_{\nu,2} \;=\; h^4\,\Omega_\nu \hat{T}_2\, e_2'''(0), \tag{4.39}$$
$$\hat{\varepsilon}_{\nu,3} \;=\; h^4\,\Omega_\nu \hat{T}_3\, e_3''(0), \tag{4.40}$$

where

$$\Omega_\nu \;:=\; h\sum_{j=2}^{\nu} R_{\nu+2-j} E^{j-2} \tag{4.41}$$

ist the solution of

$$\alpha_0\Omega_{\nu-2} + \alpha_1\Omega_{\nu-1} \;+ (\alpha_2 I - M)\Omega_\nu \;=\; hE^{\nu-2}, \quad \nu \geq 2, \tag{4.42}$$

with $\Omega_0 = \Omega_1 = 0$; see (2.17).

- **Approximation and damping properties of the discrete operator exponential**

First we need a bound for Ω_ν defined in (4.41), which follows from a simple stability argument based on Remark A.3: From (A.45) and $\|E\| \leq 1$ we immediately obtain[6]

$$\|\Omega_\nu\| \;\leq\; \frac{4}{\kappa} + \frac{3\,h}{4} \;\leq \mathcal{C} + \mathcal{O}(h)\,, \quad \nu \geq 2. \tag{4.43}$$

For sharper estimates involving powers of A, we use a representation for Ω_ν obtained from Lemma A.7, with $L = E$: Since A, E and the Ω_ν commute, Lemma A.7 yields

$$\Omega_\nu \underbrace{\big(\alpha_0 I + \alpha_1 E + (\alpha_2 I - M)E^2\big)}_{=\,T(E)\text{ from (4.36)}} \;=\; h(E^\nu - \tilde{E}_\nu)\,, \quad \nu \geq 2\,, \tag{4.44}$$

where $\tilde{E}_\nu$ is the BDF2 approximation to the operator exponential E^ν, i.e., the solution of

$$\alpha_0\tilde{E}_{\nu-2} + \alpha_1\tilde{E}_{\nu-1} \;+ (\alpha_2 I - M)\tilde{E}_\nu \;=\; 0, \quad \nu \geq 2, \quad \text{with} \;\; \tilde{E}_0 = I,\; \tilde{E}_1 = E\,. \tag{4.45}$$

From (4.44) and (4.34), (4.35) we see

$$\Omega_\nu \hat{T}_2 \;=\; hM^{-3}\big(E^\nu - \tilde{E}_\nu\big), \tag{4.46}$$
$$\Omega_\nu \hat{T}_3 \;=\; hM^{-2}\big(E^\nu - \tilde{E}_\nu\big). \tag{4.47}$$

In view of (4.39) and (4.40), estimation of ε_ν requires bounds for (4.46) and (4.47). This can be based on corresponding bounds for the scalar case, in particular Lemmas A.5 and A.6. The special case where A is selfadjoint is straightforward.

Proposition 4.1 *For A selfadjoint, the following estimates hold for $\nu \geq 2$:*

$$\|M^{-q}(E^\nu - \tilde{E}_\nu)\| \leq \frac{\mathcal{C}}{\nu^{2-q}}\,, \quad q = 0, 1, 2, \tag{4.48}$$

with κ from (4.1) and with corresponding constants $\mathcal{C} = \mathcal{C}(0)$ from Lemma A.6.

[6] Making use of $\|E\| \leq \mathrm{e}^{-h\kappa}$ together with (A.44), a sharper estimate for Ω_ν with exponential damping like in R_ν is also easy to derive, but this is not essential in the present context.

Proof: Standard spectral argument based on the scalar estimates from Lemma A.6. $\square$

Proposition 4.2 *Proposition 4.1 is also valid in the general sectorial case where A satisfies (2.12), with $\mathcal{C} = \mathcal{C}(\theta)$.*

Proof: Estimates of a type similar to (4.48) are given in [9], for $q = 0$, in a more general multistep context. In our context, $q = 1, 2$ will be relevant. For the proof of (4.48) in the general case we proceed in an analogous way using a Cauchy integral split into two parts: Consider a function $\varphi(\mu)$ ($\varphi(-\infty) = 0$) which is continuous on the sector $\mathcal{T}_{\hat\theta}$ ($\theta < \hat\theta < \frac{\pi}{2}$) and holomorphic on the interior of $\mathcal{T}_{\hat\theta}$, and which satisfies the following estimates for some constant $R > 0$ and two functions φ_1, φ_2:

$$\forall\, \mu \in \partial\mathcal{T}_{\hat\theta}, \quad |\mu| \le R: \qquad |\varphi(\mu)| \le \varphi_1(|\mu|), \tag{4.49}$$

$$\forall\, \mu \in \partial\mathcal{T}_{\hat\theta}, \quad |\mu| \ge R: \qquad |\varphi(\mu)| \le \varphi_2(|\mu|). \tag{4.50}$$

Then, using (2.14) it is easy to show (cf. [9, Lemma 1.2]) that there is constant $\mathcal{C}$ such that, for $0 < \theta < \hat\theta < \frac{\pi}{2}$,

$$\|\varphi(M)\| \le \frac{\mathcal{C}}{|\hat\theta - \theta|} \Big(\int_0^R \varphi_1(r)\, \frac{\mathrm{d}r}{r} + \int_R^\infty \varphi_2(r)\, \frac{\mathrm{d}r}{r} \Big). \tag{4.51}$$

Now, fix $\hat\theta \in (\theta, \frac{\pi}{2})$ (e.g., $\hat\theta = \frac{1}{2}(\frac{\pi}{2} - \theta)$), and let $R = \frac{1}{4\cos\hat\theta}$. In order to estimate

$$\varphi(M) := M^{-q}(E^\nu - \tilde{E}_\nu) \quad q = 0, 1, 2; \quad \nu = 2, 3, \ldots, \tag{4.52}$$

we consider its scalar analog[7] $\varphi(\mu) = \mu^{-q}(e^\nu - \tilde{e}_\nu)$ (with $e = \mathrm{e}^\mu$) for $\mu \in \partial\mathcal{T}_{\hat\theta}$ and invoke Lemma A.5 and Lemma A.6, with $\hat\theta$ playing the role of θ. With our choice $R = \frac{1}{4\cos\hat\theta}$ we have $\operatorname{Re}\mu \le [\ge] -\frac{1}{4} \iff |\mu| \ge [\le]\, R$. Estimate (A.27) (cf. the proof of Lemma A.6) shows

$$0 \le |\mu| \le R: \quad |\varphi(\mu)| \le \frac{\mathcal{C}(\theta)}{\nu^{2-q}}\, |\nu\mu|^{3-q}\, \mathrm{e}^{-\hat\xi\,|\nu\mu|} =: \varphi_1(|\mu|)\,, \quad \text{with } \hat\xi := \frac{2\ln(\frac{3}{2})}{\cos\hat\theta} > 0\,. \tag{4.53}$$

Furthermore, with $\delta(\mu) = \frac{1}{\sqrt{1 - 2\operatorname{Re}\mu}}$ from (A.3), estimate (A.23) from Lemma A.5 implies

$$|\mu| \ge R: \quad |\varphi(\mu)| \le 4\,|\mu|^{-q}(\nu-1)\,(\delta(\mu))^{\nu-1} \le 4\,|\mu|^{-q}(\nu-1)\Big(\frac{1}{\sqrt{1 + 2|\mu|\cos\hat\theta}}\Big)^{\nu-1} =: \varphi_2(|\mu|). \tag{4.54}$$

Integration over φ_1 gives

$$\int_0^R \varphi_1(r)\, \frac{\mathrm{d}r}{r} \le \int_0^\infty \varphi_1(r)\, \frac{\mathrm{d}r}{r} = \mathcal{C}\, \frac{(2-q)!}{\nu^{2-q}}\, \hat\xi^{q-3} = \frac{\mathcal{C}(\theta)}{\nu^{2-q}}\,, \tag{4.55}$$

and the integral over φ_2 can be estimated as [8]

$$\begin{aligned} \int_R^\infty \varphi_2(r)\, \frac{\mathrm{d}r}{r} &= 4(\nu-1) \int_R^\infty \Big(\frac{1}{\sqrt{1 + 2\,r\cos\hat\theta}}\Big)^{\nu-1} \frac{\mathrm{d}r}{r^{q+1}} \\ &\le \frac{4(\nu-1)}{R^{q+1}} \int_R^\infty \Big(\frac{1}{\sqrt{1 + 2\,r\cos\hat\theta}}\Big)^{\nu-1} \mathrm{d}r \le \frac{\mathcal{C}}{R^q} \sqrt{\tfrac{2}{3}}^{\,\nu} = \mathcal{C}(\theta) \sqrt{\tfrac{2}{3}}^{\,\nu} \le \frac{\mathcal{C}(\theta)}{\nu^{2-q}}\,, \end{aligned} \tag{4.56}$$

due to exponential damping. Invoking (4.51) concludes the proof. $\square$

[7] see Lemma A.4 for an explicit representation.

[8] Estimate (4.56) holds for $\nu \ge 4$. For $\nu = 2, 3$, the integral in the second line is not convergent, but $\int_0^R \varphi_1(r)\, \frac{\mathrm{d}r}{r}$ is (it can be exactly evaluated). Therefore, (4.56) holds for all $\nu \ge 2$.

- **Estimation of the remainder ε_ν**

Now we are ready to estimate the various contributions to the remainder ε_ν, see (4.38).

- **Contribution from the $\mathcal{O}(h^4)$-term in (4.38)** : The contribution from the $\mathcal{O}(h^4)$-terms can be estimated by a straightforward stability argument based on (A.45) . With the denotation introduced above, this yields

$$\varepsilon_\nu \;=\; \hat\varepsilon_{\nu,2} + \hat\varepsilon_{\nu,3} + \mathcal{O}(h^4)\,. \tag{4.57}$$

The 'critical' terms $\hat\varepsilon_{\nu,2}$ and $\hat\varepsilon_{\nu,3}$ remain to be estimated. On the basis of the above representations (4.34)–(4.47) for $\hat T_2$ and $\hat T_3$ we shall argue in different ways. With the explicit expressions (4.25),(4.26) for $e_2'''(0)$, $e_3''(0)$, the critical terms read

$$\hat\varepsilon_{\nu,2} \;=\; h^4\,\underbrace{\Omega_\nu \hat T_2\, u^V(0)}_{=:\;\hat\omega^{(i)}_{\nu,2}} - \tfrac{2}{3}\,h^4\,\underbrace{\Omega_\nu \hat T_2 A(A^{-1} f^{IV}(0))}_{=:\;\hat\omega^{(ii)}_{\nu,2}} - \tfrac{1}{3}\,h^4\,\underbrace{\Omega_\nu \hat T_2 A^2(A^{-1} f'''(0))}_{=:\;\hat\omega^{(iii)}_{\nu,2}}\,, \tag{4.58}$$

$$\hat\varepsilon_{\nu,3} \;=\; -\tfrac{1}{2}\,h^4\,\underbrace{\Omega_\nu \hat T_3\, u^V(0)}_{=:\;\hat\omega^{(i)}_{\nu,3}} + \tfrac{1}{4}\,h^4\,\underbrace{\Omega_\nu \hat T_3 A(A^{-1} f^{IV}(0))}_{=:\;\hat\omega^{(ii)}_{\nu,3}}\,. \tag{4.59}$$

The various contributions from (4.58) and (4.59) are now considered separately.

- **Estimation of $\hat\omega^{(i)}_{\nu,2}$ and $\hat\omega^{(i)}_{\nu,3}$:** A simple stability argument based on (4.43) and (4.37) yields

$$\|\hat\omega^{(i)}_{\nu,2}\| \;\le\; \|\Omega_\nu\|\,\|\hat T_2\|\,\|u^V(0)\| \;\le\; \mathcal{C} \tag{4.60}$$

$$\|\hat\omega^{(i)}_{\nu,3}\| \;\le\; \|\Omega_\nu\|\,\|\hat T_3\|\,\|u^{IV}(0)\| \;\le\; \mathcal{C}. \tag{4.61}$$

- **Estimation of $\hat\omega^{(ii)}_{\nu,2}$ and $\hat\omega^{(ii)}_{\nu,3}$:** From (4.46),(4.47) and (4.48), together with the fact that all occurring operators commute we obtain

$$\|\hat\omega^{(ii)}_{\nu,2}\| \;\le\; \|\Omega_\nu \hat T_2\, A\|\,\|A^{-1} f^{IV}(0)\| \;=\; \|M^{-2}(E^\nu - \tilde E_\nu)\|\,\|A^{-1} f^{IV}(0)\| \;\le\; \mathcal{C}(\theta) \tag{4.62}$$

$$\|\hat\omega^{(ii)}_{\nu,3}\| \;\le\; \|\Omega_\nu \hat T_3\, A\|\,\|A^{-1} f^{IV}(0)\| \;=\; \|M^{-1}(E^\nu - \tilde E_\nu)\|\,\|A^{-1} f^{IV}(0)\| \;\le\; \frac{\mathcal{C}(\theta)}{\nu}\,. \tag{4.63}$$

- **Estimation of $\hat\omega^{(iii)}_{\nu,2}$:** From (4.46) and (4.48), together with the fact that all occurring operators commute we obtain

$$\|\hat\omega^{(iii)}_{\nu,2}\| \;\le\; \|\Omega_\nu \hat T_2\, A^2\|\,\|A^{-1} f'''(0)\| \;=\; h^{-1}\|M^{-1}(E^\nu - \tilde E_\nu)\|\,\|A^{-1} f'''(0)\| \;\le\; \frac{\mathcal{C}(\theta)}{\nu h}. \tag{4.64}$$

Summarizing all these considerations eventually leads us to the following result. We formulate it in a slightly informal way; the essential point is the order reduction of ε_ν down to $\mathcal{O}(h^3)$ at the first grid points due to (4.64), which is damped out algebraically at later grid points.

Proposition 4.3 *Under the assumptions stated at the beginning of Section 4, the global error of the BDF2 scheme admits an expansion*

$$\tilde u_\nu - u(t_\nu) \;=\; h^2 e_2(t_\nu) + h^3 e_3(t_\nu) + \varepsilon_\nu, \tag{4.65}$$

where $e_2(t), e_3(t)$ are the solutions (4.21),(4.22) of the variational equations (4.17),(4.18), and where the remainder ε_ν can be estimated by

$$\|\varepsilon_\nu\| \le \Big(\mathcal{C}_0(\hat\theta) + \frac{\mathcal{C}_1(\hat\theta)}{t_\nu}\Big) h^4 + \mathcal{O}(h^4). \tag{4.66}$$

Here the $\mathcal{C}_j(\hat\theta)$ depend on derivatives of $A^{-1} f$ and are independent on ν; the $\mathcal{O}(h^4)$ term can be estimated in terms of derivatives of the solution $u(t)$ over the integration interval considered.

If f and its relevant derivatives are $\mathcal{O}(1)$, we have $\|\varepsilon_\nu\| \leq \mathcal{C}(\theta)h^4$, without order reduction. This follows easily by estimating

$$\|\hat{\omega}_{\nu,2}^{(iii)}\| \;\leq\; \|\Omega_\nu \hat{T}_2\, A\|\, \|f'''(0)\| \;=\; \|M^{-2}(E^\nu - \tilde{E}_\nu)\|\, \|f'''(0)\| \;\leq\; \mathcal{C}(\theta), \tag{4.67}$$

instead of (4.64). However, for longer expansions one will expect that transient order reduction effects at a higher level are to be expected also in this case.

From the above analysis one can also argue that the damping behavior is more of an exponential type if the stepsize h is bounded from below. We refrain from formulating this an quantitative way.

5 Numerical illustration

Via the method of lines, we apply the BDF2 scheme to

$$\partial_t u(x,t) = \partial_{xx}(u(x,t) - \mathrm{e}^{-t}\cos x) - \mathrm{e}^{-t}\cos x, \qquad u(x,0) = \cos x, \tag{5.1}$$

with periodic boundary conditions on the interval $0 \leq x \leq 2\pi$. The solution of (5.1) is $u(x,t) = \mathrm{e}^{-t}\cos x$.

We use standard second order finite differences over $[0, 2\pi]$ with mesh size $\Delta x = 2\pi/400$, and combine BDF2 with global Richardson extrapolation w.r.t. t to obtain access to the higher order terms of the error. The coarse step size is fixed, $h_0 = 1/64$, and two extrapolation steps with are performed using $h_1 = 1/128$ and $h_2 = 1/256$. The error was measured in the discrete L^2-norm at $t_\nu = \nu h_0$, $\nu = 1, 2, \ldots$. With the exact solution at hand, in Figure 1 we keep track of the order observed for the global error of BDF2, and the orders observed for the first and second step of Richardson extrapolation are shown under under "3$^{\text{rd}}$ order" and "4$^{\text{th}}$ order". A graphical illustration is also included.

ν	BDF2	3$^{\text{rd}}$ order	4$^{\text{th}}$ order
1	1.78	2.18	3.95
2	1.83	2.72	7.12
3	1.88	2.93	4.81
4	1.91	3.01	1.27
5	1.93	3.03	3.53
6	1.94	3.04	3.88
7	1.95	3.05	3.99
8	1.96	3.05	4.02
9	1.96	3.05	4.04
10	1.97	3.05	4.04

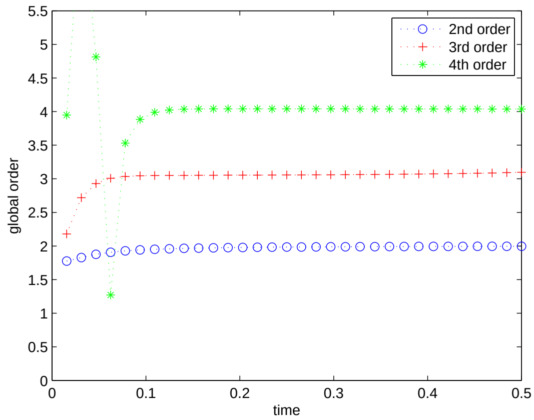

Figure 1: Observed order for BDF2 and after 1 and 2 steps of Richardson extrapolation.

The somewhat irregular transient behavior of the higher order terms is to be attributed to the order reduction effect described in Section 4.

A Auxiliary results for the BDF2 scheme

This appendix contains an orchard of estimates specific to the BDF2 scheme. Assumptions and notation are as in the main sections, cf. in particular Section 2.

A.1 The scalar case

It should be noted that the central estimates of this section, namely those from Lemma A.5, are analoga of results derived in [9] for general $A(\theta)$-stable multistep schemes. Nevertheless, we decided to provide explicit versions for the special, A-stable case BDF2 based on quantitative information about the location of characteristic roots. Our estimates, for instance (A.3), are elementary but not completely trivial.

- **Estimation of characteristic roots**

For the BDF2 scheme, the solutions of the characteristic equation

$$p(\zeta;\mu) \;=\; \alpha_0+\alpha_1\zeta+(\alpha_2-\mu)\zeta^2 \;=\; \tfrac12-2\zeta+(\tfrac32-\mu)\zeta^2 \;=\; 0 \tag{A.1}$$

are given by [9]

$$\zeta_{1,2} \;=\; \zeta_{1,2}(\mu) \;=\; \frac{2\pm\sqrt{1+2\mu}}{3-2\mu} \;=\; \frac{1}{2\mp\sqrt{1+2\mu}} \neq 0\,, \quad |\zeta_1|\geq|\zeta_2|\,. \tag{A.2}$$

Lemma A.1 *For all* $\mu\in\mathbb{C}_-$,

$$|\zeta_{1,2}(\mu)| \;\leq\; \frac{1}{\sqrt{1-2\,\mathrm{Re}\,\mu}} \;=:\; \delta(\mu)\,. \tag{A.3}$$

Proof: We estimate $|\,2\mp\sqrt{1+2\mu}\,|^{\,2}$ from below. Let us denote

$$a \;:=\; |1+2\mu|, \quad r \;:=\; \mathrm{Re}\,(1+2\mu)<1. \tag{A.4}$$

With $\frac12\arg(1+2\mu)=:\psi\in(-\frac{\pi}{2},\frac{\pi}{2})$ we have

$$r \;=\; a\cos(2\psi) \;=\; 2\,a\cos^2\!\psi-a, \qquad \mathrm{Re}\,\sqrt{1+2\mu} \;=\; \sqrt{a}\,\cos\psi, \tag{A.5}$$

hence

$$\mathrm{Re}\,\sqrt{1+2\mu} \;=\; \sigma\,\frac{\sqrt2}{2}\,\sqrt{a+r}\,, \quad \sigma \;=\; \mathrm{sign}(\mathrm{Re}\,\sqrt{1+2\mu}). \tag{A.6}$$

This yields [10]

$$\begin{aligned} |\,2\mp\sqrt{1+2\mu}\,|^{\,2} \;=\; (2\mp\mathrm{Re}\,\sqrt{1+2\mu})^{\,2}+(\mathrm{Im}\,\sqrt{1+2\mu})^{\,2} \;&=\; 4\mp4\,\mathrm{Re}\,\sqrt{1+2\mu}+a \\ &\geq\; 4-2\sqrt2\,\sqrt{a+r}+a \;\geq\; 2-r, \end{aligned} \tag{A.7}$$

hence

$$\Big|\,\frac{1}{2\mp\sqrt{1+2\mu}}\,\Big| \;\leq\; \frac{1}{\sqrt{2-r}} \;=\; \frac{1}{\sqrt{1-2\,\mathrm{Re}\,\mu}}\,, \tag{A.8}$$

which proves (A.3). □

For values of $\mu\in\mathbb{C}_-$ away from zero, estimate (A.3) reflects the strong stability of the BDF2 scheme.

Next we give a precise estimate of the principal root ζ_1 for small values of $\mathrm{Re}\,\mu$. For $\mu\to0$ it satisfies $|\zeta_1(\mu)|\approx \mathrm{e}^{\mathrm{Re}\,\mu}$ by consistency of the scheme, while $|\zeta_2(\mu)|\approx 1/3$. The quantitative estimate given in the following lemma is not asymptotically sharp for $\mu\to0$. However, it is reasonably sharp and uniformly valid for a finite range of μ- values.

Lemma A.2 *For* $\mu\in\mathbb{C}_-$ *with* $\mathrm{Re}\,\mu\geq-\frac14$,

$$|\zeta_1(\mu)| \;=\; \Big|\,\frac{1}{2-\sqrt{1+2\mu}}\,\Big| \;\leq\; \mathrm{e}^{\,\iota\,\mathrm{Re}\,\mu}, \quad \iota \;=\; 2\ln(\tfrac32) \;\approx\; 0.81. \tag{A.9}$$

Proof: (A.9) follows from (A.3) by applying the elementary estimate

$$\frac{1}{1-2\,\xi} \;\leq\; \mathrm{e}^{\,2\,\iota\,\xi}, \qquad -\tfrac14\leq\xi\leq0, \quad \iota \text{ from (A.9)}, \tag{A.10}$$

to $\xi=\mathrm{Re}\,\mu\in[-\frac14,0]$. □

Now we show that under a sectorial condition on $\mu\in\mathbb{C}_-$, the principal root $\zeta_1(\mu)$ is strictly larger in size than $\zeta_2(\mu)$, uniformly for $\mathrm{Re}\,\mu\geq-\frac14$.

[9] $\sqrt{\cdot}$ denotes the main branch of the complex root, with nonnegative real part.

[10] The last inequality in (A.7) is equivalent to the arithmetic-geometric mean inequality $2\sqrt2\,\sqrt{a+r}\;\leq\;2+(a+r)$.

Lemma A.3 *For $\mu \in \mathcal{T}_\theta$ (cf. (2.13)) with $\operatorname{Re}\mu \ge -\frac{1}{4}$,*

$$\Big| \frac{\zeta_2(\mu)}{\zeta_1(\mu)} \Big| \le \beta(\theta) \quad \textit{with} \quad \beta(\theta) < 1, \quad 0 \le \theta < \tfrac{\pi}{2}\,. \tag{A.11}$$

An explicit estimate for $\beta(\theta)$ is given in the proof below. For $\theta \to \frac{\pi}{2}$ we have $\beta(\theta) \to 1$.

Proof: We will estimate

$$\Big| \frac{\zeta_2(\mu)}{\zeta_1(\mu)} \Big|^2 = \Big| \frac{2-\sqrt{1+2\mu}}{2+\sqrt{1+2\mu}} \Big|^2 \tag{A.12}$$

from above. With the same denotation as in the proof of Lemma A.2, we have

$$\Big| \frac{2-\sqrt{1+2\mu}}{2+\sqrt{1+2\mu}} \Big|^2 = \frac{4-2\sqrt{2}\,\sqrt{a+r}+a}{4+2\sqrt{2}\,\sqrt{a+r}+a} = \frac{1-x}{1+x}, \qquad \text{with} \quad x = 2\sqrt{2}\,\frac{\sqrt{a+r}}{4+a}\,. \tag{A.13}$$

For values of μ under consideration we can derive a simple lower estimate for x by separately estimating numerator and denominator, namely:

$$\sqrt{a+r} \ge 1, \quad 4+a \le 4+m(\theta), \quad m(\theta) := \max\{1, \frac{1}{2\cos\theta}\} = \begin{cases} 1, & \theta \le \frac{\pi}{3}, \\ \frac{1}{2\cos\theta}, & \theta > \frac{\pi}{3}. \end{cases} \tag{A.14}$$

This gives

$$x \ge \frac{2\sqrt{2}}{4+m(\theta)} \quad \Rightarrow \quad \frac{1-x}{1+x} \le \frac{1-\frac{2\sqrt{2}}{4+m(\theta)}}{1+\frac{2\sqrt{2}}{4+m(\theta)}} \in \Big(\frac{1-\frac{2\sqrt{2}}{5}}{1+\frac{2\sqrt{2}}{5}}, 1 \Big), \quad 0 \le \theta < \tfrac{\pi}{2}\,. \tag{A.15}$$

This proves (A.11), with

$$\beta(\theta) = \left(\frac{1-\frac{2\sqrt{2}}{4+m(\theta)}}{1+\frac{2\sqrt{2}}{4+m(\theta)}} \right)^{\frac{1}{2}} < 1\,. \tag{A.16}$$

□

- **BDF2 approximation to the scalar exponential function**

Let $\mu \in \mathbb{C}$ be arbitrarily fixed. Analogously as in (4.28) we introduce the shortcut

$$e := \mathrm{e}^{\mu}. \tag{A.17}$$

The solution $\tilde{e}_\nu$ of the difference equation

$$\alpha_0\,\tilde{e}_{\nu-2} + \alpha_1\,\tilde{e}_{\nu-1} + \alpha_2\,\tilde{e}_\nu = \mu\,\tilde{e}_\nu, \quad \nu \ge 2, \quad \text{with } \tilde{e}_0 = 1,\ \tilde{e}_1 = e\,, \tag{A.18}$$

(the scalar version of (4.45)), is the BDF2 approximation to the exponential $e^\nu = \mathrm{e}^{\nu\mu}$, with exact initial values. For our purpose in Section 4 it is convenient to express $\tilde{e}_\nu$ in a direct way in terms of the characteristic roots (A.2). The proof of the following lemma is elementary.

Lemma A.4 *For $\nu \ge 2$ we have*

$$e^\nu - \tilde{e}_\nu = (e-\zeta_1) \sum_{j=0}^{\nu-2} \zeta_1^{\nu-2-j}\,(e^{j+1} - \zeta_2^{j+1}) = \hat{p}(e)\, e^{\nu-2} \sum_{j=0}^{\nu-2} \Big(\frac{\zeta_1}{e}\Big)^j \sum_{\ell=0}^{j} \Big(\frac{\zeta_2}{\zeta_1}\Big)^\ell. \tag{A.19}$$

Here,

$$\hat{p}(e) = (e-\zeta_1)(e-\zeta_2) = p(e)(\alpha_2-\mu)^{-1} = e^2 - \tilde{e}_2 \tag{A.20}$$

is a truncation error, namely the monic characteristic polynomial (2.5) evaluated at $\zeta = e$, satisfying[11]

$$|\hat{p}(e)| \le \tfrac{2}{9}\,|\mu|^3 \quad \textit{for all } \mu \in \mathbb{C}_-\,. \tag{A.21}$$

[11] (A.21) is reasonably sharp, and will be used, for small values of μ.

In the sequel we use the denotation $\mathcal{C}$ resp. $\mathcal{C}(\theta)$ in a generic sense, with $\mathcal{C}(\theta)$ bounded for fixed θ but $\mathcal{C}(\theta) \to \infty$ for $\theta \to \frac{\pi}{2}$.

Lemma A.5 *For $\nu \geq 2$ we have*

– *For $\mu \in \mathcal{T}_\theta$ with $\operatorname{Re}\mu \geq -\frac{1}{4}$,*

$$|e^\nu - \tilde{e}_\nu| \;\leq\; \mathcal{C}(\theta)\,|\mu|^3\,(\nu-1)\,\mathrm{e}^{(\nu-2)\iota\operatorname{Re}\mu}\,, \tag{A.22}$$

with $\iota \approx 0.81$ from (A.9).

– *For $\operatorname{Re}\mu \leq -\frac{1}{4}$,*

$$|e^\nu - \tilde{e}_\nu| \;\leq\; = \; 4\,(\nu-1)\,\delta(\mu)^{\nu-1}\,, \tag{A.23}$$

with $\delta(\mu)$ from (A.3)*, $\delta(\mu) \leq \sqrt{\frac{2}{3}} \approx 0.82$ for $\operatorname{Re}\mu \leq -\frac{1}{4}$.*

Proof:

– $\operatorname{Re}\mu \geq -\frac{1}{4},\ \mu \in \mathcal{T}_\theta:$ On basis of the second identity in (A.19) and together with (A.21), Lemmas A.2 and A.3 (with $\iota < 1$ from (A.9) and $\beta(\theta) < 1$ from (A.16)) yield the bound

$$\begin{aligned} |e^\nu - \tilde{e}_\nu| \;&\leq\; \tfrac{2}{9}\,|\mu|^3\,\mathrm{e}^{(\nu-2)\operatorname{Re}\mu}\sum_{\ell=0}^{\infty}\beta(\theta)^\ell\sum_{j=0}^{\nu-2}\mathrm{e}^{j(\iota-1)\operatorname{Re}\mu} \\ &\leq\; \mathcal{C}(\theta)\,|\mu|^3\,\mathrm{e}^{(\nu-2)\operatorname{Re}\mu}(\nu-1)\,\mathrm{e}^{(\nu-2)(\iota-1)\operatorname{Re}\mu} \;=\; \mathcal{C}(\theta)\,|\mu|^3\,(\nu-1)\,\mathrm{e}^{(\nu-2)\iota\operatorname{Re}\mu}\,, \end{aligned} \tag{A.24}$$

with $\mathcal{C}(\theta)$ depending on $\beta(\theta)$. (Note that $\mathrm{e}^{(\iota-1)\operatorname{Re}\mu} \geq 1$.)

– $\operatorname{Re}\mu \leq -\frac{1}{4}:$ On basis of the first identity in (A.19), Lemma A.1 (with $\delta(\mu)$ from (A.3)) together with the elementary estimate $|e| = |\mathrm{e}^\mu| = \mathrm{e}^{\operatorname{Re}\mu} \leq \delta(\mu)$ for $\operatorname{Re}\mu \leq 0$ yield the bound

$$|e^\nu - \tilde{e}_\nu| \;\leq\; 2\,\delta(\mu)\sum_{j=0}^{\nu-2}\delta(\mu)^{\nu-2-j}\,2\,\delta(\mu)^{j+1} \;=\; 4\,(\nu-1)\,\delta(\mu)^{\nu-1}\,. \tag{A.25}$$

This completes the proof. □

The proof of the following lemma is based on a routine argument. The estimates are uniformly valid for $\mu \in \mathcal{T}_\theta$, (they are not sharp for larger values of $|\mu|$, where exponential damping occurs in (A.23)).

Lemma A.6 *For $0 \neq \mu \in \mathcal{T}_\theta$ and $\nu \geq 2$, the following estimates hold:*

$$|\mu^{-q}(e^\nu - \tilde{e}_\nu)| \leq \frac{\mathcal{C}(\theta)}{\nu^{2-q}}\,, \quad q = 0,1,2, \tag{A.26}$$

with constants $\mathcal{C}(\theta)$ independent of μ.

Proof: Let $q \in \{0,1,2\}$. For $\operatorname{Re}\mu \geq -\frac{1}{4}$, (A.22) implies

$$|\mu^{-q}(e^\nu - \tilde{e}_\nu)| \;\leq\; \mathcal{C}(\theta)\nu^{q-2}|\nu\mu|^{3-q}\,\mathrm{e}^{-\xi\,|\nu\mu|} \;\leq\; \frac{\mathcal{C}(\theta)}{\nu^{2-q}}\,|\nu\mu|^3\,\mathrm{e}^{-\xi\,|\nu\mu|}\,, \tag{A.27}$$

with $\xi = \xi(\theta) := \frac{\iota}{\cos\theta}$ (ι from (A.9)), which implies (A.26). For $\operatorname{Re}\mu \leq -\frac{1}{4}$, (A.26) is a direct consequence of the exponential damping property (A.23). □

A.2 Some simple estimates based on G-stability

The BDF2 scheme is G-stable, cf. [7, Example 6.5]. We can easily exploit this property to obtain various stability estimates for the case $k = 2$, which are of a similar quality as for the backward Euler scheme. As before, $M = hA$. Throughout this section, we simply write R_ν for the coefficients $R_\nu = R_\nu(M)$ of the discrete resolvent.

First we assume (2.12) with $\theta \leq \frac{\pi}{2}$, i.e.

$$\operatorname{Re}\langle Av, v\rangle \leq 0 \quad \text{for all } v \in \mathcal{H}. \tag{A.28}$$

Remark A.1 *For A satisfying* (A.28),

$$\|R_\nu\| \le \tfrac{3}{2}\,, \quad \|\hat R_\nu\| \le \tfrac{9}{4}\,, \quad \nu \ge 2. \tag{A.29}$$

Proof:

– The R_ν satisfy the homogeneous difference equation (2.18) with $k=2$, and $R_0 = -\alpha_0^{-1} I = -2I,\ R_1 = 0$, $R_2 = (\alpha_2 I - M)^{-1}$. With

$$\boldsymbol{R}_\nu := \begin{pmatrix} R_\nu \\ R_{\nu-1} \end{pmatrix} :\ \mathcal{H} \to \mathcal{H}^2, \tag{A.30}$$

the recursion for the R_ν can be written as

$$\boldsymbol{R}_\nu = \boldsymbol{C}\boldsymbol{R}_{\nu-1}, \quad \nu \ge 2, \tag{A.31}$$

with the block companion operator

$$\boldsymbol{C} = \boldsymbol{C}(M) = \begin{pmatrix} -\alpha_1(\alpha_2 I - M)^{-1} & -\alpha_0(\alpha_2 I - M)^{-1} \\ I & 0 \end{pmatrix} :\ \mathcal{H}^2 \to \mathcal{H}^2. \tag{A.32}$$

Since A is assumed to be dissipative, G-stability of the BDF2 scheme implies that $\boldsymbol{C}$ is a contraction with respect to a modified inner product on $\mathcal{H}^2$. In particular, [7, Example 6.5] shows that the transformed operator

$$\tilde{\boldsymbol{C}} := \tilde{\boldsymbol{G}}\boldsymbol{C}\tilde{\boldsymbol{G}}^{-1} :\ \mathcal{H}^2 \to \mathcal{H}^2, \quad \text{with} \quad \tilde{\boldsymbol{G}} = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \otimes I, \tag{A.33}$$

satisfies[12] $\|\tilde{\boldsymbol{C}}\| \le 1$, or equivalently, $\|\boldsymbol{C}\|_{\tilde{\boldsymbol{G}}} \le 1$ w.r.t. the vector norm $\|\boldsymbol{v}\|_{\tilde{\boldsymbol{G}}} := \|\tilde{\boldsymbol{G}}\boldsymbol{v}\|$ on $\mathcal{H}^2$. Together with (A.31) we have

$$\tilde{\boldsymbol{R}}_\nu = \tilde{\boldsymbol{C}}\tilde{\boldsymbol{R}}_{\nu-1}, \quad \text{where} \quad \tilde{\boldsymbol{R}}_\nu := \tilde{\boldsymbol{G}}\boldsymbol{R}_\nu = \begin{pmatrix} 3R_\nu - R_{\nu-1} \\ R_\nu - R_{\nu-1} \end{pmatrix}. \tag{A.34}$$

Now, from G-stability and the fact that $\|(\alpha_2 I - M)^{-1}\| \le \frac{2}{3}$ we infer for arbitrary $v \in \mathcal{H}$ and $\nu \ge 2$,

$$\|\tilde{\boldsymbol{R}}_\nu v\| = \|\boldsymbol{R}_\nu v\|_{\tilde{\boldsymbol{G}}} \le \|\boldsymbol{R}_2 v\|_{\tilde{\boldsymbol{G}}} = \|\tilde{\boldsymbol{R}}_2 v\| = \left\| \begin{pmatrix} 3\,(\alpha_2 I - M)^{-1} v \\ -(\alpha_2 I - M)^{-1} v \end{pmatrix} \right\| = \tfrac{2}{3}\sqrt{10}\,\|v\|. \tag{A.35}$$

Together with the straightforward estimate

$$\|\,2R_\nu v\| \le \|(3R_\nu - R_{\nu-1})v\| + \|(R_\nu - R_{\nu-1})v\| \le \sqrt{2}\,\|\tilde{\boldsymbol{R}}_\nu v\|, \tag{A.36}$$

this shows $\|R_\nu v\| \le \frac{2}{3}\sqrt{5}\,\|v\| \le \frac{3}{2}\,\|v\|$, which proves the first inequality in (A.29).

– The second inequality in (A.29) is obtained in the same way, considering $\hat R_\nu = (a_2 I - M)R_\nu$ instead of R_ν, with $\hat R_1 = 0,\ \hat R_2 = I$. □

Our next remarks concern the strictly dissipative case with a coercive operator A,

$$\operatorname{Re}\langle Av, v\rangle \le -\kappa\langle v, v\rangle \quad \text{for all } v \in \mathcal{H}, \quad \text{with } \kappa > 0. \tag{A.37}$$

Here, the results from [7] about G-stability can easily be modified to show a damping property of the R_ν. First we note that (A.37) implies

$$\|R_2\| = \|(\tfrac{3}{2}\,I - M)^{-1}\| \le \frac{1}{\frac{3}{2} + h\kappa} \le \frac{\frac{2}{3}}{\sqrt{1 + \frac{3}{4}\,h\kappa}}\,. \tag{A.38}$$

With the denotation introduced before, we have

[12] Here, by $\|\cdot\|$ we also denote the natural extension of the [operator] norm on $\mathcal{H}$ to $\mathcal{H}^2$.

Remark A.2 *For A satisfying* (A.37),

$$\| \boldsymbol{C} \|_{\tilde{\boldsymbol{G}}} \le \frac{1}{\sqrt{1+\frac{3}{4}\,h\kappa}} < 1. \tag{A.39}$$

Proof: We consider a recursion ($M = hA$)

$$\alpha_0\, v_{\nu-2} + \alpha_1\, v_{\nu-1} + (\alpha_2 I - M) v_\nu = 0, \quad \nu \ge 2\,, \tag{A.40}$$

which is equivalent to $\boldsymbol{v}_\nu = \boldsymbol{C}\,\boldsymbol{v}_{\nu-1}$ for $\boldsymbol{v}_\nu := \begin{pmatrix} v_\nu \\ v_{\nu-1} \end{pmatrix}$. For A satisfying (A.37), taking the inner product with v_ν gives

$$\mathrm{Re}\,\langle \alpha_0\, v_{\nu-2} + \alpha_1\, v_{\nu-1} + \alpha_2\, v_\nu, v_\nu \rangle = \alpha_2\, h \langle A\, v_\nu, v_\nu \rangle \le \alpha_2\, h\kappa \langle v_\nu, v_\nu \rangle. \tag{A.41}$$

Now, exactly following the argument in [7, Example 6.5] results in the estimate

$$\|\boldsymbol{v}_\nu\|_{\tilde{\boldsymbol{G}}}^2 \le \|\boldsymbol{v}_{\nu-1}\|_{\tilde{\boldsymbol{G}}}^2 + \alpha_2\, h\kappa \|v_\nu\|^2. \tag{A.42}$$

Together with the estimate $\|v_\nu\|^2 \le \frac{1}{2}\,\|\boldsymbol{v}_\nu\|_{\tilde{\boldsymbol{G}}}^2$ (which is analogous to (A.36)) and $\alpha_2 = \frac{3}{2}$ this yields

$$\|\boldsymbol{v}_\nu\|_{\tilde{\boldsymbol{G}}} \le \frac{1}{\sqrt{1+\frac{3}{4}\,h\kappa}}\, \|\boldsymbol{v}_{\nu-1}\|_{\tilde{\boldsymbol{G}}}\,, \tag{A.43}$$

which shows (A.39). □

From (A.39) we obtain the following estimates for the R_ν. In particular, (A.45) [13] is valid uniformly in ν.

Remark A.3

– *For A satisfying* (A.37) *we have for all $\nu \ge 2$*

$$\|R_\nu\| \le \frac{3}{2}\Big(1+\tfrac{3}{4}\,h\kappa\Big)^{-\frac{\nu-1}{2}}, \qquad \|\hat{R}_\nu\| \le \frac{9}{4}\Big(1+\tfrac{3}{4}\,h\kappa\Big)^{-\frac{\nu-2}{2}}. \tag{A.44}$$

– *Furthermore,*

$$h\sum_{j=2}^{\nu} \|R_{\nu+2-j}\| \le \frac{4}{\kappa} + \frac{3\,h}{4}\,. \tag{A.45}$$

Proof:

– The first inequality in (A.44) is obtained in exactly the same way as in the proof of Remark A.1. Making use of (A.39), we simply replace (A.35) by

$$\|\tilde{\boldsymbol{R}}_\nu v\| \le \Big(1+\tfrac{3}{4}\,h\kappa\Big)^{-\frac{\nu-2}{2}}\, \|\tilde{\boldsymbol{R}}_2 v\|, \tag{A.46}$$

where $\|\tilde{\boldsymbol{R}}_2 v\| \le \sqrt{10}\,\|R_2 v\| \le \frac{2}{3}\,\sqrt{10}\,\big(1+\frac{3}{4}\,h\kappa\big)^{-\frac{1}{2}}\,\|v\|$ due to (A.38). The second inequality in (A.44) is obtained analogously.

– (A.45) follows from (A.44) by summing up norms. With $\gamma := \frac{3}{4}\,h\kappa$ we have

$$h\sum_{j=2}^{\nu} \|R_{\nu+2-j}\| \le \frac{3\,h}{2}\sum_{j=2}^{\nu} (1+\gamma)^{-\frac{\nu+1-j}{2}} \le \frac{3\,h}{2}\,\frac{1}{\sqrt{1+\gamma}\,-1} = \frac{3\,h}{2}\,\frac{1+\sqrt{1+\gamma}}{\gamma} \le \frac{3\,h}{2}\,\frac{2+\frac{\gamma}{2}}{\gamma}\,, \tag{A.47}$$

which is (A.45). □

[13] The behavior is analogous as for B-stable one-step methods like the backward Euler scheme. In the limit $\kappa \to 0$, linear growth in t_ν occurs.

A.3 A special difference equation

Lemma A.7 below is formulated for $k = 2$, for use in Section 4. However, the generalization to general k is apparent. The assumptions on the given problem are as before, see Section 2.

We consider the operator-valued difference equation with BDF2 coefficients,

$$\alpha_0\Omega_{\nu-2} + \alpha_1\Omega_{\nu-1} + (\alpha_2 I - M)\,\Omega_\nu \;=\; hL^{\nu-2}, \quad \nu \geq 2, \tag{A.48}$$

with a given, densely defined linear operator L. As before, $R_\nu = R_\nu(M)$, $M = hA$. For homogeneous initial conditions $\Omega_0 = \Omega_1 = 0$ we have (cf. (2.17))

$$\Omega_\nu \;=\; h\sum_{j=2}^{\nu} R_{\nu+2-j}L^{j-2}, \quad \nu \geq 2. \tag{A.49}$$

Lemma A.7 *For all $\nu \geq 2$, the solution Ω_ν of* (A.48) *satisfies*

$$\alpha_0\Omega_\nu + \alpha_1\Omega_\nu L + (\alpha_2 I - M)\Omega_\nu L^2 \;=\; h(L^\nu - \tilde{L}_\nu)\,, \tag{A.50}$$

where

$$\tilde{L}_\nu \;=\; -\alpha_0 R_\nu - (\alpha_0 R_{\nu-1} + \alpha_1 R_\nu)L \tag{A.51}$$

is the solution of

$$\alpha_0\tilde{L}_{\nu-2} + \alpha_1\tilde{L}_{\nu-1} + (\alpha_2 I - M)\tilde{L}_\nu \;=\; 0, \quad \textit{with} \;\; \tilde{L}_0 = I,\; \tilde{L}_1 = L\,. \tag{A.52}$$

Proof:

– First we prove (A.50),(A.51). From (A.49) we have a one-step recursion for the Ω_ν in the form

$$\Omega_\nu \;=\; \Omega_{\nu-1}L + hR_\nu, \quad \nu \geq 2\,. \tag{A.53}$$

We now proceed by induction. For $\nu = 2$ we have $\Omega_2 = h\,(\alpha_2 I - M)^{-1} = hR_2$, hence

$$\begin{aligned}\alpha_0\Omega_2 + \alpha_1\Omega_2\,L + (\alpha_2 I - M)\,\Omega_2\,L^2 \;&=\; h(\alpha_0 R_2 \;+\alpha_1 R_2\,L + L^2)\\ &=\; h(\alpha_0 R_2 + (\alpha_0 R_1 + \alpha_1 R_2)L + L^2) \;=\; h(L^2 - \tilde{L}_2)\,,\end{aligned} \tag{A.54}$$

since $R_1 = 0$. Thus, (A.50) is satisfied for $\nu = 2$.

For $\nu > 2$, the induction step $\nu-1 \;\rightarrow\; \nu$ is verified by making use of (A.53):

$$\begin{aligned}&\alpha_0\Omega_\nu + \alpha_1\Omega_\nu L + (\alpha_2 I - M)\,\Omega_\nu L^2\\ &= \underbrace{\big(\alpha_0\Omega_{\nu-1} + \alpha_1\Omega_{\nu-1}L + (\alpha_2 I - M)\,\Omega_{\nu-1}\,L^2\big)}_{= h(\alpha_0 R_{\nu-1} + (\alpha_0 R_{\nu-2} + \alpha_1 R_{\nu-1})L + L^{\nu-1}) \text{ (induction on (A.50),(A.51))}} L + h\big(\alpha_0 R_\nu + \alpha_1 R_\nu\,L + (\alpha_2 I - M)R_\nu L^2\big)\\ &= h\underbrace{\big(\alpha_0 R_{\nu-2} + \alpha_1 R_{\nu-1} + (\alpha_2 I - M)R_\nu\big)}_{=0} L^2 \;+\; h\,(\alpha_0 R_{\nu-1}\,L + L^\nu + \alpha_0 R_\nu + \alpha_1 R_\nu\,L)\\ &= h\big(\alpha_0 R_\nu + (\alpha_0 R_{\nu-1} + \alpha_1 R_\nu)L + L^\nu\big) \;=\; h(L^\nu - \tilde{L}_\nu)\end{aligned} \tag{A.55}$$

with $\tilde{L}_\nu$ from (A.51), which completes the induction for (A.50).

– For $\tilde{L}_\nu$ defined in (A.51), relation (A.52) immediately follows from (2.17). □

References

[1] W. Auzinger, *On error structures and extrapolation for stiff systems, with application in the method of lines*, Computing 44, 331–356, 1990.

[2] W. Auzinger, *Sectorial operators and normalized numerical range*, Appl. Numer. Math. 45, 367–388, 2003.

[3] W. Auzinger, W. Herfort, *A uniform quantitative stiff stability estimate for BDF schemes*, Opuscula Mathematica 26/2, 203–227, 2006.

[4] M. Crouzeix, P. A. Raviart, *Approximation des équations d'évolution linéares par des méthodes à pas multiples*, C. R. Acad. Sc. Paris, Série A 287, 367–370, 1976.

[5] N. Del Buono, A. T. Hill, *On a multistep method approximating a linear sectorial evolution equation*, IMA J. Numer. Anal. 22, 481–499, 2002.

[6] E. Hairer, S. P. Nørsett, G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*, 2nd ed., Springer-Verlag, 1993.

[7] E. Hairer, G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, 2nd ed., Springer-Verlag, 1996.

[8] K. J. in 't Hout, M. N. Spijker, *Analysis of error growth via stability regions in numerical initial value problems*, BIT 43, 363–385, 2003.

[9] M. N. Le Roux, *Semidiscretization in time for parabolic problems*, Math. Comp. 33, 919–931, 1979.

[10] R. J. LeVeque, L. N. Trefethen, *On the resolvent condition in the Kreiss matrix theorem*, BIT 24, 584–591, 1984.

[11] C. Lubich, *Convolution quadrature and discretized operational calculus I*, Numer. Math. 52, 129–145, 1988.

[12] C. Lubich, O. Nevanlinna, *On resolvent conditions and stability estimates*, BIT 31, 293–313, 1991.

[13] M. Miklavčič, *Applied Functional Analysis and Partial Differential Equations*, World Scientific, 1998.

[14] M. N. Spijker, *On a conjecture by Le Veque and Trefethen related to the Kreiss matrix theorem*, BIT 31, 551–555, 1991.

[15] M. N. Spijker, F. A. J. Straetemans, *Error growth analysis via stability regions for discretizations of initial value problems*, Appl. Numer. Math. 37, 442–464, 1997.

[16] L. N. Trefethen, M. Embree, *Spectra and Pseudospectra: The Behavior of Nonnormal Matrices and Operators*, Princeton University Press, 2005.