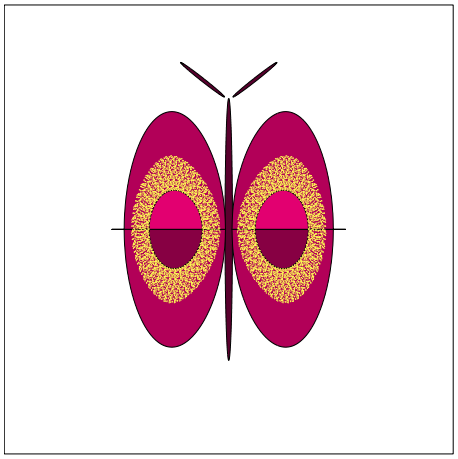
Time Integration of Differential Equations

Habilitation Thesis

Mechthild Thalhammer

Institut für Mathematik
Leopold-Franzens-Universität Innsbruck

Innsbruck, February 2006



It required several years to realise this work.
For support and company during that time
I thank

My parents

My brother

Alexander and Barbara

My colleagues and friends in Innsbruck, in particular,
Gerhard and Günther,
Anna, Margot, Michael, Norbert, Peter and Peter, Reinhard, Thomas

My colleagues and friends abroad, in particular,
Gabriela, Mari Paz, Marlis, Assyr, Cesar, Cesareo, Jitse, Stein, Volker, Will,
Antonella and Hans, Elena and Brynjulf, Evi and Ernst,
Lourdes and Julian, Myriam and Gerhard

My colleagues and friends from the choirs
Kammerchor Collegium vocale Innsbruck, Kammerchor Innsbruck,
Schütz-Obrecht-Ensemble

Susanne, Martina and Philipp, Christoph, Horst

Contents

Preface

Differential equations are fundamental in the description of dynamical processes. The areas of applications include the domains of natural science and engineering technology, finance and medical science.

For instance, hyperbolic partial differential equations such as Schrödinger type equations used in quantum physical models are presently attracting a lot of interest. Examples for parabolic equations are reaction-diffusion equations and the incompressible Navier-Stokes equation which arise in the modelling of air-currents, population models, and circulations.

For the solution of realistic models, due to their complex nature, it is in general indispensable to utilise numerical methods. Though, in particular in connection with computer-aided simulations over long times, the actual result is adulterated by the influence of the discretisation, rounding errors, and inaccuracies in the data. Therefore, this raises the question how to interprete the obtained results and how to draw significant conclusions thereof.

This habilitation thesis comprises contributions to the topic *Time Integration of Differential Equations*. Our main objective is the construction and analysis of time integration methods for stiff differential equations. In particular, our concern is to investigate the error behaviour and the qualitative properties of certain numerical method classes.

> *Around 1960, things became completely different and everyone became aware that the world was full of stiff problems.*
>
> GERMUND DAHLQUIST (1925-2005)

Primarily, the scope of applications includes nonlinear initial-boundary value problems of parabolic type that are often used in the modelling of nonlinear diffusion and heat conduction processes. According to the employed analytical framework, it is expedient to distinguish between semilinear and (fully) nonlinear problems. In this preface, to keep the presentation simple, we restrict ourselves to one space dimension.

Typically, a *semilinear* parabolic initial-boundary value problem for a real-valued function $U : [0,1] \times [0,T] \to \mathbb{R}$ comprises a partial differential equation of the form

$$\partial_t U(x,t) = \mathscr{A}(x)\, U(x,t) + \mathscr{F}\big(t, x, U(x,t), \partial_x U(x,t)\big), \qquad 0 < x < 1, \quad t > 0, \tag{1a}$$

that involves a second-order strongly elliptic differential operator $\mathscr{A} = \alpha\, \partial_{xx} + \beta\, \partial_x + \gamma$ with space-dependent coefficients $\alpha, \beta, \gamma : [0,1] \to \mathbb{R}$ satisfying suitable regularity requirements. In particular, α has to be positive and bounded away from 0. Likewise, the nonlinearity $\mathscr{F}$ is

supposed to be regular in all variables and to fulfill certain growth conditions. The differential equation is further subject to certain boundary and initial conditions. For example, we impose a homogeneous Dirichlet boundary condition and a smooth initial condition

$$U(0,t) = 0 = U(1,t), \qquad U(x,0) = U_0(x), \qquad 0 \le x \le 1, \quad 0 \le t \le T. \tag{1b}$$

The precise assumptions on the equation as well as various applications from physics, biology, and engineering that can be cast into the form (1) are found in HENRY [8].

The following *nonlinear* parabolic initial-boundary value problem arises in detonation theory and describes the displacement of a shock

$$\begin{aligned} \partial_t U(x,t) &= \ln\left(\frac{\exp\big(U(x,t)\,\partial_{xx}U(x,t)\big)-1}{\partial_{xx}U(x,t)}\right) - \frac{1}{2}\big(\partial_x U(x,t)\big)^2, \\ \partial_x U(0,t) = 0 = \partial_x U(1,t), &\quad U(x,0) = U_0(x), \qquad 0 \le x \le 1, \quad 0 \le t \le T, \end{aligned} \tag{2}$$

see LUNARDI [11] and references therein.

> *In 1971, I read the beautiful paper of Kato and Fujita on the Navier-Stokes equations and was delighted to find that, properly viewed, it looked like an ordinary differential equation, and the analysis proceeded in ways familiar for ODEs.*
>
> DAN HENRY, 1981

For the theoretical investigation of parabolic initial-boundary value problems such as (1) and (2), it is useful to consider the partial differential equation as an ordinary differential equation. Also, this approach is advantageous for the construction and analysis of time integration methods for parabolic problems. A survey of the abstract framework, which relies on the theory of sectorial operators and analytic semigroups on Banach spaces, is found in [8, 11].

For instance, in order to formulate (1) as abstract problem on a function space, one defines a linear operator $A : D \subset X \to X$ and a map $f : [0,T] \times V \to X$ through

$$\big(A\,v\big)(x) = \mathscr{A}(x)\,v(x), \quad \big(f(t,v)\big)(x) = \mathscr{F}\big(t,x,v(x),\partial_x v(x)\big), \qquad v \in \mathscr{C}_0^\infty(0,1).$$

Therewith, one obtains the following initial value problem

$$u'(t) = A\,u(t) + f\big(t,u(t)\big), \quad 0 < t \le T, \qquad u(0) \text{ given}, \tag{3}$$

for a function $u : [0,T] \to X$ where $\big(u(t)\big)(x) = U(x,t)$. The boundary condition is reflected in the domain of the unbounded linear operator A. In the above situation, a proper choice for the underlying space is the Hilbert space $X = L^2(0,1)$. Then, it follows $D = H^2(0,1) \cap H_0^1(0,1)$ and $V = H_0^1(0,1)$.

The results in [8, 11] imply that A is a sectorial operator which generates an analytic semigroup $\big(\mathrm{e}^{tA}\big)_{t\ge 0}$ on X. Within the abstract Banach space setting, it is in particular justified to represent the solution of (3) by the variation-of-constants formula

$$u(t) = \mathrm{e}^{tA}\,u(0) + \int_0^t \mathrm{e}^{(t-\tau)A} f\big(\tau,u(\tau)\big)\,\mathrm{d}\tau, \qquad 0 \le t \le T. \tag{4}$$

A basic tool in order to establish the existence and regularity of a local solution of the integral equation (4) and (3), respectively, is Banach's Fixed Point Theorem, see [8, Chapter 3].

In LUNARDI [11], it is demonstrated how the techniques applied in the semilinear case extend to nonlinear initial values problems

$$u'(t) = F\big(t, u(t)\big), \quad 0 < t \le T, \qquad u(0) \text{ given}, \tag{5}$$

that come from a parabolic initial boundary-value problem such as (2). A basic requirement is that the derivative of the function F defining the right-hand side of the differential equation is a sectorial operator. Then, by a modification of the variation-of-constants formula (4), a local solution of (5) is constructed in a space of weighted Hölder-continuous functions.

In many situations, apart from an existence and regularity theory, it is also substantial to study the qualitative behaviour of an evolution equation, that is, the geometric properties of the flow. In this respect, we refer to the monographs [8, 11], where such questions are investigated for semilinear and fully nonlinear parabolic evolution equations.

> *There are at least two ways to combat stiffness. One is to design a better computer, the other, to design a better algorithm.*
>
> HARVARD LOMAX (1922-1999)

When solving numerically a differential equation, it is desirable that the applied method possesses a favourable error behaviour and preserves well certain qualitative features of the original problem. Furthermore, in connection with stiff systems, it is essential that the numerical method has favourable stability properties. For that reason, implicit Runge-Kutta and linear multistep methods are established schemes for the time integration of partial differential equations. In particular, the RadauIIA–methods and the backward differentiation formulas are appropriate integration methods for initial value problems that originate from the spatial discretsation of a parabolic initial-boundary problem. A thorough treatment of numerical methods for stiff differential equations is given in HAIRER AND WANNER [6], see also [1, 5, 7, 10].

> *Although most of these methods*[1] *appear at the moment to be largely of theoretical interest ...*
>
> BYRON EHLE, 1968

In the past few years, exponential integrators have attracted a lot of research interest. Due to their excellent stability properties, they are particularly appealing in situations where the differential equation comes from the spatial discretisation of a partial differential equation. However, as exponential integration methods require the numerical computation of the matrix exponential and related functions, for many years, they were considered as impracticable

[1] The citation originally refers to implicit Runge-Kutta methods.

for problems of large dimension, see MOLER AND VAN LOAN [12]. Recently obtained results give further insight in subspace methods such as Krylov subspace techniques and make it feasible to compute, in an efficient manner, matrix-vector products, see [9] and references therein.

> *Theory without practice cannot survive and dies as quickly as it lives.*
>
> *He who loves practice without theory is like the sailor who boards ship without a rudder and compass and never knows where he may cast.*
>
> LEONARDO DA VINCI (1452-1519)

The present thesis is a collection of contributions aiming at a better understanding of time integration methods for singularly perturbed and abstract parabolic problems. The contributions are unified by the fundamental hypotheses on the problem classes which rely on an abstract Banach space setting of sectorial operators and further by the employed perturbation techniques. Accordingly to the considered numerical method classes, the thesis is divided into two chapters. The first chapter is devoted to established schemes such as implicit Runge-Kutta and linear multistep methods, and the second chapter is concerned with exponential integration methods. In the following, a brief survey of each chapter is given.

The main result in OSTERMANN AND THALHAMMER [13] is a convergence estimate for linearly implicit Runge-Kutta time discretisations of semilinear parabolic problems. In this work, we focus on convergence estimates for equations involving non-smooth initial values. Such error bounds are essential in view of practical examples and have applications in the study of the long-term behaviour of time discretisations. The works GONZÁLEZ, OSTERMANN, PALENCIA, AND THALHAMMER [2] and OSTERMANN AND THALHAMMER [14] are related to my doctoral thesis [18] on the time discretisation of nonlinear parabolic problems by implicit Runge-Kutta methods. In [2, 14], amongst others, the techniques used in [18] are extended to variable stepsizes. In particular, we exploit two different approaches to obtain finite time convergence bounds for implicit Runge-Kutta time discretisations with variable stepsizes. In OSTERMANN, THALHAMMER, AND KIRLINGER [16], we proceed our analysis of nonlinear parabolic problems. As special cases, the considered numerical method class includes the k-step backward differentiation formulas. The core of the work is dedicated to the derivation of stability bounds for variable stepsize linear multistep methods. The work THALHAMMER [19] is concerned with the derivation of a convergence bound for variable stepsize linear multistep methods when applied to a singularly perturbed problem. To this aim, perturbation techniques that have been used in [16] are extended to singularly perturbed systems.

In GONZÁLEZ, OSTERMANN, AND THALHAMMER [3], we construct a second-order explicit exponential integration method for nonautonomous linear problems and study its stability and convergence properties for abstract evolutions equations of parabolic type. This work together with the note [20] provide the basis for GONZÁLEZ AND THALHAMMER [4]. There, we are concerned with the construction and analysis of an explicit exponential integrator for

quasilinear parabolic problems. Such problems are of particular relevance in view of practical applications. The aim of THALHAMMER [21] is to explain the substantial order reduction that is in general encountered when a problem of parabolic type is solved numerically by means of a higher-order commutator-free exponential integrator. In GONZÁLEZ, OSTERMANN, AND WRIGHT [17], we consider a numerical method class that combines the benefits of explicit exponential Runge-Kutta and exponential Adams methods. Within this method class, it is straightforward to construct high-order schemes that possess favourable stability and convergence properties for parabolic problems.

A recent work which is closely related to the theme of Chapter 2 is included in the appendix. In OSTERMANN AND THALHAMMER [15], we are concerned with the positivity of exponential integration methods. Our main result implies that positive exponential integrators of linear multistep type obey an order two barrier.

> *... methods[2] for stiff problems, we are just beginning to explore them ...*
>
> LAWRENCE SHAMPINE, 1985

For the near future, a main objective is to understand well the qualitative behaviour of exponential integration methods for different problem classes including singularly perturbed systems, differential-algebraic problems, Hamiltonian systems, and generalised wave equations. Further, it is intended to find efficient linearisation and error control strategies which help to improve the significance of exponential integrators for practical applications.

[2] The citation originally refers to Runge-Kutta methods.

1. Implicit Runge-Kutta and Multistep Methods

1.1. Non-smooth data error estimates

Non-smooth data error estimates for linearly implicit Runge-Kutta methods

ALEXANDER OSTERMANN AND MECHTHILD THALHAMMER

IMA Journal of Numerical Analysis (2000) 20, 167-184

IMA Journal of Numerical Analysis (2000) **20**, 167–184

Non-smooth data error estimates for linearly implicit Runge–Kutta methods

ALEXANDER OSTERMANN AND MECHTHILD THALHAMMER
Institut für Mathematik und Geometrie, Universität Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria

[Received 14 January 1999 and in revised form 7 May 1999]

Linearly implicit time discretizations of semilinear parabolic equations with non-smooth initial data are studied. The analysis uses the framework of analytic semigroups which includes reaction–diffusion equations and the incompressible Navier–Stokes equations. It is shown that the order of convergence on finite time intervals is essentially one. Applications to the long-term behaviour of linearly implicit Runge–Kutta methods are given.

1. Introduction

When analysing discretizations of parabolic initial boundary value problems, it is not sufficient to consider only smooth initial data. This is partly because such initial data give solutions that keep their smoothness up to the boundaries. They thus require compatibility conditions which are often unrealistic in practical applications. Apart from that, non-smooth data error estimates are an important tool for obtaining long-term error bounds. This has been emphasized by Larsson (1992) and is also reflected in Assumption 3.2 in Stuart's survey article (Stuart 1995). The long-term behaviour of numerical solutions is closely related to the question of whether the continuous dynamics of the problem is correctly represented in its discretization. Suppose, for example, that the continuous problem has an asymptotically stable periodic orbit. Does the discrete dynamical system then possess an asymptotically stable invariant closed curve that lies close to the continuous orbit? The construction of such discrete invariant objects is usually based on fixed-point iteration, see e.g. Alouges & Debussche (1993), van Dorsselaer (1998), van Dorsselaer & Lubich (1999), Lubich & Ostermann (1996). Although the final result itself might be smooth, the single iterates are, in general, not. The whole construction thus relies on non-smooth data error estimates.

In spite of their importance, surprisingly few such estimates can be found in the literature. For time discretizations of linear parabolic problems, non-smooth data error estimates are first given by Le Roux (1979). But only until recently have these estimates been extended to more general problem classes. For semilinear parabolic problems, optimal results for implicit Runge–Kutta methods are given in Lubich & Ostermann (1996); see also the references therein. The corresponding results for multistep methods can be found in van Dorsselaer (1998).

In this paper we derive optimal error bounds for *linearly implicit Runge–Kutta methods*, applied to semilinear parabolic problems with non-smooth initial values. We work in

an abstract Banach space setting of analytic semigroups, given in Henry (1981) and in Pazy (1983). This framework includes reaction–diffusion equations and the incompressible Navier–Stokes equations. The method class is formulated in sufficiently general terms such that it comprises classical Rosenbrock methods as well as extrapolation methods based on the linearly implicit Euler scheme. The latter have proven successful for the time integration of parabolic problems, see Bornemann (1990), Lang (1995), and Nowak (1993).

The present paper is structured as follows. In Section 2 we formulate the analytical framework, and we introduce the numerical method. The main result is stated in Section 3. There we prove that linearly implicit Runge–Kutta methods, when applied to semilinear parabolic problems with non-smooth initial data, converge with order one essentially. Low-order convergence is sufficient for applications to long-term error estimates. We illustrate this in Section 4 where we show that exponentially stable solutions of parabolic problems are uniformly approximated by linearly implicit methods over arbitrarily long time intervals. This result implies stability bounds for certain splitting methods. Under natural assumptions on the nonlinearity, it is possible to improve the convergence result of Section 3. This will be elaborated in Section 5. To keep the paper independent from other work, we have formulated all auxiliary results with an outline of the proofs in Section 6.

Compared to previous work, our convergence proofs are conceptionally simple. We consider the numerical approximation u_n of a linearly implicit Runge–Kutta method to the exact solution $u(t_n)$ as a perturbation of a suitably chosen Runge–Kutta solution $\widetilde{u}_n$. Using the triangular inequality

$$\|u_n - u(t_n)\| \leqslant \|u_n - \widetilde{u}_n\| + \|\widetilde{u}_n - u(t_n)\|,$$

we have to estimate $\|u_n - \widetilde{u}_n\|$. Together with the bounds for $\|\widetilde{u}_n - u(t_n)\|$ from Lubich & Ostermann (1996), we get the desired result. For the reader's convenience, we have collected all the necessary Runge–Kutta bounds in an appendix.

We finally remark that the above approach is not restricted to non-smooth data error estimates. It can equally be used, for example, to derive the conditions for high-order convergence of linearly implicit methods at smooth solutions.

2. Analytical framework and numerical method

In this section we state the assumptions on the evolution equation. Moreover we introduce the numerical method.

2.1 *Evolution equation*

We consider a semilinear parabolic equation of the form

$$u' + Au = f(t, u), \qquad 0 < t \leqslant T \tag{2.1a}$$

$$u(0) = u_0. \tag{2.1b}$$

This abstract evolution equation is given on a Banach space $(X, |\cdot|)$. The domain of the linear operator A on X is denoted by $\mathcal{D}(A)$, and the initial value $u_0 \in V$ is chosen in an interpolation space $\mathcal{D}(A) \subset V \subset X$ which will be specified below. Our basic assumptions on the initial value problem (2.1) are that of Henry (1981).

ASSUMPTION 2.1 Let $A : \mathcal{D}(A) \subset X \to X$ be sectorial, i.e. A is a densely defined and closed linear operator on X satisfying the resolvent condition

$$\left|(\lambda I + A)^{-1}\right|_{X \leftarrow X} \leqslant \frac{M}{|\lambda - \omega|} \tag{2.2}$$

on the sector $\{\lambda \in \mathbb{C}\,;\, |\arg(\lambda - \omega)| \leqslant \pi - \varphi\}$ for $M \geqslant 1$, $\omega \in \mathbb{R}$, and $0 \leqslant \varphi < \frac{\pi}{2}$.

Under this hypothesis, the operator $-A$ is the infinitesimal generator of an analytic semigroup $\{\mathrm{e}^{-tA}\}_{t \geqslant 0}$ which renders (2.1) parabolic. In the sequel we set

$$A_a = A + aI \qquad \text{for some } a > \omega.$$

For this operator, the fractional powers are well defined. We choose $0 \leqslant \alpha < 1$ and define $V = \mathcal{D}(A_a^\alpha)$ which is a Banach space with norm $\|v\| = |A_a^\alpha v|$. Note that this definition does not depend on a, since different choices of a lead to equivalent norms.

We are now ready to give our hypothesis on the nonlinear function f.

ASSUMPTION 2.2 Let $f : [0, T] \times V \to X$ be locally Lipschitz-continuous. Thus there exists a real number $L(R, T)$ such that

$$|f(t_1, v_1) - f(t_2, v_2)| \leqslant L(|t_1 - t_2| + \|v_1 - v_2\|) \tag{2.3}$$

for all $t_i \in [0, T]$ and $\|v_i\| \leqslant R$, $i = 1, 2$.

Reaction–diffusion equations and the incompressible Navier–Stokes equations can be cast into this abstract framework. This is verified in Section 3 of Henry (1981) and in Lubich & Ostermann (1996). For a more general class of reaction–diffusion equations that is included in our framework, we refer to Section 8.4 of Pazy (1983).

We do not distinguish between a norm and its corresponding operator norm. For elements $x = (x_1, \ldots, x_s)$ in a product space, we set $|x| = \max(|x_1|, \ldots, |x_s|)$ and $\|x\| = \max(\|x_1\|, \ldots, \|x_s\|)$, respectively. The norm of linear operators from X^s to V^s is denoted by $\|\cdot\|_{V \leftarrow X}$.

2.2 *Numerical method*

In this paper linearly implicit Runge–Kutta discretizations of parabolic problems are studied. In the sequel we will review these methods in brief. For detailed descriptions, refer to the monographs by Deuflhard & Bornemann (1994), Hairer & Wanner (1996), and Strehmel & Weiner (1992).

A *linearly implicit Runge–Kutta method* with constant stepsize $h > 0$, applied to the initial value problem (2.1), yields an approximation u_n to the value of the solution u at $t_n = nh$ and is given by the internal stages

$$\begin{aligned} U'_{ni} + AU_{ni} &= f(t_n + \alpha_i h, U_{ni}) + hJ_n \sum_{j=1}^{i} \gamma_{ij} U'_{nj} + h\gamma_i g_n \\ U_{ni} &= u_n + h \sum_{j=1}^{i-1} \alpha_{ij} U'_{nj}, \qquad 1 \leqslant i \leqslant s \end{aligned} \tag{2.4a}$$

and the one-step recursion

$$u_{n+1} = u_n + h \sum_{j=1}^{s} b_j U'_{nj}. \tag{2.4b}$$

Here J_n and g_n are approximations to the derivatives of $-Au + f(t, u)$ with respect to the variables u and t

$$J_n \approx -A + D_u f(t_n, u_n), \qquad g_n \approx D_t f(t_n, u_n).$$

The real numbers α_{ij}, γ_{ij}, b_i, α_i, γ_i are the coefficients of the method. We always assume that $\gamma_{ii} > 0$ for all i.

In contrast to fully implicit Runge–Kutta methods, where the numerical approximation is given as the solution of nonlinear equations, u_{n+1} is obtained from u_n by solving only linear equations.

In order to write the numerical method more compactly, we introduce the following matrix and vector notations

$$\Gamma = (\gamma_{ij})_{1 \leqslant i,j \leqslant s}, \qquad \mathcal{A} = (a_{ij})_{1 \leqslant i,j \leqslant s}, \qquad \mathbb{1} = (1, \ldots, 1)^T \in \mathbb{R}^s, \tag{2.5a}$$

where $a_{ij} = \alpha_{ij} + \gamma_{ij}$ with $\alpha_{ij} = 0$ for $i \leqslant j$ and $\gamma_{ij} = 0$ for $i < j$. Further we set

$$\alpha = (\alpha_1, \ldots, \alpha_s)^T, \qquad \gamma = (\gamma_1, \ldots, \gamma_s)^T \tag{2.5b}$$

$$b = (b_1, \ldots, b_s)^T, \qquad c = (c_1, \ldots, c_s)^T = \mathcal{A}\mathbb{1}. \tag{2.5c}$$

The numerical scheme has *order* p if the error of the method, applied to ordinary differential equations with sufficiently differentiable right-hand side, fulfils the relation $u_n - u(t_n) = \mathcal{O}(h^p)$ for $h \to 0$, uniformly on bounded time intervals.

A linearly implicit Runge–Kutta method is $A(\vartheta)$-*stable* if the absolute value of the *stability function*,

$$\mathcal{R}(z) = 1 + zb^T (I - z\mathcal{A})^{-1} \mathbb{1}, \tag{2.6}$$

is bounded by one for all $z \in M_\vartheta = \{z \in \mathbb{C}\,;\, |\arg(-z)| \leqslant \vartheta\}$. Note that $(I - z\mathcal{A})$ is invertible in M_ϑ since all γ_{ii} are positive. The numerical method is called *strongly* $A(\vartheta)$-*stable* if in addition the absolute value of $\mathcal{R}$ at infinity, $\mathcal{R}(\infty) = 1 - b^T \mathcal{A}^{-1} \mathbb{1}$, is strictly smaller than one.

Two types of linearly implicit Runge–Kutta methods are of particular interest. *Rosenbrock methods* satisfy the conditions

$$\alpha_i = \sum_{j=1}^{i-1} \alpha_{ij}, \qquad \gamma_i = \sum_{j=1}^{i} \gamma_{ij} \tag{2.7}$$

and use the exact Jacobians

$$J_n = -A + D_u f(t_n, u_n), \qquad g_n = D_t f(t_n, u_n). \tag{2.8}$$

As a prominent example, we mention the fourth-order method RODAS from Hairer & Wanner (1996). It is strongly $A(\pi/2)$-stable and satisfies $\mathcal{R}(\infty) = 0$.

A second important class of linearly implicit methods is determined by the requirements

$$\alpha_i = \sum_{j=1}^{i-1} \alpha_{ij} + \sum_{j=1}^{i} \gamma_{ij}, \qquad \text{and} \qquad J_n = -A, \quad g_n = 0.$$

In this paper such methods are called *W-methods*. This differs from the common diction in the literature where this term is often used as a synonym for linearly implicit Runge–Kutta methods. The operator A and the nonlinearity f are not determined uniquely, since bounded parts of A, e.g., can be included into f. Therefore the assumption $J_n = -A$ is not as restrictive as it may seem at first. As an example of W-methods, we mention the extrapolated linearly implicit Euler method which is described briefly in Section 3 of Lubich & Ostermann (1995), see also Hairer & Wanner (1996). It is strongly $A(\vartheta)$-stable with $\vartheta \approx \pi/2$ and satisfies $\mathcal{R}(\infty) = 0$.

3. Non-smooth data error estimates

In Theorem 3.1 below we state the main result of this paper. We give a non-smooth data error estimate for a general class of linearly implicit methods. For smooth initial data, their convergence is studied in Lubich & Ostermann (1995), Ostermann & Roche (1993), and Schwitzer (1995).

THEOREM 3.1 Let (2.1) satisfy Assumptions 2.1 and 2.2, and let $u_0 \in V$ be such that the solution u remains bounded in V for $0 \leqslant t \leqslant T$. Apply a strongly $A(\vartheta)$-stable linearly implicit Runge–Kutta method of order at least one with $\vartheta > \varphi$ to this initial value problem, and assume that $\|J_n + A\|_{X \leftarrow V}$ as well as $|g_n|$ are uniformly bounded for $0 \leqslant t_n \leqslant T$. Then there exist constants h_0 and C such that for all stepsizes $0 < h \leqslant h_0$ the numerical solution u_n satisfies the estimate

$$\|u_n - u(t_n)\| \leqslant C\Big(t_n^{-1}h + t_n^{-\alpha}h\,|\log h|\Big) \qquad \text{for } 0 < t_n \leqslant T.$$

The constants h_0 and C depend on T and the bound of u, on the quantities appearing in Assumptions 2.1 and 2.2, and moreover on the numerical method.

This result can be applied directly to W-methods and Rosenbrock methods. For W-methods this is obvious since $J_n = -A$ and $g_n = 0$. For Rosenbrock methods we have to suppose that the first derivatives of the nonlinearity f are locally bounded. Then, due to (2.8), Theorem 3.1 is applicable.

To study the long-term dynamics of the evolution equation (2.1), apart from a non-smooth data error estimate for finite times, an error estimate for the derivative of the solution with respect to the initial value is often needed, see Stuart (1995). This derivative, evaluated at the point u_0, is a linear operator on V and is denoted here by $v(t) = Du(t; u_0)$. Consequently (u, v) satisfies the system

$$\begin{pmatrix} u' \\ v' \end{pmatrix} + \begin{pmatrix} A & 0 \\ 0 & A \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} f(t,u) \\ D_u f(t,u)v \end{pmatrix} \tag{3.1}$$

with initial value $(u_0, v_0)^T$, where v_0 is the identity on V. The derivative of the numerical solution u_n with respect to the initial value is denoted by $v_n = Du_n(u_0)$. It is just the second component of the linearly implicit Runge–Kutta solution of (3.1) at $t_n = nh$.

We are now in a position to state the following result.

COROLLARY 3.1 In addition to the assumptions of Theorem 3.1, let the Fréchet derivative $D_u f(t, u)$ be locally Lipschitz-continuous with respect to the variables t and u, and bounded as a linear operator from V to X, uniformly in t and u. Then there exist constants h_0 and C such that for $0 < h \leqslant h_0$ the estimate

$$\|v_n - v(t_n)\| \leqslant C\Big(t_n^{-1}h + t_n^{-\alpha}h\,|\log h|\Big), \qquad 0 < t_n \leqslant T$$

is satisfied. Apart from the quantities given in Theorem 3.1, the maximum stepsize h_0 and the constant C depend on the Lipschitz constants of $D_u f$.

Proof of Corollary 3.1. Obviously (3.1) satisfies Assumptions 2.1 and 2.2. In order to apply Theorem 3.1, it remains to show that $v(t)$ is bounded by a constant, uniformly for $0 \leqslant t \leqslant T$. By means of the variation-of-constants formula, v can be represented as

$$v(t) = \mathrm{e}^{-tA} + \int_0^t \mathrm{e}^{-(t-\tau)A} D_u f(\tau, u(\tau))\, v(\tau)\, \mathrm{d}\tau,$$

see Henry (1981, Lemma 3.3.2). Applying the estimates given in Lemma 6.3 (see later), the boundeness of v follows from a Gronwall inequality given in Section 1.2.1 of Henry (1981). □

Proof of Theorem 3.1. Our basic idea is to compare the numerical solution, obtained with the linearly implicit method, with the solution of a suitably chosen implicit Runge–Kutta method.

(a) First we apply a linearly implicit method to (2.1) which gives (2.4). In order to write (2.4) more compactly, we employ the following vector notation

$$\begin{aligned} U_n &= (U_{n1}, \ldots, U_{ns})^T, \quad U_n' = (U_{n1}', \ldots, U_{ns}')^T \\ F_n &= \big(f(t_n + \alpha_1 h, U_{n1}), \ldots, f(t_n + \alpha_s h, U_{ns})\big)^T. \end{aligned} \tag{3.2}$$

Together with (2.5) we get

$$U_n' + (\mathcal{I} \otimes A)U_n = F_n + (\Gamma \otimes hJ_n)U_n' + \gamma \otimes hg_n \tag{3.3a}$$

$$U_n = \mathbb{1} \otimes u_n + \big((\mathcal{A} - \Gamma) \otimes hI\big)U_n' \tag{3.3b}$$

$$u_{n+1} = u_n + (b^T \otimes hI)U_n'. \tag{3.3c}$$

Here we have used Kronecker product notation. Thus the (k, m)-th component of $\mathcal{B} \otimes A$, where A is a linear operator and $\mathcal{B}$ an arbitrary matrix with coefficients b_{ij}, is given by $b_{km}A$. For notational simplicity, we write $\mathcal{B} \otimes hA$ instead of $\mathcal{B} \otimes (hA)$. We further distinguish between the identity matrix $\mathcal{I}$ on $\mathbb{R}^s$ and the identity operator I on X or V.

Inserting (3.3b) into (3.3a) and setting $D_n = J_n + A$, we get

$$U_n' = (\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1}\big(-\mathbb{1} \otimes (Au_n) + F_n + (\Gamma \otimes hD_n)U_n' + \gamma \otimes hg_n\big).$$

Together with (3.3c) this yields the recursion

$$u_{n+1} = \mathcal{R}(-hA)\, u_n + (b^T \otimes hI)(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1} \big(F_n + (\Gamma \otimes hD_n)U_n' + \gamma \otimes hg_n\big). \quad (3.4)$$

(b) Next we compare this numerical solution with the following implicit Runge–Kutta discretization

$$\widetilde{U}_n' + (\mathcal{I} \otimes A)\widetilde{U}_n = \widetilde{F}_n \quad (3.5a)$$

$$\widetilde{U}_n = \mathbb{1} \otimes \widetilde{u}_n + (\mathcal{Q} \otimes hI)\widetilde{U}_n' \quad (3.5b)$$

$$\widetilde{u}_{n+1} = \widetilde{u}_n + (b^T \otimes hI)\widetilde{U}_n'. \quad (3.5c)$$

Here we have used the same abbreviations as in (3.2) (replacing U_n with $\widetilde{U}_n$, etc). In particular we set

$$\widetilde{F}_n = \big(f(t_n + c_1 h, \widetilde{U}_{n1}), \ldots, f(t_n + c_s h, \widetilde{U}_{ns})\big)^T \qquad \text{with} \quad c = \mathcal{Q}\mathbb{1}.$$

A similar calculation as before yields

$$\widetilde{u}_{n+1} = \mathcal{R}(-hA)\, \widetilde{u}_n + (b^T \otimes hI)(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1} \widetilde{F}_n. \quad (3.6)$$

(c) The difference between the linearly implicit solution u_n and the Runge–Kutta solution $\widetilde{u}_n$ is denoted by $e_n = u_n - \widetilde{u}_n$. In accordance with that, E_n and E_n' are defined by $E_n = U_n - \widetilde{U}_n$ and $E_n' = U_n' - \widetilde{U}_n'$, respectively. Taking the difference between (3.4) and (3.6) gives

$$e_{n+1} = \mathcal{R}(-hA)\, e_n + (b^T \otimes hI)(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1} \big(F_n - \widetilde{F}_n + (\Gamma \otimes hD_n)U_n' + \gamma \otimes hg_n\big).$$

Solving this recursion yields

$$e_{n+1} = (b^T \otimes hI) \sum_{\nu=0}^{n} (\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1} \big(\mathcal{I} \otimes \mathcal{R}(-hA)^{n-\nu}\big) \big(F_\nu - \widetilde{F}_\nu + (\Gamma \otimes hD_\nu)U_\nu' + \gamma \otimes hg_\nu\big), \quad (3.7)$$

where we have already used the fact that both methods start with the same initial value u_0. Since f is locally Lipschitz-continuous, we have

$$|F_n - \widetilde{F}_n| \leqslant L\big(h \max_{1 \leqslant i \leqslant s} |\alpha_i - c_i| + \|E_n\|\big)$$

for $\|U_n\|,\ \|\widetilde{U}_n\| \leqslant R$. We suppose for a moment that the radius R can be chosen independently of n. This will be justified at the end of the proof.

From the last equation and the uniform boundedness of D_n and g_n we get

$$\|e_{n+1}\| \leqslant Ch \sum_{\nu=0}^{n} \left\| (\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1} \left(\mathcal{I} \otimes \mathcal{R}(-hA)^{n-\nu} \right) \right\|_{V \leftarrow X} \left(\|E_\nu\| + h \left\| U'_\nu \right\| + h \right). \quad (3.8)$$

(d) We now derive several relations that are necessary to bound e_{n+1}. First we consider hU'_n. From (3.3b) and (3.5b) we get

$$(\Gamma \otimes hI)U'_n = \mathbb{1} \otimes e_n + (\mathcal{Q} \otimes hI)E'_n - E_n. \quad (3.9)$$

In order to eliminate E'_n, we multiply (3.9) by $(\mathcal{I} \otimes J_n)$ and insert it into the difference of (3.3a) and (3.5a). This yields

$$(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)E'_n = -\mathbb{1} \otimes (Ae_n) + \mathbb{1} \otimes (D_n e_n) + F_n - \widetilde{F}_n + (\mathcal{Q} \otimes hD_n)E'_n - (\mathcal{I} \otimes D_n)E_n + \gamma \otimes hg_n. \quad (3.10)$$

We multiply this identity by $(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1}$. The existence of this operator is guaranteed by Lemma 6.5. Applying (6.2) to the term involving Ae_n and (6.3) with $\rho = \alpha$ to the remaining expressions gives

$$h\|E'_n\| \leqslant C\,\|e_n\| + Ch^{1-\alpha}\|E_n\| + Ch^{2-\alpha}$$

for h sufficiently small. Together with (3.9) we get

$$h\|U'_n\| \leqslant C\,\|e_n\| + C\,\|E_n\| + Ch^{2-\alpha}. \quad (3.11)$$

It remains to express E_n in terms of e_n. Regrouping (3.9) we have

$$E_n = \mathbb{1} \otimes e_n + ((\mathcal{Q} - \Gamma) \otimes hI)E'_n - (\Gamma \otimes hI)\widetilde{U}'_n. \quad (3.12)$$

In order to estimate $h\widetilde{U}'_n$, we use (3.5b) in the form

$$(\mathcal{Q} \otimes hI)\widetilde{U}'_n = -\mathbb{1} \otimes \widetilde{u}_n + \widetilde{U}_n. \quad (3.13)$$

Each component on the right-hand side of (3.13) can be written as

$$u(t_n) - \widetilde{u}_n + \widetilde{U}_{ni} - u(t_n + c_i h) + \int_{t_n}^{t_n + c_i h} u'(\tau)\,\mathrm{d}\tau. \quad (3.14)$$

Applying the triangular inequality as well as Lemma 6.4 and Lemma A.1 gives

$$h\|\widetilde{U}'_n\| \leqslant C\left(t_n^{-1}h + t_n^{-\alpha}h\,|\log h| \right) \qquad \text{for } n \geqslant 1.$$

Therefore we finally get from (3.12)

$$\|E_n\| \leqslant C\,\|e_n\| + C\left(t_n^{-1}h + t_n^{-\alpha}h\,|\log h| \right) \qquad \text{for } n \geqslant 1. \quad (3.15)$$

Note that $\|E_0\|$ and thus $h\|U_0'\|$ are bounded by a constant.

(e) Inserting (3.11) and (3.15) into (3.8), we obtain with (6.4) and Lemma 6.1

$$\|e_n\| \leqslant Ch \sum_{\nu=1}^{n-1} t_{n-\nu}^{-\alpha} \|e_\nu\| + C\left(t_n^{-1}h + t_n^{-\alpha}h\,|\log h|\right).$$

Applying the discrete Gronwall Lemma 6.2, we get

$$\|u_n - \widetilde{u}_n\| = \|e_n\| \leqslant C\left(t_n^{-1}h + t_n^{-\alpha}h\ |\log h|\right) \tag{3.16}$$

due to linearity. The desired estimate

$$\|u_n - u(t_n)\| \leqslant \|u_n - \widetilde{u}_n\| + \|\widetilde{u}_n - u(t_n)\| \leqslant C\left(t_n^{-1}h + t_n^{-\alpha}h\ |\log h|\right)$$

finally follows with Lemma A.1.

(f) We still have to show that the numerical solution remains in a ball of radius R. Note that the exact solution as well as the Runge–Kutta solution are bounded on $[0, T]$. We take R sufficiently large and choose a smooth cut-off function

$$\chi : V \to [0, 1] \qquad \text{with} \quad \chi(v) = \begin{cases} 1 & \text{if } \|v\| \leqslant R, \\ 0 & \text{if } \|v\| \geqslant 2R. \end{cases}$$

Since $f(t, \chi(u)\cdot u)$ has a global Lipschitz constant, we infer from (3.16) that the numerical solution, obtained with this new f, is bounded by R for h sufficiently small. It thus coincides with the numerical solution, obtained with the original f. This concludes the proof of Theorem 3.1. □

4. Applications

As already mentioned in the introduction, Theorem 3.1 together with Corollary 3.1 can be used to study the question of whether the continuous dynamics of a parabolic equation is correctly represented in its discretization. A result in this direction is given in Lubich & Ostermann (1996) for Runge–Kutta discretizations of periodic orbits. The proof there carries over literally to linearly implicit Runge–Kutta methods. Note that the necessary bounds for smooth initial data are provided by Lubich & Ostermann (1995) and Schwitzer (1995). We do not give the details here.

Another immediate consequence of our non-smooth data error estimates are long-term error bounds. As an illustration, we show below that exponentially stable solutions of (2.1a) are uniformly approximated by linearly implicit Runge–Kutta methods over arbitrarily long time intervals. Our presentation follows an idea of Larsson (1992). Alternatively one might use directly the results of Stuart (1995). A close examination of their proofs shows that they are applicable despite the additional $|\log h|$ term in our error estimate.

We recall that a solution u of (2.1a) is *exponentially stable* if there exist positive constants τ and δ such that any solution v of (2.1a) with initial value $v(\tau_0) \in V$ and $\|u(\tau_0) - v(\tau_0)\| \leqslant \delta$ satisfies

$$\|u(t) - v(t)\| \leqslant \tfrac{1}{2}\|u(\tau_0) - v(\tau_0)\| \qquad \text{for } t \geqslant \tau_0 + \tau. \tag{4.1}$$

This condition holds, for example, in the neighbourhood of an asymptotically stable fixed-point due to its exponential attractivity.

THEOREM 4.1 In addition to the assumptions of Theorem 3.1, let the solution u be exponentially stable and globally bounded. Then, for any choice of $t^* > 0$, there are positive constants C and h_0 such that for all stepsizes $0 < h \leqslant h_0$ we have

$$\|u_n - u(t_n)\| \leqslant Ch\,|\log h| \qquad \text{for } t_n \in [t^*, \infty). \tag{4.2}$$

The constant C depends on t^* and on τ, given by (4.1). Moreover it depends on the quantities appearing in Assumptions 2.1 and 2.2, on the numerical method, and on the bound for the solution.

It is remarkable that (4.2) holds for quite crude approximations to the Jacobian. For example, the choice $J_n = A$ is possible without any assumption on the growth of the semigroup, i.e. on the sign of the constant ω appearing in (2.2).

Proof. Henceforth, the constants δ and τ have the same meaning as in the definition of exponentially stable solutions. Since the solution u is globally bounded in V, it stays in a ball of radius $R/2$, say. We may assume that $t^* \leqslant \tau$ and set $T = 2\tau + t^*$. Then Theorem 3.1 shows the existence of a constant $C^* = C^*(R, t^*, T)$ with

$$\|u_n - u(t_n)\| \leqslant C^* h\,|\log h| \qquad \text{for } t^* \leqslant t_n \leqslant T \text{ and } 0 < h \leqslant h_0. \tag{4.3a}$$

After a possible reduction of δ and h_0, we have $h_0 \leqslant \tau$ and

$$C^* h\,|\log h| \leqslant \delta/2 \leqslant R/4 \qquad \text{for } 0 < h \leqslant h_0. \tag{4.3b}$$

Assume for a moment that the estimate

$$\|u_n - u(t_n)\| \leqslant 2C^* h\,|\log h|\,, \qquad t^* + k\tau < t_n \leqslant t^* + (k+1)\tau$$

holds for some $k \geqslant 2$, and let m be such that $(m-1)h < \tau \leqslant mh$. Further denote by $v_n(t)$ the solution of (2.1a) with initial value $v_n(t_n) = u_n$. From Theorem 3.1 and the exponential stability (4.1), we get

$$\begin{aligned}\|u_{n+m} - u(t_{n+m})\| &\leqslant \|u_{n+m} - v_n(t_{n+m})\| + \|v_n(t_{n+m}) - u(t_{n+m})\| \\ &\leqslant C^* h\,|\log h| + \tfrac{1}{2}\|u_n - u(t_n)\| \leqslant 2C^* h\,|\log h| \leqslant \delta.\end{aligned}$$

The bound (4.2) with $C = 2C^*$ thus follows from (4.3) by induction. □

The above result can also be used to obtain stability bounds for splitting methods. As an example, we consider the linear problem

$$u' + Au = Bu, \qquad u(0) = u_0 \tag{4.4}$$

and its discretization by the linearly implicit Euler method

$$u_{n+1} = (I + hA)^{-1}(I + hB)u_n.$$

Since A and B are treated in a different way, this scheme can be interpreted as a splitting method.

We assume that the operator A satisfies Assumption 2.1 and that B is bounded as an operator from V to X. Thus $B - A$ is the infinitesimal generator of an analytic semigroup on V, see Corollary 1.4.5 of Henry (1981). We further suppose that this semigroup satisfies

$$\left\| \mathrm{e}^{-t(A-B)} \right\| \leqslant C\mathrm{e}^{-\kappa t} \qquad \text{for } t \geqslant 0 \tag{4.5}$$

with some $\kappa > 0$. We then have the following result.

COROLLARY 4.1 Under the above assumptions, for any $\widetilde{\kappa} < \kappa$, there are positive constants C and h_0 such that for $0 < h \leqslant h_0$

$$\left\| \left((I + hA)^{-1}(I + hB) \right)^n \right\| \leqslant C\mathrm{e}^{-\tilde{\kappa} nh} \qquad \text{for } n \geqslant 1.$$

The constant C depends on κ and $\widetilde{\kappa}$, on the quantities appearing in Assumption 2.1, and on $\|B\|_{X \leftarrow V}$.

This proves the stability of the above splitting scheme for sufficiently small stepsizes. We are not aware of any other proof for this result, apart from the case $\alpha = 0$ where B has to be bounded on X.

Proof. For given κ, we choose $0 \leqslant \widetilde{\kappa} < \mu < \kappa$ and consider the equation

$$w' + Aw = \widetilde{B}w \qquad \text{with} \quad \widetilde{B} = \mu I + (1 + h\mu)B.$$

For h sufficiently small, the solutions of this problem are exponentially stable, since there exists some $\varepsilon > 0$ such that

$$\left\| \mathrm{e}^{-t(A-\widetilde{B})} \right\| \leqslant C\mathrm{e}^{-\varepsilon t} \qquad \text{for } t \geqslant 0.$$

This follows from Theorem 3.2.1 of Pazy (1983). Note that $\widetilde{B}$ has a Lipschitz constant that depends on μ, h_0, $\|B\|_{X \leftarrow V}$, but not on h. Due to linearity, we obtain from Theorem 3.1 and Theorem 4.1 the estimate

$$\left\| \left((I + hA)^{-1}(I + h\widetilde{B}) \right)^n \right\| \leqslant C \qquad \text{for } n \geqslant 1.$$

The desired bound finally follows from $I + h\widetilde{B} = (1 + h\mu)(I + hB)$. □

5. Refined error estimate

Theorem 3.1 essentially yields convergence of order one. In this section we show that we can raise the order of convergence under slightly stronger assumptions on the data. To be more specific, we require that f satisfies the following property.

ASSUMPTION 5.1 Let $f : [0, T] \times V \to X$ be locally Lipschitz-continuous with respect to the norms $\|A_a^{-\beta} \cdot \|$ and $|A_a^{-\beta} \cdot |$ for some $0 < \beta < 1 - \alpha$, i.e.

$$\left| A_a^{-\beta} \big(f(t_1, v_1) - f(t_2, v_2) \big) \right| \leqslant L \left(|t_1 - t_2| + \|A_a^{-\beta}(v_1 - v_2)\| \right) \tag{5.1}$$

for all $t_i \in [0, T]$ and $v_i \in V$ with $\|v_i\| \leqslant R$, $i = 1, 2$. Further suppose that the first- and second-order derivatives of f are locally Lipschitz bounded with respect to these norms also.

Let $X^\rho = \mathcal{D}(A_a^\rho)$ for $\rho > 0$, and let $X^{-\rho}$ denote the completion of X with respect to the norm $|A_a^{-\rho} \cdot |$. We require that the approximations D_n to the Jacobian $D_u f(t_n, u_n)$ are uniformly bounded as mappings from $X^{\alpha-\beta}$ to $X^{-\beta}$, i.e.

$$\|A_a^{-\beta} D_n A_a^\beta\|_{X \leftarrow V} \leqslant C \qquad \text{for } 0 \leqslant t_n < T, \tag{5.2}$$

with the same β as above.

For the convenience of the reader, we recall that the coefficients of a Rosenbrock method of order $p = 2$ satisfy

$$b^T \mathbb{1} = 1 \qquad \text{and} \qquad b^T \mathcal{O} \mathbb{1} = b^T c = \tfrac{1}{2}, \tag{5.3a}$$

whereas general linearly implicit Runge–Kutta methods of order two further fulfil the order conditions

$$b^T \alpha = \tfrac{1}{2}, \qquad b^T \gamma = 0, \qquad b^T \Gamma \mathbb{1} = 0. \tag{5.3b}$$

We are now in a position to state the refined error estimate.

THEOREM 5.1 In addition to the assumptions of Theorem 3.1, let Assumption 5.1 and (5.2) hold. Further, suppose that the method (2.4) has order $p \geqslant 2$. Then there exist constants h_0 and C such that for all stepsizes $0 < h \leqslant h_0$ the numerical solution u_n satisfies the estimate

$$\|u_n - u(t_n)\| \leqslant C\, t_n^{-1-\beta} h^{1+\beta} \qquad \text{for } 0 < t_n \leqslant T.$$

The constants h_0 and C depend on the constants appearing in Assumption 5.1 and in (5.2), as well as on the quantities specified in Theorem 3.1.

Proof. This proof is an extension of the proof of Theorem 3.1. We thus concentrate on those aspects that go beyond that proof.

(a) We have to estimate the difference $F_n - \widetilde{F}_n$ in (3.7) more carefully. Taylor series expansion gives

$$F_n - \widetilde{F}_n = (\alpha - c) \otimes h D_t f(t_n, u_n) + \big(\mathcal{I} \otimes D_u f(t_n, u_n)\big) E_n + \Delta_n. \tag{5.4}$$

We note for later use that the remainder Δ_n is bounded by

$$\big\|\big(\mathcal{I} \otimes A_a^{-\beta}\big)\, \Delta_n\big\| \leqslant C\, \|e_n\| + C\, t_n^{\beta-1} h^{1+\beta}. \tag{5.5}$$

This follows from Assumption 5.1, the preliminary bound $\|e_n\| \leqslant C$, and

$$\begin{aligned} h\, \big\|\big(\mathcal{I} \otimes A_a^{-\beta}\big)\, E_n'\big\| &\leqslant C\Big(h^\beta \|e_n\| + t_n^{\beta-1} h^{1+\beta}\Big) \\ h\, \big\|\big(\mathcal{I} \otimes A_a^{-\beta}\big)\, U_n'\big\| + \big\|\big(\mathcal{I} \otimes A_a^{-\beta}\big)\, E_n\big\| &\leqslant C\Big(\|e_n\| + t_n^{\beta-1} h\Big). \end{aligned} \tag{5.6}$$

The boundedness of e_n is an immediate consequence of Theorem 3.1, whereas (5.6) is obtained in a similar way to the bound for $h\|E_n'\|$. Using (5.1) and (5.2), we get from (3.10)

$$h\|(\mathcal{I} \otimes A_a^{-\beta}) E_n'\| \leqslant C\left(h^\beta \|e_n\| + h^{1-\alpha} \|(\mathcal{I} \otimes A_a^{-\beta}) E_n\| + h^{2-\alpha} |A^{-\beta} g_n|\right),$$

and since $\alpha + \beta < 1$, this implies

$$h\|(\mathcal{I} \otimes A_a^{-\beta})E_n'\| \leqslant C\left(h^\beta\|e_n\| + h^\beta\|(\mathcal{I} \otimes A_a^{-\beta})E_n\| + h^{1+\beta}\right). \tag{5.7}$$

Further, a direct estimate of (3.12) gives

$$\|(\mathcal{I} \otimes A_a^{-\beta})E_n\| \leqslant C\Big(\|e_n\| + h\|(\mathcal{I} \otimes A_a^{-\beta})E_n'\| + h\|(\mathcal{I} \otimes A_a^{-\beta})\widetilde{U}_n'\|\Big).$$

We thus have to bound $h(\mathcal{I} \otimes A_a^{-\beta})\widetilde{U}_n'$. Using Lemma 6.4 and (A.4), we obtain from (3.13)

$$h\|(\mathcal{I} \otimes A_a^{-\beta})\widetilde{U}_n'\| \leqslant C\, t_n^{\beta-1}h.$$

Reinserting this bound into the above estimates together with (3.9) finally gives (5.6).

(b) We first give the proof for Rosenbrock methods. Recall that in this case, the identities $D_n = D_u f(t_n, u_n)$ and $g_n = D_t f(t_n, u_n)$ as well as $\alpha + \gamma = c$ hold. The latter follows from (2.7) and (2.5). From (5.4) we obtain with (3.9)

$$F_n - \widetilde{F}_n + (\Gamma \otimes hD_n)U_n' + \gamma \otimes hg_n = \mathbb{1} \otimes (D_n e_n) + (\mathcal{A} \otimes hD_n)E_n' + \Delta_n. \tag{5.8}$$

We now insert this relation into (3.7) and start to estimate the recursion more carefully. For this we denote the left-hand side of (5.8) by x and the operator on the right-hand side of (3.7) by B. Using

$$|A_a^\alpha Bx| \leqslant |A_a^{\alpha+\beta}B| \cdot |A_a^{-\beta}x|$$

together with (5.5), (5.6) and Lemma 6.5, we obtain

$$\|e_n\| \leqslant Ch\sum_{\nu=1}^{n-1} t_{n-\nu}^{-\alpha-\beta}\|e_\nu\| + C\, t_n^{\beta-1}h^{1+\beta}. \tag{5.9}$$

The discrete Gronwall Lemma 6.2 and the corresponding bound (A.3) for Runge–Kutta methods finally yield the desired result.

(c) For general linearly implicit Runge–Kutta methods, the identities $D_n = D_u f(t_n, u_n)$ and $g_n = D_t f(t_n, u_n)$ are not necessarily valid. Instead, we have to use the additional order conditions (5.3), combined with an elimination process. We illustrate this with the term $(\alpha - c) \otimes hD_t f(t_n, u_n)$ from (5.4). Inserted in (3.7), it gives

$$(b^T \otimes hI)\sum_{\nu=0}^{n}(\mathcal{I} \otimes I + \mathcal{A} \otimes hA)^{-1}\left(\mathcal{I} \otimes \mathcal{R}(-hA)^{n-\nu}\right)(\alpha - c) \otimes hD_t f(t_\nu, u_\nu) \tag{5.10}$$

where a direct estimate would only give order one. We first split

$$\mathcal{R}(-hA)^n = r^n + \left(\mathcal{R}(-hA)^n - r^n\right) \qquad \text{with } r = \mathcal{R}(\infty).$$

The term with r^n can be estimated as in the proof of Theorem 3.1. Since $|r| < 1$, we get an additional factor $h^{1-\alpha}$ and hence the desired factor h^β. For the second term, we use the identity

$$(b^T \otimes I)(\mathcal{I} \otimes I + \mathcal{A} \otimes hA)^{-1} = b^T \otimes I - (b^T\mathcal{A} \otimes hA)(\mathcal{I} \otimes I + \mathcal{A} \otimes hA)^{-1}$$

together with the order conditions (5.3). This yields

$$\begin{aligned}&\big\|(b^T \otimes I)(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1}\big((\alpha - c) \otimes \big(\mathcal{R}(-hA)^n - r^n\big)\big)\big\|_{V \leftarrow X} \\ &\qquad \leqslant Ch^\beta \big|\big(b^T \mathcal{Q} \otimes (hA_a)^{1-\beta}\big)(\mathcal{I} \otimes I + \mathcal{Q} \otimes hA)^{-1}\big| \cdot \big|A_a^{\alpha+\beta}\big(\mathcal{R}(-hA)^n - r^n\big)\big|.\end{aligned}$$

An application of Lemma 6.5 thus shows that (5.10) gives a contribution of order $h^{1+\beta}$. The other terms in (3.7) are treated similarly and we again obtain (5.9). This concludes the proof. □

6. Lemmas for Section 3 and Section 5

In this section we collect several results that we have used in the proofs of Theorem 3.1, Corollary 3.1, and Theorem 5.1. We start with a discrete convolution of weakly singular functions.

LEMMA 6.1 For $n \in \mathbb{N}$ and $h > 0$, let $t_n = nh$. Then the following relation holds for $0 \leqslant \rho < 1$

$$h \sum_{\nu=1}^{n-1} t_{n-\nu}^{-\rho} t_\nu^{-\sigma} \leqslant \begin{cases} C t_n^{1-\rho-\sigma} & \text{for } 0 \leqslant \sigma < 1, \\ C t_n^{-\rho} \, |\log h| & \text{for } \sigma = 1, \\ C t_n^{1-\rho-\sigma} n^{\sigma-1} & \text{for } \sigma > 1. \end{cases}$$

Proof. We interpret the left-hand side as a Riemann-sum and estimate it by the corresponding integral. □

An integrable function $\varepsilon : [0, T] \to \mathbb{R}$ with the property

$$0 \leqslant \varepsilon(t) \leqslant a \int_0^t (t - \tau)^{-\rho} \varepsilon(\tau) \, \mathrm{d}\tau + b \, t^{-\sigma} \qquad \text{for } 0 \leqslant \rho, \sigma < 1$$

fulfils the estimate $0 \leqslant \varepsilon(t) \leqslant C t^{-\sigma}$, see Section 1.2.1 of Henry (1981). We next formulate a discrete version of this Gronwall lemma.

LEMMA 6.2 For $h > 0$ and $T > 0$, let $0 \leqslant t_n = nh \leqslant T$. Further assume that the sequence of non-negative numbers ε_n satisfies the inequality

$$\varepsilon_n \leqslant ah \sum_{\nu=1}^{n-1} t_{n-\nu}^{-\rho} \varepsilon_\nu + b \, t_n^{-\sigma}$$

for $0 \leqslant \rho < 1$ and $a, b \geqslant 0$. Then the following estimate holds

$$\varepsilon_n \leqslant \begin{cases} C b \, t_n^{-\sigma} & \text{for } 0 \leqslant \sigma < 1, \\ C b \big(t_n^{-1} + t_n^{-\rho} \, |\log h|\big) & \text{for } \sigma = 1, \end{cases}$$

where the constant C depends on ρ, σ, a, and on T.

Proof. This can be shown by using similar arguments as in the proof of Theorem 1.5.5 in Brunner & van der Houwen (1986). We omit the details. □

For the remainder of this section, we suppose that the assumptions of Theorem 3.1 are fulfiled. In particular we have $0 \leqslant \alpha < 1$.

LEMMA 6.3 The analytic semigroup e^{-tA} satisfies the bound

$$\left|A_a^\rho \mathrm{e}^{-tA}\right| \leqslant C\mathrm{e}^{-at}t^{-\rho} \qquad \text{for } t > 0 \text{ and } \rho \geqslant 0.$$

Proof. This is Theorem 1.4.3 of Henry (1981). □

LEMMA 6.4 Let u denote the solution of (2.1) with initial value $u_0 \in V$, and let $0 \leqslant \rho \leqslant 1$. Then the derivative of u with respect to t satisfies the estimate

$$\|A_a^{-\rho}u'(t)\| \leqslant Ct^{\rho-1} \qquad \text{for } 0 < t \leqslant T.$$

Proof. For $\alpha - \rho \geqslant 0$ this bound is given in Theorem 3.5.2 of Henry (1981). In the remaining case, it follows from the identity

$$\begin{aligned} A_a^{\alpha-\rho}u'(t) = & -A_a^{1-\rho}\mathrm{e}^{-tA}\cdot A_a^\alpha u_0 - \int_0^t A_a^{1+\alpha-\rho}\mathrm{e}^{-(t-\tau)A}f(\tau,u(\tau))\,\mathrm{d}\tau \\ & + A_a^{\alpha-\rho}f(t,u(t)) \end{aligned}$$

and Lemma 6.3. □

We close this section with some estimates for the numerical discretization.

LEMMA 6.5 Under the assumptions of Theorem 3.1, the following bounds hold for $0 \leqslant \rho < 1$ and $0 < nh \leqslant T$

$$\left|A_a^\rho\left(\mathcal{R}(-hA)^n - \mathcal{R}(\infty)^n\right)\right| \leqslant Ct_n^{-\rho}, \tag{6.1}$$

$$\left\|(\mathcal{I}\otimes hA)(\mathcal{I}\otimes I + \mathcal{Q}\otimes hA)^{-1}\right\| \leqslant C, \tag{6.2}$$

$$\left|(\mathcal{I}\otimes A_a^\rho)(\mathcal{I}\otimes I + \mathcal{Q}\otimes hA)^{-1}\right| \leqslant Ch^{-\rho}, \tag{6.3}$$

$$\left|(\mathcal{I}\otimes A_a^\rho)(\mathcal{I}\otimes I + \mathcal{Q}\otimes hA)^{-1}\left(\mathcal{I}\otimes \mathcal{R}(-hA)^n\right)\right| \leqslant Ct_n^{-\rho}. \tag{6.4}$$

Proof. These estimates are standard. They follow from the resolvent condition (2.2) and the interpolation result (see Theorem 1.4.4 in Henry 1981)

$$\left|A_a^\rho(\lambda I + A)^{-1}\right| \leqslant C\left|A(\lambda I + A)^{-1}\right|^\rho \cdot \left|(\lambda I + A)^{-1}\right|^{1-\rho}$$

together with the Cauchy integral formula. Similar bounds are given in Lemma 2.3 of Lubich & Ostermann (1993), and in Section 3 of Nakaguchi & Yagi (1997). Note that (6.1) can also be derived from the proof of Theorem 3.5 in Lubich & Nevanlinna (1991). □

Acknowledgement

We thank the anonymous referee for his comments that helped to improve the presentation of this paper.

REFERENCES

ALOUGES, F. & DEBUSSCHE, A. 1993 On the discretization of a partial differential equation in the neighborhood of a periodic orbit. *Numer. Math.* **65**, 143–175.

BORNEMANN, F. 1990 An adaptive multilevel approach to parabolic equations. I. General theory and 1D implementation. *Impact Comput. Sci. Eng.* **2**, 279–317.

BRUNNER, H. & VAN DER HOUWEN, P. J. 1986 *The numerical solution of Volterra equations*. Amsterdam: North-Holland.

DEUFLHARD, P. & BORNEMANN, F. 1994 *Numerische Mathematik II. Integration Gewöhnlicher Differentialgleichungen*. Berlin: de Gruyter.

HAIRER, E. & WANNER, G. 1996 *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, 2nd revised edn. Berlin: Springer.

HENRY, D. 1981 *Geometric Theory of Semilinear Parabolic Equations* (*LNM 840*). Berlin: Springer.

LANG, J. 1995 Two-dimensional fully adaptive solutions of reaction–diffusion equations. *Appl. Numer. Math.* **18**, 223–240.

LARSSON, S. 1992 Nonsmooth data error estimates with applications to the study of the long-time behavior of finite element solutions of semilinear parabolic problems. *Report 1992–36*, Department of Mathematics, Chalmers University, Göteborg.

LE ROUX, M.-N. 1979 Semidiscretization in time for parabolic problems. *Math. Comput.* **33**, 919–931.

LUBICH, CH. & NEVANLINNA, O. 1991 On resolvent conditions and stability estimates. *BIT* **31**, 293–313.

LUBICH, CH. & OSTERMANN, A. 1993 Runge–Kutta methods for parabolic equations and convolution quadrature. *Math. Comput.* **60**, 105–131.

LUBICH, CH. & OSTERMANN, A. 1995 Linearly implicit time discretization of non-linear parabolic equations. *IMA J. Numer. Anal.* **15**, 555–583.

LUBICH, CH. & OSTERMANN, A. 1996 Runge–Kutta time discretization of reaction–diffusion and Navier–Stokes equations: nonsmooth-data error estimates and applications to long-time behaviour. *Appl. Numer. Math.* **22**, 279–292.

NAKAGUCHI, E. & YAGI, A. 1997 Error estimates of implicit Runge–Kutta methods for quasilinear abstract equations of parabolic type in Banach spaces. *Preprint*, Osaka University.

NOWAK, U. 1993 Adaptive Linienmethoden für parabolische Systeme in einer Raumdimension. *Technical Report TR 93-14*, Konrad-Zuse-Zentrum Berlin.

OSTERMANN, A. & ROCHE, M. 1993 Rosenbrock methods for partial differential equations and fractional orders of convergence. *SIAM J. Numer. Anal.* **30**, 1084–1098.

PAZY, A. 1983 *Semigroups of Linear Operators and Applications to Partial Differential Equations*. New York: Springer.

SCHWITZER, F. 1995 W-methods for semilinear parabolic equations. *Appl. Numer. Math.* **18**, 351–366.

STREHMEL, K. & WEINER, R. 1992 *Linear-implizite Runge–Kutta Methoden und ihre Anwendungen*. Stuttgart: Teubner.

STUART, A. 1995 Perturbation theory for infinite dimensional dynamical systems. *Theory and Numerics of Ordinary and Partial Differential Equations* (*Advances in Numerical Analysis IV*) (M. Ainsworth, J. Levesley, W. A. Light and M. Marletta, eds). Oxford: Clarendon.

VAN DORSSELAER, J. 1998 Inertial manifolds under multistep discretization. *Preprint*, Universität Tübingen.

VAN DORSSELAER, J. & LUBICH, CH. 1999 Inertial manifolds of parabolic differential equations under higher-order discretizations. *IMA J. Numer. Anal.* **19**, 455.

Appendix A: Error estimates for Runge–Kutta methods

The present analysis relies strongly on non-smooth data error estimates for Runge–Kutta methods. For the convenience of the reader, we recall the convergence results from Lubich & Ostermann (1996).

For a given linearly implicit Runge–Kutta method of order p, we consider the corresponding *Runge–Kutta* discretization of (2.1)

$$\begin{aligned} &\widetilde{U}'_{ni} + A\widetilde{U}_{ni} = f(t_n + c_i h, \widetilde{U}_{ni}) \\ &\widetilde{U}_{ni} = \widetilde{u}_n + h\sum_{j=1}^{s} a_{ij}\widetilde{U}'_{nj}, \qquad \widetilde{u}_{n+1} = \widetilde{u}_n + h\sum_{j=1}^{s} b_j \widetilde{U}'_{nj} \end{aligned} \tag{A.1}$$

with the coefficients a_{ij}, b_i, c_i as in (2.5). This diagonally implicit Runge–Kutta method enjoys the following properties: It has order p, since the order conditions for Runge–Kutta methods form a subset of those for linearly implicit methods. Due to $c = \mathcal{O}\mathbb{1}$, its stage order q is at least one. Moreover it has the same stability function and thus the same linear stability properties as the underlying linearly implicit method. The existence of the Runge–Kutta solution for $A(\vartheta)$-stable methods follows from Theorem 2.1 in Lubich & Ostermann (1996).

In Section 3 we have used the subsequent convergence result.

LEMMA A.1 Under the assumptions of Theorem 3.1, the following estimate holds for $0 < h \leqslant h_0$ and $0 < t_n \leqslant T$

$$\|\widetilde{u}_n - u(t_n)\| + \|\widetilde{U}_{ni} - u(t_n + c_i h)\| \leqslant C\left(t_n^{-1}h + t_n^{-\alpha}h\,|\log h|\right).$$

For $n = 0$ the same bound holds as for $n = 1$. The constants C and h_0 depend on the quantities specified in Theorem 3.1.

Proof. This result is a sharper version of Theorem 2.1 in Lubich & Ostermann (1996). It follows from (4.15) of *loc. cit.* with $r = \min(p, q+1) \geqslant 1$. Note that the first iterate of the fixed-point iteration is not given correctly there. In the fourth line above formula (4.15) of *loc. cit.*, it should read

$$U_{ni}^{(1)} = X_{ni} + Y_{ni} + d_{ni} \qquad \text{with} \qquad d_n = h\sum_{\nu=0}^{n} W_{n-\nu}\big(F_\nu(U_\nu^{(0)}) - G_\nu\big).$$

Using the Lipschitz condition for f, the bound then follows from Lemmas 4.2 and 4.3 of *loc. cit.* □

Under the assumptions of Theorem 5.1, a refinement of Lemma A.1 is possible. For this we note that the function $g(t) = f(t, u(t))$ satisfies

$$|A_a^{-\beta} g'(t)| \leqslant K t^{\beta-1}, \qquad 0 < t \leqslant T. \tag{A.2}$$

This follows from Assumption 5.1 and Lemma 6.4. We are now in a position to state this refinement.

LEMMA A.2 Under the assumptions of Theorem 5.1, the following estimates hold for $0 < h \leqslant h_0$ and $0 < t_n \leqslant T$

$$\|\widetilde{u}_n - u(t_n)\| + \|\widetilde{U}_{ni} - u(t_n + c_i h)\| \leqslant C\left(t_n^{-2} h^2 + t_n^{-\alpha-\beta} h^{1+\beta}\right), \tag{A.3}$$

$$\left\|A_a^{-\beta}\left(\widetilde{u}_n - u(t_n)\right)\right\| + \left\|A_a^{-\beta}\left(\widetilde{U}_{ni} - u(t_n + c_i h)\right)\right\| \leqslant C\, t_n^{\beta-1} h. \tag{A.4}$$

For $n = 0$ the same bounds hold as for $n = 1$. The constants C and h_0 depend on the quantities specified in Theorem 5.1.

Proof. This lemma is a sharper version of Theorem 2.3 of Lubich & Ostermann (1996). The bound (A.3) follows essentially from Lemma 4.4 of *loc. cit.* There, a similar result is proved under an additional assumption on $g''(t)$ which enters the estimate of $E_h g_\delta(t)$. Since we use here only information on $g'(t)$, we have to estimate this term differently. We proceed as in the proof of Lemma 4.3 of *loc. cit.* and split the integral

$$\int_0^t \left\|E_h \mathbf{1}(t-\tau) g_\delta'(\tau)\right\| \mathrm{d}\tau \leqslant \int_0^{t/2} \left\|E_h \mathbf{1}(t-\tau)\right\|_{X^\alpha \leftarrow X^{-\beta}} \left|A_a^{-\beta} g_\delta'(\tau)\right| \mathrm{d}\tau + \int_{t/2}^t \left\|E_h \mathbf{1}(t-\tau)\right\|_{X^\alpha \leftarrow X} \left|g_\delta'(\tau)\right| \mathrm{d}\tau.$$

The desired result

$$\|E_h g_\delta(t_n)\| \leqslant C t_n^{-\alpha-\beta} h^{1+\beta}$$

then follows from (A.2) and the bounds

$$\|E_h \mathbf{1}(t)\|_{X^\alpha \leftarrow X} \leqslant C \min\left(t^{-1-\alpha} h^2, h^{1-\alpha}\right)$$

$$\|E_h \mathbf{1}(t)\|_{X^\alpha \leftarrow X^{-\beta}} \leqslant C \min\left(t^{-1-\alpha-\beta} h^2, h^{1-\alpha-\beta}\right)$$

for $0 \leqslant t \leqslant T$. Since $r = \min(p, q+1) \geqslant 2$, we obtain (A.3) as in Lubich & Ostermann (1996).

In order to verify (A.4), we consider the Runge–Kutta discretization of

$$x' + Ax = 0, \qquad x(0) = u_0.$$

The proof of Theorem 1.2 in Le Roux (1979) shows that

$$\left\|A_a^{-\beta}\left(\widetilde{x}_n - x(t_n)\right)\right\| + \left\|A_a^{-\beta}\left(\widetilde{X}_{ni} - x(t_n + c_i h)\right)\right\| \leqslant C\, t_n^{\beta-2} h^2,$$

where x_n denotes the Runge–Kutta approximation to $x(t_n)$ and $\widetilde{X}_{ni}$ the corresponding stage values. With this bound at hand, the desired result then follows as in Lemma A.1. □

1.2. Time discretization of nonlinear parabolic problems

Backward Euler discretization of fully nonlinear parabolic problems

CÉSAREO GONZÁLEZ, ALEXANDER OSTERMANN, CESAR PALENCIA, AND MECHTHILD THALHAMMER

Mathematics of Computation (2001) 71, 125-145

MATHEMATICS OF COMPUTATION
Volume 71, Number 237, Pages 125–145
S 0025-5718(01)01330-8
Article electronically published on July 22, 2001

BACKWARD EULER DISCRETIZATION OF FULLY NONLINEAR PARABOLIC PROBLEMS

C. GONZÁLEZ, A. OSTERMANN, C. PALENCIA, AND M. THALHAMMER

Abstract. This paper is concerned with the time discretization of nonlinear evolution equations. We work in an abstract Banach space setting of analytic semigroups that covers fully nonlinear parabolic initial-boundary value problems with smooth coefficients. We prove convergence of variable stepsize backward Euler discretizations under various smoothness assumptions on the exact solution. We further show that the geometric properties near a hyperbolic equilibrium are well captured by the discretization. A numerical example is given.

1. Introduction

Within the past several years, nonlinear evolution equations of parabolic type have attracted a lot of interest, both in theory and applications. This is due to the fact that such equations are increasingly used for the description of processes involving nonlinear diffusion or heat conduction. As examples we mention reaction-diffusion equations that arise in combustion modeling, the Bellman equations from stochastic control and the nonlinear Cahn-Hilliard equation from pattern formation in phase transitions. Further examples are semilinear problems with moving boundaries, such as the Stefan problem that describes the melting of ice.

The knowledge about stability and convergence for time discretizations of nonlinear parabolic problems has also increased considerably. For Runge-Kutta discretizations of semilinear problems, asymptotically sharp error bounds are given in [9]. Optimal convergence results for quasilinear problems in Hilbert spaces can be found in [10], whereas the papers [5] and [13] deal with stability and convergence of quasilinear problems in Banach spaces. Convergence of linearly implicit Runge-Kutta methods for nonlinear parabolic problems is studied in [11]; corresponding results for multistep discretizations can be found in [1] and [8]. For the fully nonlinear situation, however, not that much is known. A reason for this might be that the analytical frameworks for fully nonlinear equations are often quite involved.

The present paper is based on a new and simple framework, given in [12], that extends ideas from the semilinear case to the fully nonlinear one. This is done as

Received by the editor January 6, 2000.
2000 *Mathematics Subject Classification.* Primary 65M12, 65M15; Secondary 35K55, 35R35, 65L06, 65L20.
Key words and phrases. Nonlinear parabolic problems, time discretization, backward Euler method, convergence estimates, stability bounds, asymptotic stability, hyperbolic equilibrium.
The authors acknowledge financial support from Acciones Integradas Hispano-Austríacas 1998/99.

follows. Consider a parabolic evolution equation

$$u' = F(u), \quad t > 0, \qquad u(0) \text{ given}, \tag{1.1}$$

on a Banach space X. The nonlinearity F is defined on an open subset $\mathcal{D}$ of a second Banach space $D \subset X$ and takes values in X. By linearizing F around a state $u^* \in \mathcal{D}$, equation (1.1) takes the form of a semilinear problem

$$u' = Au + f(u), \quad t > 0, \qquad u(0) \in \mathcal{D},$$

where A is a bounded operator from D to X. Under the assumption that A generates an analytic semigroup, we have a (formal) representation of the solution u by the variation-of-constants formula

$$u(t) = \mathrm{e}^{tA}u(0) + \int_0^t \mathrm{e}^{(t-\tau)A} f(u(\tau))\, \mathrm{d}\tau, \qquad 0 \le t \le T. \tag{1.2}$$

Since $f(u(t))$ is only defined for $u(t) \in \mathcal{D}$, we have to consider the semiflow in D. But as the analytic semigroup $\mathrm{e}^{tA} : X \to D$ behaves like Ct^{-1}, the integral on the right-hand side might not exist in D. Consequently (1.2) cannot be used directly.

This is quite different to the semilinear case where intermediate spaces V between X and D are considered. There, under the assumption that the function f is locally Lipschitz continuous from V to X, a unique local solution can be constructed by a fixed-point iteration relying on formula (1.2) (see [7] and [15]).

It turns out that the following slight modification of the variation-of-constants formula

$$u(t) = \mathrm{e}^{tA}u(0) + \int_0^t \mathrm{e}^{(t-\tau)A}\big(f(u(\tau)) - f(u(t))\big)\, \mathrm{d}\tau + \int_0^t \mathrm{e}^{\tau A}\, \mathrm{d}\tau\, f(u(t)) \tag{1.3}$$

is the basic tool for the analysis of fully nonlinear equations. Within the space of α-Hölder continuous functions this relation has a precise meaning and is used to prove existence and uniqueness of a local solution (see [12, Section 8]).

The aim of the present paper is to derive existence and convergence results for time discretizations of (1.1). To keep this exposition in a reasonable length and to avoid technical details, we restrict our attention to the backward Euler method, but we allow variable stepsizes. The extension to strongly $A(\phi)$-stable Runge-Kutta methods with constant stepsizes will be given in [17]. To our knowledge, this is the first paper that provides rigorous error bounds for variable stepsize discretizations of nonlinear parabolic problems. The proofs are based on a global representation of the numerical method by means of a discrete variation-of-constants formula similar to (1.3).

In Section 2 we give the precise assumptions on the initial value problem (1.1) and we present two examples of nonlinear parabolic initial-boundary value problems that fit into this analytical framework. Besides, we introduce spaces of α-Hölder continuous sequences on which our discrete framework is based.

Section 3 deals with the existence and uniqueness of the numerical solution, and with convergence. More precisely, we prove in Theorem 3.3 the expected convergence of order one for constant stepsize discretizations of sufficiently smooth solutions. For variable stepsizes and/or less regular solutions, we show convergence of reduced order.

In Section 4 we study the question whether the dynamics of the analytical problem is well captured by the discretization. As an illustration, we consider exponentially stable equilibria and show that the numerical solution locally exists for

all positive times and decays exponentially towards the equilibrium. A numerical experiment that is in line with our theoretical result is presented.

The auxiliary results for Sections 3 and 4 are finally given in Section 5.

2. Analytical framework and examples

In this section we give the precise hypotheses for (1.1) and further introduce some notation that will be used throughout the paper.

We work in the analytical framework given by [12]. Let $(X, \|\cdot\|)$ and $(D, \|\cdot\|_D)$ be two Banach spaces with D densely embedded in X, and denote by $\mathcal{D}$ an open subset of D. We consider the abstract initial value problem

$$u'(t) = F(u(t)), \quad t > 0, \qquad u(0) \in \mathcal{D}. \tag{2.1}$$

Derivatives with respect to the argument of a function are henceforth denoted by a prime. Our assumptions on the nonlinearity F are the following.

Assumption 2.1. We assume that the function $F : \mathcal{D} \to X$ is Fréchet differentiable and that its derivative $F' : \mathcal{D} \to L(D, X)$ has the following properties.

(i) F' is locally Lipschitz continuous; i.e., for each $u^* \in \mathcal{D}$ there exist $R > 0$ and $L > 0$ such that

$$\big\|F'(v) - F'(w)\big\|_{D\to X} \le L\|v - w\|_D, \tag{2.2}$$

for all v, $w \in \mathcal{D}$ with $\|v - u^*\|_D \le R$ and $\|w - u^*\|_D \le R$.

(ii) For every $u^* \in \mathcal{D}$ the operator $F'(u^*)$ is sectorial; i.e., there exist $\theta \in (0, \pi/2)$, $\omega_0 \in \mathbb{R}$ and $M > 0$ such that if $z \in \mathbb{C}$ and $|\arg(z - \omega_0)| \le \pi - \theta$, then $z - F'(u^*)$ has a bounded inverse in X and

$$\big\|(z - F'(u^*))^{-1}\big\|_{X\to X} \le \frac{M}{|z - \omega_0|}. \tag{2.3}$$

(iii) For every $u^* \in \mathcal{D}$ the graph-norm of $F'(u^*)$ is equivalent to the norm of D.

Under these assumptions, it is known that (2.1) has a locally unique solution (see [12, Theorem 8.1.1]). This solution $u \in C([0, \delta], D) \cap C^1([0, \delta], X)$ has the regularity property $u \in C^\alpha_\alpha((0, \delta], D)$ for arbitrary $0 < \alpha < 1$. For the convenience of the reader we recall the definition of the space C^α_α. For a Banach space $(B, \|\cdot\|_B)$, $C^\alpha_\alpha((0, \delta], B)$ is the space of all bounded functions $v : (0, \delta] \to B$ such that $t \mapsto t^\alpha v(t)$

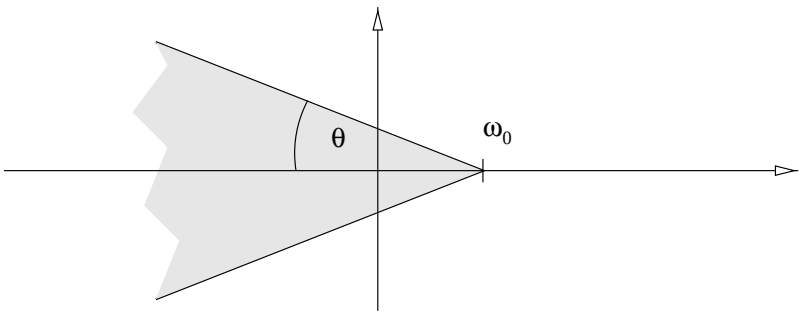


Figure 1. Condition (2.3) holds for all z *outside* the shaded cone.

is α-Hölder continuous in $(0,\delta]$. This space is endowed with the norm

$$\|v\|_{C^\alpha_\alpha((0,\delta],B)} = \sup_{0<t\le\delta} \|v(t)\|_B + \sup_{0<s<t\le\delta} \frac{\|v(t)-v(s)\|_B}{(t-s)^\alpha}\, s^\alpha.$$

We next give two nonlinear initial-boundary value problems that fit into our framework. More examples can be found in [12].

Example 2.2 (Combustion of a solid fuel, [3, Section 6.7])**.** Let $U(t,x)$ denote the temperature of a combusting solid fuel at position $x \in [0,1]$ and time $t \ge 0$. A model for the evolution of U is given by the nonlinear initial-boundary value problem

$$\partial_t U(t,x) = \partial_x\Big(k\big(\partial_x U(t,x)\big)\partial_x U(t,x)\Big) + \varphi(U(t,x)), \quad 0<x<1, \quad t>0, \tag{2.4}$$

with homogeneous Neumann boundary conditions $\partial_x U(t,0) = \partial_x U(t,1) = 0$ for all $t > 0$ and initial condition $U(0,x) = U_0(x)$ for $0 < x < 1$. We assume that the diffusion coefficient k is twice differentiable, with bounded second derivative, and that it satisfies the uniform ellipticity condition

$$k(y) + yk'(y) \ge \kappa > 0 \qquad \text{for all } y \in \mathbb{R}. \tag{2.5}$$

We further suppose that φ has a locally Lipschitz continuous derivative and that the initial value U_0 is twice continuously differentiable and satisfies the compatibility conditions $U_0'(0) = U_0'(1) = 0$.

Choosing $X = C([0,1])$ and $D = \{v \in C^2([0,1]) : v'(0) = v'(1) = 0\}$ allows us to write (2.4) in the abstract form (2.1) with $u(t) = U(t,\cdot)$ and

$$F(v) = \big(k(v')v'\big)' + \varphi(v).$$

The smoothness assumptions on k and φ immediately imply condition (i) of Assumption 2.1, and the ellipticity condition (2.5) implies (ii) and (iii) there.

Equally, it can be shown that our assumptions are satisfied for the Banach spaces $X = L^2(0,1)$ and $D = \{v \in H^2(0,1) : v'(0) = v'(1) = 0\}$. This follows from the well-known embedding $H^1(0,1) \subset C([0,1])$.

Example 2.3 (Semilinear problem with moving boundary)**.** We consider the semilinear parabolic problem

$$\partial_t V(t,y) = \partial_{yy} V(t,y) + \varphi\big(V(t,y), \partial_y V(t,y)\big), \quad 0<y<b(t), \quad t>0, \tag{2.6a}$$

with homogeneous Dirichlet boundary conditions $V(t,0) = V(t,b(t)) = 0$ for $t > 0$ and initial condition $V(0,y) = V_0(y)$ for $0 < y < b(t)$. Here the position of the right boundary $b(t)$ is determined by the ordinary differential equation

$$\partial_t b(t) = \psi\big(b(t), V(t,b(t)), \partial_y V(t,b(t))\big), \quad t>0, \qquad b(0) = 1. \tag{2.6b}$$

We assume that φ and ψ have locally Lipschitz continuous derivatives and that V_0 is twice continuously differentiable with $V_0(0) = V_0(1) = 0$.

The famous Stefan problem that models the melting of ice is of this form by taking $\varphi = 0$ and $\psi(p,q,r) = -\beta r$ with a positive constant β (see [16, Section 15.4]).

Changing the variables $U(t,x) = V(t,b(t)x)$ transforms problem (2.6) to the interval $0 \le x \le 1$, and we obtain the nonlinear system

$$\begin{aligned} \partial_t U(t,x) &= \frac{\partial_{xx} U(t,x)}{b(t)^2} + \varphi\left(U(t,x), \frac{\partial_x U(t,x)}{b(t)}\right) + \frac{x\,\partial_t b(t)}{b(t)}\,\partial_x U(t,x), \\ \partial_t b(t) &= \psi\left(b(t), U(t,1), \frac{\partial_x U(t,1)}{b(t)}\right), \quad 0 < x < 1, \quad t > 0, \end{aligned} \tag{2.7}$$

with boundary conditions $U(t,0) = U(t,1) = 0$ for $t > 0$ and initial conditions $U(0,x) = V_0(x)$ for $0 < x < 1$ and $b(0) = 1$.

We choose $X = C([0,1]) \times \mathbb{R}$ and $D = \big\{v \in C^2([0,1]) : v(0) = v(1) = 0\big\} \times \mathbb{R}$, and since the projection $P : C([0,1]) \to \mathbb{R} : v \mapsto v(1)$ is continuous, the conditions (i), (ii), and (iii) of Assumption 2.1 are again easily verified.

We finish this section by introducing some notation. The aim of the paper is the analysis of backward Euler discretizations of (2.1) which are given as sequences $u_0, u_1, \dots, u_N$ in $\mathcal{D}$, corresponding to a grid $0 = t_0 < t_1 < \cdots < t_N \le T$. This motivates the consideration of the following discrete norms and seminorms in X^N:

$$\begin{gathered} \mu(\mathbf{v}) = \sup_{1\le n\le N} \|v_n\|, \qquad \lambda_\alpha(\mathbf{v}) = \sup_{1\le k<n\le N} \frac{\|v_n - v_k\|}{(t_n - t_k)^\alpha}\, t_k^\alpha, \\ |||\mathbf{v}|||_\alpha = \mu(\mathbf{v}) + \lambda_\alpha(\mathbf{v}), \end{gathered} \tag{2.8a}$$

for $\mathbf{v} = (v_n)_{n=1}^N \in X^N$ and $0 < \alpha < 1$. Analogously, we denote

$$\begin{gathered} \mu_D(\mathbf{v}) = \sup_{1\le n\le N} \|v_n\|_D, \qquad \lambda_{D,\alpha}(\mathbf{v}) = \sup_{1\le k<n\le N} \frac{\|v_n - v_k\|_D}{(t_n - t_k)^\alpha}\, t_k^\alpha, \\ |||\mathbf{v}|||_{D,\alpha} = \mu_D(\mathbf{v}) + \lambda_{D,\alpha}(\mathbf{v}), \end{gathered} \tag{2.8b}$$

for $\mathbf{v} \in D^N$. Further we define μ_β for $0 \le \beta \le 1$ and $\mathbf{v} \in X_\beta^N$ through

$$\mu_\beta(\mathbf{v}) = \sup_{1\le n\le N} \|v_n\|_\beta. \tag{2.9}$$

Here $(X_\beta, \|\cdot\|_\beta)$ denotes the real interpolation space $(X,D)_{\beta,\infty}$ between X and D (see [12, Section 1.2.1]). Note that $|||\cdot|||_\alpha$ and $|||\cdot|||_{D,\alpha}$ are discrete versions of the norms $\|\cdot\|_{C_\alpha^\alpha((0,\delta],X)}$ and $\|\cdot\|_{C_\alpha^\alpha((0,\delta],D)}$, respectively.

3. Convergence analysis of the backward Euler solution

In this section we study the backward Euler method for discretizing (2.1) in time. We show that a unique numerical solution exists for finite times, provided that the maximal stepsize is chosen sufficiently small. We further derive convergence estimates under various smoothness assumptions on the exact solution.

We first consider a local situation for which more precise estimates can be obtained. For this, it is convenient to linearize (2.1) around the initial value $u(0)$. This gives the (formally) semilinear problem

$$u' = Au + f(u), \quad t > 0, \qquad u(0) \in \mathcal{D}, \tag{3.1}$$

where $A = F'(u(0))$ and $f(u) = F(u) - Au$ for $u \in \mathcal{D}$. In view of (2.2), there exist $R > 0$ and $L > 0$ such that

$$\|f(v) - f(w)\| \le L\varrho\, \|v - w\|_D, \tag{3.2}$$

for all $\|v - u(0)\|_D \le \varrho \le R$ and $\|w - u(0)\|_D \le \varrho \le R$ (see proof of Lemma 5.2).

Since the backward Euler method is invariant under linearization, the numerical approximation u_n to $u(t_n)$ is given by the recursion

$$\frac{u_n - u_{n-1}}{h_n} = Au_n + f(u_n), \qquad 1 \le n \le N, \tag{3.3}$$

with $t_n = t_{n-1} + h_n$ for $1 \le n \le N$ and $t_0 = 0$. Here $h_n > 0$ denotes the stepsize which is chosen according to accuracy requirements. The starting value $u_0 \in \mathcal{D}$ is allowed to be different from $u(0)$.

We remark that, due to (2.3) with $u^* = u(0)$ and (3.2), the nonlinear equation (3.3) has a unique solution $u_n \in \mathcal{D}$ for stepsizes h_n satisfying $h_n\omega_0 < 1$, as long as $\|u_{n-1} - u(0)\|_D \le \varrho$ for a certain $\varrho > 0$. In fact, (3.3) can be solved by standard fixed-point iteration (see Lemma 5.2). Let us point out, however, that already after one single step we can only expect

$$\|u_1 - u(0)\|_D \le C\varrho,$$

where $C > 1$. Thus, after a finite number of steps, independently of the stepsizes, the validity of (3.2) is no longer guaranteed. Therefore, this step-by-step approach is not suited to construct the numerical solution on a finite time interval $[0, T]$.

In order to overcome this difficulty, we adopt a global approach relying on the discrete variation-of-constants formula

$$u_n = r(t_n, 0)u_0 + \sum_{k=1}^{n} h_k\, r(t_n, t_{k-1})\, f(u_k), \tag{3.4}$$

where the discrete transition operator $r(t_n, t_k)$ is defined by

$$r(t_n, t_k) = (1 - h_nA)^{-1} \cdots (1 - h_{k+1}A)^{-1}, \qquad 0 \le k < n \le N, \tag{3.5}$$

and $r(t_k, t_k) = 1$. Note that this operator is well defined for

$$\max_{1\le k\le N} h_k \le \overline{h}, \quad \text{if} \quad \overline{h}\,\omega_0 < 1. \tag{3.6}$$

The numerical solution of (3.1) can be constructed by fixed-point iteration in (3.4). This is based on the fact that the nonlinear operator

$$\Phi : \mathcal{B} \subset \mathcal{D}^N \longrightarrow \mathcal{D}^N : \mathbf{v} \longmapsto \Phi(\mathbf{v}) = \mathbf{r}u_0 + \mathcal{K}(f(\mathbf{v})), \tag{3.7a}$$

with $\mathbf{r} = (r(t_n, 0))_{n=1}^N$, $f(\mathbf{v}) = (f(v_n))_{n=1}^N$ for $\mathbf{v} = (v_n)_{n=1}^N \in \mathcal{D}^N$, and

$$\mathcal{K}(\mathbf{w}) = \Big(\sum_{k=1}^{n} h_k\, r(t_n, t_{k-1})\, w_k\Big)_{n=1}^N \quad \text{for } \mathbf{w} = (w_n)_{n=1}^N \in X^N, \tag{3.7b}$$

is a contraction for a suitably chosen subset $\mathcal{B}$. Unfortunately, it turns out that the interval of existence is limited by the fact that $|||\Phi(\mathbf{u}_0) - \mathbf{u}_0|||_{D,\alpha}$ has to be sufficiently small, for $\mathbf{u}_0 = (u_0, \dots, u_0)^N$. Thus, nothing can be said about the size of t_N in this approach. This kind of difficulty also appears when constructing the continuous solution (see [12, Theorem 8.1.1]).

However, the global approach based on the convolution operator in (3.7b) turns out to be useful in order to derive preliminary convergence estimates. Eventually, these estimates can be used to establish the existence of the numerical solution for finite times.

Assume for a moment that the backward Euler approximations $u_0, u_1, \dots, u_N$ to the solution exist. We set $\widetilde{u}_n = u(t_n)$ and denote the errors by $e_n = \widetilde{u}_n - u_n$.

Inserting the exact solution into the numerical scheme defines the defects d_n by

$$\frac{\widetilde{u}_n - \widetilde{u}_{n-1}}{h_n} = A\widetilde{u}_n + f(\widetilde{u}_n) + d_n, \qquad 1 \le n \le N.$$

Subtracting (3.3) from this identity gives the error recursion

$$\mathbf{e} = \mathbf{r}e_0 + \mathcal{K}\big(f(\widetilde{\mathbf{u}}) - f(\mathbf{u})\big) + \mathcal{K}(\mathbf{d}), \tag{3.8}$$

where $\mathbf{e} = (e_1, e_2, \dots, e_N)^T \in \mathcal{D}^N$, etc.

Let C_3 and R be the constants provided by Lemma 5.3 for $u^* = u(0)$. We will show below that after a possible reduction of T, we may assume that there exists $0 < \varrho \le R$ such that

$$\begin{gathered}\mu_D(\widetilde{\mathbf{u}} - \mathbf{u}(0)) \le \varrho, \qquad \mu_D(\mathbf{u} - \mathbf{u}(0)) \le \varrho, \\ C_3C_5(2\varrho + \lambda_{D,\alpha}(\widetilde{\mathbf{u}})) \le \gamma < 1,\end{gathered} \tag{3.9}$$

where $0 < \alpha < 1$ is chosen and C_5 is the constant appearing in Lemma 5.5. Taking norms in (3.8) and using Lemmas 5.3, 5.4 and 5.5 yields

$$|||\mathbf{e}|||_{D,\alpha} \le \frac{1}{1-\gamma}\Big(C_4\|e_0\|_D + |||\mathcal{K}(\mathbf{d})|||_{D,\alpha}\Big). \tag{3.10}$$

Depending on our requirements on the analytical solution, we obtain different bounds for $|||\mathcal{K}(\mathbf{d})|||_{D,\alpha}$ and consequently different error estimates (see Theorems 3.1 and 3.2 below). We finally point out that because of

$$\|e_n\|_D \le |||\mathbf{e}|||_{D,\alpha}$$

these theorems also provide error estimates in D.

Theorem 3.1. *Let $u : [0,T] \to D$ be a solution of (2.1) with $u'' \in C^\alpha_\alpha((0,T],X)$ and assume that*

$$C_3C_5\Big(2\|u - u(0)\|_{L^\infty([0,T],D)} + \|u\|_{C^\alpha_\alpha((0,T],D)}\Big) < 1, \tag{3.11}$$

where C_3 and C_5 are the constants provided by Lemmas 5.3 and 5.5 for $u^* = u(0)$. Suppose that either

(a) the stepsizes $h_n = h$ are constant, or

(b) the stepsizes verify $h_n \ge \sigma h_{n-1}$, $2 \le n \le N$, for some $\sigma > 0$.

Then there exist constants $h^* > 0$, $\varrho_0 > 0$ and $C > 0$ such that the backward Euler solution u_n exists for stepsizes satisfying $0 < h_n \le h^*$ and for initial values u_0 with $\|u_0 - u(0)\|_D \le \varrho_0$, as long as $t_n \le T$. Further, we have the error bounds

$$|||\mathbf{e}|||_{D,\alpha} \le C\left(\|e_0\|_D + h\,\|u''\|_{C^\alpha_\alpha((0,T],X)}\right) \tag{3.12}$$

for constant stepsizes, and

$$|||\mathbf{e}|||_{D,\alpha} \le C\Big(\|e_0\|_D + \max_{1\le m\le n\le N}\Big(\big(h_n^{1-\alpha} + h_{m+1}^{1-\alpha}\big)M_n + h_m^{1-\alpha}M_{m,n}\Big)\Big) \tag{3.13}$$

with

$$M_n = \|u''\|_{L^\infty([t_{n-1},t_n],X)}, \qquad M_{m,n} = \sup_{t_{m-1}<s<t\le t_n} \frac{\|u''(t) - u''(s)\|}{(t-s)^\alpha}\, s^\alpha$$

for variable stepsizes, respectively. The constant C depends on α, T and on C_5 of Lemma 5.5. For variable stepsizes, it further depends on σ.

Proof. Set $\varrho_1 = \|u - u(0)\|_{L^\infty([0,T],D)}$ and choose $\varrho_1 < \varrho \le R$ such that

$$C_3 C_5 \left(2\varrho + \|u\|_{C^\alpha_\alpha((0,T],D)}\right) < 1.$$

In the first part of the proof we show the validity of the error estimates (3.12) and (3.13) under the assumptions that the numerical solution $(u_n)_{n=1}^N$ is defined as long as $t_N \le T$ and that

$$\|u_n - u(0)\|_D \le \varrho, \qquad n \le N. \tag{3.14}$$

In the second part we justify these assumptions.

(i) In view of (3.10) and Lemma 5.5, we have to estimate $|||\mathbf{d}|||_\alpha$. Taylor series expansion shows that the defects are given by

$$d_n = h_n \int_0^1 \tau u''(t_n - \tau h_n)\,\mathrm{d}\tau.$$

This immediately yields

$$\mu(\mathbf{d}) \le 1/2 \max_{1\le n\le N} h_n M_n.$$

For estimating $\lambda_{D,\alpha}(\mathbf{d})$ we first write for $m < n$

$$\begin{aligned} d_n - d_m = {} & h_m \int_0^1 \tau\big(u''(t_n - \tau h_n) - u''(t_m - \tau h_m)\big)\,\mathrm{d}\tau \\ & + (h_n - h_m) \int_0^1 \tau u''(t_n - \tau h_n)\,\mathrm{d}\tau. \end{aligned} \tag{3.15}$$

For constant stepsizes the second term in (3.15) drops and the estimate

$$\frac{\|d_n - d_m\|}{(t_n - t_m)^\alpha}\, t_m^\alpha \le h\, M_{m,n} \int_0^1 \tau\Big(\frac{m}{m-\tau}\Big)^\alpha \mathrm{d}\tau \le \frac{M_{m,n}}{1-\alpha}\, h$$

proves the first part of the theorem.

For variable stepsizes, one has

$$\begin{aligned} \frac{\|d_n - d_m\|}{(t_n - t_m)^\alpha}\, t_m^\alpha \le {} & h_m M_{m,n} \Big(\frac{t_n - t_m + h_m}{t_n - t_m}\Big)^\alpha \int_0^1 \tau\Big(\frac{t_m}{t_m - \tau h_m}\Big)^\alpha \mathrm{d}\tau \\ & + \frac{|h_n - h_m|}{(t_n - t_m)^\alpha} t_m^\alpha\, \frac{M_n}{2}. \end{aligned} \tag{3.16}$$

Due to our assumptions on the stepsize sequence, we have

$$\Big(\frac{t_n - t_m + h_m}{t_n - t_m}\Big)^\alpha \le \Big(1 + \frac{1}{\sigma}\Big)^\alpha$$

and

$$\frac{|h_n - h_m|}{(t_n - t_m)^\alpha} t_m^\alpha \le \Big(h_n^{1-\alpha} + \frac{1}{\sigma} h_m^{1-\alpha}\Big) T^\alpha.$$

The remaining term in (3.16) is bounded as follows:

$$h_m \int_0^1 \tau\Big(\frac{t_m}{t_m - \tau h_m}\Big)^\alpha \mathrm{d}\tau \le h_m^{1-\alpha} T^\alpha \int_0^1 \tau\Big(\frac{h_m}{t_m - \tau h_m}\Big)^\alpha \mathrm{d}\tau \le \frac{h_m^{1-\alpha} T^\alpha}{1-\alpha}.$$

Inserting these bounds into (3.16) gives the required bound for $\lambda_\alpha(\mathbf{d})$.

(ii) It remains to show that the backward Euler solution exists and that (3.14) holds. The idea of the proof is simple and standard for nonlinear equations: as long as u_{n-1} remains sufficiently close to $\widetilde{u}_{n-1}$, (3.3) can be solved for u_n and the above error estimate ensures that u_n is close enough to $\widetilde{u}_n$ as well. Repeating this

process proves the desired result. However, we have to pay some attention to the parameters involved.

For simplicity, we give the proof for constant stepsizes only. Set $\varrho_2 = \varrho - \varrho_1$, and let $0 < \varrho_* < \varrho^* \le \varrho_2$ and $h^* \le \underline{h}$ denote the thresholds provided by Lemma 5.2, applied to ϱ_2 and u^*, where u^* varies in the compact set formed by the values $u(t)$, $0 \le t \le T$. After a possible reduction of h^*, we can choose $\varrho_0 > 0$ such that

$$C\big(\varrho_0 + h^* \|u''\|_{C^\alpha_\alpha((0,T],X)}\big) \le \varrho_*/2,$$

where C is the constant from (3.12). Since u is uniformly continuous, we can further assume that $\|u(t_n) - u(t_{n-1})\|_D \le \varrho_*/2$ for $h \le h^*$.

Suppose by induction that u_n exists and that (3.14) is satisfied for $n \le m$. Then, due to (3.12) and the above choice of parameters, the bound

$$\|e_m\|_D \le C\big(\|e_0\|_D + h\,\|u''\|_{C^\alpha_\alpha((0,T],X)}\big) \le \varrho_*/2 \qquad \text{for } h \le h^*$$

implies

$$\|u_m - \widetilde{u}_{m+1}\|_D \le \|e_m\|_D + \|\widetilde{u}_m - \widetilde{u}_{m+1}\|_D \le \varrho_*.$$

An application of Lemma 5.2 with $u^* = \widetilde{u}_{m+1}$ and $w = u_m$ shows that u_{m+1} exists and $\|e_{m+1}\|_D \le \varrho^*$. Consequently, the estimate

$$\|u_{m+1} - u(0)\|_D \le \|e_{m+1}\|_D + \|\widetilde{u}_{m+1} - u(0)\|_D \le \varrho_2 + (\varrho - \varrho_2) = \varrho$$

follows. This yields (3.14) and concludes the proof. □

In practice it might be difficult to know whether u'' belongs to $C^\alpha_\alpha((0,T],X)$. This limitation is overcome in the next theorem where we impose the natural condition

$$Au(0) + f(u(0)) \in X_\beta, \tag{3.17}$$

for some $\alpha < \beta \le 1$. Note that, in actual applications, X_β is often a Sobolev space that does not depend on the boundary conditions for β sufficiently small. Hence, if the initial value is sufficiently smooth, this condition is easily seen to be satisfied.

It is also known that under (3.17) the exact solution of (2.1) has the additional regularity properties $u' \in L^\infty([0,T],X_\beta) \cap C^\beta([0,T],X)$ (see [12, Theorem 8.1.3]).

Theorem 3.2. *Let $u : [0,T] \to \mathcal{D}$ be a solution of* (2.1) *such that* (3.11) *is satisfied and assume that* (3.17) *holds for some $0 < \alpha < \beta \le 1$.*

Then there exist constants $h^* > 0$, $\varrho_0 > 0$ and $C > 0$ such that, for arbitrary stepsizes $0 < h_n \le h^*$ and for initial values u_0 with $\|u_0 - u(0)\|_D \le \varrho_0$, the backward Euler solution u_n is defined as long as $t_n \le T$ and we have

$$|||\mathbf{e}|||_{D,\alpha} \le C\Big(\|e_0\|_D + \max_{1\le n\le N} h_n^{\beta-\alpha} I_n^{\alpha/\beta} J_n^{1-\alpha/\beta}\Big) \tag{3.18}$$

with

$$I_n = \|u'\|_{L^\infty([t_{n-1},t_n],X_\beta)}, \qquad J_n = \|u'\|_{C^\beta([t_{n-1},t_n],X)}.$$

The constant C depends on α, β, T and on C_6 of Lemma 5.6.

Proof. We follow the arguments of the proof of Theorem 3.1. Therefore, it is sufficient to establish the validity of (3.18) under the assumptions that the numerical solution $(u_n)_{n=1}^N$ exists as long as $t_N \le T$ and that (3.9) holds for some $\varrho \le R$.

In view of (3.10) and Lemma 5.6, we have to estimate $\mu_\alpha(\mathbf{d})$. For the defects d_n, we use the representation

$$d_n = \int_0^1 \big(u'(t_n - \tau h_n) - u'(t_n)\big)\,\mathrm{d}\tau$$

to obtain the estimates

$$\|d_n\| \le \frac{h_n^\beta}{1+\beta} J_n \le h_n^\beta J_n \quad \text{and} \quad \|d_n\|_\beta \le 2I_n.$$

By a standard interpolation argument

$$\|d_n\|_\alpha \le \|d_n\|^{1-\alpha/\beta} \|d_n\|_\beta^{\alpha/\beta},$$

we get

$$\mu_\alpha(\mathbf{d}) \le 2 \max_{1\le n\le N} h^{\beta-\alpha} I_n^{\alpha/\beta} J_n^{1-\alpha/\beta},$$

which yields the desired result. □

The previous theorems are local in nature. By applying them recursively, we obtain pointwise convergence estimates in D for finite times. In the following theorem, the number α has the same meaning as in the local results before.

Theorem 3.3. *Let $u : [0,T] \to \mathcal{D}$ be a solution of (2.1) and let $0 < \alpha < 1$. Assume that either*

(a) *$u'' \in C_\alpha^\alpha((0,T],X)$ and the stepsizes $h_n = h$ are constant, or*

(b) *$u'' \in C_\alpha^\alpha((0,T],X)$ and $h_n \ge \sigma h_{n-1}$, $2 \le n \le N$, for some $\sigma > 0$, or*

(c) *$Au(0) + f(u(0)) \in X_\beta$ for some $0 < \alpha < \beta \le 1$.*

Then there exist constants $h^* > 0$, $\delta > 0$ and $C > 0$ such that the backward Euler solution u_n exists for stepsizes satisfying $0 < h_n \le h^*$ and for initial values u_0 with $\|u_0 - u(0)\|_D \le \delta$, as long as $t_n \le T$. For $0 \le n \le N$, we have the error bounds

(a) $\|e_n\|_D \le C\Big(\|e_0\|_D + h\,\|u''\|_{C_\alpha^\alpha((0,t_N],X)}\Big)$, *or*

(b) $\|e_n\|_D \le C\Big(\|e_0\|_D + \max\limits_{1\le m\le k\le N} \Big(\big(h_k^{1-\alpha} + h_{m+1}^{1-\alpha}\big) M_k + h_m^{1-\alpha} M_{m,k}\Big)\Big)$, *or*

(c) $\|e_n\|_D \le C\Big(\|e_0\|_D + \max\limits_{1\le m\le N} h_m^{\beta-\alpha} I_m^{\alpha/\beta} J_m^{1-\alpha/\beta}\Big)$,

respectively.

Proof. We only give the proof of the first result. The remaining statements follow in a similar way.

Since u is continuous, there are constants $R > 0$ and $L > 0$ such that (2.2) is uniformly satisfied for u^* varying in the set formed by the values $u(t)$, $0 \le t \le T$. Moreover, by Lemma 5.8, there exists a partition $0 = T_0 < T_1 < \cdots < T_J = T$ of $[0,T]$ such that

$$C_3 C_5 \Big(2\|u_{T_j} - u(T_{j-1})\|_{L^\infty([0,H_j],D)} + \|u_{T_j}\|_{C_\alpha^\alpha((0,H_j],D)}\Big) < 1, \qquad 1 \le j \le J,$$

with $H_j = T_j - T_{j-1}$ and $u_{T_j}(t) = u(T_{j-1} + t)$, $0 \le t \le H_j$. Here C_3 and C_5 are the constants provided by Lemmas 5.3 and 5.5 for $u^* = u(T_j)$ and R. Notice that these constants only depend on R and L.

Therefore, we deduce from Theorem 3.1, applied piece-by-piece, that there exist positive constants h^* and $\widetilde{C}$, such that for $0 < h \le h^*$

$$\|e_n\|_D \le C^j \|e_0\|_D + \widetilde{C} h \, \|u''\|_{C^\alpha_\alpha((0,T_j],X)}, \qquad 0 \le t_n \le T_j.$$

After a possible reduction of δ and h^*, this estimate shows that $\|e_n\|_D \le \varrho_0$, with ϱ_0 given by Theorem 3.1. Notice that J is independent of $0 < h \le h^*$. The desired error estimate thus follows by recursion. □

4. Behaviour near an asymptotically stable equilibrium

In this section we study the long-term behaviour of time discretizations of (2.1). To keep our exposition in a reasonable length, we restrict our attention to hyperbolic equilibria. For these the *principle of linearized stability* holds, which means that the dynamical behaviour near such an equilibrium $\overline{u}$ is fully determined by the linearized equation

$$v' = F'(\overline{u})(v - \overline{u})$$

(see [12, Section 9.1]). We show that a similar property holds for the backward Euler discretization of (2.1). Further, numerical simulations that illustrate our theoretical result are given.

For notational simplicity, we concentrate on the asymptotically stable case. Let $\overline{u} \in \mathcal{D}$ be an equilibrium of (2.1), i.e., $F(\overline{u}) = 0$, and assume that the sectorial operator

$$A = F'(\overline{u}) \qquad \text{is asymptotically stable, i.e., } \omega_0 < 0. \tag{4.1}$$

The number ω_0 is defined in (2.3) (see also Figure 1). In this situation, it is well known that $\overline{u}$ is asymptotically stable and attracts all solutions in a sufficiently small neighbourhood of $\overline{u}$ with exponential speed. More precisely, it is shown in [12, Theorem 9.1.2] that for each $\omega < |\omega_0|$ there are constants $\delta_0 > 0$ and $C > 0$ such that the solution of (2.1) exists for all positive times and satisfies

$$\|u(t) - \overline{u}\|_D \le C \cdot \mathrm{e}^{-\omega t} \|u(0) - \overline{u}\|_D, \qquad \text{for all } t \ge 0, \tag{4.2}$$

whenever the initial value satisfies $\|u(0) - \overline{u}\|_D \le \delta_0$.

The following theorem gives the corresponding result for the backward Euler discretization. Note that any equilibrium of (2.1) is also an equilibrium of the backward Euler discretization.

Theorem 4.1. *Let $\overline{u}$ be an equilibrium of* (2.1) *and assume that* (4.1) *holds. Then, for any choice of $\omega < |\omega_0|$, there are positive constants $\overline{h}$, δ and C such that the following holds. The backward Euler solution $(u_n)_{n=1}^\infty$ of* (2.1) *exists for all stepsize sequences satisfying $0 < h_n \le \overline{h}$ and for all initial values u_0 with $\|u_0 - \overline{u}\|_D \le \delta$, and we have*

$$\|u_n - \overline{u}\|_D \le C \cdot \mathrm{e}^{-\omega t_n} \|u_0 - \overline{u}\|_D, \qquad \textit{for all } n \ge 0. \tag{4.3}$$

Note that the constant C depends on ω, but not on the particular choice of the stepsize sequence.

Demanding that the numerical solution decays towards the equilibrium nearly as fast as the exact solution imposes a severe restriction on the maximal stepsize. This restriction is overcome in the following theorem, where exponentially fast convergence is obtained, if the stepsizes remain bounded.

Theorem 4.2. *Let $\overline{u}$ be an equilibrium of* (2.1) *and assume that* (4.1) *holds. Then, for any $\overline{h} > 0$, there are constants $0 < \omega < |\omega_0|$, $\delta > 0$ and $C > 0$ such that the backward Euler solution $(u_n)_{n=1}^{\infty}$ of* (2.1) *exists for all stepsize sequences satisfying $0 < h_n \le \overline{h}$ and for all initial values u_0 with $\|u_0 - \overline{u}\|_D \le \delta$, and* (4.3) *holds.*

Proof of Theorem 4.1. We linearize (2.1) around the equilibrium $\overline{u}$ and construct the backward Euler solution by fixed-point iteration. In order to capture the decaying behaviour of the solution we use exponentially weighted norms. For $\omega > 0$ and sequences $\mathbf{v} = (v_n)_{n=1}^{\infty}$ in D, we modify (2.8b) in the following way:

$$\mu_{D,\omega}(\mathbf{v}) = \sup_{1\le n<\infty} \mathrm{e}^{\omega t_n}\|v_n\|_D, \qquad \lambda_{D,\alpha,\omega}(\mathbf{v}) = \sup_{1\le k<n<\infty} \mathrm{e}^{\omega t_k}\frac{\|v_n - v_k\|_D}{(t_n - t_k)^{\alpha}}\, t_k^{\alpha},$$
$$|||\mathbf{v}|||_{D,\alpha,\omega} = \mu_{D,\omega}(\mathbf{v}) + \lambda_{D,\alpha,\omega}(\mathbf{v}),$$

as well as the corresponding norm $|||\cdot|||_{\alpha,\omega}$ based on $\|\cdot\|$. A crucial observation is that Lemmas 5.3, 5.4 and 5.5 have an extension to these exponentially weighted norms for $0 < \omega < |\omega_1| < |\omega_0|$ with ω_1 as in Lemma 5.1. The gap $\omega - |\omega_1|$ is needed to bound the powers of t_n that are encountered.

With these preparations, we are ready to give the proof. Let $\mathcal{B}$ denote the ball

$$\mathcal{B} = \{\mathbf{v} \in \mathcal{D}^{\infty} : |||\mathbf{v} - \overline{\mathbf{u}}|||_{D,\alpha,\omega} \le \varrho\}.$$

We define Φ as in (3.7a) with $N = \infty$. Using the above-mentioned extensions of Lemmas 5.3 and 5.5 with $u^* = \overline{u}$ proves

$$|||\Phi(\mathbf{v}) - \Phi(\mathbf{w})|||_{D,\alpha,\omega} \le 3\varrho\, C_3 C_5 |||\mathbf{v} - \mathbf{w}|||_{D,\alpha,\omega}$$

for $\mathbf{v}, \mathbf{w} \in \mathcal{B}$. This shows that Φ is a contraction on $\mathcal{B}$ with contraction factor $1/2$ for ϱ sufficiently small.

It remains to show that Φ maps $\mathcal{B}$ onto $\mathcal{B}$ if u_0 lies sufficiently close to $\overline{u}$. Since

$$\Phi(\overline{\mathbf{u}}) = \mathbf{r}u_0 + (1 - \mathbf{r})\overline{u},$$

we have for all $\mathbf{v} \in \mathcal{B}$

$$\begin{aligned} |||\Phi(\mathbf{v}) - \overline{\mathbf{u}}|||_{D,\alpha,\omega} &\le |||\Phi(\mathbf{v}) - \Phi(\overline{\mathbf{u}})|||_{D,\alpha,\omega} + |||\Phi(\overline{\mathbf{u}}) - \overline{\mathbf{u}}|||_{D,\alpha,\omega} \\ &\le 1/2\ |||\mathbf{v} - \overline{\mathbf{u}}|||_{D,\alpha,\omega} + |||\mathbf{r}(u_0 - \overline{u})|||_{D,\alpha,\omega}. \end{aligned}$$

The last term is estimated by the first part of Lemma 5.4:

$$|||\mathbf{r}(u_0 - \overline{u})|||_{D,\alpha,\omega} \le C_4\|u_0 - \overline{u}\|_D.$$

For δ satisfying $2\delta C_4 \le \varrho$, we thus have $\Phi(\mathcal{B}) \subset \mathcal{B}$.

This proves the existence of a unique fixed-point $\mathbf{u}$, which is the searched backward Euler solution. Using further

$$|||\Phi(\mathbf{u}) - \overline{\mathbf{u}}|||_{D,\alpha,\omega} \le 1/2\ |||\Phi(\mathbf{u}) - \overline{\mathbf{u}}|||_{D,\alpha,\omega} + |||\mathbf{r}(u_0 - \overline{u})|||_{D,\alpha,\omega}$$

yields

$$|||\mathbf{u} - \overline{\mathbf{u}}|||_{D,\alpha,\omega} \le 2C_4\|u_0 - \overline{u}\|_D.$$

In particular, we get

$$\sup_{1\le n<\infty} \mathrm{e}^{\omega t_n}\|u_n - \overline{u}\|_D \le 2C_4\|u_0 - \overline{u}\|_D,$$

which proves the assertion of the theorem. □

Proof of Theorem 4.2. The proof is very similar to the preceding one and therefore omitted. It relies on the stability bounds given in Lemma 5.1, part (b). □

TABLE 1. Numerically observed contraction factors

h	1.0000	0.5000	0.2500	0.1250	0.0625	0.03125
ω_h	0.4055	0.4463	0.4711	0.4849	0.4922	0.4960

We close this section with a numerical example that illustrates Theorem 4.1.

Example 4.3 (Combustion of a solid fuel)**.** We take up Example 2.2 and specify the functions k, φ and U_0 as follows:

$$k(y) = y^2 + 1, \qquad \varphi(y) = -y(y-1/2)(y-1), \qquad U_0(x) = 1/2 + 2x^2(1-x)^2.$$

Note that the initial condition U_0 is compatible with the boundary conditions and that k satisfies the ellipticity condition (2.5) with $\kappa = 1$. The problem has three equlibria $\overline{u} = 1$, $\overline{u} = 0$, and $\overline{u} = 1/2$. The first two are asymptotically stable with $\omega_0 = -1/2$. Due to our choice of U_0, we expect convergence to $\overline{u} = 1$.

The partial differential equation (2.4) is discretized in space by standard finite differences on an equidistant grid with meshwidth $1/200$ and in time by the backward Euler method with constant stepsize h. For different values of h, the integration is performed up to $t = 40$. The numerical approximations ω_h to ω are displayed in Table 1. The results are in complete agreement with Theorem 4.1.

5. LEMMAS

In this section we collect the auxiliary results that are needed in the proofs of the previous theorems. Throughout the section we set

$$f(u) = F(u) - Au, \qquad \text{where} \quad A = F'(u^*) \tag{5.1}$$

for some $u^* \in \mathcal{D}$ and denote by $\omega_0 \in \mathbb{R}$ the constant from (2.3) that corresponds to u^*. We fix $\kappa > \omega_0$ and $\overline{h} > 0$ such that $\overline{h}\,\omega_0 < 1$ and consider arbitrary grid points $0 = t_0 < t_1 < \cdots < t_N$ that satisfy $h_n = t_n - t_{n-1} \le \overline{h}$. There is no restriction on the maximal stepsize for $\omega_0 \le 0$.

For the discrete transition operators (3.5), we have the following stability bounds.

Lemma 5.1. (*a*) *For any $\omega_1 > \omega_0$ there exist constants $h^* > 0$ and $C_1 > 0$ such that for $0 \le \nu \le 1$*

$$\|(\kappa - A)^\nu\, r(t_n, t_k)\|_{X \to X} \le C_1\, \frac{\mathrm{e}^{\omega_1 (t_n - t_k)}}{(t_n - t_k)^\nu}, \qquad 0 \le k < n, \tag{5.2}$$

whenever the stepsizes are bounded by h^*.

(*b*) *Let $\omega_0 < 0$ and $h^* > 0$. Then there exist constants $\omega_0 < \omega_1 < 0$ and $C_1 > 0$ such that* (5.2) *holds, whenever the stepsizes are bounded by h^*.*

Similar bounds are given in [2, 4, 14]. We note for later use that (5.2) also holds for $1 \le \nu \le 2$ if $k < n - 1$.

Proof. The estimate (5.2) is a consequence of the stability bounds

$$\|\mathrm{e}^{tA}\|_{X \to X} \le C\, \mathrm{e}^{\omega_0 t}, \qquad \|A\mathrm{e}^{tA}\|_{X \to X} \le \frac{C}{t}\, \mathrm{e}^{\omega_0 t}, \qquad t > 0,$$

for the analytic semigroup. Using the representation

$$r(t_n, t_k) = \int_0^\infty \cdots \int_0^\infty \mathrm{e}^{-s_{k+1} - \ldots - s_n}\, \mathrm{e}^{(s_{k+1} h_{k+1} + \ldots + s_n h_n) A}\, \mathrm{d}s_{k+1} \cdots \mathrm{d}s_n$$

shows

$$\|r(t_n,t_k)\|_{X\to X} \le C \prod_{j=k+1}^{n} (1-h_j\omega_0)^{-1},$$
$$\|Ar(t_n,t_k)\|_{X\to X} \le C \int_{-\infty}^{\omega_0} \prod_{j=k+1}^{n} (1-h_j\omega)^{-1}\mathrm{d}\omega.$$

For $\omega \le \omega_0 \le 0$, we have

$$1-h_j\omega \ge (1-h_j\omega_0)\big(1+ch_j(\omega_0-\omega)\big) \tag{5.3}$$

with $c = 1-h^*\omega_0$. Arguing as in [4] shows

$$\|Ar(t_n,t_k)\|_{X\to X} \le C\frac{1-h^*\omega_0}{t_n-t_k} \prod_{j=k+1}^{n} (1-h_j\omega_0)^{-1}.$$

For $\omega_0 > 0$ we use an idea from [6] and eliminate the small steps by

$$(1-h_j\omega)(1-h_k\omega) \ge 1-(h_j+h_k)\omega,$$

until (5.3) is again satisfied. Part (a) of the lemma then follows from standard estimates and interpolation.

In order to verify (b) we note that the function

$$\omega_1 = \omega_1(H) = -\frac{\log(1-H\omega_0)}{H} \tag{5.4}$$

is monotonically increasing for $\omega_0 < 0$ with $\omega_1(0) = \omega_0$ and $\omega_1(\infty) = 0$. Hence,

$$(1-h_j\omega_0)^{-1} \le \mathrm{e}^{\omega_1 h_j}$$

with ω_1 given by (5.4) for $H = \max h_j \le h^*$. □

We note for later use that the identity

$$r(t_n,t_k) - 1 = \sum_{l=k+1}^{n} h_l\, A\, r(t_n,t_{l-1})$$

together with Lemma 5.1 implies for $0 \le k < n \le N$ and $0 < \nu \le 1$ the bound

$$\big\|(r(t_n,t_k)-1)\ (\kappa-A)^{-\nu}\big\|_{X\to X} \le \frac{C_1}{\nu}\ \mathrm{e}^{\omega_1^+(t_n-t_k)}\ (t_n-t_k)^{\nu} \tag{5.5}$$

with $\omega_1^+ = \max(\omega_1, 0)$. For simplicity, we make no notational difference between the constants in (5.2) and (5.5).

Lemma 5.2. *Let $u^* \in \mathcal{D}$ and $R_0 > 0$. Then there exist $0 < \varrho_* < \varrho^* \le R_0$ and $h^* > 0$ such that, for $w \in \mathcal{D}$ with $\|w-u^*\|_D \le \varrho_*$ and for $0 < h \le h^*$, the equation*

$$\frac{v-w}{h} = Av + f(v) \tag{5.6}$$

possesses a unique solution $v \in \mathcal{D}$ with $\|v-u^*\|_D \le \varrho^*$. Moreover, the quantities ϱ_*, ϱ^* and h^* can be chosen uniformly for u^* belonging to a relatively compact subset of $\mathcal{D}$.

Proof. We first note that, due to (2.2), there exist $R > 0$ and $L > 0$ such that

$$\|f'(v) - f'(w)\|_{D\to X} \le L\|v - w\|_D, \tag{5.7}$$

for all $v, w \in \mathcal{D}$ with $\|v - u^*\|_D \le R$ and $\|w - u^*\|_D \le R$. Therefore, since $f'(u^*) = 0$, we also have

$$\|f'(v)\|_{D\to X} \le L\varrho, \tag{5.8}$$

for all $v \in \mathcal{D}$ such that $\|v - u^*\|_D \le \varrho \le R$. This implies that

$$\|f(v) - f(w)\| \le L\varrho\,\|v - w\|_D, \tag{5.9}$$

for all $v, w \in \mathcal{D}$ with $\|v - u^*\|_D \le \varrho \le R$ and $\|w - u^*\|_D \le \varrho \le R$.

Equation (5.6) is equivalent to $v = g(v)$, where g is defined by

$$g(v) = (1 - hA)^{-1}w + h(1 - hA)^{-1}f(v).$$

We solve (5.6) by fixed-point iteration in the set $\mathcal{B} = \{v \in \mathcal{D} : \|v - u^*\|_D \le \varrho^*\}$. For this, we have to show that g is contractive and maps $\mathcal{B}$ onto $\mathcal{B}$, for ϱ_* and ϱ^* sufficiently small.

By the equivalence of $\|\cdot\|_D$ with the graph-norm of A, there exists $\overline{M}$ such that

$$\|h(1 - hA)^{-1}\|_{X\to D} \le \overline{M}, \qquad \|(1 - hA)^{-1}\|_{D\to D} \le \overline{M} \qquad \text{for } 0 < h \le \overline{h}.$$

On the one hand, we have

$$\|g(v) - g(\widetilde{v})\|_D \le \overline{M}\,L\,\varrho^*\|v - \widetilde{v}\|_D,$$

for $h \le \overline{h}$ and v, $\widetilde{v} \in \mathcal{D}$ with $\|v - u^*\|_D \le \varrho^*$ and $\|\widetilde{v} - u^*\|_D \le \varrho^*$. On the other hand, it holds

$$\begin{aligned} g(v) - u^* = (1 - hA)^{-1}\big(w - u^*\big) + h(1 - hA)^{-1}(Au^* + f(u^*)) \\ + h(1 - hA)^{-1}\,(f(v) - f(u^*)), \end{aligned}$$

so that

$$\|g(v) - u^*\|_D \le \overline{M}\|w - u^*\|_D + \|h(1 - hA)^{-1}(Au^* + f(u^*))\| + \overline{M}L\varrho^*\|v - u^*\|_D.$$

In view of these bounds, if we choose ϱ_* and ϱ^* such that $\overline{M}\varrho_* \le \varrho^*/3$ and $\overline{M}L\varrho^* \le 1/3$, then g is a contraction on $\mathcal{B}$. Moreover, since $h(1 - hA)^{-1} \to 0$ strongly as an operator from X to D, we can select $h^* \le \overline{h}$ such that, for $0 < h \le h^*$,

$$\|h(1 - hA)^{-1}(Au^* + f(u^*))\|_D \le \varrho^*/3.$$

Thus, g maps $\mathcal{B}$ into $\mathcal{B}$, and the fixed-point theorem provides the existence of a unique solution v of (5.6).

Since F' is locally Lipschitz continuous, R_0 and L can be taken uniformly for u^* in a compact set. Moreover, the equivalence of $\|\cdot\|_D$ with the graph-norm of $F'(u^*)$ is also uniform on the compact set. With this, the statement of the lemma follows easily. □

Lemma 5.3. *For $0 < \alpha < 1$ there exist constants $C_3 > 0$ and $R > 0$ such that*

$$|||f(\mathbf{v}) - f(\mathbf{w})|||_\alpha \le C_3\Big(2\varrho + \lambda_{D,\alpha}(\mathbf{w})\Big)|||\mathbf{v} - \mathbf{w}|||_{D,\alpha} \tag{5.10}$$

for all $\mathbf{v}$ and $\mathbf{w}$ in the set $\mathcal{V} = \{\,\widetilde{\mathbf{v}} \in \mathcal{D}^N : \mu_D(\widetilde{\mathbf{v}} - \mathbf{u}^*) \le \varrho\,\}$ whenever $0 < \varrho \le R$.

Proof. Choose R as in (2.2) and let $\mathbf{v}$, $\mathbf{w} \in \mathcal{V}$ for some $0 < \varrho \le R$. In view of (5.9), we have

$$\mu(f(\mathbf{v}) - f(\mathbf{w})) \le L\varrho\,\mu_D(\mathbf{v} - \mathbf{w}), \tag{5.11}$$

and thus it remains to bound $\lambda_\alpha(f(\mathbf{v}) - f(\mathbf{w}))$. We set $G_n = f'(\sigma v_n + (1-\sigma)w_n)$, $0 \le \sigma \le 1$, and write for $m < n$

$$\begin{aligned}\big(f(v_n) - f(w_n)\big) - \big(f(v_m) - f(w_m)\big) &= \int_0^1 \big(G_n(v_n - w_n) - G_m(v_m - w_m)\big)\,\mathrm{d}\sigma \\ &= \int_0^1 G_n\big((v_n - w_n) - (v_m - w_m)\big)\,\mathrm{d}\sigma + \int_0^1 \big(G_n - G_m\big)\,(v_m - w_m)\,\mathrm{d}\sigma.\end{aligned}$$

Using (5.8), we can estimate the first term on the right-hand side by

$$\int_0^1 \big\|G_n\,\big((v_n - w_n) - (v_m - w_m)\big)\big\|\,\mathrm{d}\sigma \le L\varrho\,\lambda_{D,\alpha}(\mathbf{v} - \mathbf{w})(t_n - t_m)^\alpha t_m^{-\alpha}.$$

Due to (5.7), the remaining term can be bounded as follows:

$$\begin{aligned}&\int_0^1 \big\|\big(G_n - G_m\big)\,(v_m - w_m)\big\|\,\mathrm{d}\sigma \\ &\quad\le L\int_0^1 \Big(\|w_n - w_m\|_D + \sigma\|(v_n - w_n) - (v_m - w_m)\|_D\Big)\mathrm{d}\sigma\,\|v_m - w_m\|_D \\ &\quad\le L\Big(\lambda_{D,\alpha}(\mathbf{w}) + 1/2\;\lambda_{D,\alpha}(\mathbf{v} - \mathbf{w})\Big)\mu_D(\mathbf{v} - \mathbf{w})(t_n - t_m)^\alpha t_m^{-\alpha}.\end{aligned}$$

The above estimates readily give

$$\lambda_\alpha(f(\mathbf{v}) - f(\mathbf{w})) \le L\Big(2\varrho\,\lambda_{D,\alpha}(\mathbf{v} - \mathbf{w}) + \lambda_{D,\alpha}(\mathbf{w})\mu_D(\mathbf{v} - \mathbf{w})\Big),$$

and this inequality combined with (5.11) proves (5.10). □

In Lemmas 5.4, 5.5 and 5.6 below we establish certain estimates involving $|||\cdot|||_{D,\alpha}$. As $\|\cdot\|_D$ is equivalent to the graph-norm of A, the norm

$$|||\mathbf{v}|||_\alpha + |||A\mathbf{v}|||_\alpha, \qquad \mathbf{v} \in D^N,$$

is equivalent to $|||\cdot|||_{D,\alpha}$ as well, for all $0 < \alpha < 1$. Since the required estimates for $|||\cdot|||_\alpha$ are usually obtained more easily (and in a similar way) than the corresponding estimates for $|||A\cdot|||_\alpha$, we give for simplicity the proofs only for $|||A\cdot|||_\alpha$. Henceforth, C denotes a generic constant that possibly depends on C_1 and on constants that arise from changing between equivalent norms.

Lemma 5.4. *Let $0 < \alpha < 1$.*

(a) There exists a constant $C_4 > 0$ such that for every $v \in D$

$$|||\mathbf{r}\,v|||_{D,\alpha} \le C_4\|v\|_D.$$

The constant C_4 depends on t_N, but it is otherwise independent of the grid. If ω_0 is nonnegative, then C_4 is bounded for finite times. If ω_0 is negative, then C_4 can be chosen independently of t_N.

(b) For $x \in X$ we have

$$\lim_{t_N \to 0} |||(\mathbf{r} - 1)\,x|||_\alpha = 0.$$

The convergence is uniform on relatively compact subsets of X.

Proof. In order to prove the first statement of the lemma, we have to estimate $A\big(r(t_n,0)-r(t_m,0)\big)\,v$. Using the identity

$$r(t_n,0)-r(t_m,0)=\big(r(t_n,t_m)-1\big)\,r(t_m,0)$$

we obtain

$$\begin{aligned}\|A(r(t_n,0)-r(t_m,0))v\| \le C\,&\|(r(t_n,t_m)-1)(\kappa-A)^{-\alpha}\|_{X\to X}\\ &\times\|(\kappa-A)^{\alpha}r(t_m,0)\|_{X\to X}\|v\|_D.\end{aligned}$$

With the help of (5.2) and (5.5) the right-hand side can be bounded by

$$\frac{C}{\alpha}\,\mathrm{e}^{\omega_1^+ t_N}\|v\|_D(t_n-t_m)^{\alpha}t_m^{-\alpha},$$

which proves the first result.

To show the second statement of the lemma we choose $\widetilde{x}\in D$. From the identity

$$r(t_k,0)-r(t_{k-1},0)=h_k\,A\,r(t_k,0)$$

we get

$$\begin{aligned}\|(r(t_n,0)-r(t_m,0))\,x\| &\le \sum_{k=m+1}^{n} h_k\left(\|A\,r(t_k,0)\;(x-\widetilde{x})\,\|+\|r(t_k,0)\,A\,\widetilde{x}\|\right)\\ &\le C\,t_m^{-\alpha}\sum_{k=m+1}^{n} h_k\,t_k^{-1+\alpha}\,\mathrm{e}^{\omega_1 t_k}\,\|x-\widetilde{x}\|+C\sum_{k=m+1}^{n} h_k\,\mathrm{e}^{\omega_1 t_k}\,\|\widetilde{x}\|_D\\ &\le C\Big(\|x-\widetilde{x}\|+t_N\;\|\widetilde{x}\|_D\Big)\,\mathrm{e}^{\omega_1^+ t_N}\;(t_n-t_m)^{\alpha}\,t_m^{-\alpha}.\end{aligned}$$

Since D is dense in X, the second statement of the lemma follows. □

Lemma 5.5. *For $0<\alpha<1$ there exists a constant $C_5>0$ such that if $|||\mathbf{v}|||_\alpha<\infty$ we have*

$$|||\mathcal{K}(\mathbf{v})|||_{D,\alpha}\le C_5|||\mathbf{v}|||_\alpha.$$

The constant C_5 depends on t_N, but it is otherwise independent of the grid. If ω_0 is nonnegative, then C_5 is bounded for finite times. If ω_0 is negative, then C_5 can be chosen independently of t_N.

Proof. Analogously to the modified variation-of-constants formula (1.3) and with the help of

$$h_k A\,r(t_n,t_{k-1})=r(t_n,t_{k-1})-r(t_n,t_k), \tag{5.12}$$

we split $\mathcal{K}(\mathbf{v})$ such that $A\mathcal{K}(\mathbf{v})=\mathbf{a}+\mathbf{b}$ where

$$\begin{aligned}a_n&=\sum_{k=1}^{n} h_k A\,r(t_n,t_{k-1})\;(v_k-v_n)\,,\\ b_n&=\sum_{k=1}^{n} h_k A\,r(t_n,t_{k-1})\,v_n=\big(r(t_n,0)-1\big)v_n.\end{aligned}$$

According to this we have to estimate the four terms $\mu(\mathbf{a})$, $\lambda_\alpha(\mathbf{a})$, $\mu(\mathbf{b})$, and $\lambda_\alpha(\mathbf{b})$.

(i) Using (5.2) we get

$$\begin{aligned}\|a_n\| &\le C_1\lambda_\alpha(\mathbf{v})\sum_{k=1}^{n} h_k\,\frac{(t_n-t_k)^\alpha}{(t_n-t_{k-1})\,t_k^\alpha}\,\mathrm{e}^{\omega_1(t_n-t_{k-1})}\\ &\le CB(\alpha,1-\alpha)\,\mathrm{e}^{\omega_1^+ t_N}\,\lambda_\alpha(\mathbf{v}),\end{aligned}$$

where B denotes the beta function. The last bound follows from comparing with the integral in a similar way as in Lemma 5.7.

(ii) In order to estimate $\lambda_\alpha(\mathbf{a})$ we use the identity

$$\begin{aligned}a_n - a_m = &\sum_{k=m+1}^{n} h_k\,A\,r(t_n,t_{k-1})\;(v_k-v_n) + \sum_{k=1}^{m} h_k\,A\,r(t_n,t_{k-1})\;(v_m-v_n)\\ &+\sum_{k=1}^{m} h_k\,A\Big(r(t_n,t_{k-1})-r(t_m,t_{k-1})\Big)\;(v_k-v_m)\\ = S_1+S_2+S_3.\end{aligned}$$

We take norms and use again (5.2) and (5.5). This gives

$$\begin{aligned}\|S_1\| &\le C_1\,\mathrm{e}^{\omega_1^+ t_N}\,\lambda_\alpha(\mathbf{v})\sum_{k=m+1}^{n}\frac{h_k}{(t_n-t_{k-1})^{1-\alpha}t_k^\alpha}\\ &\le \alpha^{-1}\,C_1\,\mathrm{e}^{\omega_1^+ t_N}\lambda_\alpha(\mathbf{v})\,(t_n-t_m)^\alpha\,t_m^{-\alpha},\end{aligned}$$

and, together with (5.12),

$$\begin{aligned}\|S_2\| &\le \big\|r(t_n,t_m)\,\big(r(t_m,0)-1\big)\,(v_m-v_n)\big\|\\ &\le C_1\,(1+C_1)\,\mathrm{e}^{\omega_1^+ t_N}\,\lambda_\alpha(\mathbf{v})\,(t_n-t_m)^\alpha\,t_m^{-\alpha}.\end{aligned}$$

With the help of Lemma 5.7 we get

$$\begin{aligned}\|S_3\| &\le \sum_{k=1}^{m} h_k\sum_{l=m+1}^{n} h_l\,\big\|A^2 r(t_l,t_{k-1})\,(v_k-v_m)\big\|\\ &\le C_1\lambda_\alpha(\mathbf{v})\sum_{k=1}^{m} h_k\sum_{l=m+1}^{n} h_l\,\mathrm{e}^{\omega_1(t_l-t_{k-1})}\,\frac{(t_m-t_k)^\alpha}{(t_l-t_{k-1})^2 t_k^\alpha}\\ &\le \frac{4C_1}{\alpha(1-\alpha)}\;\mathrm{e}^{\omega_1^+ t_N}\,\lambda_\alpha(\mathbf{v})\,(t_n-t_m)^\alpha\,t_m^{-\alpha}\end{aligned}$$

which proves the estimate for $\lambda_\alpha(\mathbf{a})$.

(iii) The stability bound (5.2) for the transition operator immediately gives

$$\mu(\mathbf{b}) \le \Big(1+C_1\,\mathrm{e}^{\omega_1^+ t_N}\Big)\;\mu(\mathbf{v}).$$

(iv) For the estimate of $\lambda_\alpha(\mathbf{b})$ we write

$$\begin{aligned}b_n - b_m &= \big(r(t_n,0)-1\big)\,v_n - \big(r(t_m,0)-1\big)\,v_m\\ &= \big(r(t_n,t_m)-1\big)(\kappa-A)^{-\alpha}(\kappa-A)^\alpha\,r(t_m,0)\,v_n\\ &\quad+\big(r(t_m,0)-1\big)\;(v_n-v_m)\,.\end{aligned}$$

A further application of (5.2) and (5.5) yields

$$\|b_n-b_m\| \le \left(\frac{C_1^2}{\alpha}\,\mathrm{e}^{\omega_1^+ t_N}\,\mu(\mathbf{v}) + \Big(1+C_1\,\mathrm{e}^{\omega_1^+ t_N}\Big)\;\lambda_\alpha(\mathbf{v})\right)(t_n-t_m)^\alpha t_m^{-\alpha}.$$

This finally concludes the proof of the lemma. □

The following lemma is used in the proof of Theorem 3.2. For the definition of the norm μ_α, we refer to (2.9).

Lemma 5.6. *For $0 < \alpha < 1$ there exists a constant $C_6 > 0$ such that*

$$|||\mathcal{K}(\mathbf{v})|||_{D,\alpha} \le C_6\, \mu_\alpha(\mathbf{v}) \qquad \text{for } \mathbf{v} \in \mathcal{D}^N.$$

The constant C_6 depends on t_N, but it is otherwise independent of the grid.

Proof. Using (5.2) we have for $\mathbf{w} = \mathcal{K}(\mathbf{v})$

$$\begin{aligned}\|w_n\|_D &\le C \sum_{k=1}^{n} h_k \, \big\|(\kappa - A)^{1-\alpha} r(t_n, t_{k-1})\big\|_{X\to X} \, \|(\kappa - A)^\alpha v_k\| \\ &\le C\, \mathrm{e}^{\omega_1^+ t_N}\, \mu_\alpha(\mathbf{v}) \sum_{k=1}^{n} \frac{h_k}{(t_n - t_{k-1})^{1-\alpha}}\end{aligned}$$

and further, by comparing the sum with the corresponding integral,

$$\mu_D(\mathbf{w}) \le \frac{C\, t_N^\alpha}{\alpha}\, \mathrm{e}^{\omega_1^+ t_N} \mu_\alpha(\mathbf{v}).$$

In order to estimate $\lambda_{D,\alpha}(\mathbf{w})$, we split $A(w_n - w_m) = S_1 + S_2$ where

$$S_1 = \sum_{k=m+1}^{n} h_k\, A\, r(t_n, t_{k-1})\, v_k$$

and, due to $r(t_l, t_{k-1}) - r(t_{l-1}, t_{k-1}) = h_l A\, r(t_l, t_k)$, we get

$$S_2 = \sum_{k=1}^{m} h_k\, A\big(r(t_n, t_{k-1}) - r(t_m, t_{k-1})\big)\, v_k = \sum_{k=1}^{m} h_k \sum_{l=m+1}^{n} h_l\, A^2 r(t_l, t_{k-1})\, v_k.$$

As before, we premultiply v_k with $(\kappa - A)^\alpha$ and use (5.2) and the corresponding integrals to estimate S_1 and S_2 by

$$\begin{aligned}\|S_1\| &\le \frac{C\, t_N^\alpha}{\alpha}\, \mathrm{e}^{\omega_1^+ t_N} \mu_\alpha(\mathbf{v})\, (t_n - t_m)^\alpha t_m^{-\alpha}, \\ \|S_2\| &\le \frac{C\, t_N^\alpha}{\alpha(1-\alpha)}\, \mathrm{e}^{\omega_1^+ t_N} \mu_\alpha(\mathbf{v})\, (t_n - t_m)^\alpha t_m^{-\alpha}.\end{aligned}$$

This concludes the proof of the lemma. □

Lemma 5.7. *The inequality*

$$\sum_{k=1}^{m} h_k \sum_{l=m+1}^{n} h_l\, \frac{(t_m - t_k)^\alpha}{(t_l - t_{k-1})^2\, t_k^\alpha} \le \frac{4}{\alpha(1-\alpha)}\, (t_n - t_m)^\alpha t_m^{-\alpha}$$

holds for $1 \le m < n \le N$.

Proof. By comparing with the corresponding integral, we get

$$\sum_{l=m+1}^{n} \frac{h_l}{(t_l - t_{k-1})^2} \le \frac{t_n - t_m}{(t_m - t_{k-1})(t_n - t_{k-1})} \qquad \text{for } 1 \le k \le m.$$

We thus have to estimate

$$(t_n - t_m) \sum_{k=1}^{m} \frac{h_k}{(t_m - t_{k-1})^{1-\alpha}\, t_k^\alpha\, (t_n - t_{k-1})}.$$

For this we consider the function

$$G(s) = \frac{1}{(t_m - s)^{1-\alpha}\, s^{\alpha}\, (t_n - s)}, \qquad 0 < s < t_m.$$

Let t^* be the point where G attains its minimum, and let $1 \le p \le m$ be the index such that $t_{p-1} \le t^* \le t_p$. We split the sum into three parts (from 1 to $p-1$, the term with $k = p$, and from p to m) and compare each part with a corresponding integral. By means of the variable change $\sigma t_m = s$, we get

$$(t_n - t_m) \sum_{k=1}^{m} \frac{h_k}{(t_m - t_{k-1})^{1-\alpha}\, t_k^{\alpha}\, (t_n - t_{k-1})} \le \theta \int_0^1 \frac{\mathrm{d}\sigma}{(1-\sigma)^{1-\alpha}\, \sigma^{\alpha}\, (\theta + 1 - \sigma)},$$

where $\theta = (t_n - t_m)/t_m$. The elementary estimates

$$\theta \int_0^{1/2} \frac{\mathrm{d}\sigma}{(1-\sigma)^{1-\alpha}\, \sigma^{\alpha}\, (\theta + 1 - \sigma)} \le \frac{2\,\theta}{2\,\theta + 1} \frac{1}{1-\alpha} \le \frac{2\,\theta^{\alpha}}{1-\alpha}$$

and

$$\theta \int_{1/2}^1 \frac{\mathrm{d}\sigma}{(1-\sigma)^{1-\alpha}\, \sigma^{\alpha}\, (\theta + 1 - \sigma)} \le 2^{\alpha} \theta^{\alpha} \int_0^{\infty} \frac{\mathrm{d}\tau}{(1+\tau)\tau^{1-\alpha}} \le \frac{\theta^{\alpha}}{\alpha} + \frac{2\,\theta^{\alpha}}{1-\alpha},$$

where we used $1 - \sigma = \theta\tau$, finally prove the lemma. □

Lemma 5.8. *Let $u : [0,T] \to \mathcal{D}$ be a solution of (2.1). Then, for every $\sigma > 0$, there exists a partition $0 = T_0 < T_1 \cdots < T_J = T$ such that*

$$2\|u_{T_j} - u(T_{j-1})\|_{L^{\infty}([0,H_j],D)} + \|u_{T_j}\|_{C_{\alpha}^{\alpha}((0,H_j],D)} \le \sigma, \qquad 1 \le j \le J, \tag{5.13}$$

where $H_j = T_j - T_{j-1}$ and $u_{T_j}(t) = u(T_{j-1} + t)$, $0 \le t \le H_j$.

Proof. Choose $R > 0$ and $L > 0$ as in (2.2) for $u^* = u(0)$. From the proof of Theorem 8.1.1 in [12], it follows that there exist constants $C > 0$ and $0 < H_1 \le T$ such that

$$\begin{aligned} &2\|u - u(0)\|_{L^{\infty}([0,H_1],D)} + \|u\|_{C_{\alpha}^{\alpha}((0,H_1],D)} \\ &\qquad \le C \big\| \big(\mathrm{e}^{(\cdot)A} - 1\big)(Au(0) + f(u(0))) \big\|_{C_{\alpha}^{\alpha}((0,H_1],X)}. \end{aligned} \tag{5.14}$$

The constant C depends on L and the bound is valid as long as $\|u(t) - u(0)\|_D \le R$ for $0 \le t \le H_1$. The right-hand side of (5.14) tends to 0 as H_1 goes to 0. Thus, after a possible reduction of H_1, we get (5.13) with $j = 1$. We go on with this construction, and since the constants L and R can be taken uniformly, the final time T is reached after a finite number of steps. □

References

1. G. Akrivis, M. Crouzeix and C. Makridakis, *Implicit-explicit multistep methods for quasilinear parabolic equations.* Numer. Math. **82** (1999), 521-541. MR **2000e:**65075
2. N. Bakaev, *On variable stepsize Runge-Kutta approximations of a Cauchy problem for the evolution equation.* BIT **38** (1998), 462-485. MR **99i:**65069
3. A. Belleni-Morante and A.C. McBride, *Applied Nonlinear Semigroups. An Introduction.* John Wiley & Sons, Chichester, 1998. MR **99m:**47083
4. K. Eriksson, C. Johnson and S. Larsson, *Adaptative finite element methods for parabolic problems VI: Analytic semigroups.* SIAM J. Numer. Anal. **35** (1998), 1315-1325. MR **99d:**65281
5. C. González and C. Palencia, *Stability of Runge-Kutta methods for quasilinear parabolic problems.* Math. Comp. **69** (2000), 609–628. MR **2000i:**65130
6. E. Hairer and M. Zennaro, *On error growth functions of Runge-Kutta methods.* Appl. Numer. Math. **22** (1996), 205-216. MR **97j:**65116

7. D. Henry, *Geometric Theory of Semilinear Parabolic Equations.* LNM 840, Springer, 1981. MR **83j:**35084
8. M.-N. Le Roux, *Méthodes multipas pour des équations paraboliques non linéaires.* Numer. Math. **35** (1980), 143-162. MR **81i:**65075
9. Ch. Lubich and A. Ostermann, *Runge-Kutta methods for parabolic equations and convolution quadrature.* Math. Comp. **60** (1993), 105-131. MR **93d:**65082
10. Ch. Lubich and A. Ostermann, *Runge-Kutta approximation of quasilinear parabolic equations.* Math. Comp. **64** (1995), 601-627. MR **95g:**65122
11. Ch. Lubich and A. Ostermann, *Linearly implicit time discretization of non-linear parabolic equations.* IMA J. Numer. Anal. **15** (1995), 555-583. MR **96g:**65085
12. A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems.* Birkhäuser, Basel, 1995. MR **96e:**47039
13. E. Nakaguchi and A. Yagi, *Error estimates of implicit Runge-Kutta methods for quasilinear abstract equations of parabolic type.* Japan. J. Math. (N.S.) **25** (1999), 181-226. MR **2000d:**65164
14. C. Palencia, *On the stability of variable stepsize rational approximations of holomorphic semigroups.* Math. Comp. **62** (1994), 93-103. MR **94c:**47066
15. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer, New York, 1983. MR **85g:**47061
16. M.E. Taylor, *Partial Differential Equations III. Nonlinear Equations.* Springer, New York, 1996. MR **98k:**35001
17. M. Thalhammer, *Runge-Kutta time discretization of nonlinear parabolic equations.* Thesis, Universität Innsbruck, 2000.

Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, E-47011 Valladolid, Spain
E-mail address: cesareo@mac.cie.uva.es

Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, Technikerstrasse 13, A-6020 Innsbruck, Austria
Current address: Section de mathématiques, Université de Genève, rue du Lièvre 2–4, CH-1211 Genève 24, Switzerland
E-mail address: Alexander.Ostermann@uibk.ac.at, Alexander.Ostermann@math.unige.ch

Departamento de Matemática Aplicada y Computación, Universidad de Valladolid, Facultad de Ciencias, E-47011 Valladolid, Spain
E-mail address: palencia@mac.cie.uva.es

Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, Technikerstrasse 13, A-6020 Innsbruck, Austria
E-mail address: Mechthild.Thalhammer@uibk.ac.at

1.3. Runge-Kutta methods for nonlinear parabolic equations

Convergence of Runge-Kutta methods for nonlinear parabolic equations

ALEXANDER OSTERMANN AND MECHTHILD THALHAMMER

Applied Numerical Mathematics (2002) 42, 367-380

ELSEVIER

Applied Numerical Mathematics 42 (2002) 367–380

APPLIED NUMERICAL MATHEMATICS

www.elsevier.com/locate/apnum

Convergence of Runge–Kutta methods for nonlinear parabolic equations

Alexander Ostermann *, Mechthild Thalhammer

Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, Technikerstrasse 13, A-6020 Innsbruck, Austria

Abstract

In this paper, we study time discretizations of fully nonlinear parabolic differential equations. Our analysis uses the fact that the linearization along the exact solution is a uniformly sectorial operator. We derive smooth and nonsmooth-data error estimates for the backward Euler method, and we prove convergence for strongly $A(\vartheta)$-stable Runge–Kutta methods. For the latter, the order of convergence for smooth solutions is essentially determined by the stage order of the method. Numerical examples illustrating the convergence estimates are presented. © 2001 IMACS. Published by Elsevier Science B.V. All rights reserved.

Keywords: Fully nonlinear parabolic problems; Time discretization; Backward Euler method; Runge–Kutta methods; Convergence estimates; Non-smooth data error estimates

1. Introduction

The aim of the present paper is to derive existence and convergence results for Runge–Kutta time discretizations of the abstract differential equation

$$u'(t) = f\big(t, u(t)\big), \quad u(0) = u_0. \tag{1}$$

The precise assumptions on the nonlinearity f are given in Section 2 below. Our interest in this abstract initial value problem stems from the fact that fully nonlinear parabolic initial-boundary value problems can be cast in this form. Such problems arise in various fields of applications as for example in combustion theory, differential geometry, and stochastic control theory. Moreover, semilinear problems with free boundaries may be reduced to this form.

* Corresponding author.
E-mail addresses: Alexander.Ostermann@uibk.ac.at (A. Ostermann), Mechthild.Thalhammer@uibk.ac.at (M. Thalhammer).

PII: S0168-9274(01)00161-1

The existence and regularity theory for fully nonlinear parabolic problems has been developed in recent years and is summarized in the monograph [12]. Whereas the literature on numerical discretizations of semilinear and quasilinear parabolic problems is quite rich, see, e.g., [1,8,10,11,14], not that much is known for the fully nonlinear case. We are aware of the following two references only: In [3] the convergence of a full discretization, based on the forward Euler method and standard finite differences is studied. Due to the stiffness of the problem, this involves a severe restriction on the admissible stepsizes. The second reference is our recent paper [6], where we took up the analytical framework of [12] to obtain convergence results for variable stepsize backward Euler discretizations of (1).

In the present paper, we consider a slightly different approach that avoids the complicated weighted Hölder norms encountered in [12,6]. The main idea is to linearize the problem along the exact solution $u(t)$ to get

$$u'(t) = A(t)u(t) + g\big(t, u(t)\big), \quad u(0) = u_0. \tag{2}$$

Note that Runge–Kutta methods are invariant under this linearization. Since the Fréchet derivative of g with respect to the second variable vanishes along the exact solution, techniques from the semilinear case like the variation-of-constants formula can be used. Consequently, stability bounds for discretizations of the nonautonomous problem

$$w'(t) = A(t)w(t) \tag{3}$$

are indispensable. For Runge–Kutta methods with constant stepsizes, such results have been provided recently by [5].

The paper is organized as follows: In Section 2 we give the precise assumptions that render the initial value problem (1) parabolic. We also present an example from detonation theory that fits into this analytical framework.

Section 3 is devoted to the existence and convergence of backward Euler approximations. We show that the expected order 1 is attained for smooth solutions on bounded time intervals. For nonsmooth initial data, the order of convergence is still one on compact time intervals that are bounded away from $t = 0$. However, an order reduction takes place for $t \to 0$, see Theorem 5 below. For the convenience of the reader and for the sake of completeness, we have also included a new and short proof of the above mentioned stability result.

In Section 4, we prove the convergence of strongly $A(\vartheta)$-stable Runge–Kutta discretizations under the assumption that the exact solution is sufficiently smooth. The attained order of convergence turns out to be $\min(p, q+1)$, where p and q denote the order and the stage order of the method, respectively. This order reduction is expected, since it appears already for semilinear problems, see [10].

In Section 5, we explain how our results carry over to variable stepsizes.

A numerical experiment is finally presented in Section 6. We illustrate therein our convergence results for the backward Euler method with constant stepsizes at the aforementioned detonation problem. We have also performed more realistic calculations using the 3-stage Radau IIA method. This was partly done to obtain a good approximation to the exact solution in the above experiment. We used the variable stepsize implementation RADAU5 by Hairer and Wanner [7] that gave very reliable results in all tests.

2. Problem class and example

In our subsequent analysis of time discretizations of (1), we use a simplified version of the analytical framework given in [12]. For the convenience of the reader, we resume the precise hypotheses for (1) in this section. More details are found in Lunardi's monograph [12].

Let $(X, |\cdot|)$ and $(D, \|\cdot\|)$ be two Banach spaces with D densely embedded in X, and denote by $\mathcal{D}$ an open subset of D. We consider the abstract initial value problem

$$u'(t) = f\big(t, u(t)\big), \quad t > 0, \qquad u(0) = u_0 \in \mathcal{D}, \tag{4}$$

where the right-hand side satisfies the following assumption.

Assumption 1. The function $f : [0, T] \times \mathcal{D} \to X$ is twice continuously Fréchet differentiable and its Fréchet derivative $D_2 f(t, v)$ with respect to the second variable is sectorial in X. Moreover, the graph-norm of $D_2 f(t, v)$ is equivalent to the norm of D for all $0 \leqslant t \leqslant T$ and for all $v \in \mathcal{D}$.

We further impose the following condition on the initial value. For a definition of the real interpolation space $(X, D)_{\alpha,\infty}$, we refer to [12, Section 1.2] and [16].

Assumption 2. The initial value $u_0 \in \mathcal{D}$ satisfies $f(0, u_0) \in (X, D)_{\alpha,\infty}$ for some $0 < \alpha < 1$.

Under these assumptions, the existence of a locally unique solution of (4) can be shown. Since the regularity properties of this solution are essential for our analysis, we collect them in the following lemma.

Lemma 3. *Under the above assumptions and after a possible reduction of T, problem* (4) *has a unique solution u which is twice differentiable on* $(0, T]$ *and satisfies*

$$u \in C^{\alpha}\big([0, T], D\big) \cap C^{1+\alpha}\big([0, T], X\big),$$
$$t^{1-\alpha} u' \in B\big([0, T], D\big), \quad \text{and} \quad t^{1-\alpha} u'' \in B\big([0, T], X\big).$$

We note that the size of T in general depends on u_0.

As usual, $C^{\alpha}([0, T], D)$ denotes the Banach space of α-Hölder continuous functions on $[0, T]$ with values in D, and $B([0, T], D)$ denotes the corresponding space of bounded functions. Both spaces are endowed with the usual norms.

Proof of Lemma 3. The existence and α-Hölder continuity of u and its derivative is proved in [12, Theorem 8.1.3]. The boundedness of $t^{1-\alpha} u'(t)$ in D is a consequence of [13, Theorem 2.2], and that of $t^{1-\alpha} u''(t)$ in X finally follows from the identity

$$u''(t) = D_1 f\big(t, u(t)\big) + D_2 f\big(t, u(t)\big) u'(t), \quad 0 < t \leqslant T,$$

together with $D_2 f(t, u(t)) \in C([0, T], L(D, X))$. □

We close this section with an example of a nonlinear initial-boundary value problem from detonation theory. More examples that fit into our framework can be found in [6,12] and references therein.

Example 4 *Displacement of a shock, see* [4,12]. The following fully nonlinear problem arises in detonation theory and describes the displacement of a shock

$$\begin{aligned} &\partial_t U(t,x) = \log\left(\frac{\exp(aU(t,x)\partial_{xx}U(t,x))-1}{a\partial_{xx}U(t,x)}\right) - \frac{1}{2}\big(\partial_x U(t,x)\big)^2, \\ &\partial_x U(t,0) = \partial_x U(t,1) = 0, \quad U(0,x) = U_0(x), \quad 0 < x < 1,\ t > 0. \end{aligned} \tag{5}$$

Here a denotes a positive constant.

Choosing $X = C([0,1])$ and $D = \{v \in C^2([0,1])\colon\ v'(0) = v'(1) = 0\}$ allows us to write (5) in the abstract form (4) with $u(t) = U(t,\cdot)$ and

$$f(t,v) = \log\left(\frac{\exp(avv'')-1}{av''}\right) - \frac{1}{2}\big(v'\big)^2. \tag{6}$$

Note that the right-hand side of (6) is analytic, if we restrict the domain to the set

$$\mathcal{D} = \big\{v \in D\colon\ v(x) > 0 \text{ for } 0 \leqslant x \leqslant 1\big\}.$$

It is verified in [12, Section 8.5.1] that problem (5) enters our framework for $U_0 \in \mathcal{D}$.

We finally remark that in the present example

$$(X, D)_{\alpha,\infty} = \begin{cases} C^{2\alpha}\big([0,1]\big), & \alpha < \frac{1}{2}, \\ C_0^{2\alpha}\big([0,1]\big), & \alpha > \frac{1}{2}, \end{cases} \tag{7}$$

where

$$C_0^{1+\gamma}\big([0,1]\big) = \big\{v \in C^{1+\gamma}\big([0,1]\big)\colon\ v'(0) = v'(1) = 0\big\}$$

for $\gamma \geqslant 0$. This follows from [12, Theorem 3.1.30 and Proposition 2.2.2]. For a smooth function in $\mathcal{D}$ that does not necessarily satisfy unnatural boundary conditions, we can thus take any α smaller than $1/2$.

3. Backward Euler discretization

In this section we give two convergence results for the backward Euler discretization of the initial value problem (4). We decided to treat the backward Euler method separately from general Runge–Kutta methods for the following two reasons: Firstly, this method is of great importance in applications and secondly, the proofs are much less involved than for general Runge–Kutta methods. Therefore, the underlying ideas can be perceived more easily.

Let $h > 0$ denote the stepsize. The backward Euler approximation u_{n+1} to the exact solution u of (4) at $t_{n+1} = (n+1)h$ is given by the recursion

$$\frac{u_{n+1} - u_n}{h} = f(t_{n+1}, u_{n+1}), \quad n \geqslant 0. \tag{8}$$

Our first convergence result can be seen as an error bound in terms of the data. Note that the imposed assumptions can easily be checked in applications.

Theorem 5 Error estimate in terms of the data. *Under Assumptions* 1 *and* 2, *and for* T *as in Lemma* 3, *there exists* $H > 0$ *such that for all stepsizes* $0 < h \leqslant H$ *the following holds. The backward Euler solution*

of (4) *is well-defined in a neighbourhood of the exact solution, and the difference between numerical and exact solution is bounded by*

$$\|u_n - u(t_n)\| \leqslant Ct_n^{\alpha-1}h\big(1 + |\log h|\big), \quad 0 < nh \leqslant T. \tag{9}$$

The constant C in general depends on T, but is independent of n and h.

In situations where it is known in advance that the exact solution has more smoothness, the above bound can be sharpened. We have the following result.

Theorem 6 Error estimate in terms of the solution. *Let Assumption* 1 *hold, and assume that the exact solution u of* (4) *satisfies $u \in C^\beta([0,T], D)$ for some $\beta > 0$, and $u'' \in B([0,T], X)$. Then, there exists $H > 0$ such that for all stepsizes $0 < h \leqslant H$ the following holds. The backward Euler solution of* (4) *is well-defined in a neighbourhood of the exact solution, and the difference between numerical and exact solution is bounded by*

$$\|u_n - u(t_n)\| \leqslant Ch\big(1 + |\log h|\big), \quad 0 \leqslant nh \leqslant T. \tag{10}$$

The constant C in general depends on T, but is independent of n and h.

Our main technique for proving both theorems is to linearize (4) along the exact solution. Setting

$$A(t) = D_2 f\big(t, u(t)\big) \quad \text{and} \quad g(t,v) = f(t,v) - A(t)v, \tag{11a}$$

we arrive at the formally semilinear problem

$$u'(t) = A(t)u(t) + g\big(t, u(t)\big), \quad t > 0. \tag{11b}$$

Due to our assumptions and Lemma 3, we know that

$$A \in C^\alpha\big([0,T], L(D,X)\big). \tag{12}$$

Since the backward Euler method is invariant under the above linearization, we obtain from (8) the following representation of the numerical solution

$$\frac{u_{n+1} - u_n}{h} = A(t_{n+1})u_{n+1} + g(t_{n+1}, u_{n+1}), \quad n \geqslant 0. \tag{13}$$

In order to analyze this recursion, stability bounds are all-important. Henceforth, we write $A_n = A(t_n)$ for short, and we use the following notation for the discrete evolution operators

$$R(t_n, t_j) = (I - hA_n)^{-1} \cdots (I - hA_{j+1})^{-1}, \quad 0 \leqslant j < n,$$

with $R(t_n, t_n) = I$. Due to Assumption 1, these operators are well-defined and bounded for h sufficiently small. Moreover, we have the following stability estimates.

Lemma 7. *Under condition* (12)*, there exists $H > 0$ such that for all stepsizes $0 < h \leqslant H$ we have*

$$\big\|R(t_n, t_j)\big\|_{D \leftarrow X} \leqslant C\big(t_{n-j}^{-1} + |\log h| t_{n-j}^{\alpha-1}\big), \quad 0 < t_j < t_n \leqslant T. \tag{14}$$

The constant C in general depends on T, but is independent of n and h.

A slightly stronger estimate that avoids the $|\log h|$ term follows from [5, Theorem 1.1]. In order to keep this section self-contained, and since our proof of (14) is very short, we decided to give it at the end of this section. We remark that under the condition

$$\left\| A_0^{\varepsilon}(A_{j+1} - A_0)A_0^{-1-\varepsilon} \right\| \leqslant Ct_{j+1}^{\alpha} \quad \text{for some } \varepsilon > 0,$$

the $|\log h|$ term does not appear in our proof. This condition is often satisfied in applications.

We are now in the position to prove the two theorems.

Proof of Theorem 5. Inserting the exact solution $\hat{u}_n = u(t_n)$ into the numerical scheme (13) gives

$$\frac{\hat{u}_{n+1} - \hat{u}_n}{h} = A_{n+1}\hat{u}_{n+1} + g\bigl(t_{n+1}, \hat{u}_{n+1}\bigr) + \delta_{n+1}. \tag{15}$$

This recursion differs from (13) by the defects

$$\delta_{n+1} = \int_0^1 \bigl(u'(t_n + \tau h) - u'(t_{n+1})\bigr)\,\mathrm{d}\tau.$$

As a direct consequence of Lemma 3, the defects are bounded by

$$|\delta_1| \leqslant Ch^{\alpha} \quad \text{and} \quad |\delta_{n+1}| \leqslant Cht_n^{\alpha-1}, \quad n \geqslant 1, \tag{16}$$

where the constants depend on the first and second derivatives of u.

The backward Euler solution of (4) is constructed by fixed-point iteration. Let N be defined by $Nh \leqslant T < (N+1)h$, and let

$$\mathcal{D}_h = \left\{ v = (v_n)_{n=1}^N \in \mathcal{D}^N \colon \sup_{1\leqslant n\leqslant N} t_n^{1-\alpha} \bigl\| v_n - u(t_n) \bigr\| \leqslant c_0 h^{\gamma} \right\} \tag{17a}$$

with suitably chosen constants $c_0 > 0$ and $1 - \alpha < \gamma < 1$. For h sufficiently small, this is a closed subset of the space D^N, endowed with the weighted norm

$$\|v\|_{\infty} = \sup_{1\leqslant n\leqslant N} t_n^{1-\alpha} \|v_n\|, \quad v \in D^N. \tag{17b}$$

We consider the mapping $\Phi : \mathcal{D}_h \to D^N$, defined by

$$\bigl(\Phi(v)\bigr)_n = R(t_n, 0)u_0 + h\sum_{j=0}^{n-1} R(t_n, t_j)g(t_{j+1}, v_{j+1}).$$

Our aim is to show that Φ is a contraction on $\mathcal{D}_h$. By construction, the fixed-point of Φ is the searched backward Euler solution.

From the definition of g, we deduce

$$g(t_j, v_j) - g(t_j, w_j) = \int_0^1 \bigl(D_2 f\bigl(t_j, \tau v_j + (1-\tau)w_j\bigr) - A_j\bigr)\,\mathrm{d}\tau \cdot (v_j - w_j),$$

which implies the bound

$$\bigl|g(t_j, v_j) - g(t_j, w_j)\bigr| \leqslant c_0 L h^{\gamma+\alpha-1} \|v_j - w_j\|. \tag{18}$$

Note that the Lipschitz constant L of $D_2 f$ can be chosen here independently of j. We next make use of the relations

$$h\sum_{j=1}^{n-1} t_{n-j}^{\beta-1} t_j^{\alpha-1} \leqslant \begin{cases} Ct_n^{\alpha-1}|\log n|, & \beta = 0, \\ Ct_n^{\alpha+\beta-1}, & 0 < \beta < 1, \end{cases} \tag{19}$$

that are obtained in a standard way by comparing the sum with the corresponding integral. Together with (14) and (18), we get

$$\begin{aligned} \big\|\big(\Phi(v)\big)_n - \big(\Phi(w)\big)_n\big\| &\leqslant h\sum_{j=0}^{n-1} \big\|R(t_n, t_j)\big\|_{D\leftarrow X} \big|g(t_{j+1}, v_{j+1}) - g(t_{j+1}, w_{j+1})\big| \\ &\leqslant c_0 c_1 L t_n^{\alpha-1}\big(1 + |\log h|\big) h^{\gamma+\alpha-1}\|v - w\|_\infty, \end{aligned}$$

where c_1 is a constant that depends on the stability constant of Lemma 7, and on T. This proves that Φ is contractive

$$\big\|\Phi(v) - \Phi(w)\big\|_\infty \leqslant \kappa\|v - w\|_\infty$$

with an h-independent factor $\kappa < 1$ for h sufficiently small.

In order to verify that Φ maps $\mathcal{D}_h$ onto $\mathcal{D}_h$, we exploit

$$\big\|\Phi(v) - \hat{u}\big\|_\infty \leqslant \kappa\big\|v - \hat{u}\big\|_\infty + \big\|\hat{u} - \Phi\big(\hat{u}\big)\big\|_\infty \leqslant \kappa c_0 h^\gamma + \big\|\hat{u} - \Phi\big(\hat{u}\big)\big\|_\infty.$$

It thus remains to show that

$$\big\|\hat{u} - \Phi\big(\hat{u}\big)\big\|_\infty \leqslant (1 - \kappa)c_0 h^\gamma. \tag{20}$$

With the help of (14), (16), and (19), we obtain

$$t_n^{\alpha-1}\big\|\hat{u}_n - \Phi\big(\hat{u}\big)_n\big\| = t_n^{\alpha-1}\left\|h\sum_{j=0}^{n-1} R(t_n, t_j)\delta_{j+1}\right\| \leqslant Ch\big(1 + |\log h|\big). \tag{21}$$

The desired bound (20) can thus be achieved for $\gamma < 1$.

Since Φ is a contraction on $\mathcal{D}_h$, the numerical solution $u^* = (u_n)_{n=1}^N$ exists as the unique fixed-point of Φ. Moreover, we have the preliminary convergence result

$$\big\|u_n - u(t_n)\big\| \leqslant c_0 t_n^{\alpha-1} h^\gamma, \quad 0 < nh \leqslant T.$$

In order to show the convergence estimate (9), we use again (21)

$$\begin{aligned} t_n^{1-\alpha}\big\|u_n - \hat{u}_n\big\| &\leqslant \big\|u^* - \hat{u}\big\|_\infty \leqslant \big\|\Phi\big(u^*\big) - \Phi\big(\hat{u}\big)\big\|_\infty + \big\|\hat{u} - \Phi\big(\hat{u}\big)\big\|_\infty \\ &\leqslant \kappa\big\|u^* - \hat{u}\big\|_\infty + Ch\big(1 + |\log h|\big). \end{aligned}$$

Since $\kappa < 1$, this implies (9) and concludes our proof. □

Proof of Theorem 6. This proof is very similar to the preceding one. It is essentially obtained by setting $\alpha = 1$ there. We omit the details. □

Proof of Lemma 7. Since we are working on an equidistant grid, it is sufficient to consider the case $j = 0$. The idea of the proof consists in comparing the time-dependent operator $R(t_n, 0)$ with the frozen

operator $(I - hA_0)^{-n}$. For the latter, stability estimates are well-established, see [9, Estimate (3.31)]. We will use below that

$$\left\| A_0(I - hA_0)^{-n} \right\|_{X \leftarrow X} \leqslant Ct_n^{-1}, \quad 0 < nh \leqslant T, \tag{22}$$

holds with a constant C that depends on T, but not on n or h. Let

$$\Delta_j = A_0\big(R(t_n, t_{n-j}) - (I - hA_0)^{-j}\big), \quad 1 \leqslant j \leqslant n.$$

Expanding Δ_n into a telescopic sum and using the resolvent identity

$$(I - hA_{j+1})^{-1} - (I - hA_0)^{-1} = h(I - hA_{j+1})^{-1}(A_{j+1} - A_0)(I - hA_0)^{-1}$$

gives the recursion

$$\begin{aligned}
\Delta_n &= h \sum_{j=0}^{n-1} A_0 R(t_n, t_j)(A_{j+1} - A_0)(I - hA_0)^{-j-1} \\
&= h \sum_{j=0}^{n-1} \Delta_{n-j} \cdot (A_{j+1} - A_0)A_0^{-1} \cdot A_0(I - hA_0)^{-j-1} \\
&\quad + h \sum_{j=0}^{n-1} A_0(I - hA_0)^{j-n} \cdot (A_{j+1} - A_0)A_0^{-1} \cdot A_0(I - hA_0)^{-j-1}.
\end{aligned} \tag{23}$$

Taking norms in (23), and using (22) and (19), we arrive at

$$\|\Delta_n\|_{X \leftarrow X} \leqslant Ch \sum_{j=0}^{n-1} t_{j+1}^{\alpha-1} \|\Delta_{n-j}\|_{X \leftarrow X} + Ct_n^{\alpha-1}\big(1 + |\log h|\big).$$

Solving this Gronwall-type inequality and using once more (22) proves the desired result. □

4. Runge–Kutta discretizations

In this section we generalize the convergence result of Theorem 6 to general Runge–Kutta methods. We show below that, under certain smoothness assumptions on the exact solution and stability requirements on the method, the convergence behaviour on finite time intervals is essentially governed by the stage order of the numerical method.

An s-stage *Runge–Kutta method* applied to (4) with stepsize $h > 0$, is given by the scheme

$$\begin{aligned}
U'_{ni} &= f(t_n + c_i h, U_{ni}), \quad U_{ni} = u_n + h \sum_{j=1}^{s} a_{ij} U'_{nj}, \quad 1 \leqslant i \leqslant s, \\
u_{n+1} &= u_n + h \sum_{i=1}^{s} b_i U'_{ni}, \quad n \geqslant 0,
\end{aligned} \tag{24}$$

where $a_{ij}, b_i, c_i \in \mathbb{R}$ are the coefficients of the method.

In the sequel we introduce the basic notions of order and stability. For details we refer to the monograph [7]. Recall that the Runge–Kutta method (24) has *order* p if the error fulfills the relation

$u_n - u(t_n) = \mathcal{O}(h^p)$ for $h \to 0$, uniformly on bounded time intervals, whenever the method is applied to an ordinary differential equation with sufficiently smooth right-hand side; the method has *stage order* q whenever the internal stages satisfy $U_{0i} - u(c_i h) = \mathcal{O}(h^q)$ as $h \to 0$ for all $1 \leqslant i \leqslant s$. We always assume $p \geqslant 1$.

For specifying the stability requirements on the numerical method, it is useful to introduce the matrix and vector notation

$$\mathcal{A} = (a_{ij})_{i,j=1}^{s}, \qquad \mathbb{1} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^s, \qquad b = (b_1, \ldots, b_s)^{\mathrm{T}}.$$

Then the *stability function* of (24) is defined through

$$R(z) = 1 + z b^{\mathrm{T}} (I - z\mathcal{A})^{-1} \mathbb{1}.$$

The Runge–Kutta method is $A(\vartheta)$-*stable* if $I - z\mathcal{A}$ is invertible on the sector $M_\vartheta = \{z \in \mathbb{C}\colon |\arg(-z)| \leqslant \vartheta\}$ and if $|R(z)| \leqslant 1$ holds for all $z \in M_\vartheta$; the method is called *strongly* $A(\vartheta)$-*stable* if additionally $\mathcal{A}$ is invertible and the module of R at infinity, $R(\infty) = 1 - b^{\mathrm{T}} \mathcal{A}^{-1} \mathbb{1}$, is strictly smaller than one.

Our analysis is in the lines of Section 3 and uses the fact that the derivative $A(t) = D_2 f(t, u(t))$ along the exact solution is uniformly sectorial on $[0, T]$. This follows from the Hölder continuity of u. Thus there are constants $M > 0$, $a \in \mathbb{R}$ and $0 < \varphi < \pi/2$ such that the resolvent estimate

$$\big|\big(\lambda - A(t)\big)^{-1}\big|_{X \leftarrow X} \leqslant \frac{M}{|\lambda - a|} \quad \text{for } \big|\arg(\lambda - a)\big| \leqslant \pi - \varphi \tag{25}$$

uniformly holds for $0 \leqslant t \leqslant T$.

Now we are ready to state the convergence result for Runge–Kutta methods.

Theorem 8 Error estimate in terms of the solution. *Let Assumption* 1 *hold and apply a Runge–Kutta method of order* p *and stage order* q *to* (4)*. Assume further that the exact solution has the regularity properties* $u^{(r)} \in B([0, T], D)$ *and* $u^{(r+1)} \in B([0, T], X)$ *with* $r = \min(p, q + 1)$*, and that the method is strongly* $A(\vartheta)$*-stable with* $\vartheta > \varphi$*, where* φ *is given by* (25)*. Then there exists* $H > 0$ *such that for* $0 < h \leqslant H$ *the numerical solution* u_n *and the internal stages* U_{ni} *of the Runge–Kutta method exist for all* n *with* $0 \leqslant nh \leqslant T$ *and satisfy*

$$\big\|u_n - u(t_n)\big\| + \max_{1 \leqslant i \leqslant s} \big\|U_{ni} - u(t_n + c_i h)\big\| \leqslant C h^r \big(1 + |\log h|\big), \quad 0 \leqslant nh \leqslant T.$$

The constant C *in general depends on* T*, but not on* n *or* h*.*

Although the requirement of strong stability excludes the Gauss–Legendre methods, the assumptions of Theorem 8 are still satisfied by many interesting classes of Runge–Kutta methods: The s-stage Radau IIA methods satisfy the assumptions with $p = 2s - 1$ and $q = s$, the s-stage Lobatto IIIC methods with $p = 2s - 2$ and $q = s - 1$. Both classes are strongly $A(\pi/2)$-stable with $R(\infty) = 0$, see [7, Chapter IV.5].

Proof of Theorem 8. For simplicity, we give the proof only for the case where $R(\infty) = 0$ and henceforth suppose $c_i \in [0, 1]$ for all $1 \leqslant i \leqslant s$. For a more general proof, we refer to [15].

In order to write the Runge–Kutta scheme more compactly, it is useful to introduce some notation

$$U_n = (U_{n1}, \ldots, U_{ns})^{\mathrm{T}}, \quad f_{n+1}(U_n) = \big(f(t_n + c_i h, U_{ni})\big)_{i=1}^{s}, \quad \text{etc.}$$

With the help of these abbreviations, (24) takes the form

$$U_n' = f_{n+1}(U_n), \qquad U_n = \mathbb{1} u_n + h\mathcal{A} U_n', \qquad u_{n+1} = u_n + h b^{\mathrm{T}} U_n'. \tag{26}$$

Here, the matrix $\mathcal{A}$ is considered as a linear operator on X^s and the ith component of $\mathcal{A}U'_n$ is thus given by $\sum_{j=1}^s a_{ij}U'_{nj}$.

Our analysis follows the ideas of Section 3 and relies on the consideration of the formally semilinear equation (11). Let

$$A_{n+1} = \operatorname{diag}\bigl(A(t_n + c_1 h), \ldots, A(t_n + c_s h)\bigr).$$

Due to the resolvent condition (25) and the $A(\vartheta)$-stability of the method, the operators

$$J_{n+1} = (I - h\mathcal{A}A_{n+1})^{-1} \quad \text{and} \quad K_{n+1} = (I - hA_{n+1}\mathcal{A})^{-1}$$

are well-defined and bounded for h sufficiently small.

In this notation, the stages are given by

$$U_n = J_{n+1}\mathbb{1}u_n + hJ_{n+1}\mathcal{A}g_{n+1}(U_n), \tag{27a}$$

and the Runge–Kutta solution has the representation

$$u_{n+1} = R(hA_{n+1})u_n + hb^{\mathrm{T}}K_{n+1}g_{n+1}(U_n), \quad n \geqslant 0,$$

with the stability function

$$R(hA_{n+1}) = 1 + hb^{\mathrm{T}}A_{n+1}(I - h\mathcal{A}A_{n+1})^{-1}\mathbb{1}.$$

Solving this recursion for u_n yields furthermore

$$u_n = R(t_n, 0)u_0 + h\sum_{j=0}^{n-1} R(t_n, t_{j+1})b^{\mathrm{T}}K_{j+1}g_{j+1}(U_j), \quad n \geqslant 0, \tag{27b}$$

where

$$R(t_n, t_j) = R(hA_n)\cdots R(hA_{j+1}), \quad 0 \leqslant j < n, \qquad R(t_n, t_n) = I$$

denote the discrete transition operators. Due to the validity of (12), they satisfy the stability estimate

$$\bigl\|R(t_n, t_j)\bigr\|_{D\leftarrow X} \leqslant Ct_{n-j}^{-1}, \quad 0 < t_j < t_n \leqslant T, \tag{28}$$

for sufficiently small stepsizes $0 < h \leqslant H$, see [5, Theorem 1.1]. The constant C depends on T, but not on h or n.

Inserting the exact solution $\hat{u}_n = u(t_n)$ and $\widehat{U}_n = (u(t_n + c_i h))_{i=1}^s$ into the Runge–Kutta scheme (26) yields

$$\widehat{U}'_n = f_{n+1}\bigl(\widehat{U}_n\bigr), \quad \widehat{U}_n = \mathbb{1}\hat{u}_n + h\mathcal{A}\widehat{U}'_n + \Delta_n, \quad \hat{u}_{n+1} = \hat{u}_n + hb^{\mathrm{T}}\widehat{U}'_n + \delta_{n+1}, \tag{29}$$

where the defects are given by

$$\delta_{n+1} = h^{k+1}\int_0^1 \frac{(1-\tau)^{k-1}}{k!}\left((1-\tau)u^{(k+1)}(t_n + \tau h) - k\sum_{j=1}^s b_j c_j^k u^{(k+1)}(\tau_{nj})\right)\mathrm{d}\tau,$$

$$\Delta_{ni} = h^r\int_0^1 \frac{(1-\tau)^{r-2}}{(r-1)!}\left((1-\tau)c_i^r u^{(r)}(\tau_{ni}) - (r-1)\sum_{j=1}^s a_{ij}c_j^{r-1}u^{(r)}(\tau_{nj})\right)\mathrm{d}\tau,$$

with $k = r - 1$ or $k = r$ and $\tau_{ni} = t_n + \tau c_i h$. Consequently we have

$$|\delta_{n+1}| \leqslant Ch^{r+1}, \qquad \|\delta_{n+1}\| \leqslant Ch^r, \qquad \|\Delta_{ni}\| \leqslant Ch^r \tag{30}$$

with constants depending on the method and the derivatives of u of order r and $r+1$.

For the construction of the internal stages we use a fixed-point iteration Ψ based on (27). It maps a sequence $V = (V_n)_{n=0}^N$ in $\mathcal{D}$ to another sequence $\Psi(V)$ with components

$$\begin{aligned}\big(\Psi(V)\big)_n &= J_{n+1}\mathbb{1}R(t_n,0)u_0 + hJ_{n+1}\mathcal{Q}g_{n+1}(V_n)\\ &\quad + h\sum_{j=0}^{n-1} J_{n+1}\mathbb{1}R(t_n,t_{j+1})b^{\mathrm{T}}K_{j+1}g_{j+1}(V_j).\end{aligned} \tag{31}$$

For some $c_0 > 0$ and $0 < \gamma < 1$ we choose the set

$$\mathcal{D}_h = \big\{V = (V_n)_{n=0}^N \in \mathcal{D}^{(N+1)s}\colon\ \big\|V - \widehat{U}\big\|_\infty \leqslant c_0 h^\gamma\big\}$$

as domain of Ψ and endow it with the norm

$$\|V\|_\infty = \sup_{0\leqslant n\leqslant N} \|V_n\|, \quad \text{where } \|V_n\| = \max_{1\leqslant i\leqslant s} \|V_{ni}\|.$$

Here, N is defined through $(N+1)h \leqslant T < (N+2)h$.

We will show next that Ψ is contractive with contraction factor $\kappa < 1$ for sufficiently small stepsizes. For this, we use the corresponding estimate to (18)

$$\big|g_{j+1}(V_j) - g(t_{j+1}, W_j)\big| \leqslant c_0 L h^\gamma \|V_j - W_j\|. \tag{32}$$

With the help of the stability result (28) and (32), we thus receive

$$\big\|\big(\Psi(V)\big)_n - \big(\Psi(W)\big)_n\big\| \leqslant c_0 c_1 L\big(1 + |\log h|\big)h^\gamma \|V - W\|_\infty,$$

with c_1 depending on the quantity C from (28). This proves the contractivity of Ψ for sufficiently small h.

From formula (29) and the definition of Ψ we further get

$$\begin{aligned}\big\|\widehat{U}_n - \big(\Psi\big(\widehat{U}\big)\big)_n\big\| &\leqslant \sum_{j=0}^{n-1}\big\|J_{n+1}\mathbb{1}R(t_n,t_{j+1})\big\|_{D\leftarrow X}|\delta_{j+1}| + \|J_{n+1}\|_{D\leftarrow D}\|\Delta_n\|\\ &\quad + h\sum_{j=0}^{n-1}\big\|J_{n+1}\mathbb{1}R(t_n,t_{j+1})b^{\mathrm{T}}K_{j+1}\big\|_{D\leftarrow X}|A_{j+1}\Delta_j|.\end{aligned}$$

Applying the bounds (28) and (30) yields

$$\big\|\widehat{U} - \Psi\big(\widehat{U}\big)\big\|_\infty \leqslant Ch^r\big(1 + |\log h|\big). \tag{33}$$

An argument similar to that in the proof of Theorem 5 thus shows $\Psi(\mathcal{D}_h) \subset \mathcal{D}_h$.

The convergence estimate for the internal steps now follows directly from the contractivity of Ψ and (33)

$$\big\|U_n - \widehat{U}_n\big\| \leqslant \big\|U - \widehat{U}\big\|_\infty \leqslant \frac{1}{1-\kappa}\big\|\widehat{U} - \Psi\big(\widehat{U}\big)\big\|_\infty \leqslant Ch^r\big(1 + |\log h|\big). \tag{34}$$

In order to estimate the error between the numerical and the exact solution, we use the relation

$$u_{n+1} - \hat{u}_{n+1} = \big(1 - b^{\mathrm{T}}\mathcal{Q}^{-1}\mathbb{1}\big)\big(u_n - \hat{u}_n\big) + b^{\mathrm{T}}\mathcal{Q}^{-1}\big(U_n - \widehat{U}_n + \Delta_n\big) - \delta_{n+1}.$$

Due to our assumption $R(\infty) = 0$, the desired result follows at once from (30) and (34). □

5. Variable stepsizes

In order to keep the presentation as simple as possible, we have focused our attention in the previous sections to constant stepsizes. This limitation, however, is not necessary and the results there hold for variable stepsize sequences as well. The reason for this is quite simple: the techniques employed in our proofs are either based on fixed-point iteration or rely on the comparison of Riemann-sums with their corresponding integrals. Obviously, their use is not limited to constant stepsizes.

Although the generalization to variable stepsizes is straightforward, we briefly describe how the variable stepsize version of our stability lemma comes about. For this, we need some additional notation. Let $t_0 = 0 < t_1 < \cdots < t_N$ be the given grid and denote by

$$h_n = t_n - t_{n-1}, \quad 1 \leqslant n \leqslant N,$$

the corresponding stepsizes. As in Section 3, we define the discrete evolution operators

$$R(t_n, t_j) = (I - h_n A_n)^{-1} \cdots (I - h_{j+1} A_{j+1})^{-1}, \quad 0 \leqslant j < n \leqslant N,$$

as well as their counterparts with frozen arguments

$$r(t_n, t_j) = (I - h_n A_j)^{-1} \cdots (I - h_{j+1} A_j)^{-1}, \quad 0 \leqslant j < n \leqslant N.$$

Further, let

$$\Delta_{nj} = A_j \big(R(t_n, t_j) - r(t_n, t_j) \big), \quad 0 \leqslant j < n \leqslant N.$$

The main idea is again to compare the time-dependent operator $R(t_n, t_j)$ with the frozen operator $r(t_n, t_j)$. For the latter, we have the stability estimate [6, Lemma 5.1]

$$\big\| A_j r(t_n, t_j) \big\|_{X \leftarrow X} \leqslant C (t_n - t_j)^{-1}, \quad 0 \leqslant j < n \leqslant N,$$

where the constant C depends on t_N, but not on n and j. In the same way as in the proof of Lemma 7, by using the telescopic identity and the estimate

$$\sum_{k=j}^{n-1} h_{k+1} (t_n - t_k)^{-1} (t_{k+1} - t_j)^{\alpha-1} \leqslant C (t_n - t_j)^{\alpha-1} \big(1 + |\log h_n| \big)$$

we arrive at

$$\|\Delta_{nj}\|_{X \leftarrow X} \leqslant C \sum_{k=j}^{n-1} h_{k+1} (t_{k+1} - t_j)^{\alpha-1} \|\Delta_{nk}\|_{X \leftarrow X} + C (t_n - t_j)^{\alpha-1} \big(1 + |\log h_n| \big).$$

Applying a discrete Gronwall lemma thus gives the desired result. For a similar Gronwall-type inequality, we refer to [2, Lemma 4.4].

We finally remark that our variable stepsize estimates are valid without any additional condition on the stepsize sequence.

6. Numerical examples

The numerical examples given below illustrate our convergence results for the backward Euler method.

We consider again the nonlinear initial-boundary value problem (5). It is noteworthy that it has an unstable equilibrium $U = 1$ which is hyperbolic under the generic condition $a\pi^2 n^2 \neq 2$ for all $n \in \mathbb{N}$. In the following we choose $a = 1$ and consider various initial values that satisfy the requirements of Theorems 5 and 6.

Example 9. The smooth and positive function

$$U_0(x) = \frac{x^3}{3} - \frac{x^2}{2} + 1, \quad 0 \leqslant x \leqslant 1,$$

satisfies the Neumann boundary conditions and thus lies in $\mathcal{D}$. Since the composition $f(0, U_0)$ is analytic, it further fulfills $f(0, U_0) \in (X, D)_{\alpha,\infty}$ for every $0 < \alpha < 1/2$, see (7). Therefore, Theorem 5 is applicable.

Example 10. The polynomial

$$U_0(x) = -20x^7 + 70x^6 - 84x^5 + 35x^4 + 1$$

is positive for all $x \in [0, 1]$. Moreover, the derivatives of U_0 up to order 3 vanish at the boundary, which implies $U_0 \in \mathcal{D}$ and $f(0, U_0) \in D$. Therefore, Theorem 8.1.1 of [12] applied to

$$u'' = D_1 f(t, u) + D_2 f(t, u)u', \quad u'(0) = f(0, U_0)$$

guarantees that $u' \in C([0, T], D)$ and in particular $u'' \in B([0, T], X)$. Thus the requirements of Theorem 6 hold.

Example 11. For a constant initial value, the solution $U(t, x)$ depends on t only. Along such a solution, problem (5) reduces to the simple ordinary differential equation

$$w' = \log w, \quad w(0) = U_0,$$

and we get $U(t, x) = w(t)$. In our experiment, we integrated the original problem with $U_0 = 5$.

We discretized problem (5) in space by standard finite differences on an equidistant grid with meshwidth $\Delta x = 10^{-4}$, and in time by the backward Euler method, respectively. For the different initial values, the integration was performed up to $T = 1$ with stepsizes $h = H/2^j$ where $H = 0.2$ and $0 \leqslant j \leqslant 7$. We emphasize that the implementation of the right-hand side (6) as well as the approximation to its Jacobian requires some care.

In order to determine the errors, we compared the results with more precise approximations that have been obtained with the code RADAU5. This code is a variable stepsize implementation of the 3-stage

Table 1
Numerically observed orders of convergence at $T = 1$

Stepsize h	1/5	1/10	1/20	1/40	1/80	1/160	1/320
Example 9	1.167	1.074	1.036	1.018	1.009	1.005	1.002
Example 10	1.238	1.203	1.180	1.151	1.114	1.076	1.045
Example 11	1.008	1.004	1.002	1.001	1.001	1.000	1.000

Radau IIA method, see [7]. From the quotients of the errors, the numerical orders of convergence were computed in a standard way. The results are given in Table 1. As expected, the numbers approach one as h decreases.

Acknowledgement

This research was partly supported by the Austrian Science Fund FWF under Grant P13754-MAT and by the Swiss National Science Foundation under Grant 2000-0056577.

References

[1] G. Akrivis, M. Crouzeix, C. Makridakis, Implicit–explicit multistep methods for quasilinear parabolic equations, Numer. Math. 82 (1999) 521–541.
[2] N.Yu. Bakaev, On variable stepsize Runge–Kutta approximations of a Cauchy problem for the evolution equation, BIT 38 (1998) 462–485.
[3] G. Barles, P.E. Souganidis, Convergence of approximation schemes for fully nonlinear second order equations, Asymptotic Anal. 4 (1991) 271–283.
[4] C.-M. Brauner, J. Buckmaster, J. Dold, C. Schmidt-Lainé, On an evolution equation arising in detonation theory, in: M. Onofri, A. Tesei (Eds.), Fluid Dynamical Aspects of Combustion Theory, Pitman Res. Notes in Math., Vol. 223, Longman, Harlow, 1991, pp. 196–210.
[5] C. González, C. Palencia, Stability of Runge–Kutta methods for abstract time-dependent parabolic problems: the Hölder case, Math. Comp. 68 (1999) 73–89.
[6] C. González, A. Ostermann, C. Palencia, M. Thalhammer, Backward Euler discretization of fully nonlinear parabolic problems, Math. Comp. 71 (2002) 125–145.
[7] E. Hairer, G. Wanner, Solving Ordinary Differential Equations II. Stiff and Differential–Algebraic Problems, 2nd rev. edn., Springer, Berlin, 1996.
[8] M.-N. Le Roux, Méthodes multipas pour des équations paraboliques non linéaires, Numer. Math. 35 (1980) 143–162.
[9] Ch. Lubich, O. Nevanlinna, On resolvent conditions and stability estimates, BIT 31 (1991) 293–313.
[10] Ch. Lubich, A. Ostermann, Runge–Kutta methods for parabolic equations and convolution quadrature, Math. Comp. 60 (1993) 105–131.
[11] Ch. Lubich, A. Ostermann, Runge–Kutta approximation of quasilinear parabolic equations, Math. Comp. 64 (1995) 601–627.
[12] A. Lunardi, Analytic Semigroups and Optimal Regularity in Parabolic Problems, Birkhäuser, Basel, 1995.
[13] A. Lunardi, C^∞ regularity for fully nonlinear abstract evolution equations, in: A. Favini, E. Obrecht (Eds.), Differential Equations in Banach spaces, Lecture Notes in Math., Vol. 1223, Springer, Berlin, 1986.
[14] E. Nakaguchi, A. Yagi, Error estimates of implicit Runge–Kutta methods for quasilinear abstract equations of parabolic type, Japan. J. Math. (N.S.) 25 (1999) 181–226.
[15] M. Thalhammer, Runge–Kutta time discretization of nonlinear parabolic equations, Ph.D. Thesis, Universität Innsbruck, Innsbruck, 2000.
[16] H. Triebel, Interpolation Theory, Function Spaces, Differential Operators, North-Holland, Amsterdam, 1978.

1.4. Stability of linear multistep methods

Stability of linear multistep methods and applications to nonlinear parabolic problems
ALEXANDER OSTERMANN, MECHTHILD THALHAMMER, AND GABRIELA KIRLINGER
Applied Numerical Mathematics (2004) 48, 389-407

Available online at www.sciencedirect.com
SCIENCE DIRECT®
Applied Numerical Mathematics 48 (2004) 389–407
APPLIED NUMERICAL MATHEMATICS
www.elsevier.com/locate/apnum

Stability of linear multistep methods and applications to nonlinear parabolic problems

A. Ostermann [a], M. Thalhammer [a,*], G. Kirlinger [b]

[a] *Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria*
[b] *Institut für Angewandte und Numerische Mathematik, Technische Universität Wien, Wiedner Hauptstraße 8-10/115, A-1040 Wien, Austria*

Abstract

In the present paper, stability and convergence properties of linear multistep methods are investigated. The attention is focused on parabolic problems and variable stepsizes. Under weak assumptions on the method and the stepsize sequence an asymptotic stability result is shown. Further, stability bounds for linear nonautonomous parabolic problems with Hölder continuous operator are given. With the help of these results, convergence estimates for semilinear and fully nonlinear parabolic problems are derived.

Keywords: Linear multistep methods; Stability; Variable stepsizes; Semilinear parabolic problems; Fully nonlinear parabolic problems; Convergence; Geometric properties

1. Introduction

In this paper, we study the stability and convergence properties of linear multistep methods applied to nonlinear parabolic problems. Our analysis admits variable stepsizes and is based on an abstract framework of sectorial operators and analytic semigroups in Banach spaces.

Stability results for variable stepsize multistep discretizations generally require that the ratios of two subsequent steps are bounded from below and above, i.e., $\Omega_1 \leqslant h_n/h_{n-1} \leqslant \Omega_2$ with appropriate Ω_1 and Ω_2. For ordinary differential equations, two conceptually different types of stability estimates are found in literature, see [9, Section III.5].

* Corresponding author.
E-mail addresses: alexander.ostermann@uibk.ac.at (A. Ostermann), mechthild.thalhammer@uibk.ac.at (M. Thalhammer), g.schranz-kirlinger@tuwien.ac.at (G. Kirlinger).

0168-9274/$30.00
doi:10.1016/j.apnum.2003.10.004

The first approach ensures stability *independently* of the chosen stepsize sequence. To our knowledge, Grigorieff [8] was the first who analyzed BDF discretizations of ordinary differential equations in this way. Recently, stability bounds that depend only weakly on the stepsize sequence have been derived for BDF discretizations of parabolic problems in a Hilbert space setting by Becker [2], Calvo and Grigorieff [3], and Emmrich [5]. Notwithstanding the merits of this approach, it has its shortcomings as well, since it gives, in general, quite disappointing values for Ω_1 and Ω_2. For BDF5, e.g., one obtains the stringent condition $0.997 \leqslant h_n/h_{n-1} \leqslant 1.003$ in order to ensure zero-stability without further restrictions on the stepsize sequence.

The second approach allows the stability factor to *depend* on the stepsize sequence. To obtain (practical) stability, however, it must be guaranteed that this factor remains bounded by a (reasonable) constant. For ordinary differential equations, this approach was used by Gear and Tu [6]. Under the assumption that the stepsize sequence depends smoothly on the local errors, they obtained favourable convergence results. More recently, this direction has been further exploited for linear parabolic problems in a series of papers by Palencia [14,15] and Palencia and García-Archilla [16]. The results of Palencia, however, are not sufficient to obtain convergence for nonlinear problems. Our motivation for the present paper was to derive the missing stability estimates and to develop a convergence theory of multistep methods for nonlinear parabolic problems.

The present paper is structured as follows: In Section 2, we first introduce the analytical framework, based on the theory of sectorial operators in Banach spaces, and we specify the requirements on the numerical method. We then derive our main stability results for asymptotically stable analytic semigroups. This is the key for proving asymptotic stability of multistep discretizations. In Section 3, we extend the stability results of Section 2 to arbitrary sectorial operators, and then in Section 4 to linear nonautonomous parabolic problems with Hölder continuous operator. In Section 5, we apply the stability results to semilinear parabolic problems, and we derive a convergence result for finite times. In Section 6, we give applications to fully nonlinear parabolic problems. We study the long-term behaviour of multistep discretizations nearby an asymptotically stable equilibrium, and we state a convergence result for smooth solutions on compact time intervals. Corresponding results for Runge–Kutta methods are found in our papers [7,13,18].

Throughout this paper, we employ the following notation. For normed spaces Y and Z, the space $L(Y, Z)$ comprises all linear operators from Y to Z. It is endowed with the usual operator norm denoted by $\|\cdot\|_{Z\leftarrow Y}$. For an integer $k \geqslant 1$, the norm on the product space Y^k is defined by $\|y\|_{Y^k} = \max\{\|y_i\|_Y\colon\ 1 \leqslant i \leqslant k\}$ for $y = (y_1, \ldots, y_k)^{\mathrm{T}} \in Y^k$. In order to simplify the notation, we usually dismiss the dimensions in the operator norm and write $\|\cdot\|_{Z\leftarrow Y}$ instead of $\|\cdot\|_{Z^k\leftarrow Y^k}$ for short. We recall that for an arbitrary matrix B with coefficients b_{ij} and a linear operator A, the (i, j)th component of the Kronecker product $B \otimes A$ equals $b_{ij}A$. We further distinguish between the identity operator I and the identity matrix $\mathcal{I}$ on $\mathbb{R}^k$.

Henceforth, C denotes a generic constant with possibly different values at different occurrences.

2. Asymptotic stability for time-independent operators

In this section, we derive fundamental stability estimates for linear multistep methods with variable stepsizes. Our results substantially rely on the papers [15] and [16].

We study the abstract initial value problem on a Banach space $(X, \|\cdot\|_X)$

$$u'(t) = Au(t), \quad t > 0, \qquad u(0) \text{ given}, \tag{1}$$

where A is a densely defined and closed linear operator on X. The domain D of A is endowed with the graph norm $\|\cdot\|_D$. Our main assumption on A is the following, cf. [12] or [10].

HA1. *We suppose that $A \in L(D, X)$ is sectorial on X, i.e., for some constants $a \in \mathbb{R}$, $M \geqslant 1$ and $\varphi \in (0, \pi/2)$, the resolvent of A fulfills the condition*

$$\left\|(\lambda I - A)^{-1}\right\|_{X \leftarrow X} \leqslant \frac{M}{|\lambda - a|}, \quad \lambda \in \mathbb{C} \setminus S_\varphi(a), \tag{2}$$

on the complement of the sector $S_\varphi(a) = \{\lambda \in \mathbb{C}\colon |\arg(a - \lambda)| \leqslant \varphi\} \cup \{a\}$.

Let $(h_n)_{n\geqslant 0}$ denote the sequence of positive time steps with corresponding ratios $\omega_n = h_n/h_{n-1}$, $n \geqslant 1$, and set $\boldsymbol{\omega}_n = (\omega_n, \ldots, \omega_{n+k-2})$. The associated grid points are denoted by $t_n = h_0 + h_1 + \cdots + h_{n-1}$. Throughout the paper, we use the following assumption on the stepsize sequence.

HS1. *We assume that there exists $\Omega > 1$ such that the stepsize ratios satisfy $\Omega^{-1} \leqslant \omega_n \leqslant \Omega$ for all $n \geqslant 1$.*

We first draw some conclusions from this hypothesis that are all-important for our stability results. Let $(h_n)_{n\geqslant 0}$ be a stepsize sequence satisfying HS1. For the subsequence $h_{k-1}, h_k, \ldots, h_{k+j-2}$ of length j, consider the associated sequence of ordered stepsizes $h_{\pi(1)} \leqslant h_{\pi(2)} \leqslant \cdots \leqslant h_{\pi(j)}$ and set

$$\tau_\varkappa^{(j)} = h_{\pi(1)} + \cdots + h_{\pi(\varkappa)} \quad \text{for } 0 \leqslant \varkappa \leqslant j. \tag{3a}$$

From the identity

$$t_{n+k-1} = h_{n+k-2} + \cdots + h_{j+k-1} + h_{\pi(j)} + \cdots + h_{\pi(\varkappa+1)} + \tau_\varkappa^{(j)} + t_{k-1},$$

with the help of HS1 and the obvious estimates $h_{j+k-1} \leqslant \Omega h_{\pi(j)}$ and $h_{\pi(\varkappa+1)} \leqslant \Omega\tau_\varkappa^{(j)}$, we get the useful relation

$$t_{n+k-1} - t_{k-1} \leqslant C\Omega^{n-\varkappa}\tau_\varkappa^{(j)} \quad \text{for } 1 \leqslant \varkappa \leqslant j \leqslant n. \tag{3b}$$

We further note for later use that

$$\begin{aligned} &Ct_{n+k-1} \leqslant t_{n+k-1} - t_{k-1} \leqslant t_{n+k-1}, \quad n \geqslant 1, \\ &t_{n+k-1} - t_{k-1} \leqslant C\Omega^n h_{k-1}. \end{aligned} \tag{3c}$$

The numerical approximation u_{n+k} to the solution of (1) at time t_{n+k} by a linear multistep method is given recursively by

$$\sum_{i=0}^{k} \alpha_{ni} u_{n+i} = h_{n+k-1} A \sum_{i=0}^{k} \beta_{ni} u_{n+i}, \quad n \geqslant 0. \tag{4}$$

This relation involves the coefficients α_{ni} and β_{ni}, $0 \leqslant i \leqslant k$, that may depend on $\boldsymbol{\omega}_{n+1}$, and on the starting values $u_0, u_1, \ldots, u_{k-1}$. For more information on variable stepsize linear multistep methods, we refer to the monograph [9].

In order to write the numerical scheme (4) in compact vector form, we denote

$$U_n = (u_n, u_{n+1}, \ldots, u_{n+k-1})^{\mathrm{T}},$$

for $n \geqslant 0$. Further, we introduce the functions

$$\begin{aligned} &J_{n+1}(z) = (\alpha_{nk} - \beta_{nk} z)^{-1}, \\ &s_{n+1,i}(z) = -J_{n+1}(z) \cdot (\alpha_{ni} - \beta_{ni} z), \quad 0 \leqslant i \leqslant k-1. \end{aligned} \tag{5}$$

Here, the first index indicates the dependence on $\boldsymbol{\omega}_{n+1}$. Then, the companion matrix of the method is given by

$$r_{n+1}(z) = \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \dots & 0 & 1 \\ s_{n+1,0}(z) & s_{n+1,1}(z) & \dots & \dots & s_{n+1,k-1}(z) \end{pmatrix}.$$

For constant time steps, we denote the companion matrix by $r(z)$ for short.

The above notation allows us to rewrite (4) as

$$U_n = \prod_{j=1}^{n} r_j(h_{j+k-2} A)\, U_0, \quad n \geqslant 0. \tag{6}$$

We note that the factors arising in the product do not commute, in general.

Throughout the paper, we require the following stability assumption for constant stepsizes.

HM1. *We assume that the linear multistep method* (4) *is $A(\varphi)$-stable and strictly stable at* 0 *and infinity. Thus, $\lambda = 1$ is the only eigenvalue of the companion matrix at* 0 *with modulus one, and the spectral radius of the companion matrix at infinity, $\varrho = \varrho(r(\infty))$, is less than one.*

The following hypothesis is needed for variable stepsizes.

HM2. *We assume that the coefficients α_{ni} and β_{ni} in* (4) *are bounded for all stepsize sequences satisfying* HS1. *We further require that the rational functions $J_n(z)$ and $s_{ni}(z)$ in* (5) *remain bounded for $z \in S_\varphi(0)$.*

If the k-step method is consistent of order p, the principal eigenvalue $\lambda_1(z)$ of $r(z)$ fulfills

$$\lambda_1(z) = \mathrm{e}^z + \mathcal{O}\big(z^{p+1}\big), \quad z \to 0,$$

see [11]. Note that this relation implies

$$\lambda_1(z) = \mathrm{e}^{z+\eta(z)} \quad \text{with } \eta(z) = \mathcal{O}\big(z^{p+1}\big) \text{ for } z \to 0,$$

which is important for the results we have in mind.

HM3. *We assume that the linear multistep method* (4) *has order $p \geqslant 1$.*

Example. The k-step BDF methods, for $1 \leqslant k \leqslant 6$, satisfy HM1–2 with $\varrho = 0$ for any $\Omega > 0$, and HM3 with $p = k$, see [9].

Now we are ready to state the stability result. The corresponding results for Runge–Kutta methods are given in [1]. In order to simplify the notation we introduce the abbreviation

$$C^{\Delta}_{\omega,n} = C \prod_{j=1}^{n+k-2} \bigl(1 + \Delta|\omega_j - 1|\bigr)^2 \leqslant C\mathrm{e}^{2\Delta \sum_{j=1}^{n+k-2} |\omega_j - 1|}. \tag{7}$$

Theorem 1. *Consider the linear multistep discretization* (4) *of Eq.* (1), *and assume that* HA1 *with* $a < 0$, HS1, *and* HM1–3 *hold. If* $\mu \geqslant 1$ *satisfies* $\varrho\Omega^{\mu} < 1$, *then the following bound is valid for all* $n \geqslant 1$

$$\left\| \prod_{j=1}^{n} r_j(h_{j+k-2}A) \right\|_{D\leftarrow D} \leqslant \frac{C^{\Delta}_{\omega,n}}{1 + t^{\mu}_{n+k-1}}.$$

If in addition $\mu \geqslant 2$, *we have for all* $n \geqslant 1$

$$\left\| \prod_{j=1}^{n} r_j(h_{j+k-2}A) - \prod_{j=1}^{n} r_j(\infty) \right\|_{D\leftarrow X} \leqslant \frac{C^{\Delta}_{\omega,n}}{t_{n+k-1} + t^{\mu}_{n+k-1}}.$$

Recall that under hypothesis HA1 with $a < 0$, the semigroup e^{tA} decays exponentially fast to 0, since for any $\tilde{a} > a$

$$\bigl\|\mathrm{e}^{tA}\bigr\|_{D\leftarrow D} + t\bigl\|\mathrm{e}^{tA}\bigr\|_{D\leftarrow X} \leqslant C\mathrm{e}^{\tilde{a}t}, \quad t \geqslant 0.$$

Let the stepsize sequence be such that $t_n \to \infty$ for $n \to \infty$. If the quotient

$$\frac{C^{\Delta}_{\omega,n}}{1 + t^{\mu}_{n+k-1}} \to 0, \quad \text{for } n \to \infty, \tag{8}$$

then the numerical method is asymptotically stable. For *practical* purposes, however, it is essential that the quotient in (8) remains bounded by a *reasonable* constant for all n. This is achieved, for example, in the following situation.

Example. Suppose that the size of the gridpoints grows exponentially fast $t_n \geqslant Cq^n$ with some $1 < q \leqslant \Omega$ and that for some $\mu \geqslant 2$

$$\bigl(1 + \Delta(\Omega - 1)\bigr)^2 < q^{\mu} \quad \text{and} \quad \varrho\Omega^{\mu} < 1. \tag{9}$$

Then, the numerical method is asymptotically stable, and the constant in (8) is reasonable, if $|a|h_0$ is not too small. Note that (9) holds for μ sufficiently large, if $\varrho = 0$.

Proof of Theorem 1. Our proof is strongly based on the work of Palencia. An application of a matrix version of [15, Lemma 1 and Theorem 2] to the shifted operator $A - aI$ shows that the following bound holds

$$\bigl\|g(A)\bigr\| \leqslant C\|g\|_{\varphi,a} + C\|g\|_{\varphi,a} \log^{+}\left(\frac{N_{\varphi,a}(g)}{\|g\|_{\varphi,a}}\right),$$

for any holomorphic mapping g defined on some neighbourhood of $S_{\varphi}(a)$ taking values in the space of complex $k \times k$ matrices. We here denote

$$\|g\|_{\varphi,a} = \sup\bigl\{\bigl\|g(\lambda)\bigr\|\colon \lambda \in S_{\varphi}(a)\bigr\},$$
$$N_{\varphi,a}(g) = \bigl\|g(a)\bigr\| + \bigl\|g(\infty)\bigr\| + \sqrt{Z_{\varphi,a}(g) I_{\varphi,a}(g)},$$

with

$$Z_{\varphi,a}(g) = \sup\{\|(\lambda - a)^{-1}(g(\lambda) - g(a))\|\colon \lambda \in S_\varphi(a)\},$$
$$I_{\varphi,a}(g) = \sup\{\|(\lambda - a)(g(\lambda) - g(\infty))\|\colon \lambda \in S_\varphi(a)\}.$$

The first part of the theorem follows by applying this bound to the function

$$g(\lambda) = \prod_{j=1}^{n} r_j(h_{j+k-2}\lambda), \tag{10}$$

and Lemma 4 below. More precisely, we use the inequality $x \log^+(y/x) \leqslant \log^+(y/b) + b/\mathrm{e}$ which holds for $y \geqslant 0$ and $x, b > 0$. We recall here that $\log^+ x = \max(0, \log x)$. Setting $x = \|g\|_{\varphi,a}$, $y = N_{\varphi,a}(g)$ and

$$b = \frac{1}{1 + t_{n+k-1}^{\mu}} \prod_{j=1}^{n+k-2} \bigl(1 + \Delta|\omega_j - 1|\bigr)^2,$$

then yields the first estimate of the theorem with the additional factor $1 + \log^+ t_{n+k-1}$. This factor, however, can be omitted by slightly increasing μ. The second part of the theorem follows in the same way by using the function

$$G(\lambda) = \lambda\left(\prod_{j=1}^{n} r_j(h_{j+k-2}\lambda) - \prod_{j=1}^{n} r_j(\infty)\right), \tag{11}$$

and Lemma 5. □

The auxiliary results that are needed in the above proof are collected in the remainder of this section. First, we study the behaviour of the companion matrix for constant time steps. We remark that $r(z)$ satisfies an estimate of the form

$$\|r(z_1) - r(z_2)\| \leqslant C \min\bigl(|z_1 - z_2|, |z_1^{-1} - z_2^{-1}|\bigr), \quad z_1, z_2 \in S_\varphi(0).$$

The following lemma is an extension of [16, Theorem A.1]. For a related decomposition of the companion matrix, see [4].

Lemma 1. *Let $r(z)$ be the companion matrix of a linear multistep method satisfying* HM1 *and* HM3. *Then there exists a map T defined on $S_\varphi(0)$ with values in the space of complex $k \times k$ matrices with the following properties: For any $0 < \varrho < \delta < 1$, there exist a neighbourhood $B = \{z \in \mathbb{C}\colon |z| \leqslant \sigma\}$ of the origin and a constant $0 < c < 1$ such that the following estimates hold*

$$\begin{aligned} &\|T(z)r(z)T(z)^{-1}\| \leqslant \mathrm{e}^{c\,\mathrm{Re}\,z}, && z \in \Sigma_0 = B \cap S_\varphi(0), \\ &\|T(z)r(z)T(z)^{-1}\| \leqslant \delta, && z \in \Sigma_\infty = S_\varphi(0) \setminus \Sigma_0. \end{aligned} \tag{12a}$$

Furthermore, we have for all $z, z_1, z_2 \in S_\varphi(0)$

$$\begin{aligned} &\|T(z)\| \leqslant C, \qquad \|T(z)^{-1}\| \leqslant C, \\ &\|T(z_1) - T(z_2)\| \leqslant C \min\bigl(|z_1 - z_2|, |z_1^{-1} - z_2^{-1}|\bigr). \end{aligned} \tag{12b}$$

Proof. The lemma is a consequence of [16, Theorem A.1]. In order to show the additional estimates (12a), we choose a Lipschitz-continuous map ψ that coincides with the exponential $\psi(z) = \mathrm{e}^{cz}$ near the origin and satisfies $\psi(z) = \delta$ in a neighborhood of ∞ in such a way that $\varrho\big(\psi(z)^{-1} r(z)\big) < 1$ holds on $S_\varphi(0) \setminus \{0\}$. Then, the result follows from an application of Theorem A.1 in *loc.cit.* □

In order to study the product $g(\lambda)$ it is useful to introduce the map

$$\Phi : S_\varphi(0) \to \mathbb{C} : z \mapsto \Phi(z) = \begin{cases} \mathrm{e}^{c\,\mathrm{Re}\,z} & \text{if } z \in \Sigma_0, \\ \delta & \text{if } z \in \Sigma_\infty, \end{cases}$$

which essentially captures the behaviour of $r(z)$, see (12a).

Lemma 2. *Under the assumptions of Lemma* 1 *and* HM2*, it holds*

$$\left\| \prod_{j=1}^{n} r_j(h_{j+k-2}\lambda) \right\| \leqslant C^{\Delta}_{\omega,n} \cdot \prod_{j=1}^{n} \Phi(h_j \lambda), \quad \lambda \in S_\varphi(a).$$

Proof. The proof is very close to that of [16, Lemma 3.2]. Replacing relation (32) of *loc.cit.* with (12a) and tracing its effects, yields the result. In particular, the very form of $C^{\Delta}_{\omega,n}$ in (7) as a product is obtained. □

Lemma 3. *Let the stepsize sequence satisfy* HS1*, and let $\gamma > 0$ and $\mu > 0$ be such that $\gamma \Omega^\mu \leqslant 1$. Then, for $c > 0$, there exists a constant C such that*

$$\gamma^{n-m} \mathrm{e}^{-c\tau_m^{(j)}} \leqslant \frac{C}{1 + t_{n+k-1}^{\mu}} \quad \textit{for all } \ 0 \leqslant m \leqslant j \leqslant n$$

with $\tau_m^{(j)}$ given by (3a)*.*

Proof. From (3) we obtain $t_{n+k-1}^{\mu} \gamma^{n-m} \leqslant C(\gamma\Omega^\mu)^{n-m} (\tau_m^{(j)})^\mu$, and the assertion follows at once from the uniform boundedness of $s^\mu \mathrm{e}^{-cs}$ for positive s. □

We are now ready to derive the desired estimates for the function $g(\lambda)$, defined in (10).

Lemma 4. *Under the assumptions of the theorem, it holds*

$$\sup_{\lambda \in S_\varphi(a)} \big\| g(\lambda) \big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1 + t_{n+k-1}^{\mu}},$$

$$\sup_{\lambda \in S_\varphi(a)} \big\| (\lambda - a)^{-1} \big(g(\lambda) - g(a) \big) \big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1 + t_{n+k-1}^{\mu}}\, t_{n+k-1},$$

$$\sup_{\lambda \in S_\varphi(a)} \big\| (\lambda - a) \big(g(\lambda) - g(\infty) \big) \big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1 + t_{n+k-1}^{\mu-1}}\, t_{n+k-1}^{-1}.$$

Proof. It is convenient to employ the following abbreviations

$$r_j = r_j(h_{j+k-2}\lambda), \qquad \varrho_j = r_j(\infty). \tag{13}$$

We choose $\varrho < \delta < 1$ such that $\delta\Omega^{\mu} < 1$. Let $\varkappa$ denote the number of indices $k-1 \leqslant m \leqslant n-1$ such that $|h_m\lambda| < \sigma$. From Lemma 2 and (3b), we get

$$\big\|g(\lambda)\big\| \leqslant C^{\Delta}_{\omega,n}\delta^{n-\varkappa}\mathrm{e}^{c\tau_{\varkappa}^{(n)}\operatorname{Re}\lambda}.$$

Since $\operatorname{Re}\lambda \leqslant a < 0$, the first assertion of the lemma follows at once from (3) and Lemma 3. For the second bound, we use the telescopic identity

$$g(\lambda) - g(a) = \sum_{j=1}^{n}\prod_{l=j+1}^{n} r_l(h_{l+k-2}a)\big(r_j - r_j(h_{j+k-2}a)\big)\prod_{i=1}^{j-1} r_i,$$

and the estimate $\|r_j - r_j(h_{j+k-2}a)\| \leqslant Ch_{j+k-2}|\lambda - a|$. This yields

$$\big\|(\lambda-a)^{-1}\big(g(\lambda)-g(a)\big)\big\| \leqslant C^{\Delta}_{\omega,n}\sum_{j=1}^{n} h_{j+k-2}\delta^{n-\varkappa_j-1}\mathrm{e}^{c\tau_{\varkappa_j}^{(j-1)}a},$$

with $0 \leqslant \varkappa_j \leqslant n-1$, and the same arguments as before yield the desired bound. The last estimate follows in a similar way from

$$g(\lambda) - g(\infty) = \sum_{j=1}^{n}\prod_{l=j+1}^{n}\varrho_l(r_j - \varrho_j)\prod_{i=1}^{j-1} r_i.$$

For fixed $j \geqslant 2$, let $\varkappa = \varkappa(j)$ denote the number of indices $k-1 \leqslant m \leqslant j+k-3$ such that $|h_m\lambda| < \sigma$. Using $\|(\lambda-a)(r_j-\varrho_j)\| \leqslant Ch_{j+k-2}^{-1}$ for $\varkappa(j) = 0$ and $|\lambda-a|\tau_{\varkappa}\,\mathrm{e}^{c\tau_{\varkappa}\operatorname{Re}\lambda} \leqslant C$ for $\varkappa(j) \geqslant 1$ yields the desired result. □

We next study the behaviour of $G(\lambda)$, defined in (11).

Lemma 5. *Under the assumptions of the theorem, it holds*

$$\sup_{\lambda\in S_{\varphi}(a)}\big\|G(\lambda)\big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1+t_{n+k-1}^{\mu-1}}\,t_{n+k-1}^{-1},$$

$$\sup_{\lambda\in S_{\varphi}(a)}\big\|(\lambda-a)^{-1}\big(G(\lambda)-G(a)\big)\big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1+t_{n+k-1}^{\mu}}(1+t_{n+k-1}),$$

$$\sup_{\lambda\in S_{\varphi}(a)}\big\|(\lambda-a)\big(G(\lambda)-G(\infty)\big)\big\| \leqslant \frac{C^{\Delta}_{\omega,n}}{1+t_{n+k-1}^{\mu-2}}\,t_{n+k-1}^{-2}.$$

Proof. Since $G(\lambda) = \lambda(g(\lambda) - g(\infty))$, the first estimate follows in the same way as that in the previous lemma. For the second estimate, we use the identity

$$G(\lambda) - G(a) = (\lambda-a)\big(g(\lambda)-g(a)\big) + a\big(g(\lambda)-g(a)\big),$$

and the previous lemma. In order to show the last relation, we define with (13) the analytic function $\psi_j(\lambda) = \lambda(r_j - \varrho_j)$ which is bounded at infinity by Ch_{j+k-2}^{-1}. Using

$$G(\infty) = \sum_{j=1}^{n}\prod_{l=j+1}^{n}\varrho_l\psi_j(\infty)\prod_{i=1}^{j-1}\varrho_i,$$

and the telescopic identity, we get

$$G(\lambda) - G(\infty) = \sum_{j=1}^{n} \prod_{l=j+1}^{n} \varrho_l \big(\psi_j(\lambda) - \psi_j(\infty)\big) \prod_{i=1}^{j-1} \varrho_i + \sum_{j=1}^{n} \prod_{l=j+1}^{n} \varrho_l (r_j - \varrho_j)\lambda \left(\prod_{i=1}^{j-1} r_i - \prod_{i=1}^{j-1} \varrho_i \right).$$

Expanding the last term again with the telescopic identity, the desired bound now follows as in the previous lemma. □

3. Stability on compact time intervals

In this section, we derive stability estimates for (4) on compact time intervals $[0, T]$. We first give an extension of Theorem 1 to nonnegative a. We make use of the following hypothesis which is familiar from the convergence analysis of linear multistep methods for ODEs, see [9, Theorem III.5.7].

HS2. *We assume that the stability factors $C^{\Delta}_{\omega,n}$ in* (7) *are uniformly bounded by a constant for all $n \geqslant 0$.*

We emphasize that the size of the constant in HS2 may depend on the length of the considered time interval.

Theorem 2. *Consider the linear multistep discretization* (4) *of Eq.* (1) *on the interval* $[0, T]$*, and assume that* HA1, HS1–2, HM1–3 *hold and that $\varrho\Omega^2 < 1$. Then, there exist positive constants H and C such that for $0 < h_j \leqslant H$ the following bounds are valid for all $n \geqslant 1$ with $t_n \leqslant T$*

$$\left\| \prod_{j=1}^{n} r_j(h_{j+k-2}A) \right\|_{D \leftarrow D} \leqslant C, \qquad \left\| \prod_{j=1}^{n} r_j(h_{j+k-2}A) - \prod_{j=1}^{n} r_j(\infty) \right\|_{D \leftarrow X} \leqslant \frac{C}{t_{n+k-1}}.$$

The constant C depends on the constants that appear in our assumptions and on T, but it is independent of n.

Proof. Our proof relies on a smart idea of Palencia [14, Section 3]. Since our assumptions on the stepsize sequence here are different, we shortly comment on the necessary modifications. For $b > a \geqslant 0$, let $f_j = (1 + \Delta|\omega_j - 1|)^{-2}$ and

$$\tilde{r}_j = f_j \cdot r_j\big(h_{j+k-2}(A - bI)\big) \quad \text{and} \quad \hat{r}_j = f_j \cdot r_j(h_{j+k-2}A).$$

Note that the (k, m)-entry of $\hat{r}_j - \tilde{r}_j$ is given by

$$h_{j+k-2} b f_j \cdot (\beta_{j-1,k}\alpha_{j-1,m} - \alpha_{j-1,k}\beta_{j-1,m})\, J_j(h_{j+k-2}A) J_j\big(h_{j+k-2}(A - bI)\big). \tag{14}$$

Therefore, there exists a constant C such that

$$\big\|\tilde{r}_j - \hat{r}_j\big\| \leqslant C \cdot h_{j+k-2}.$$

The first assertion of the theorem now follows at once from the telescopic identity

$$\prod_{j=1}^{n} \hat{r}_j = \sum_{j=1}^{n} \prod_{l=j+1}^{n} \tilde{r}_l \big(\hat{r}_j - \tilde{r}_j\big) \prod_{i=1}^{j-1} \hat{r}_i + \prod_{j=1}^{n} \tilde{r}_j,$$

and a discrete Gronwall lemma. To obtain the second bound, we write

$$(\mathcal{I}\otimes A)\left(\prod_{j=1}^{n}\hat{r}_j-\prod_{j=1}^{n}\tilde{r}_j\right)=\sum_{j=1}^{n}\prod_{l=j+1}^{n}\tilde{r}_l\big(\hat{r}_j-\tilde{r}_j\big)(\mathcal{I}\otimes A)\left(\prod_{i=1}^{j-1}\hat{r}_i-\prod_{i=1}^{j-1}\tilde{r}_i\right)$$
$$+(\mathcal{I}\otimes A)\sum_{j=1}^{n}\prod_{l=j+1}^{n}\tilde{r}_l\big(\hat{r}_j-\tilde{r}_j\big)\prod_{i=1}^{j-1}\tilde{r}_i.$$

To bound the inhomogeneity, we use again (14). Due to

$$\left\|\mathcal{I}\otimes(A-bI)^{\theta}\prod_{j=2}^{n}\hat{r}_j\cdot\big(\mathcal{I}\otimes J_1\big(h_{k-1}(A-bI)\big)\big)\right\|_{X\leftarrow X}\leqslant\frac{C}{t_{n+k-1}^{\theta}},\quad 0\leqslant\theta\leqslant 1,$$

which follows from Theorem 1 by interpolation, the inhomogeneity is seen to be bounded by

$$\sum_{j=1}^{n}h_{j+k-2}\,(t_{n+k-1}-t_{j+k-2})^{-1/2}\,t_{j+k-2}^{-1/2}\leqslant C.$$

The desired result now follows again with a discrete Gronwall lemma. □

The following lemma is a discrete version of the well-known identity

$$A\int_0^t \mathrm{e}^{\tau A}\,\mathrm{d}\tau=\mathrm{e}^{tA}-I.$$

We denote again $r_j=r_j(h_{j+2-k}A)$, $J_j=J_j(h_{j+2-k}A)$, and further

$$e_k=(0,\dots,0,1)^{\mathrm{T}}\quad\text{and}\quad\mathbb{1}=(1,\dots,1)^{\mathrm{T}}\in\mathbb{R}^k. \tag{15}$$

Recall that a linear multistep method is consistent of order 0, if $\alpha_{j0}+\cdots+\alpha_{j,k-1}=0$ for all j.

Lemma 6. *Assume that* HA1, HS1, HM1–2 *hold, and that the multistep method is consistent of order* 0. *We then have*

$$h_{j+k-1}(\beta_{j0}+\cdots+\beta_{j,k-1})(e_k\otimes AJ_{j+1})=(r_{j+1}-I)\mathbb{1},$$

and in particular

$$\sum_{j=1}^{n}h_{j+k-2}\left(\prod_{l=j+1}^{n}r_l\right)\big(e_k\otimes(\beta_{j-1,0}+\cdots+\beta_{j-1,k-1})AJ_j\big)=\left(\prod_{j=1}^{n}r_j-I\right)\mathbb{1}.$$

The proof is straightforward and therefore omitted.

4. Stability for time-dependent operators

In this section, we consider the time-dependent problem

$$u'(t)=A(t)u(t),\quad 0<t\leqslant T,\qquad u(0)\text{ given}, \tag{16}$$

where $A:[0,T]\to L(D,X)$ for some $T>0$. Our basic assumptions on the operator $A(t)$ rely on [12].

HA2. *We assume that the operator $A(t)$ satisfies* HA1 *uniformly in t. In particular, $A(t)$ is supposed to have a fixed domain D.*

The following assumption concerning the Hölder continuity of A is motivated by the framework considered in [13].

HA3. *We suppose that $A\in\mathcal{C}^\alpha([0,T],L(D,X))$ for some $0<\alpha\leqslant 1$, i.e., there exists a constant $L>0$ such that*

$$\big\|A(t)-A(s)\big\|_{X\leftarrow D}\leqslant L(t-s)^\alpha,\quad \text{for all } 0\leqslant s<t\leqslant T.$$

A linear k-step method, applied to (16) takes the form

$$\sum_{i=0}^{k}\alpha_{ni}u_{n+i}=h_{n+k-1}\sum_{i=0}^{k}\beta_{ni}A(t_{n+i})u_{n+i},\quad n\geqslant 0. \tag{17}$$

Again, it is convenient to work with $U_n=(u_n,u_{n+1},\dots,u_{n+k-1})^{\mathrm{T}}$. For this purpose, we denote

$$s_{n+1,i}(z,w)=-(\alpha_{nk}-\beta_{nk}z)^{-1}(\alpha_{ni}-\beta_{ni}w),$$

and we define the companion matrix of the method through

$$r_{n+1}(z_0,z_1,\dots,z_k)=\begin{pmatrix}0 & 1 & 0 & \dots & 0\\ \vdots & 0 & 1 & \ddots & \vdots\\ \vdots & \vdots & \ddots & \ddots & 0\\ 0 & 0 & \dots & 0 & 1\\ s_{n+1,0}(z_k,z_0) & s_{n+1,1}(z_k,z_1) & \dots & \dots & s_{n+1,k-1}(z_k,z_{k-1})\end{pmatrix}.$$

Further, we set

$$s_{ni}(z)=s_{ni}(z,z)\quad\text{and}\quad r_n(z)=r_n(z,\dots,z),\quad n\geqslant 1,$$

which makes our new notation compatible with that of the previous sections. Besides, for integers $j\geqslant 0$, we set $A_j=A(t_j)$, and we write for short

$$R_{n+1}=r_{n+1}(h_{n+k-1}A_n,\dots,h_{n+k-1}A_{n+k}),\quad n\geqslant 0.$$

This allows us to rewrite the numerical method (17) as

$$U_n=R_nR_{n-1}\cdots R_1U_0,\quad n\geqslant 0. \tag{18}$$

We are now in a position to give the stability result for (17). Henceforth, we denote $h_{\max}=\max\{h_j:\ 0\leqslant t_j\leqslant T\}$.

Theorem 3. *Consider the linear multistep discretization* (17) *of Eq.* (16) *on the interval* $[0,T]$, *and assume that* HA2–3, HS1–2, HM1–3 *hold and that $\varrho\Omega^2<1$. Then, there exist positive constants H and C such that for $0<h_j\leqslant H$ the following bounds are valid for all $n\geqslant 1$ with $t_n\leqslant T$*

$$\left\|\prod_{j=1}^{n} R_j\right\|_{X\leftarrow X} + \left\|\prod_{j=1}^{n} R_j\right\|_{D\leftarrow D} \leqslant C, \tag{19a}$$

$$\left\|\prod_{j=2}^{n} R_j \cdot \bigl(\mathcal{I}\otimes J_1(h_{k-1}A_k)\bigr)\right\|_{D\leftarrow X} \leqslant C\left(\frac{1}{t_{n+k-1}} + \frac{|\log h_{\max}|}{t_{n+k-1}^{1-\alpha}}\right). \tag{19b}$$

The constant C depends on the constants that appear in our assumptions and on T, but it is independent of n.

Proof. The main idea for proving the theorem is to compare R_n with the frozen operator $r_n = r_n(h_{n+k-2}A)$. To show the first estimate in the norm of D, we choose $A = A_{n+k-1}$ and use the telescopic identity and the bounds of Theorem 2 to get

$$\left\|\prod_{j=1}^{n} R_j\right\|_{D\leftarrow D} \leqslant C\sum_{j=1}^{n}(t_{n+k-1}-t_{j+k-2})^{-1}\|R_j - r_j\|_{X\leftarrow D}\left\|\prod_{i=1}^{j-1} R_i\right\|_{D\leftarrow D} + C.$$

Due to HA3, we have

$$\|R_j - r_j\|_{X\leftarrow D} \leqslant C\cdot h_{j+k-2}(t_{n+k-1}-t_{j+k-2})^{\alpha},$$

and the application of a discrete Gronwall lemma yields the desired result. The corresponding estimate in the norm of X is obtained in a similar way from

$$\prod_{j=m}^{n} R_j = \sum_{j=m}^{n}\prod_{l=j+1}^{n} R_l(R_j - r_j)\prod_{i=m}^{j-1} r_i + \prod_{j=m}^{n} r_j, \quad 1\leqslant m\leqslant n,$$

by choosing $A = A_{m+k-1}$. A preliminary estimate for (19b) is obtained with the same choice of A from the identity

$$\prod_{j=m+1}^{n} R_j - \prod_{j=m+1}^{n} r_j = \sum_{j=m+1}^{n}\left(\prod_{l=j+1}^{n} R_l - \prod_{l=j+1}^{n} r_l\right)(R_j - r_j)\prod_{i=m+1}^{j-1} r_i + \sum_{j=m+1}^{n}\prod_{l=j+1}^{n} r_l(R_j - r_j)\prod_{i=m+1}^{j-1} r_i.$$

Multiplying this relation from the right with $\mathcal{I}\otimes J_m(h_{m+k-2}A_{m+k-1})$ shows that the inhomogeneity, as an operator from X to D, is bounded by

$$\sum_{j=m+1}^{n} h_{j+k-2}\,(t_{n+k-1}-t_{j+k-2})^{-1}\,(t_{j+k-2}-t_{m+k-2})^{\alpha-1} \leqslant C\cdot\bigl(1+|\log h_{n+k-2}|\bigr).$$

With the help of a discrete Gronwall lemma, we thus get a preliminary bound for (19b) with $\log h_{n+k-2}$ in place of $\log h_{\max}$. In a similar way, we get a bound with $\log h_{k-1}$ instead.

It remains to show the sharper estimate with $\log h_{\max}$. For this, let $k-1 \leqslant m \leqslant n+k-2$ be an index with $h_m = h_{\max}$. Depending on the size of t_m, we distinguish two cases. If $2t_m \geqslant t_{n+k-1} - t_{k-1}$, we write with $J_1 = J_1(h_{k-1}A_k)$

$$\left\| \prod_{j=2}^{n} R_j \cdot (\mathcal{I} \otimes J_1) \right\|_{D \leftarrow X} \leqslant \|R_n \cdots R_{m+1}\|_{D \leftarrow D} \left\| R_m \cdots R_2 (\mathcal{I} \otimes J_1) \right\|_{D \leftarrow X}$$

and use (19a) and the preliminary bound from above to obtain the desired estimate. If $2t_m \leqslant t_{n+k-1} - t_{k-1}$, we use the identity

$$\begin{aligned} \left\| \prod_{j=2}^{n} R_j \cdot (\mathcal{I} \otimes J_1) \right\|_{D \leftarrow X} &\leqslant \left\| R_n \cdots R_{m+1} (\mathcal{I} \otimes J_m) \right\|_{D \leftarrow X} \\ &\quad \times \left\| \big(\mathcal{I} \otimes (\alpha_{m-1,k} - \beta_{m-1,k} h_{m+k-2} A_{m+k-1})\big) R_m \cdots R_2 (\mathcal{I} \otimes J_1) \right\|_{X \leftarrow X}. \end{aligned}$$

Expressing $R_m \cdots R_2 - \varrho_m \cdots \varrho_2$ through the telescopic identity then yields as before the desired result. □

5. Applications to semilinear parabolic problems

As a first application of our stability results, we study the behaviour of time discretizations for semilinear parabolic problems by linear multistep methods with variable stepsizes. In the following, we briefly sketch a convergence result for finite times.

For our purposes, it is useful to employ an abstract formulation of the parabolic initial-boundary value problem as an initial value problem on a Banach space $(X, \|\cdot\|_X)$

$$u'(t) = Au(t) + f\big(u(t)\big), \quad t > 0, \qquad u(0) \text{ given.} \tag{20}$$

Here, the linear operator $A : D \to X$ is assumed to be sectorial. We further suppose that the map $f : \mathcal{O} \subset X_\theta \to X : v \mapsto f(v)$ defined on some open subset of an interpolation space $X_\theta = [X, D]_\theta$, $0 \leqslant \theta < 1$, is Fréchet differentiable and that its Fréchet derivative $Df(v)$ satisfies a local Lipschitz condition. Reaction–diffusion equations and the incompressible Navier–Stokes equations fit into this analytical framework, see [10,12,17].

As linear multistep methods are invariant under linearization, we may assume without loss of generality that Eq. (20) is already linearized around $u(0)$. Consequently, f satisfies

$$\left\| f(v) - f(w) \right\|_X \leqslant L\varrho \|v - w\|_{X_\theta} \tag{21}$$

for all $v, w \in X_\theta$ with $\|v - u(0)\|_{X_\theta} \leqslant \varrho$ and $\|w - u(0)\|_{X_\theta} \leqslant \varrho$.

Applying a linear k-step method (4) of order p to Eq. (20) yields

$$U_{n+1} = r_{n+1} U_n + h_{n+k-1} (\mathcal{I} \otimes J_{n+1}) f(U_{n+1}), \quad n \geqslant 0. \tag{22}$$

Here, we make use of the notation introduced in Section 2. In particular, we have $U_n = (u_n, u_{n+1}, \ldots, u_{n+k-1})^{\mathrm{T}}$, $r_{n+1} = r_{n+1}(h_{n+k-1}A)$ and $J_{n+1} = (\alpha_{nk} - h_{n+k-1}\beta_{nk}A)^{-1}$. Furthermore, with $e_k = (0, \ldots, 0, 1)^{\mathrm{T}} \in \mathbb{R}^k$, we denote

$$f(U_{n+1}) = \sum_{i=0}^{k} \beta_{ni} e_k \otimes f(u_{n+i}).$$

Solving (22), we receive the discrete variation-of-constants formula

$$U_n = \prod_{j=1}^{n} r_j U_0 + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} r_j (\mathcal{I} \otimes J_m) f(U_m), \quad n \geqslant 0. \tag{23}$$

We now carry out a fixed point iteration based on this relation. That is, for finite sequences $V = (V_n)_{n=0}^{N}$ belonging to a ball around the constant sequence $U(0)$ with components $U(0) = \mathbb{1} \otimes u(0)$

$$\mathcal{V} = \Big\{ V = (V_n)_{n=0}^{N} \colon \big\| V - U(0) \big\|_{X_\theta,\infty} = \max_{0 \leqslant n \leqslant N} \big\| e^{-\gamma t_{n+k-1}} \big(V_n - U(0) \big) \big\|_{X_\theta} \leqslant \varrho \Big\},$$

we define a map $\Psi : \mathcal{V} \to \mathcal{V}$ through

$$\Psi(V)_n = \prod_{j=1}^{n} r_j U_0 + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} r_j (\mathcal{I} \otimes J_m) f(V_m), \quad n \geqslant 0.$$

We remark that under the requirements of Theorem 2 the estimate

$$\left\| \prod_{j=1}^{n} r_j \right\|_{X_\theta \leftarrow X_\theta} + (t_{n+k-1} - t_{m+k-2})^\theta \left\| \prod_{j=m+1}^{n} r_j (\mathcal{I} \otimes J_m) \right\|_{X_\theta \leftarrow X} \leqslant C, \tag{24}$$

follows easily by interpolation. Using moreover (21), it is straightforward to show that Ψ is a contraction with contraction factor $\kappa < 1$ for stepsizes sufficiently small and exponent $\gamma > 0$ large enough, since

$$\kappa = CL\varrho \max_{0 \leqslant n \leqslant N} \sum_{m=1}^{n} h_{m+k-2} \frac{\mathrm{e}^{-\gamma (t_{n+k-1} - t_{m+k-1})}}{(t_{n+k-1} - t_{m+k-2})^\theta},$$

with the constant C from (24). Moreover, Ψ maps $\mathcal{V}$ to $\mathcal{V}$ if U_0 lies sufficiently close to $U(0)$. Hence, an application of Banach's fixed point theorem proves the existence of the numerical solution. Besides, the vector $\widehat{U}_n = (\hat{u}_n, \hat{u}_{n+1}, \ldots, \hat{u}_{n+k-1})^{\mathrm{T}}$ comprising the exact solution $\hat{u}_n = u(t_n)$ satisfies (23) with additional defects D_m

$$\widehat{U}_n = \prod_{j=1}^{n} r_j \, \widehat{U}_0 + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} r_j \, (\mathcal{I} \otimes J_m) \big(f\big(\widehat{U}_m\big) + D_m \big), \quad n \geqslant 0.$$

Provided that the $(p+1)$st order derivative of u remains bounded, we have $\|D_m\|_X \leqslant C h_{m+k-2}^p$. Therefore, due to the fact that

$$\begin{aligned} \big\| U_n - \widehat{U}_n \big\|_{X_\theta} &\leqslant \big\| U - \widehat{U} \big\|_{X_\theta,\infty} \leqslant \frac{1}{1-\kappa} \big\| \Psi\big(\widehat{U}\big) - \widehat{U} \big\|_{X_\theta,\infty} \\ &\leqslant C \big\| U_0 - \widehat{U}_0 \big\|_{X_\theta} + C \sum_{m=1}^{N} \frac{h_{m+k-2}}{(t_{n+k-1} - t_{m+k-2})^\theta} \| D_m \|_X, \end{aligned}$$

the desired convergence estimate follows.

Theorem 4. *In the above situation, apply a linear k-step method* (4) *of order p to Eq.* (20). *Assume further that the requirements of Theorem* 2 *are satisfied and that the derivative $u^{(p+1)}(t)$ of the true*

solution remains bounded in X for $t \in [0,T]$. Then, for initial values $u_0, u_1, \ldots, u_{k-1} \in X_\theta$ that lie sufficiently close to $u(0)$ and for stepsize sequences $(h_j)_{j\geqslant 0}$ with $0 < h_j \leqslant h_{\max}$ small enough, the associated numerical solution fulfills the relation

$$\|u_n - u(t_n)\|_{X_\theta} \leqslant C \max_{0\leqslant i\leqslant k-1} \|u_i - u(t_i)\|_{X_\theta} + C \sum_{m=k}^{n} \frac{h_{m-1}^{p+1}}{(t_n - t_{m-1})^\theta},$$

as long as $0 \leqslant t_n \leqslant T$. The constant C depends on the constants that appear in our assumptions and on T, but it is independent of n.

6. Applications to fully nonlinear parabolic problems

In this section, we study variable stepsize linear multistep time discretizations of fully nonlinear parabolic problems. As in the preceding section, we employ an abstract formulation of the partial differential equation and we work within the setting of sectorial operators. Our assumptions on the equation

$$u'(t) = F\big(t, u(t)\big), \quad t > 0, \qquad u(0) \text{ given}, \tag{25}$$

are mainly that of [12]. For nonlinear initial-boundary value problems that can be cast in this analytical framework, see also [7] and [13].

In the following, we specify two illustrations. First, we give a result on the dynamical behaviour nearby a stable equilibrium of the equation, and secondly, a convergence result for finite time intervals.

6.1. Asymptotically stable stationary solutions

We consider an autonomous equation on a Banach space $(X, \|\cdot\|_X)$

$$u'(t) = F\big(u(t)\big), \quad t > 0, \tag{26}$$

with right side $F : \mathcal{D} \subset D \to X$ defined on some open subset $\mathcal{D}$ of another densely embedded Banach space $D \subset X$. Our assumptions on (26) are that of [12], see also [7]. Thus, the Fréchet derivative $DF : \mathcal{D} \to L(D,X)$ satisfies a local Lipschitz condition. Further, for any $v \in \mathcal{D}$, the linear operator $DF(v)$ is sectorial and its graph norm is equivalent to the norm $\|\cdot\|_D$ in D. We suppose that $\bar{u} \in \mathcal{D}$ is an asymptotically stable equilibrium point of Eq. (26), that is, $F(\bar{u}) = 0$, and the sectorial operator $A = DF(\bar{u})$ fulfills the resolvent estimate (2) with $a < 0$.

Linearizing the right side of (26) around the equilibrium point $\bar{u}$ yields a formally semilinear problem

$$u'(t) = Au(t) + G\big(u(t)\big), \quad t > 0, \tag{27}$$

with map G defined through $G(v) = F(v) - Av$ for $v \in \mathcal{D}$. For a linear k-step method (4) applied to (27), in accordance with the notation of Sections 2 and 5, we thus receive the following relation with $G_j = G(u_j)$

$$U_{n+1} = r_{n+1}U_n + h_{n+k-1}e_k \otimes \left(J_{n+1} \sum_{i=0}^{k} \beta_{ni} G_{n+i} \right), \quad n \geqslant 0.$$

We represent the numerical solution by means of a discrete version of a modification of the variation-of-constants formula

$$U_n = \prod_{j=1}^{n} r_j U_0 + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} r_j e_k \otimes \left(J_m \sum_{i=0}^{k} \beta_{m-1,i} G_{n+k-1} \right) + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} r_j e_k \otimes \left(J_m \sum_{i=0}^{k} \beta_{m-1,i} (G_{m+i-1} - G_{n+k-1}) \right). \tag{28}$$

This relation remains well-defined in a space of weighted α-Hölder continuous sequences for some $0 < \alpha < 1$, that is, the set

$$\mathcal{C}_\alpha^\alpha(D) = \left\{ V = (V_n)_{n\geqslant 0} \colon V_n \in \mathcal{D}^k,\ |||V|||_D = \sup_{n\geqslant 0} \|V_n\|_D + \sup_{0<m<n} t_m^\alpha (t_n - t_m)^{-\alpha} \|V_n - V_m\|_D < \infty \right\},$$

endowed with the norm $|||\cdot|||_D$. With the help of Lemma 6, we are able to bound the second term on the right side of formula (28).

For our situation, it is known that if the initial value lies close to the equilibrium point $\bar{u}$, then the true solution decays against $\bar{u}$ exponentially fast. Permitting increasing stepsizes, a similar result holds true for linear multistep methods if stability estimates of the form

$$\begin{aligned} \left\| \prod_{j=1}^{n} r_j \right\|_{D\leftarrow D} &\leqslant \frac{C}{1 + t_{n+k-1}^\eta}, \\ \left\| \prod_{j=m+1}^{n} r_j (\mathcal{I} \otimes J_m) \right\|_{D\leftarrow X} &\leqslant \frac{C}{t_{n+k-1} - t_{m+k-2} + (t_{n+k-1} - t_{m+k-2})^{\eta+1}}, \end{aligned} \tag{29}$$

with exponent $\eta > 0$ hold for $n \geqslant 1$, see Theorem 1 and the subsequent discussion.

Theorem 5. *Under the above requirements on F, let $\bar{u}$ be an asymptotically stable equilibrium point of* (26)*. Apply a linear k-step method with stepsizes $(h_j)_{j\geqslant 0}$ such that* (29) *is valid. Then, for $0 < \nu < \eta$, there exist constants $\delta > 0$ and $C > 0$ such that for all initial values $u_0, u_1, \ldots, u_{k-1} \in \mathcal{D}$ with $\|u_i - \bar{u}\|_D \leqslant \delta$, $0 \leqslant i \leqslant k-1$, the numerical solution $(u_n)_{n\geqslant k}$ satisfies the estimate*

$$\|u_n - \bar{u}\|_D \leqslant \frac{C}{1 + t_n^\nu} \max_{0\leqslant i\leqslant k-1} \|u_i - \bar{u}\|_D, \qquad n \geqslant 0.$$

Proof. We shortly indicate the proof of Theorem 5. For a precise explanation of the employed techniques, we refer to [7] and [18]. For constructing the numerical solution, we use the ideas of the preceding section. We carry out a fixed point iteration relying on (28) in a subset of $\mathcal{C}_\alpha^\alpha(D)$. In order to capture the decaying behaviour of the numerical solution, we introduce appropriate weights. More precisely, for $0 < \nu < \eta$, we define the norm

$$|||V|||_{\nu,D} = \sup_{n\geqslant 0} \left\| \left(1 + t_n^\nu\right) V_n \right\|_D + \sup_{0<m<n} t_m^\alpha (t_n - t_m)^{-\alpha} \left\| \left(1 + t_n^\nu\right) V_n - \left(1 + t_m^\nu\right) V_m \right\|_D,$$

and set $\mathcal{V} = \{V = (V_n)_{n\geqslant 0}\colon\ |||V - \overline{U}|||_{v,D} \leqslant \varrho\}$ where $\overline{U}$ denotes the constant sequence with components equal to $\mathbb{1} \otimes \bar{u}$. Then the iteration based on (28) turns out to be a contraction on $\mathcal{V}$ provided that ϱ and δ are chosen sufficiently small. □

6.2. Convergence for finite times

Another approach that avoids the technicalities which arise in connection with the modified variation-of-constants formula and the consideration of the sequence space $\mathcal{C}^{\alpha}_{\alpha}(D)$ is based on a slightly stronger setting. This framework is presented in [13].

We consider an initial value problem of the form

$$u'(t) = F\big(t, u(t)\big), \quad t > 0, \qquad u(0) \text{ given}, \tag{30}$$

where the right-hand side function $F : [0, T] \times \mathcal{D} \to X : (t, v) \mapsto F(t, v)$ is defined on an open subset $\mathcal{D} \subset D$ of a densely embedded Banach space $D \subset X$. We suppose that F is twice continuously Fréchet differentiable and that its Fréchet derivative $D_2F(t, v)$ with respect to the second variable is a sectorial operator in X. Moreover, we assume that the graph-norm of $D_2F(t, v)$ is equivalent to the norm of D for all $0 \leqslant t \leqslant T$ and for all $v \in \mathcal{D}$. In view of our convergence result, we further suppose that the true solution of (30) is differentiable. As a consequence, the hypotheses HA2–3 are satisfied with $\alpha = 1$ on $[0, T]$. Linearizing around $u(t)$ leads to the equation

$$u'(t) = A(t)u(t) + G\big(t, u(t)\big), \quad t > 0,$$

involving the time-dependent sectorial operator $A(t) = D_2F(t, u(t))$. Here, the nonlinearity G is defined by $G(t, v) = F(t, v) - A(t)v$ for $(t, v) \in [0, T] \times \mathcal{D}$. In the present situation, the discrete variation-of-constants formula is still meaningful. For a linear multistep method (17), we receive the following relation

$$U_n = \prod_{j=1}^{n} R_j\, U_0 + \sum_{m=1}^{n} h_{m+k-2} \prod_{j=m+1}^{n} R_j(\mathcal{I} \otimes J_m)\, G(U_m), \quad n \geqslant 0. \tag{31}$$

Here, we use the abbreviations introduced in Section 4. In particular, we let $A_n = A(t_n)$, $R_{n+1} = r_{n+1}(h_{n+k-1}A_n, \ldots, h_{n+k-1}A_{n+k})$ and $J_{n+1} = (\alpha_{nk} - h_{n+k-1}\beta_{nk}A_{n+k})^{-1}$. Besides, we set

$$G(U_{n+1}) = \sum_{i=0}^{k} \beta_{ni}\, e_k \otimes G(t_{n+i}, u_{n+i}).$$

Following [13], we employ as in Section 5 a fixed point iteration based on (31). We define the fixed point operator Ψ on a tube around the true solution $\widehat{U}_n = (u(t_n), \ldots, u(t_{n+k-1}))^{\mathrm{T}}$

$$\mathcal{V} = \Big\{V = (V_n)_{n=0}^{N}\colon\ \big\|V - \widehat{U}\big\|_{D,\infty} = \max_{0\leqslant n\leqslant N} \big\|V_n - \widehat{U}_n\big\|_D \leqslant \varrho h_{\max}^{p/2}\Big\}.$$

By means of the stability estimates from Theorem 3 it follows that Ψ is a contraction and maps $\mathcal{V}$ to $\mathcal{V}$ if for all $n \leqslant N$

$$C\, h_{\max}^{p/2} \sum_{m=k}^{n} \left(\frac{h_{m-1}}{t_n - t_{m-1}} + \frac{h_{m-1}|\log h_{\max}|}{(t_n - t_{m-1})^{1-\alpha}} \right) < 1,$$

with a constant C depending on the stability constant, the Lipschitz constant of F, the bound on the $(p+1)$st-order derivative of the true solution, on the coefficients of the method, and on ϱ. In particular, this bound is satisfied if

$$\bigl(1+|\log h_{\min}|\bigr)\, h_{\max}^{p/2} < \gamma, \tag{32}$$

with γ sufficiently small. We remark that this is essentially a condition on the maximal stepsize $h_{\max}$.

We are now prepared to state the convergence result for finite time intervals.

Theorem 6. *In the above situation and under the assumptions of Theorem* 3, *apply a linear k-step method* (17) *of order p to Eq.* (30). *Suppose further that the derivative $u^{(p+1)}(t)$ of the true solution remains bounded in X for $t \in [0,T]$. Then, provided that the stepsize sequence $(h_j)_{j\geqslant 0}$ satisfies* (32) *with γ sufficiently small, the following bound is valid. For initial values $u_0, u_1, \ldots, u_{k-1}$ in $\mathcal{D}$ with $\|u_i - u(t_i)\|_D$ sufficiently small, the associated numerical solution fulfills the relation*

$$\bigl\|u_n - u(t_n)\bigr\|_D \leqslant C \max_{0\leqslant i\leqslant k-1} \bigl\|u_i - u(t_i)\bigr\|_D + C\sum_{m=k}^{n}\left(\frac{h_{m-1}^{p+1}}{t_n - t_{m-1}} + \frac{h_{m-1}^{p+1}|\log h_{\max}|}{(t_n - t_{m-1})^{1-\alpha}}\right),$$

for all $0 \leqslant t_n \leqslant T$. The constant C depends on the constants that appear in our assumptions and on T, but it is independent of n.

References

[1] N. Bakaev, A. Ostermann, Long-term stability of variable stepsize approximations of semigroups, Math. Comp. 71 (2002) 1545–1567.

[2] J. Becker, A second order backward difference method with variable steps for a parabolic problem, BIT 38 (1998) 644–662.

[3] M. Calvo, R.D. Grigorieff, Time discretisation of parabolic problems with the variable 3-step BDF, BIT 42 (2002) 689–701.

[4] A. Eder, G. Kirlinger, A normal form for multistep companion matrices, Math. Models Methods Appl. Sci. 11 (2001) 57–70.

[5] E. Emmrich, Stability and error of the variable two-step BDF for semilinear parabolic problems, Extended version of Preprint No. 703, TU Berlin, Berlin, 2001.

[6] C.W. Gear, K.W. Tu, The effect of variable mesh size on the stability of multistep methods, SIAM J. Numer. Anal. 11 (1974) 1025–1043.

[7] C. González, A. Ostermann, C. Palencia, M. Thalhammer, Backward Euler discretization of fully nonlinear parabolic problems, Math. Comp. 71 (2002) 125–145.

[8] R.D. Grigorieff, Stability of multistep-methods on variable grids, Numer. Math. 42 (1983) 359–377.

[9] E. Hairer, S.P. Nørsett, G. Wanner, Solving Ordinary Differential Equations I. Nonstiff Problems, second revised ed., Springer, Berlin, 1993.

[10] D. Henry, Geometric Theory of Semilinear Parabolic Equations, in: Lecture Notes in Math., vol. 840, Springer, Berlin, 1981.

[11] A.T. Hill, E. Süli, Upper semicontinuity of attractors for linear multistep methods approximating sectorial evolution equations, Math. Comp. 64 (1995) 1097–1122.

[12] A. Lunardi, Analytic Semigroups and Optimal Regularity in Parabolic Problems, Birkhäuser, Basel, 1995.

[13] A. Ostermann, M. Thalhammer, Convergence of Runge–Kutta methods for nonlinear parabolic problems, Appl. Numer. Math. (2002) 367–380.

[14] C. Palencia, A stability result for sectorial operators in Banach spaces, SIAM J. Numer. Anal. 30 (1993) 1373–1384.

[15] C. Palencia, On the stability of variable stepsize rational approximations of holomorphic semigroups, Math. Comp. 62 (1994) 93–103.

[16] C. Palencia, B. García-Archilla, Stability of linear multistep methods for sectorial operators in Banach spaces, Appl. Numer. Math. 445 (1993) 503–520.
[17] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer, New York, 1983.
[18] M. Thalhammer, Runge–Kutta Time Discretization of Fully Nonlinear Parabolic Problems, Thesis, Universität Innsbruck, Innsbruck, 2000.

1.5. Multistep methods for singularly perturbed problems

On the convergence behaviour of variable stepsize multistep methods for singularly perturbed problems

MECHTHILD THALHAMMER

BIT Numerical Mathematics (2004) 44, 343-361

BIT Numerical Mathematics **44:** 343–361, 2004.

ON THE CONVERGENCE BEHAVIOUR OF VARIABLE STEPSIZE MULTISTEP METHODS FOR SINGULARLY PERTURBED PROBLEMS ⋆

MECHTHILD THALHAMMER[1]

[1] *Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, Technikerstraße 13, A-6020 Innsbruck, Austria. email: Mechthild.Thalhammer@uibk.ac.at*

Abstract.

In this note, we investigate the convergence behaviour of linear multistep discretizations for singularly perturbed systems, emphasising the features of variable stepsizes. We derive a convergence result for $A(\varphi)$-stable linear multistep methods and specify a refined error estimate for backward differentiation formulas. Important ingredients in our convergence analysis are stability bounds for non-autonomous linear problems that are obtained by perturbation techniques.

AMS subject classification (2000): 65L05, 65L06, 65L20.

Key words: singular perturbation problems, linear multistep methods, backward differentiation formulas, variable stepsizes, stability, convergence.

1 Introduction.

In this paper, we analyse the convergence and stability behaviour of variable stepsize linear multistep methods applied to singularly perturbed systems. Singular perturbation problems arise in various applications such as chemical kinetics and fluid mechanics, see for example [7, 8, 5] and references therein. Another illustration modelling oscillations in electric circuits is the well-known unforced Van der Pol equation [12, 13]

$$\ddot{z}(\tau) + \mu\big(z^2(\tau) - 1\big)\dot{z}(\tau) + z(\tau) = 0, \quad \mu \gg 1.$$

By rescaling the independent variable $\tau = \mu t$ and introducing a new function y, this nonlinear differential equation takes the usual form of a first order system

$$\begin{cases} y'(t) = -z(t), \\ \varepsilon z'(t) = y(t) + z(t) - \dfrac{1}{3}z^3(t), \end{cases} \qquad \text{where } \varepsilon = \frac{1}{\mu^2} \ll 1.$$

⋆ Received July 2003. Accepted March 2004. Communicated by Timo Eirola.

In this note, more generally, we consider a singularly perturbed system of nonlinear differential equations involving a small parameter $0 < \varepsilon \leq \varepsilon_0$

$$\begin{cases} y'(t) = f\big(y(t), z(t)\big), \\ \varepsilon z'(t) = g\big(y(t), z(t)\big). \end{cases}$$

A basic assumption is that the solutions y and z are bounded and have bounded derivatives with bounds independent of the parameter ε for $\varepsilon \in (0, \varepsilon_0]$. This requirement can be achieved by choosing the initial values on the existent invariant manifold, e.g. In this situation, it is shown by Lubich [6, Theorem 3] that a strongly stable linear k-step method of order p, applied with constant time step $h > 0$ sufficiently small, satisfies the following error estimate on bounded time intervals $t_n = nh \leq T$ for some $0 < \gamma < 1$ if $h \geq \varepsilon$

$$\begin{aligned} &\|y_n - y(t_n)\| + \|z_n - z(t_n)\| \\ &\quad \leq C \max_{0 \leq i \leq k-1} \|y_i - y(t_i)\| + C\big(h + \gamma^n\big) \max_{0 \leq i \leq k-1} \|z_i - z(t_i)\| + \\ &\qquad + Ch^p \int_0^{t_n} \big\|y^{(p+1)}(\tau)\big\| \, \mathrm{d}\tau + \varepsilon \, Ch^p \max_{0 \leq \tau \leq t_n} \big\|z^{(p+1)}(\tau)\big\|. \end{aligned}$$

In consideration of practical implementations, our objective is to extend this convergence estimate to variable stepsizes. In this regard, main techniques are a linearization of the right-hand side of the singularly perturbed equation along the exact solution and a fixed-point iteration based on a discrete variation-of-constants formula. Thereto, essential tools are stability bounds for non-autonomous linear problems. For proving the needed stability estimates, we employ perturbation techniques related to [9] where variable stepsize linear multistep discretizations of parabolic equations are analysed. As in [9], following an approach used by [3] and later by [10], our stability estimates involve a stability factor which depends on the stepsize sequence. As a consequence, stability is obtained under the requirement that the considered stepsize sequence varies smoothly. Moreover, an essential ingredient is a decomposition of the companion matrix of a variable stepsize linear multistep method specified in [10] and further investigated in [9].

The contents of the present paper are as follows. In Section 2, we first introduce the problem and numerical method classes and give the precise assumptions on the singularly perturbed system, the linear multistep method, and the stepsize sequence. Besides, we collect some useful relations for the solution and the right-hand side of the singular perturbation problem. Section 3 is devoted to the derivation of the necessary stability results stated in Theorem 3.3. The main idea is to relate the original equation to a less involved problem. The desired bounds then follow from a telescopic identity and a Gronwall lemma. In Section 4, we finally prove the analogue of Lubich's convergence estimate for variable stepsizes.

2 Problem and numerical discretization.

In this section, we introduce the problem and numerical method class under consideration. We specify the general scheme of a variable stepsize linear multistep method for a singular perturbation problem and further state the precise hypotheses on the problem, the method, and the stepsize sequence. For our purposes, it is useful to write the differential equation and its discretization in compact vector notation. Auxiliary results for the solution and the right-hand side of the differential equation are given in Sections 2.3 and 2.4.

2.1 Singular perturbation problem.

We consider a singularly perturbed system of ordinary differential equations involving a small parameter $0 < \varepsilon \leq \varepsilon_0$

$$\begin{cases} y'(t) = f\big(y(t), z(t)\big), & y(0) \text{ given}, \\ \varepsilon z'(t) = g\big(y(t), z(t)\big), & z(0) \text{ given}, \end{cases} \tag{2.1}$$

with solution $(y(t), z(t))^{\mathrm{T}} \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ defined on some finite time interval $[0, T]$. For notational simplicity, the dependence of y and z on ε is omitted.

In many cases, it is convenient to employ a compact vector notation of (2.1). For that reason, we set $u = (y, z)^{\mathrm{T}}$ and denote the function defining the right-hand side of the differential equation by $F = (F_1, F_2)^{\mathrm{T}} = (f, g)^{\mathrm{T}}$. Therewith, the above initial value problem writes as

$$I_\varepsilon u'(t) = F\big(u(t)\big), \quad u(0) \text{ given}. \tag{2.2}$$

Here, I_ε denotes a diagonal matrix of dimension $m_1 + m_2$ with entries 1 or ε, respectively. The minimum regularity assumption on the function F is as follows.

HP 1. *Assume that F is differentiable and that its first derivative DF is locally Lipschitz-continuous.*

A basic concept for our proof of the convergence estimate stated in Theorem 4.1 is a linearization of the right-hand side of (2.2) along the exact solution. This yields the equation

$$I_\varepsilon u'(t) = F\big(u(t)\big) = A(t)u(t) + G\big(t, u(t)\big), \quad u(0) \text{ given}, \tag{2.3}$$

with time-dependent matrix $A = (A_{ij})_{1 \leq i,j \leq 2}$ where $A_{ij}(t) = D_j F_i(u(t))$. For some $v = (v_1, v_2)^{\mathrm{T}} \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, the nonlinear function $G = (G_1, G_2)^{\mathrm{T}}$ is given by $G(t, v) = F(v) - A(t)v$, that is, $G_i(t, v) = F_i(v) - A_{i1}(t)v_1 - A_{i2}(t)v_2$.

Clearly, $A(t)$ is uniformly bounded for $0 \leq t \leq T$. The following assumption on the diagonal element A_{22} is essential for our stability and convergence analysis in Sections 3 and 4.

HP 2. *Suppose that for every $0 \leq t \leq T$ all eigenvalues $\lambda(t)$ of $A_{22}(t)$ have negative real parts $\Re(\lambda(t)) \leq a_{22} < 0$.*

By multiplying Equation (2.3) with the inverse of I_ε, we alternatively obtain

$$u'(t) = \mathscr{F}\big(u(t)\big) = \mathscr{A}(t)u(t) + \mathscr{G}\big(t, u(t)\big), \quad u(0) \text{ given}, \tag{2.4}$$

where $\mathscr{F}(v) = I_\varepsilon^{-1}F(v)$, $\mathscr{A}(t) = I_\varepsilon^{-1}A(t)$ and $\mathscr{G}(t,v) = I_\varepsilon^{-1}G(t,v)$.

2.2 Variable stepsize linear multistep method.

In this section, we specify our hypotheses on the linear multistep method applied to the initial value problem (2.4).

Let $(h_j)_{j\geq 0}$ be a sequence of positive time steps with ratios $\omega_j = h_j/h_{j-1}$, $j \geq 1$, and associated grid points $t_j = h_0 + h_1 + \cdots + h_{j-1}$, $j \geq 0$. For given starting values $u_0, u_1, \ldots, u_{k-1}$, the numerical approximation u_{n+k} to the value of the exact solution at time t_{n+k}, $n \geq 0$, is determined by a linear k-step method, that is, u_{n+k} is given recursively by a relation of the form

$$\begin{aligned} \sum_{i=0}^{k} \alpha_{ni} u_{n+i} &= h_{n+k-1} \sum_{i=0}^{k} \beta_{ni} \mathscr{F}(u_{n+i}) \\ &= h_{n+k-1} \sum_{i=0}^{k} \beta_{ni} \big(\mathscr{A}(t_{n+i})u_{n+i} + \mathscr{G}(t_{n+i}, u_{n+i})\big), \quad n \geq 0, \end{aligned} \tag{2.5}$$

where the coefficients of the method α_{ni} and β_{ni}, $0 \leq i \leq k$, depend on the quantities $\omega_{n+1}, \omega_{n+2}, \ldots, \omega_{n+k-1}$. Throughout, the components of the numerical solution value u_j are denoted by $u_j = (y_j, z_j)^{\mathrm{T}} \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$, $j \geq 0$.

For the analysis, we employ a compact notation of the multistep method (2.5) as a one-step method for the vector $U_n = (u_n, u_{n+1}, \ldots, u_{n+k-1})^{\mathrm{T}} \in \mathbb{R}^{k\cdot m}$ comprising k consecutive numerical approximations. We introduce complex functions $s_j(z) = (\alpha_{j-1,k} - \beta_{j-1,k}z)^{-1}$ and $c_{ji}(z,\tilde{z}) = -s_j(z)(\alpha_{j-1,i} - \beta_{j-1,i}\tilde{z})$ for $j \geq 1$ and $0 \leq i \leq k-1$. Therewith, the companion matrix of the method equals

$$r_j(z_0, z_1, \ldots, z_k) = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ \vdots & 0 & 1 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \ldots & 0 & 1 \\ c_{j0}(z_k, z_0) & c_{j1}(z_k, z_1) & \ldots & \ldots & c_{j,k-1}(z_k, z_{k-1}) \end{pmatrix}.$$

If $z_i = z$ for all $0 \leq i \leq k$, we set $r_j(z) = r_j(z, z, \ldots, z)$. The index j indicates the dependence of r_j on the coefficients $\alpha_{j-1,i}$ and $\beta_{j-1,i}$ of the method and thus on $\omega_j, \omega_{j+1}, \ldots, \omega_{j+k-2}$. For constant stepsizes, i.e. $\omega_j = 1$ for all $j \geq 1$, we omit the index and write r for short. Further, we let $\mathscr{A}_j = \mathscr{A}(t_j)$ for $j \geq 0$ and define $\mathscr{J}_j = s_j(h_{j+k-2}\mathscr{A}_{j+k-1})$ and

$$\mathscr{R}_j = r_j(h_{j+k-2}\mathscr{A}_{j-1}, h_{j+k-2}\mathscr{A}_j, \ldots, h_{j+k-2}\mathscr{A}_{j+k-1})$$

for $j \geq 1$. Besides, with $e_k = (0, \ldots, 0, 1)^{\mathrm{T}} \in \mathbb{R}^k$, we denote

$$\mathscr{G}(U_j) = \sum_{i=0}^{k} \beta_{j-1,i} e_k \otimes \mathscr{G}(t_{j+i-1}, u_{j+i-1}).$$

We recall that for matrices $B = (b_{ij})_{ij}$ and M, the (i,j)-th component of the Kronecker product $B \otimes M$ equals $b_{ij}M$.

With the above notation, the numerical scheme (2.5) becomes

$$U_{n+1} = \mathscr{R}_{n+1}U_n + h_{n+k-1}\mathscr{J}_{n+1}\mathscr{G}(U_{n+1}), \quad n \geq 0.$$

Solving this recursion yields the following relation

$$U_n = \prod_{i=1}^{n} \mathscr{R}_i U_0 + \sum_{j=1}^{n} h_{j+k-2} \prod_{i=j+1}^{n} \mathscr{R}_i \mathscr{J}_j \mathscr{G}(U_j), \quad n \geq 0, \tag{2.6}$$

a representation of the numerical solution by means of a discrete *variation-of-constants formula*.

Our hypotheses on the stepsize sequence and the linear multistep scheme rely on [9]. As in [6], we further suppose that the stepsizes are bounded from below by the parameter ε. For the definition of the notions of order and $A(\varphi)$-stability of a variable stepsize linear multistep method, we refer to [4, 5].

The following assumption on the stepsize ratios is fulfilled by classical step size selection procedures such as the differential/algebraic system solver DASSL based on backward differentiation formulas, see [11].

HS 1. *Suppose $h_j \geq \varepsilon$ for all $j \geq 0$. Assume further that there exists $\Omega > 1$ such that the stepsize ratios $\omega_j = h_j/h_{j-1}$ fulfill $\Omega^{-1} \leq \omega_j \leq \Omega$ for $j \geq 1$.*

The stability factors of the linear multistep method are of the form

$$C_j = D_1 \prod_{i=1}^{j} \big(1 + D_2|\omega_i - 1|\big)^2$$

with positive constants D_1 and D_2, see [9, Theorem 1]. In order to obtain meaningful stability and convergence estimates, we need these quantities to be bounded by a moderate constant.

HS 2. *Assume that the stability factors C_j of the linear multistep method are uniformly bounded by a constant for all $j \geq 1$ such that $t_j \leq T$.*

Besides, we suppose that the following stability requirement is satisfied for constant stepsizes. The angle $0 < \varphi < \pi/2$ is chosen in such a way that for some $a \in \mathbb{R}$ the spectrum of $A(t)$ is contained in the interior of the sector $S_\varphi(a) = \{\lambda \in \mathbb{C} : |\arg(a - \lambda)| \leq \varphi\} \cup \{a\}$ for all $0 \leq t \leq T$.

HM 1. *Assume that the linear multistep method (2.5) is $A(\varphi)$-stable and strictly stable at zero and infinity. Thus, $\lambda = 1$ is the only eigenvalue of the companion matrix at zero with modulus one, and the spectral radius of the companion matrix at infinity, $\sigma = \sigma(r(\infty))$, is less than one.*

Moreover, we make use of the following hypothesis for variable stepsizes.

HM 2. *Suppose that the coefficients α_{ji} and β_{ji} of the linear multistep scheme* (2.5) *are bounded for all stepsize sequences satisfying* HS1. *Assume further that the rational functions $s_j(z)$ and $c_{ji}(z)$ remain bounded for $z \in S_\varphi(0)$.*

We close this section with an assumption concerning the order of the method.

HM 3. *Assume that the multistep method* (2.5) *is consistent of order $p \geq 1$.*

2.3 Exact solution.

Throughout the paper, we employ the abbreviation $\hat{u}_j = u(t_j)$ for the value of the solution of (2.4) at time t_j, $j \geq 0$. Inserting the solution into the numerical scheme (2.5)

$$\begin{aligned}\sum_{i=0}^{k} \alpha_{ni} \hat{u}_{n+i} &= h_{n+k-1} \sum_{i=0}^{k} \beta_{ni} \big(\mathscr{F}(\hat{u}_{n+i}) + \delta_{n+1}\big) \\ &= h_{n+k-1} \sum_{i=0}^{k} \beta_{ni} \big(\mathscr{A}(t_{n+i})\hat{u}_{n+i} + \mathscr{G}(t_{n+i}, \hat{u}_{n+i}) + \delta_{n+1}\big), \quad n \geq 0,\end{aligned}$$

defines the defect δ_{n+1} at t_{n+k}. In vector notation, we have

$$\widehat{U}_{n+1} = \mathscr{R}_{n+1} \widehat{U}_n + h_{n+k-1} \mathscr{J}_{n+1} \big(\mathscr{G}(\widehat{U}_{n+1}) + \Delta_{n+1}\big), \quad n \geq 0,$$

with $\widehat{U}_n = (\hat{u}_n, \hat{u}_{n+1}, \ldots, \hat{u}_{n+k-1})^{\mathrm{T}}$ and

$$\Delta_{n+1} = \sum_{i=0}^{k} \beta_{ni} e_k \otimes \delta_{n+1}.$$

As a consequence, we receive the analogue of (2.6) for the exact solution

$$\widehat{U}_n = \prod_{i=1}^{n} \mathscr{R}_i \, \widehat{U}_0 + \sum_{j=1}^{n} h_{j+k-2} \prod_{i=j+1}^{n} \mathscr{R}_i \mathscr{J}_j \big(\mathscr{G}(\widehat{U}_j) + \Delta_j\big), \quad n \geq 0. \tag{2.7}$$

Moreover, provided that the solution $u = (y, z)^{\mathrm{T}}$ of (2.4) is sufficiently smooth, the bounds

$$\begin{aligned}\big\|\delta_j^{(1)}\big\| &\leq C h_{j+k-2}^{p-1} \int_{t_{j-1}}^{t_{j+k-1}} \big\|y^{(p+1)}(\tau)\big\| \, \mathrm{d}\tau, \\ \big\|\delta_j^{(2)}\big\| &\leq C h_{j+k-2}^{p} \max_{t_{j-1} \leq \tau \leq t_{j+k-1}} \big\|z^{(p+1)}(\tau)\big\|,\end{aligned} \tag{2.8}$$

for the components $\delta_j^{(i)} \in \mathbb{R}^{m_i}$ of δ_j, $i = 1, 2$, follow by means of a Taylor series expansion. Here, $\|\cdot\|$ denotes an arbitrary norm on $\mathbb{R}^{m_i}$. For notational simplicity, we do not consider different norms on $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$.

2.4 Nonlinearity.

In the following, we state an auxiliary estimate for the nonlinear function G defined in (2.3). For simplicity, as indicated above, we endow $\mathbb{R}^{m_1}$ and $\mathbb{R}^{m_2}$ with the same norm $\|\cdot\|$ and define the norm on the product space $\mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$ through $\|v\| = \|v_1\| + \|v_2\|$ for $v = (v_1, v_2)^{\mathrm{T}} \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$.

LEMMA 2.1. *Under hypothesis* HP1*, there exists a constant* $L > 0$ *such that*

$$\|G_i(t,v) - G_i(t,w)\| \le L\varrho\|v - w\|, \quad i = 1, 2,$$

for all $v, w \in \mathbb{R}^m$ *satisfying* $\|v - u(t)\| \le \varrho$ *and* $\|w - u(t)\| \le \varrho$.

PROOF. Fix $t \in [0,T]$ and consider a ball of radius $\varrho > 0$ around the value of the solution $u(t) = (y(t), z(t))^{\mathrm{T}}$. Due to the fact that DF is locally Lipschitz continuous by HP1, there exists $C > 0$ such that

$$\big\|D_jF_i(v) - D_jF_i(w)\big\| \le C\|v - w\|, \quad i, j = 1, 2,$$

for all $v, w \in \mathbb{R}^m$ with $\|v - u(t)\| \le \varrho$ and $\|w - u(t)\| \le \varrho$. From the identity

$$\begin{aligned} G_i(t,v) - G_i(t,w) &= F_i(v) - F_i(w) - A_{i1}(t)(v_1 - w_1) - A_{i2}(t)(v_2 - w_2) \\ &= \int_0^1 \big(D_1F_i(\sigma v_1 + (1-\sigma)w_1, v_2) - D_1F_i(y(t), z(t))\big)(v_1 - w_1)\,\mathrm{d}\sigma + \\ &\quad + \int_0^1 \big(D_2F_i(w_1, \sigma v_2 + (1-\sigma)w_2) - D_2F_i(y(t), z(t))\big)(v_2 - w_2)\,\mathrm{d}\sigma \end{aligned}$$

the desired estimate follows with $L = 2C$. □

3 Stability estimates.

Throughout this section, we make use of the hypotheses and notation introduced in Section 2.

We next derive stability bounds for the linear multistep discretization (2.5) of (2.4). Hence, it suffices to consider the associated linear equation

$$u'(t) = \mathscr{A}(t)u(t), \tag{3.1}$$

where the linear multistep approximation simplifies to $U_n = \mathscr{R}_n\mathscr{R}_{n-1}\cdots\mathscr{R}_1U_0$. So, we study $\mathscr{R}_n\mathscr{R}_{n-1}\cdots\mathscr{R}_\ell X$ for arbitrary $X \in \mathbb{R}^{k\cdot m}$ and $1 \le \ell \le n$, and, in view of formula (2.6), also $h_{j+k-2}\mathscr{R}_n\mathscr{R}_{n-1}\cdots\mathscr{R}_{j+1}e_k \otimes \mathscr{J}_j(I_\varepsilon^{-1}x)$ for some $x \in \mathbb{R}^m$ and $1 \le j \le n$. Our basic idea is to compare $\mathscr{R}_i$ with the companion matrix

$$\mathscr{T}_i = r_i(h_{i+k-2}\mathscr{L}_{i-1}, h_{i+k-2}\mathscr{L}_i, \ldots, h_{i+k-2}\mathscr{L}_{i+k-1})$$

that corresponds to the lower triangular matrix $\mathscr{L}(t)$ resulting from $\mathscr{A}(t)$. In other words, we relate (3.1) to the partly coupled problem

$$u'(t) = \mathscr{L}(t)u(t), \quad \text{where } \mathscr{L}(t) = \begin{pmatrix} A_{11}(t) & 0 \\ \frac{1}{\varepsilon}A_{21}(t) & \frac{1}{\varepsilon}A_{22}(t) \end{pmatrix}. \tag{3.2}$$

In order to prove the necessary stability results for (3.2), as a first step, we consider in Section 3.1 the fully decoupled system

$$u'(t) = \mathscr{D}(t)u(t) \quad \text{with } \mathscr{D}(t) = \begin{pmatrix} A_{11}(t) & 0 \\ 0 & \frac{1}{\varepsilon}A_{22}(t) \end{pmatrix}. \tag{3.3}$$

In this special case, the desired stability estimates for the associated companion matrix

$$\mathscr{S}_i = r_i(h_{i+k-2}\mathscr{D}_{i-1}, h_{i+k-2}\mathscr{D}_i, \dots, h_{i+k-2}\mathscr{D}_{i+k-1})$$

are a consequence of the results given in [9].

3.1 The decoupled problem.

For studying the stability behaviour of the linear multistep method (2.5) applied to the decoupled equation (3.3), it is useful to consider each component of the numerical solution $u_n = (y_n, z_n)^{\mathrm{T}}$ separately, that is, we henceforth identify $U_n = (u_n, u_{n+1}, \dots, u_{n+k-1})^{\mathrm{T}}$ with the reordered vector $U_n = (Y_n, Z_n)^{\mathrm{T}}$ where $Y_n = (y_n, y_{n+1}, \dots, y_{n+k-1})^{\mathrm{T}}$ and $Z_n = (z_n, z_{n+1}, \dots, z_{n+k-1})^{\mathrm{T}}$. Hence, in this new order of components, for $X = (X_1, X_2)^{\mathrm{T}} \in \mathbb{R}^{k\cdot m_1} \times \mathbb{R}^{k\cdot m_2}$, we receive

$$\prod_{i=\ell}^{n} \mathscr{S}_i X = \prod_{i=\ell}^{n} \begin{pmatrix} R_i & 0 \\ 0 & S_i \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \prod_{i=\ell}^{n} R_i X_1 & 0 \\ 0 & \prod_{i=\ell}^{n} S_i X_2 \end{pmatrix}. \tag{3.4}$$

Here, $R_i = r_i\big(h_{i+k-2}A_{11}(t_{i-1}), h_{i+k-2}A_{11}(t_i), \dots, h_{i+k-2}A_{11}(t_{i+k-1})\big)$ and

$$S_i = r_i\bigg(\frac{h_{i+k-2}}{\varepsilon}A_{22}(t_{i-1}), \frac{h_{i+k-2}}{\varepsilon}A_{22}(t_i), \dots, \frac{h_{i+k-2}}{\varepsilon}A_{22}(t_{i+k-1})\bigg)$$

denote the companion matrix associated with the first and the second component, respectively. The following result provides an estimate for (3.4). For later use, we further introduce $J_j = s_j(h_{j+k-2}A_{11}(t_{j+k-1}))$ and

$$K_{\varepsilon,j} = \frac{h_{j+k-2}}{\varepsilon}K_j = \frac{h_{j+k-2}}{\varepsilon}s_j\bigg(\frac{h_{j+k-2}}{\varepsilon}A_{22}(t_{j+k-1})\bigg).$$

An application of the integral formula of Cauchy as indicated in the proof of Lemma 3.1 shows the boundedness of J_j and $K_{\varepsilon,j}$, see also Remark 3.1 below.

LEMMA 3.1. *Under* HP2 *assume that the linear multistep discretization* (2.5) *of Equation* (3.3) *satisfies* HS1–2, HM1–3, *and further* $\sigma\Omega^2 < 1$. *Then, there exist* $H > 0$ *and* $C > 0$ *such that for any stepsize sequence* $(h_j)_{j\geq 0}$ *with* $0 < h_j \leq H$ *the following estimate holds with some* $0 < \gamma < 1$ *for* $n \geq 1$ *as long as* $t_{n+k-1} \leq T$

$$\bigg\| \prod_{i=\ell}^{n} \mathscr{S}_i X \bigg\| \leq C\|X_1\| + C\gamma^{n-\ell+1}\|X_2\|, \quad 1 \leq \ell \leq n.$$

In particular, the constant C *does not depend on* n, h_j *and* ε.

PROOF. Our proof is substantially based on the results and techniques from [9]. Owing to (3.4), it is sufficient to treat each component separately. For the first one, the bound $\|R_n R_{n-1} \cdots R_\ell\| \le C$ is a simple special case of [9, Theorem 3]. In order to estimate the second component, we compare $S_n S_{n-1} \cdots S_\ell$ with the frozen product $S_n^* S_{n-1}^* \cdots S_\ell^*$ where

$$S_i^* = r_i\Big(\frac{h_{i+k-2}}{\varepsilon} A_{22}^*\Big)$$

with fixed $A_{22}^* = A_{22}(t^*)$ for some $0 \le t^* \le T$. A telescopic identity for the difference $\Delta S_n = S_n S_{n-1} \cdots S_\ell - S_n^* S_{n-1}^* \cdots S_\ell^*$ yields

$$\Delta S_n = \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} S_i^* \,(S_j - S_j^*)\, \Delta S_{j-1} + \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} S_i^* \,(S_j - S_j^*) \prod_{i=\ell}^{j-1} S_i^*. \tag{3.5}$$

We recall that all eigenvalues of A_{22}^* are strictly negative by hypothesis HP2. With the help of Cauchy's integral formula, we thus receive the representation

$$\prod_{i=\ell}^{n} S_i^* = \frac{1}{2\pi\iota} \int_\Gamma \prod_{i=\ell}^{n} r_i\Big(\frac{h_{i+k-2}\lambda}{\varepsilon}\Big)(\lambda I - A_{22}^*)^{-1}\, \mathrm{d}\lambda \tag{3.6}$$

with a finite path $\Gamma \subset \mathbb{C}_{<0}$ contained in the negative complex plane that encircles the eigenvalues of A_{22}^*. By [9, Lemma 2], for some $0 < \gamma < 1$, it holds

$$\left\| \prod_{i=\ell}^{n} r_i\Big(\frac{h_{i+k-2}\lambda}{\varepsilon}\Big) \right\| \le C\gamma^{n-\ell+1}.$$

Using that $(\lambda I - A_{22}^*)^{-1}$ is bounded, we therefore obtain from (3.6)

$$\left\| \prod_{i=\ell}^{n} S_i^* \right\| \le C\gamma^{n-\ell+1}. \tag{3.7}$$

In addition, a comparison of S_i with S_i^* shows the boundedness of S_i. Consequently, by estimating (3.5), after slightly increasing $\gamma < 1$, we have

$$\|\Delta S_n\| \le C \sum_{j=\ell}^{n} \gamma^{n-j+1} \|\Delta S_{j-1}\| + C\gamma^{n-\ell+1}.$$

Now, a Gronwall inequality yields the bound $\|\Delta S_n\| \le C\gamma^{n-\ell+1}$, and, finally, another application of (3.7) gives the desired result. □

We next summarize some useful relations for the quantities J_j, $R_n R_{n-1} \cdots R_\ell$, $K_{\varepsilon,j}$, and $S_n S_{n-1} \cdots S_\ell$, see (3.4) and below. Note that the specified estimates for $R_n R_{n-1} \cdots R_\ell$, and $S_n S_{n-1} \cdots S_\ell$ are a direct consequence of Lemma 3.1.

Further, the boundedness of J_j follows in an easy way from Cauchy's integral formula. As $\frac{h_{j+k-2}}{\varepsilon}(1+\frac{h_{j+k-2}}{\varepsilon})^{-1} \leq 1$, an estimation of

$$K_{\varepsilon,j} = \frac{1}{2\pi\iota}\int_\Gamma \frac{h_{j+k-2}}{\varepsilon}\left(\alpha_{j-1,k} - \beta_{j-1,k}\frac{h_{j+k-2}\lambda}{\varepsilon}\right)^{-1}\big(\lambda I - A_{22}(t_{j+k-1})\big)^{-1}\mathrm{d}\lambda$$

shows that $K_{\varepsilon,j}$ is bounded.

REMARK 3.1. In the situation of Lemma 3.1, for any $1 \leq \ell, j \leq n$, the bounds

$$\|J_j\| \leq C, \quad \left\|\prod_{i=\ell}^{n} R_i\right\| \leq C, \quad \|K_{\varepsilon,j}\| \leq C, \quad \text{and} \quad \left\|\prod_{i=\ell}^{n} S_i\right\| \leq C\gamma^{n-\ell+1}$$

are valid with constants $C > 0$ and $0 < \gamma < 1$.

We close this section with a remark on BDF-methods where $\beta_{ji} = 0$ for all $0 \leq i \leq k-1$ and $j \geq 0$ and thus $\sigma = 0$. In particular, the condition $\sigma\Omega^2 < 1$ of Lemma 3.1 is fulfilled for any $\Omega > 1$. Here, a further investigation of $S_nS_{n-1}\cdots S_1$ shows that the sharper estimate

$$\left\|\prod_{i=1}^{n} S_i\right\| \leq C\varepsilon\frac{\gamma^n}{h_{k-1}} \tag{3.8}$$

is valid for $n \geq k$.

3.2 The partly coupled problem.

We are now ready to estimate the linear multistep approximation of the partly coupled equation (3.2). Following the lines of the previous section, we employ an alternative representation of $U_n = \mathscr{T}_n\mathscr{T}_{n-1}\cdots\mathscr{T}_1U_0$ by determining successively Y_n and Z_n. For the y-component, it clearly holds $Y_n = R_nR_{n-1}\cdots R_1Y_0$. Thus, by applying the discrete variation-of-constants formula to the z-component and inserting the above representation for Y_j, Z_n writes as

$$Z_n = \prod_{i=1}^{n} S_iZ_0 + \sum_{j=1}^{n}\prod_{i=j+1}^{n} S_ie_k \otimes K_{\varepsilon,j}\big(B_j + \widetilde{B}_jR_j\big)\prod_{i=1}^{j-1} R_iY_0$$

with $B_j = (\beta_{j-1,0}A_{21}(t_{j-1}), \frac{1}{2}\beta_{j-1,1}A_{21}(t_j), \ldots, \frac{1}{2}\beta_{j-1,k-1}A_{21}(t_{j+k-2}))$ and also $\widetilde{B}_j = (\frac{1}{2}\beta_{j-1,1}\,A_{21}(t_j), \ldots, \frac{1}{2}\beta_{j-1,k-1}A_{21}(t_{j+k-2}), \beta_{j-1,k}A_{21}(t_{j+k-1}))$ denoting a bounded matrix of dimension $m_2 \times k\cdot m_1$. In particular, for $k = 1$, let $B_j = \beta_{j-1,0}A_{21}(t_{j-1})$ and $\widetilde{B}_j = \beta_{j-1,1}A_{21}(t_j)$. Henceforth, we identify the transfer operator $\mathscr{T}_n\mathscr{T}_{n-1}\cdots\mathscr{T}_\ell$ with

$$\prod_{i=\ell}^{n}\mathscr{T}_i = \begin{pmatrix} \prod_{i=\ell}^{n} R_i & 0 \\ P_{n\ell} & \prod_{i=\ell}^{n} S_i \end{pmatrix},$$

where the quantity $P_{n\ell}$ is defined through

$$P_{n\ell} = \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} S_i \, e_k \otimes K_{\varepsilon,j} \left(B_j + \widetilde{B}_j R_j \right) \prod_{i=\ell}^{j-1} R_i.$$

With the help of Remark 3.1, it is easy to see that $P_{n\ell}$ is bounded by a constant, and, consequently, we obtain

$$\begin{aligned} \left\| \prod_{i=\ell}^{n} \mathscr{T}_i X \right\| &\leq \left(\left\| \prod_{i=\ell}^{n} R_i \right\| + \|P_{n\ell}\| \right) \|X_1\| + \left\| \prod_{i=\ell}^{n} S_i \right\| \|X_2\| \\ &\leq C\|X_1\| + C\gamma^{n-\ell+1}\|X_2\|. \end{aligned}$$

This proves the following result.

LEMMA 3.2. *In the situation of Lemma* 3.1*, the linear multistep discretization* (2.5) *of Equation* (3.2) *fulfills the estimate*

$$\left\| \prod_{i=\ell}^{n} \mathscr{T}_i X \right\| \leq C\|X_1\| + C\gamma^{n-\ell+1}\|X_2\|, \quad 1 \leq \ell \leq n,$$

with constant C independent of n, h_j and ε.

We note for later use that for BDF-methods the refined estimate

$$\left\| \prod_{i=1}^{n} \mathscr{T}_i X \right\| \leq C\|X_1\| + C\varepsilon \frac{\gamma^n}{h_{k-1}} \|X_2\|, \quad n \geq k, \tag{3.9}$$

follows from (3.8).

3.3 The coupled problem.

In the following, we derive stability estimates for the linear multistep approximation $U_n = \mathscr{R}_n \mathscr{R}_{n-1} \cdots \mathscr{R}_1 U_0$ of the original coupled equation (3.1). As in the preceding sections, we henceforth identify $U_n = (u_n, u_{n+1}, \ldots, u_{n+k-1})^{\mathrm{T}}$ comprising the solution values $u_n = (y_n, z_n)^{\mathrm{T}}$ with the reordered vector $U_n = (Y_n, Z_n)^{\mathrm{T}}$ where $Y_n = (y_n, y_{n+1}, \ldots, y_{n+k-1})^{\mathrm{T}}$ and $Z_n = (z_n, z_{n+1}, \ldots, z_{n+k-1})^{\mathrm{T}}$. Accordingly to that new order, we interprete elements $X = (X_1, X_2)^{\mathrm{T}} \in \mathbb{R}^{k \cdot m_1} \times \mathbb{R}^{k \cdot m_2}$. Further, let $x = (x_1, x_2)^{\mathrm{T}} \in \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$.

In order to prove stability results for the transfer operator $\mathscr{R}_n \mathscr{R}_{n-1} \cdots \mathscr{R}_\ell$ of (3.1), we make use of the fact that bounds for $\mathscr{T}_n \mathscr{T}_{n-1} \cdots \mathscr{T}_\ell$ are provided by Lemma 3.2. Thus, it suffices to study $\Delta\mathscr{R}_n = \mathscr{R}_n \mathscr{R}_{n-1} \cdots \mathscr{R}_\ell - \mathscr{T}_n \mathscr{T}_{n-1} \cdots \mathscr{T}_\ell$. For estimating this difference, our basic tool is the telescopic identity

$$\Delta\mathscr{R}_n = \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} \mathscr{T}_i (\mathscr{R}_j - \mathscr{T}_j) \, \Delta\mathscr{R}_{j-1} + \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} \mathscr{T}_i \, (\mathscr{R}_j - \mathscr{T}_j) \prod_{i=\ell}^{j-1} \mathscr{T}_i \tag{3.10}$$

combined with a Gronwall inequality. Therewith, we are able to establish the following stability bounds.

THEOREM 3.3. *In the situation of Lemma* 3.1*, the linear multistep discretization* (2.5) *of Equation* (3.1) *satisfies the relations*

$$\begin{aligned}\left\|\prod_{i=\ell}^{n}\mathscr{R}_i X\right\| &\le C\|X_1\| + C\big(h_{\ell+k-2}+\gamma^{n-\ell+1}\big)\|X_2\|,\\ h_{j+k-2}\left\|\prod_{i=j+1}^{n}\mathscr{R}_i e_k\otimes \mathscr{J}_j\big(I_\varepsilon^{-1}x\big)\right\| &\le Ch_{j+k-2}\|x_1\| + C\big(h_{j+k-2}+\gamma^{n-j}\big)\|x_2\|,\end{aligned}$$

for $1\le \ell, j\le n$ *with constants* C *independent of* n, h_j *and* ε.

PROOF. In order to estimate (3.10), we first indicate the derivation of a needful relation for the difference $\mathscr{R}_j - \mathscr{T}_j$ of the companion matrices associated with the fully and partly coupled equations (3.1) and (3.2). By definition, it holds

$$\mathscr{R}_j - \mathscr{T}_j = \begin{pmatrix} 0 & 0 & \dots & 0\\ \Delta c_{j0} & \Delta c_{j1} & \dots & \Delta c_{j,k-1}\end{pmatrix}$$

where the entries Δc_{ji} are given by

$$\begin{aligned}\Delta c_{ji} &= c_{ji}(h_{j+k-2}\mathscr{A}_{j+k-1}, h_{j+k-2}\mathscr{A}_{j+i-1}) -\\ &\quad - c_{ji}(h_{j+k-2}\mathscr{L}_{j+k-1}, h_{j+k-2}\mathscr{L}_{j+i-1}),\quad 0\le i\le k-1,\end{aligned}$$

see beginning of Section 2.2. With the help of the quantity

$$\mathscr{K}_j = s_j(h_{j+k-2}\mathscr{L}_{j+k-1}) = \begin{pmatrix} J_j & 0\\ \beta_{j-1,k}K_{\varepsilon,j}A_{21}(t_{j+k-1})J_j & K_j\end{pmatrix}$$

which is bounded according to Remark 3.1, Δc_{ji} also writes as

$$\begin{aligned}\Delta c_{ji} &= -\alpha_{j-1,i}\big(\mathscr{J}_j-\mathscr{K}_j\big) + h_{j+k-2}\beta_{j-1,i}\big(\mathscr{J}_j-\mathscr{K}_j\big)\mathscr{A}_{j+i-1} +\\ &\quad + h_{j+k-2}\beta_{j-1,i}\mathscr{K}_j\big(\mathscr{A}_{j+i-1}-\mathscr{L}_{j+i-1}\big).\end{aligned}$$

A straightforward calculation shows the identity $\mathscr{J}_j = (I+h_{j+k-2}\mathscr{B}_j)\mathscr{K}_j$ where $D_j = (I - h_{j+k-2}\beta_{j-1,k}^2 K_{\varepsilon,j}A_{21}(t_{j+k-1})J_jA_{12}(t_{j+k-1}))^{-1}$ and thus

$$\mathscr{B}_j = \begin{pmatrix} 0 & B_{j1}\\ 0 & B_{j2}\end{pmatrix} = \begin{pmatrix} 0 & \beta_{j-1,k}J_jA_{12}(t_{j+k-1})D_j\\ 0 & \beta_{j-1,k}^2K_{\varepsilon,j}A_{21}(t_{j+k-1})J_jA_{12}(t_{j+k-1})D_j\end{pmatrix}$$

is bounded for $h_{j+k-2}\le H$ sufficiently small, see again Remark 3.1. Thus, we obtain the relation $\Delta c_{ji} = h_{j+k-2}(\mathscr{B}_{ji}\mathscr{K}_j + \widetilde{\mathscr{B}}_{ji})$ with $\mathscr{B}_{ji} = -\alpha_{j-1,i}\mathscr{B}_j$ and $\widetilde{\mathscr{B}}_{ji} = \beta_{j-1,i}(h_{j+k-2}\mathscr{B}_j\mathscr{K}_j\mathscr{A}_{j+i-1} + \mathscr{K}_j(\mathscr{A}_{j+i-1}-\mathscr{L}_{j+i-1}))$ bounded. Finally, this yields the identity

$$\mathscr{R}_j - \mathscr{T}_j = h_{j+k-2}\mathscr{C}_j,\quad \text{where } \mathscr{C}_j = \begin{pmatrix} 0 & 0 & \dots & 0\\ \mathscr{C}_{j0} & \mathscr{C}_{j1} & \dots & \mathscr{C}_{j,k-1}\end{pmatrix} \tag{3.11}$$

comprises the bounded entries $\mathscr{C}_{ji} = \mathscr{B}_{ji}\mathscr{K}_j + \widetilde{\mathscr{B}}_{ji}$.

Now, a thorough investigation of the second term in (3.10) and a further application of Remark 3.1 shows that

$$\left\| \sum_{j=\ell}^{n} \prod_{i=j+1}^{n} \mathscr{T}_i \, (\mathscr{R}_j - \mathscr{T}_j) \prod_{i=\ell}^{j-1} \mathscr{T}_i X \right\| \leq C\|X_1\| + C\big(h_{\ell+k-2} + \gamma^{n-\ell+1}\big)\|X_2\|,$$

and, altogether, we obtain

$$\begin{aligned} \|\Delta\mathscr{R}_n X\| \;\leq\; & C \sum_{j=\ell}^{n} h_{j+k-2}(1+\gamma^{n-j})\|\Delta\mathscr{R}_{j-1}X\| + \\ & + C\|X_1\| + C\big(h_{\ell+k-2} + \gamma^{n-\ell+1}\big)\|X_2\|. \end{aligned} \tag{3.12}$$

For estimating $\Delta\mathscr{R}_n X$, we next split $X = (X_1, 0)^{\mathrm{T}} + (0, X_2)^{\mathrm{T}}$ and replace (3.12) with the following two inequalities. On the one hand, it holds

$$\big\|\Delta\mathscr{R}_n(X_1, 0)^{\mathrm{T}}\big\| \leq C \sum_{j=\ell}^{n} h_{j+k-2} \big\|\Delta\mathscr{R}_{j-1}(X_1, 0)^{\mathrm{T}}\big\| + C\|X_1\| \tag{3.13a}$$

and, on the other hand, we have

$$\begin{aligned} \big\|\Delta\mathscr{R}_n(0, X_2)^{\mathrm{T}}\big\| \;\leq\; & C \sum_{j=\ell}^{n} h_{j+k-2}(1+\gamma^{n-j}) \big\|\Delta\mathscr{R}_{j-1}(0, X_2)^{\mathrm{T}}\big\| + \\ & + C\big(h_{\ell+k-2} + \gamma^{n-\ell+1}\big)\|X_2\|. \end{aligned} \tag{3.13b}$$

Now, at each time, the desired bound results from a discrete Gronwall-type inequality, see [1, 2], e.g. In fact, for (3.13a), the estimate

$$\big\|\Delta\mathscr{R}_n(X_1, 0)^{\mathrm{T}}\big\| \leq C\|X_1\| \tag{3.14a}$$

follows at once from a standard Gronwall inequality. In view of (3.13b), we consider a sequence $(\xi_j)_{j\geq \ell-1}$ of positive numbers satisfying a relation of the following form involving constants $a, b > 0$

$$\xi_n = a \sum_{j=\ell}^{n} h_{j+k-2}(1+\gamma^{n-j})\xi_{j-1} + b\big(h_{\ell+k-2} + \gamma^{n-\ell+1}\big).$$

Due to the fact that this identity is reducible to the recursion

$$\xi_{n+1} = (2ah_{n+k-1} + \gamma)\xi_n + a(1-\gamma) \sum_{j=\ell-1}^{n-1} h_{j+k-1}\xi_j + b(1-\gamma)h_{\ell+k-2},$$

we further obtain $\xi_n \leq Cb(h_{\ell+k-2} + \gamma^{n-\ell+1})$ which proves

$$\big\|\Delta\mathscr{R}_n(0, X_2)^{\mathrm{T}}\big\| \leq C\big(h_{\ell+k-2} + \gamma^{n-\ell+1}\big)\|X_2\|. \tag{3.14b}$$

Therefore, by combining (3.14) and Lemma 3.2, the first bound of Theorem 3.3 follows.

It remains to derive the second bound of the theorem. An easy calculation shows the identity

$$\begin{aligned} &h_{j+k-2}\,\mathscr{J}_j\big(I_\varepsilon^{-1}x\big) \\ &\quad= \begin{pmatrix} h_{j+k-2}J_jx_1+ \\ h_{j+k-2}\beta_{j-1,k}K_{\varepsilon,j}A_{21}(t_{j+k-1})J_jx_1+K_{\varepsilon,j}x_2 \end{pmatrix} + \\ &\qquad+h_{j+k-2}\begin{pmatrix} h_{j+k-2}B_{j1}\beta_{j-1,k}K_{\varepsilon,j}A_{21}(t_{j+k-1})J_jx_1+B_{j1}K_{\varepsilon,j}x_2 \\ h_{j+k-2}B_{j2}\beta_{j-1,k}K_{\varepsilon,j}A_{21}(t_{j+k-1})J_jx_1+B_{j2}K_{\varepsilon,j}x_2 \end{pmatrix}. \end{aligned}$$

Thus, the first bound applied with $X = h_{j+k-2}\,\mathscr{J}_j(I_\varepsilon^{-1}x)$ and $\ell = j+1$ proves the desired result. □

In view of our convergence estimate for BDF-methods, the first relation of Theorem 3.3 with $\ell = 1$ is replaced by

$$\left\|\prod_{i=1}^{n}\mathscr{R}_iX\right\| \le C\|X_1\| + C\varepsilon\Big(1+\frac{\gamma^n}{h_{k-1}}\Big)\|X_2\|, \quad n\ge k. \tag{3.15}$$

This sharper bound is obtained by modifying slightly the proof of Theorem 3.3. In the present situation, formula (3.11) holds for $\mathscr{C}_{ji} = \mathscr{B}_{ji}\mathscr{K}_j$. As a consequence, $\mathscr{C}_{ji}$ is of the form

$$\mathscr{C}_{ji} = \begin{pmatrix} C_{11} & C_{12}K_j \\ C_{21} & C_{22}K_j \end{pmatrix} = \begin{pmatrix} C_{11} & \frac{\varepsilon}{h_{j+k-2}}\,C_{12}K_{\varepsilon,j} \\ C_{21} & \frac{\varepsilon}{h_{j+k-2}}\,C_{22}K_{\varepsilon,j} \end{pmatrix}$$

with bounded matrices C_{11}, C_{12}, C_{21}, and C_{22}. Following the above proof of Theorem 3.3 and tracing the z-component together with (3.9) then yields the refined stability bound for BDF-methods.

4 Convergence result.

In this section, we state our convergence estimate for variable stepsize linear multistep methods applied to singular perturbation problems of the form (2.4). For some function ϕ denote

$$\|\phi\|_{1,[\tau_1,\tau_2]} = \int_{\tau_1}^{\tau_2}\|\phi(\tau)\|\,\mathrm{d}\tau \quad\text{and}\quad \|\phi\|_{\infty,[\tau_1,\tau_2]} = \max_{\tau_1\le\tau\le\tau_2}\|\phi(\tau)\|.$$

Then, the following result holds, provided that the $(p+1)$-st order derivatives of the solution $u = (y,z)^{\mathrm{T}}$ of (2.4) remain bounded, precisely, if the bounds

$$\big\|y^{(p+1)}\big\|_{1,[t_{j-k},t_j]} \le C \quad\text{and}\quad \big\|z^{(p+1)}\big\|_{\infty,[t_{j-k},t_j]} \le C \tag{4.1}$$

are valid for every $t_j \le T$ with constants $C > 0$ not depending on the parameter ε for $\varepsilon \in (0,\varepsilon_0]$.

THEOREM 4.1. *Under* HP2 *assume that the solution* $u = (y, z)^{\mathrm{T}}$ *of the initial value problem* (2.4) *is sufficiently often differentiable and that its derivatives fulfill* (4.1). *Assume further that the linear multistep discretization* (2.5) *of* (2.4) *satisfies* HS1–2, HM1–3, *and* $\sigma\Omega^2 < 1$. *Then, there exist* $H > 0$, $d > 0$ *and* $C > 0$ *such that for any stepsize sequence* $(h_j)_{j\geq 0}$ *with* $0 < h_j \leq H$ *and for initial values satisfying* $\|y_i - y(t_i)\| + \|z_i - z(t_i)\| \leq d$ *for each* $0 \leq i \leq k-1$ *the estimate*

$$\begin{aligned}\|u_n - u(t_n)\| \;\leq\; & C \max_{0\leq i\leq k-1} \|y_i - y(t_i)\| + C\big(h_{k-1} + \gamma^n\big) \max_{0\leq i\leq k-1} \|z_i - z(t_i)\| + \\ & + C \sum_{j=k}^{n} h_{j-1}^p \big\|y^{(p+1)}\big\|_{1,[t_{j-k},t_j]} + \\ & + \varepsilon C \sum_{j=k}^{n} \big(h_{j-1} + \gamma^{n-j}\big) h_{j-1}^p \big\|z^{(p+1)}\big\|_{\infty,[t_{j-k},t_j]}\end{aligned}$$

is valid with some $0 < \gamma < 1$ *for all* $n \geq k$ *as long as* $t_n \leq T$. *Especially, the constant* C *does not depend on* n, $(h_j)_{j\geq 0}$ *and* ε.

REMARK 4.1. Note that for BDF-methods, due to $\sigma = 0$, the requirement $\sigma\Omega^2 < 1$ is satisfied for any $\Omega > 1$. Here, a refined convergence estimate is valid, namely, the factor $h_{k-1} + \gamma^n$ multiplying the z-component of the errors in the starting values, is replaced with $\varepsilon(1 + \frac{\gamma^n}{h_{k-1}})$. This result follows at once from the proof of Theorem 4.1 by estimating (4.3) with the help of relation (3.15).

REMARK 4.2. In particular, for constant or bounded stepsizes $h_j \leq h$, $j \geq 0$, we receive the convergence estimate

$$\begin{aligned}\|u_n - u(t_n)\| \;\leq\; & C \max_{0\leq i\leq k-1} \|y_i - y(t_i)\| + C\big(h + \gamma^n\big) \max_{0\leq i\leq k-1} \|z_i - z(t_i)\| + \\ & + Ch^p \big\|y^{(p+1)}\big\|_{1,[0,t_n]} + \varepsilon\, Ch^p \big\|z^{(p+1)}\big\|_{\infty,[0,t_n]}, \quad t_n \leq T,\end{aligned}$$

which is in accordance with the bound from [6, Theorem 3] for a constant stepsize linear multistep method.

PROOF OF THEOREM 4.1. For constructing the linear multistep solution (2.5) of (2.4), we carry out a fixed-point iteration based on the discrete variation-of-constants formula (2.6). Thereto, we first introduce some useful notation.

For $N \in \mathbb{N}$ such that $t_{N+k-1} \leq T$, consider a sequence $\mathbf{V} = (V_n)_{n=0}^N$ comprising the vectors $V_n = (v_n, v_{n+1}, \ldots, v_{n+k-1})^{\mathrm{T}} \in \mathbb{R}^{k\cdot m}$ with entries

$$v_j = \big(v_j^{(1)}, v_j^{(2)}\big)^{\mathrm{T}} \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}, \quad j \geq 0.$$

In particular, we denote by $\mathbf{U} = (U_n)_{n=0}^N$ and $\widehat{\mathbf{U}} = (\widehat{U}_n)_{n=0}^N$ the sequences that comprise the numerical and exact solution values. We recall the notation $u_n = (y_n, z_n)^{\mathrm{T}}$ for the components of the numerical solution. As in Section 3, we henceforth identify the associated vector $U_n = (u_n, u_{n+1}, \ldots, u_{n+k-1})^{\mathrm{T}}$ with

$U_n = (Y_n, Z_n)^{\mathrm{T}} = (y_n, \dots, y_{n+k-1}, z_n, \dots, z_{n+k-1})^{\mathrm{T}}$. Likewise, we employ the notation $\hat{u}_n = (\hat{y}_n, \hat{z}_n)^{\mathrm{T}}$ for the values of the exact solution and define $\widehat{U}_n = (\hat{u}_n, \hat{u}_{n+1}, \dots, \hat{u}_{n+k-1})^{\mathrm{T}} = (\widehat{Y}_n, \widehat{Z}_n)^{\mathrm{T}}$. In accordance with the preceding sections, we further set

$$\|v_n\| = \big\|v_n^{(1)}\big\| + \big\|v_n^{(2)}\big\| \quad \text{for } v_n = \big(v_n^{(1)}, v_n^{(2)}\big)^{\mathrm{T}} \in \mathbb{R}^m = \mathbb{R}^{m_1} \times \mathbb{R}^{m_2}$$

and define the vector norm through

$$\|V_n\| = \max_{0 \le i \le k-1} \|v_{n+i}\|.$$

In order to guarantee the contraction property of the fixed-point iteration for a reasonable time, we introduce additional weights in the sequence norm and set

$$\|\mathbf{V}\|_{\infty,\mu} = \max_{0 \le n \le N} \|V_n\|_\mu, \quad \text{where } \|V_n\|_\mu = \mathrm{e}^{-\mu t_n} \|V_n\|$$

for some exponent $\mu > 0$ sufficiently large.

With the help of these abbreviations, we are in the position to define the fixed point iteration Φ on a ball around $\widehat{\mathbf{U}}$ by means of formula (2.6)

$$\begin{aligned} \Phi\colon \mathscr{V} &= \big\{\mathbf{V} = (V_n)_{n=0}^N : \|\mathbf{V} - \widehat{\mathbf{U}}\|_{\infty,\mu} \le \varrho\big\} \longrightarrow \mathscr{V} : \mathbf{V} \longmapsto \Phi(\mathbf{V}), \\ \Phi(\mathbf{V})_n &= \prod_{i=1}^n \mathscr{R}_i U_0 + \sum_{j=1}^n h_{j+k-2} \prod_{i=j+1}^n \mathscr{R}_i \mathscr{J}_j \mathscr{G}(V_j). \end{aligned}$$

It remains to verify that the function Φ defining the iteration is a contraction on $\mathscr{V}$. On the one hand, we have for sequences $\mathbf{V}, \mathbf{W} \in \mathscr{V}$

$$\begin{aligned} \big(\Phi(\mathbf{V}) - \Phi(\mathbf{W})\big)_n &= \sum_{j=1}^n h_{j+k-2} \prod_{i=j+1}^n \mathscr{R}_i \, e_k \otimes \mathscr{J}_j\big(I_\varepsilon^{-1} x\big), \quad \text{where} \\ x &= \sum_{i=0}^k \beta_{j-1,i}\big(G(t_{j+i-1}, v_{j+i-1}) - G(t_{j+i-1}, w_{j+i-1})\big). \end{aligned}$$

An application of Lemma 2.1 shows

$$\big\|G_\ell(t_{j+i-1}, v_{j+i-1}) - G_\ell(t_{j+i-1}, w_{j+i-1})\big\| \le C\mathrm{e}^{\mu t_j} \varrho \, \|\mathbf{V} - \mathbf{W}\|_{\infty,\mu}, \quad \ell = 1, 2.$$

Together with the second estimate from Theorem 3.3, this yields

$$\big\|\big(\Phi(\mathbf{V}) - \Phi(\mathbf{W})\big)_n\big\|_\mu \le C\varrho \sum_{j=1}^n \mathrm{e}^{-\mu(t_n - t_j)}\big(h_{j+k-2} + \gamma^{n-j}\big)\|\mathbf{V} - \mathbf{W}\|_{\infty,\mu},$$

and, furthermore,

$$\|\Phi(\mathbf{V}) - \Phi(\mathbf{W})\|_{\infty,\mu} \le \kappa \|\mathbf{V} - \mathbf{W}\|_{\infty,\mu},$$

that is, Φ is contractive with contraction factor

$$\kappa = C\varrho \max_{0\le n\le N} \sum_{j=1}^{n} \mathrm{e}^{-\mu(t_n - t_j)} \big(h_{j+k-2} + \gamma^{n-j}\big) < 1$$

for ϱ sufficiently small. We note that the size of $C > 0$ is moderate for μ large, whereas for $\mu = 0$ the above relation becomes

$$\kappa = C\varrho\left(T + \frac{1}{1-\gamma}\right) < 1.$$

As a consequence, this condition considerably restricts the size of T.

We next prove that Φ maps $\mathscr{V}$ to $\mathscr{V}$, that is,

$$\big\|\Phi(\mathbf{V}) - \widehat{\mathbf{U}}\big\|_{\infty,\mu} \le \varrho \quad \text{whenever} \quad \|\mathbf{V} - \widehat{\mathbf{U}}\|_{\infty,\mu} \le \varrho.$$

By means of the contraction property of Φ on $\mathscr{V}$, we obtain

$$\begin{aligned} \big\|\Phi(\mathbf{V}) - \widehat{\mathbf{U}}\big\|_{\infty,\mu} &\le \big\|\Phi(\mathbf{V}) - \Phi(\widehat{\mathbf{U}})\big\|_{\infty,\mu} + \big\|\Phi(\widehat{\mathbf{U}}) - \widehat{\mathbf{U}}\big\|_{\infty,\mu} \\ &\le \kappa\varrho + \big\|\Phi(\widehat{\mathbf{U}}) - \widehat{\mathbf{U}}\big\|_{\infty,\mu}. \end{aligned}$$

Thus, it suffices to show that the quantity

$$\begin{aligned} \big(\Phi(\widehat{\mathbf{U}}) - \widehat{\mathbf{U}}\big)_n &= \prod_{i=1}^{n} \mathscr{R}_i\big(U_0 - \widehat{U}_0\big) - \sum_{j=1}^{n} h_{j+k-2} \prod_{i=j+1}^{n} \mathscr{R}_i\, e_k \otimes \mathscr{J}_j I_\varepsilon^{-1} x \\ &\quad \text{with } x = \sum_{i=0}^{k} \beta_{j-1,i} I_\varepsilon \delta_j \end{aligned}$$

is small enough, see (2.7). With the help of Theorem 3.3 and the bound (2.8) for the defects, it follows

$$\begin{aligned} (4.2)\ \big\|\Phi(\widehat{\mathbf{U}})_n - \widehat{\mathbf{U}}_n\big\|_\mu \le\ & C\mathrm{e}^{-\mu t_n}\Bigg(\big\|Y_0 - \widehat{Y}_0\big\| + \big(h_{k-1} + \gamma^n\big)\big\|Z_0 - \widehat{Z}_0\big\| + \\ & + \sum_{j=1}^{n} h_{j+k-2}^{p} \big\|y^{(p+1)}\big\|_{1,[t_{j-1},t_{j+k-1}]} + \\ & + \varepsilon \sum_{j=1}^{n} h_{j+k-2}^{p} \big(h_{j+k-2} + \gamma^{n-j}\big) \times \\ & \times \big\|z^{(p+1)}\big\|_{\infty,[t_{j-1},t_{j+k-1}]}\Bigg). \end{aligned}$$

Taking the maximum over $0 \le n \le N$ finally gives

$$\big\|\Phi(\widehat{\mathbf{U}}) - \widehat{\mathbf{U}}\big\|_{\infty,\mu} \le (1-\kappa)\varrho$$

if the errors of the initial values and $h_j \le H$ are sufficiently small. Altogether, an application of Banach's Fixed Point Theorem yields the existence of the numerical solution $\mathbf{U}$ as unique fixed point of Φ.

In order to estimate $E_n = \|U_n - \widehat{U}_n\|$, we employ the following representation obtained by formulas (2.6) and (2.7)

$$U_n - \widehat{U}_n = \prod_{i=1}^{n} \mathscr{R}_i\big(U_0 - \widehat{U}_0\big) + \sum_{j=1}^{n} h_{j+k-2} \prod_{i=j+1}^{n} \mathscr{R}_i \mathscr{J}_j\big(\mathscr{G}(U_j) - \mathscr{G}(\widehat{U}_j) - \Delta_j\big). \tag{4.3}$$

Now, the above considerations together with the bounds from Theorem 3.3, Lemma 2.1, and (2.8) show that the error satisfies

$$\begin{aligned} E_n \le\ & C\|Y_0 - \widehat{Y}_0\| + C\big(h_{k-1} + \gamma^n\big)\|Z_0 - \widehat{Z}_0\| + \\ & + C\sum_{j=1}^{n-1} \big(h_{j+k-2} + \gamma^{n-j}\big)E_j + C\sum_{j=1}^{n} h_{j+k-2}^{p} \big\|y^{(p+1)}\big\|_{1,[t_{j-1},t_{j+k-1}]} + \\ & + \varepsilon\, C\sum_{j=1}^{n} h_{j+k-2}^{p}\big(h_{j+k-2} + \gamma^{n-j}\big)\big\|z^{(p+1)}\big\|_{\infty,[t_{j-1},t_{j+k-1}]}. \end{aligned}$$

Hence, the desired convergence estimate for $\|u_n - \hat{u}_n\| \le E_{n-k+1}$ follows from a Gronwall lemma, see proof of Theorem 3.3. □

Acknowledgement.

I thank Gabriela Schranz-Kirlinger for her invitation and many fruitful discussions. Also, I am grateful to Alexander Ostermann for valuable suggestions and comments relating to the present work.

REFERENCES

1. R. Agarwal, *Difference Equations and Inequalities. Theory, Methods and Applications*, Dekker, New York, 1992.
2. H. Brunnner and P. J. van der Houwen, *The numerical solution of Volterra equations*, CWI Monographs 3, North-Holland, Amsterdam, 1986.
3. C. W. Gear and K. W. Tu, *The effect of variable mesh size on the stability of multistep methods*, SIAM J. Numer. Anal. 11 (1974), pp. 1025–1043.
4. E. Hairer, S. P. Nørsett and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*, 2nd rev. edn, Springer, Berlin, 1993.
5. E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*, 2nd rev. edn, Springer, Berlin, 1996.
6. Ch. Lubich, *On the convergence of multistep methods for nonlinear stiff differential equations*, Numer. Math. 58 (1991), pp. 839–853.
7. W. L. Miranker, *Numerical Methods for Stiff Equations and Singular Perturbation Problems*, Mathematics and Its Applications 5, Reidel, Dordrecht, 1981.
8. R. E. O'Malley, Jr., *Singular Perturbation Methods for Ordinary Differential Equations*, Applied Mathematical Sciences 89, Springer, Berlin, 1991.

9. A. Ostermann, M. Thalhammer, and G. Kirlinger, *Stability of linear multistep methods and applications to nonlinear parabolic problems*, Appl. Numer. Math. 48 (2004), pp. 389–407.
10. C. Palencia and B. García-Archilla, *Stability of linear multistep methods for sectorial operators in Banach spaces*, Appl. Numer. Math. 445 (1993), pp. 503–520.
11. L. R. Petzold, *A description of DASSL: A differential/algebraic system solver*, in Scientific Computing, R. S. Stepleman et al., eds., North-Holland, Amsterdam, 1983, pp. 65–68.
12. B. van der Pol, *A theory of the amplitude of free and forced triode vibrations*, Radio Review 1 (1920), pp. 701–710/754–762.
13. B. van der Pol, *On relaxation oscillations I*, Phil. Mag. 2 (1926), pp. 978–992.

2. Explicit Exponential Integrators

2.1. A Magnus integrator for nonautonomous problems

A second-order Magnus integrator for nonautonomous parabolic problems
CÉSAREO GONZÁLEZ, ALEXANDER OSTERMANN, AND MECHTHILD THALHAMMER
Journal of Computational and Applied Mathematics (2006) 189, 142-156

Available online at www.sciencedirect.com
SCIENCE DIRECT
ELSEVIER
Journal of Computational and Applied Mathematics 189 (2006) 142–156
JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS
www.elsevier.com/locate/cam

A second-order Magnus-type integrator for nonautonomous parabolic problems

C. González[a], A. Ostermann[b], M. Thalhammer[b,*]

[a] *Departamento de Matemática Aplicada y Computación, Universidad de Valladolid, E-47011 Valladolid, Spain*
[b] *Institut für Mathematik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria*

Received 21 July 2004; received in revised form 16 March 2005

Abstract

We analyse stability and convergence properties of a second-order Magnus-type integrator for linear parabolic differential equations with time-dependent coefficients, working in an analytic framework of sectorial operators in Banach spaces. Under reasonable smoothness assumptions on the data and the exact solution, we prove a second-order convergence result without unnatural restrictions on the time stepsize. However, if the error is measured in the domain of the differential operator, an order reduction occurs, in general. A numerical example illustrates and confirms our theoretical results.

Keywords: Linear parabolic problems; Time-dependent coefficients; Magnus integrators; Exponential integrators; Stability; Convergence

1. Introduction

In this paper, we are concerned with the numerical solution of nonautonomous linear differential equations

$$u'(t) = A(t)u(t) + b(t), \quad 0 < t \leqslant T, \quad u(0) = u_0. \tag{1}$$

* Corresponding author.
E-mail addresses: cesareo@mac.uva.es (C. González), alexander.ostermann@uibk.ac.at (A. Ostermann), mechthild.thalhammer@uibk.ac.at (M. Thalhammer).

0377-0427/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2005.04.036

In particular, we are interested in analysing the situation where (1) constitutes an abstract parabolic problem on a Banach space. The precise assumptions on the operator family $A(t)$, $0 \leqslant t \leqslant T$, are given in Section 2.

For linear matrix differential equations $y'(t) = A(t)y(t)$ with possibly noncommuting matrices $A(t)$, Magnus [11] has constructed the solution in the form $y(t) = \exp(\Omega(t))y(0)$ with a matrix $\Omega(t)$ depending on iterated integrals of $A(t)$, see also [5, Section IV.7]. Only recently, this Magnus expansion has been exploited numerically by approximating the arising integrals by quadrature methods, see [9,16] within the context of geometric integration and [1] in connection with the time-dependent Schrödinger equation.

As the convergence of the Magnus expansion is only guaranteed if $\|\Omega(t)\| < \pi$, stiff problems with large or even unbounded $\|A(t)\|$ seemed to be excluded. However, in an impressing paper [8], Hochbruck and Lubich give error bounds for Magnus integrators applied to time-dependent Schrödinger equations, solely working with matrix commutator bounds. The aim of the present paper is to derive the corresponding result for a second-order Magnus-type integrator applied to linear parabolic differential equations with time-dependent coefficients, exploiting the temporal regularity of the exact solution. For that purpose, we employ an abstract formulation of the partial differential equation and work within the framework of sectorial operators and analytic semigroups in Banach spaces.

The paper is organised as follows. In Section 2, we state the main assumptions on the problem and its numerical discretisation. Our numerical scheme for (1) is a mixed method that integrates the homogeneous part by a second-order Magnus integrator and the inhomogeneity by the exponential midpoint rule. In Section 3, we first study the stability properties of the Magnus integrator. The given stability bounds form the basis for the convergence results specified in Section 4. Under the main assumption that the data and the exact solution are sufficiently smooth in time, the actual order of convergence depends on the chosen norm in which the error is measured as well as on the boundary values of a certain function, depending itself on the data of the problem. For instance, for a second-order strongly elliptic differential operator with smooth coefficients, we obtain second-order convergence with respect to the L^p-norm for $1 < p < \infty$. However, if the error is measured in the domain of the differential operator, an order reduction down to $1 + 1/(2p)$ is encountered, in general. These theoretical results are illustrated and confirmed by a numerical experiment given in Section 5.

Throughout the paper, $C > 0$ denotes a generic constant.

2. Equation and numerical method

In the sequel, we introduce the basic assumptions on (1) and specify the numerical scheme. For a detailed treatise of time-dependent evolution equations we refer to [10,15]. The monographs [6,14] delve into the theory of sectorial operators and analytic semigroups.

We first consider abstract initial value problems of the form (1) with $b=0$. Our fundamental requirement on the map A defining the right-hand side of the equation is the following.

Hypothesis 1. Let $(X, \|\cdot\|_X)$ and $(D, \|\cdot\|_D)$ be Banach spaces with D densely embedded in X. We suppose that the closed linear operator $A(t)\colon D \to X$ is uniformly sectorial for $0 \leqslant t \leqslant T$. Thus, there exist constants $a \in \mathbb{R}$, $0 < \phi < \pi/2$, and $M_1 \geqslant 1$ such that $A(t)$ satisfies the following resolvent condition

on the complement of the sector $S_\phi(a) = \{\lambda \in \mathbb{C} : |\arg(a-\lambda)| \leqslant \phi\} \cup \{a\}$

$$\|(\lambda I - A(t))^{-1}\|_{X \leftarrow X} \leqslant \frac{M_1}{|\lambda - a|} \quad \text{for any } \lambda \in \mathbb{C} \backslash S_\phi(a). \tag{2}$$

Besides, we assume that the graph norm of $A(t)$ and the norm in D are equivalent, i.e., for every $0 \leqslant t \leqslant T$ and for all $x \in D$ the estimate

$$C_v^{-1}\|x\|_D \leqslant \|x\|_X + \|A(t)x\|_X \leqslant C_v\|x\|_D \tag{3}$$

holds with some constant $C_v \geqslant 1$.

We remark that for any linear operator $F : X \to D$ relation (3) implies

$$\|A(t)F\|_{X \leftarrow X} \leqslant C_v\|F\|_{D \leftarrow X} \quad \text{and} \quad \|F\|_{D \leftarrow X} \leqslant C_v(1 + \|A(t)F\|_{X \leftarrow X}). \tag{4}$$

As a consequence, for fixed $0 \leqslant s \leqslant T$, the sectorial operator $A(s)$ generates an analytic semigroup $(\mathrm{e}^{tA(s)})_{t \geqslant 0}$ which satisfies the bound

$$\|\mathrm{e}^{tA(s)}\|_{X \leftarrow X} + \|\mathrm{e}^{tA(s)}\|_{D \leftarrow D} + \|t\mathrm{e}^{tA(s)}\|_{D \leftarrow X} \leqslant M_2 \quad \text{for } 0 \leqslant t \leqslant T \tag{5}$$

with some constant $M_2 \geqslant 1$, see e.g., [10].

In view of our convergence and stability results it is essential that $A(t)$ is Hölder-continuous with respect to t.

Hypothesis 2. We assume $A \in C^\alpha([0, T], L(D, X))$ for some $0 < \alpha \leqslant 1$, i.e., the following estimate is valid with a constant $M_3 > 0$

$$\|A(t) - A(s)\|_{X \leftarrow D} \leqslant M_3(t-s)^\alpha \tag{6}$$

for all $0 \leqslant s \leqslant t \leqslant T$.

The nonautonomous problem (1) with $b = 0$ is discretised by a Magnus integrator which is of classical order 2. For this, let $t_j = jh$ be the grid points associated with a constant stepsize $h > 0$, $j \geqslant 0$. Then, for some initial value $u_0 \in X$, the numerical approximation u_{n+1} to the true solution at time t_{n+1} is defined recursively by

$$u_{n+1} = \mathrm{e}^{hA_n}u_n, \quad n \geqslant 0 \quad \text{where } A_n = A\left(t_n + \frac{h}{2}\right). \tag{7}$$

This method was studied for time-dependent Schrödinger equations in [8].

We next extend (7) to initial value problems (1) with an additional inhomogeneity $b : [0, T] \to X$. Motivated by the time-invariant case, we approximate the inhomogeneity by the exponential midpoint rule. This yields the recursion

$$u_{n+1} = \mathrm{e}^{hA_n}u_n + h\varphi(hA_n)b_n, \quad n \geqslant 0 \quad \text{with } b_n = b\left(t_n + \frac{h}{2}\right), \tag{8}$$

where the linear operator $\varphi(hA_n)$ is given by

$$\varphi(hA_n) = \frac{1}{h}\int_0^h \mathrm{e}^{(h-\tau)A_n}\,\mathrm{d}\tau. \tag{9}$$

The competitiveness of the numerical scheme (8) relies on an efficient calculation of the exponential and the related function (9). More precisely, the product of a matrix exponential and a vector has to be computed. It has been shown in [2,7] that Krylov methods prove to be excellent for this aim.

We note for later use that the estimates (4) and (5) imply

$$\|\varphi(hA_n)\|_{X\leftarrow X} + \|\varphi(hA_n)\|_{D\leftarrow D} + \|h\varphi(hA_n)\|_{D\leftarrow X} \leqslant M_4 \tag{10}$$

with some constant $M_4 \geqslant 1$.

In the following example we show that linear parabolic problems with time-dependent coefficients enter our abstract framework.

Example 1. Let $\Omega \in \mathbb{R}^d$ be a bounded domain with smooth boundary. We consider the linear parabolic initial-boundary value problem

$$\frac{\partial U}{\partial t}(x,t) = \mathscr{A}(x,t)U(x,t) + f(x,t), \quad x \in \Omega, \quad 0 < t \leqslant T \tag{11a}$$

with homogeneous Dirichlet boundary conditions and initial condition

$$U(x,0) = U_0(x), \quad x \in \Omega. \tag{11b}$$

Here, $\mathscr{A}(x,t)$ is a second-order strongly elliptic differential operator

$$\mathscr{A}(x,t) = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(\alpha_{ij}(x,t)\frac{\partial}{\partial x_j}\right) + \sum_{i=1}^{d} \beta_i(x,t)\,\frac{\partial}{\partial x_i} + \gamma(x,t). \tag{11c}$$

We require that the time-dependent coefficients α_{ij}, β_i, and γ are smooth functions of the variable $x \in \overline{\Omega}$ and Hölder-continuous with respect to t. For $1 < p < \infty$ and $\psi \in C_0^\infty(\Omega)$, we set $(A_p(t)\psi)(x) = \mathscr{A}(x,t)\psi(x)$ and consider $A_p(t)$ as an unbounded operator on $L^p(\Omega)$. It is well-known that this operator satisfies Hypotheses 1 and 2 with

$$X = L^p(\Omega) \quad \text{and} \quad D_p = W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega), \tag{11d}$$

see [14, Section 7.6, 15, Section 5.2].

Our aim is to analyse the convergence behaviour of (8) for parabolic problems (1). Section 3 is concerned with the derivation of the needed stability results.

3. Stability

In order to study the stability properties of the Magnus integrator (8), it suffices to consider the homogeneous equation under discretisation. Resolving recursion (7) yields

$$u_{n+1} = \prod_{i=0}^{n} \mathrm{e}^{hA_i} u_0 \quad \text{for } n \geqslant 0.$$

Here, for noncommutative operators F_i on a Banach space the product is defined by

$$\prod_{i=m}^{n} F_i = \begin{cases} F_n F_{n-1} \cdots F_m & \text{if } n \geqslant m, \\ I & \text{if } n < m. \end{cases}$$

In the sequel, we derive bounds for the discrete evolution operator

$$\prod_{i=m}^{n} \mathrm{e}^{hA_i} \quad \text{for } n > m \geqslant 0 \tag{12}$$

in different norms. In Theorem 1, for notational simplicity, we do not distinguish the appearing constants.

Theorem 1 (*Stability*). *Under Hypotheses* 1–2 *the bounds*

$$\left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} \right\|_{X \leftarrow X} \leqslant M_5 \quad \text{and} \quad \left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} \right\|_{D \leftarrow X} \leqslant M_5 (t_{n+1} - t_m)^{-1} (1 + (1 + |\log h|)(t_{n+1} - t_m)^{\alpha})$$

are valid for $0 \leqslant t_m < t_n \leqslant T$ *with constant* $M_5 \geqslant 1$ *not depending on n and h.*

Proof. For proving the above stability bounds, our techniques are close to that used in [13]. The needed auxiliary estimates are given in Lemma 1 at the end of this section.

The main idea is to compare the discrete evolution operator (12) with the frozen operator

$$\prod_{i=m}^{n} \mathrm{e}^{hA_m} = \mathrm{e}^{(t_{n+1} - t_m) A_m},$$

where (5) applies directly. Therefore, it remains to estimate the difference

$$\Delta_m^n = \prod_{i=m}^{n} \mathrm{e}^{hA_i} - \prod_{i=m}^{n} \mathrm{e}^{hA_m}.$$

From a telescopic identity, it follows

$$\Delta_m^n = \sum_{j=m+1}^{n-1} \Delta_{j+1}^n (\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) \mathrm{e}^{(t_j - t_m) A_m} + \sum_{j=m+1}^{n} \mathrm{e}^{(t_{n+1} - t_{j+1}) A_m} (\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) \mathrm{e}^{(t_j - t_m) A_m}. \tag{13}$$

(i) We first estimate Δ_m^n as operator from X to X. An application of Lemma 1 and relation (5) yields

$$\begin{aligned} \|\Delta_m^n\|_{X \leftarrow X} \leqslant & \sum_{j=m+1}^{n-1} \|\Delta_{j+1}^n\|_{X \leftarrow X} \|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) \mathrm{e}^{(t_j - t_m) A_m}\|_{X \leftarrow X} \\ & + \sum_{j=m+1}^{n} \|\mathrm{e}^{(t_{n+1} - t_{j+1}) A_m}\|_{X \leftarrow X} \|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) \mathrm{e}^{(t_j - t_m) A_m}\|_{X \leftarrow X} \\ \leqslant & Ch \sum_{j=m+1}^{n-1} \|\Delta_{j+1}^n\|_{X \leftarrow X} (t_j - t_m)^{-1+\alpha} + Ch \sum_{j=m+1}^{n} (t_j - t_m)^{-1+\alpha} \end{aligned}$$

with some constant $C > 0$ depending on M_2 and M_6. Interpreting the second sum as a Riemann-sum and bounding it by the corresponding integral shows

$$\|\Delta_m^n\|_{X\leftarrow X} \leqslant Ch \sum_{j=m+1}^{n-1} \|\Delta_{j+1}^n\|_{X\leftarrow X}(t_j - t_m)^{-1+\alpha} + C,$$

where the constant additionally depends on T, see also [13]. A Gronwall-type inequality implies

$$\|\Delta_m^n\|_{X\leftarrow X} \leqslant C, \tag{14}$$

and, with the help of (5), the desired estimate for the discrete evolution operator follows:

$$\left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} \right\|_{X\leftarrow X} \leqslant \|\Delta_m^n\|_{X\leftarrow X} + \|\mathrm{e}^{(t_{n+1}-t_m)A_m}\|_{X\leftarrow X} \leqslant M_5.$$

(ii) For estimating $\|\Delta_m^n\|_{D\leftarrow X}$, we consider (13) and apply once more Lemma 1 and relation (5)

$$\begin{aligned}
\|\Delta_m^n\|_{D\leftarrow X} \leqslant & \sum_{j=m+1}^{n-1} \|\Delta_{j+1}^n\|_{D\leftarrow X} \|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_j-t_m)A_m}\|_{X\leftarrow X} \\
& + \sum_{j=m+1}^{n-1} \|\mathrm{e}^{(t_{n+1}-t_{j+1})A_m}\|_{D\leftarrow X} \|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_j-t_m)A_m}\|_{X\leftarrow X} \\
& + \|(\mathrm{e}^{hA_n} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_n-t_m)A_m}\|_{D\leftarrow X} \\
\leqslant & Ch \sum_{j=m+1}^{n-1} \|\Delta_{j+1}^n\|_{D\leftarrow X}(t_j - t_m)^{-1+\alpha} + Ch \sum_{j=m+1}^{n-1} (t_{n+1} - t_{j+1})^{-1}(t_j - t_m)^{-1+\alpha} \\
& + C(t_n - t_m)^{-1+\alpha}.
\end{aligned}$$

We estimate the Riemann-sum by the corresponding integral and apply a Gronwall inequality, see [12]. This yields

$$\|\Delta_m^n\|_{D\leftarrow X} \leqslant C(1 + |\log h|)(t_{n+1} - t_m)^{-1+\alpha}.$$

Together with (5) we finally obtain the desired result. □

The following auxiliary result is needed in the proof of Theorem 1.

Lemma 1. *In the situation of Theorem* 1, *the estimates*

$$\|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_j-t_m)A_m}\|_{X\leftarrow X} \leqslant M_6 h(t_j - t_m)^{-1+\alpha} \quad \textit{and}$$
$$\|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_j-t_m)A_m}\|_{D\leftarrow X} \leqslant M_6(t_j - t_m)^{-1+\alpha}$$

are valid for $0 \leqslant t_m < t_j \leqslant T$ *with some constant* $M_6 > 0$ *not depending on n and h.*

Proof. For proving Lemma 1, we employ standard techniques, see e.g., [10, Proof of Prop. 2.1.1].

Let Γ be a path surrounding the spectrum of the sectorial operators A_j and A_m. By means of the integral formula of Cauchy, the representation

$$\begin{aligned}(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m})\mathrm{e}^{(t_j - t_m)A_m} &= \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda}((\lambda - hA_j)^{-1} - (\lambda - hA_m)^{-1})\, \mathrm{e}^{(t_j - t_m)A_m}\, \mathrm{d}\lambda \\ &= \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda}(\lambda - hA_j)^{-1} h(A_j - A_m)(\lambda - hA_m)^{-1} \mathrm{e}^{(t_j - t_m)A_m}\, \mathrm{d}\lambda \qquad (15)\end{aligned}$$

follows. The main tools for estimating this relation are the resolvent bound (2), estimate (5) and the Hölder property (6). We omit the details. □

4. Convergence

In the following, we analyse the convergence behaviour of the Magnus integrator (8) for (1). For that purpose, we next derive a representation of the global error.

We consider the initial value problem (1) on a subinterval $[t_n, t_{n+1}]$ and rewrite the right-hand side of the equation as follows:

$$u'(t) = A(t)u(t) + b(t) = A_n u(t) + b_n + g_n(t),$$

where the map g_n is defined by

$$g_n(t) = (A(t) - A_n)u(t) + b(t) - b_n \quad \text{for } t_n \leqslant t \leqslant t_{n+1}. \qquad (16)$$

Consequently, by the variation-of-constants formula, we obtain the following representation of the exact solution:

$$u(t_{n+1}) = \mathrm{e}^{hA_n} u(t_n) + \int_0^h \mathrm{e}^{(h-\tau)A_n}(b_n + g_n(t_n + \tau))\, \mathrm{d}\tau. \qquad (17)$$

On the other hand, the numerical solution is given by relation (8), see also (9). Let $e_{n+1} = u_{n+1} - u(t_{n+1})$ denote the error at time t_{n+1} and δ_{n+1} the corresponding defect

$$\delta_{n+1} = \int_0^h \mathrm{e}^{(h-\tau)A_n} g_n(t_n + \tau)\, \mathrm{d}\tau. \qquad (18)$$

By taking the difference of (8) and (17), we thus obtain

$$e_{n+1} = \mathrm{e}^{hA_n} e_n - \delta_{n+1}, \quad n \geqslant 0, \quad e_0 = 0.$$

Resolving this error recursion finally yields

$$e_n = -\sum_{j=0}^{n-1} \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \delta_{j+1}, \quad n \geqslant 1, \quad e_0 = 0.$$

For the subsequent convergence analysis, it is useful to employ an expansion of the defects which we derive in the following.

Provided that the map g_n is twice differentiable on (t_n, t_{n+1}), we obtain from a Taylor series expansion

$$g_n(t_n+\tau) = \left(\tau - \frac{h}{2}\right) g_n'\left(t_n + \frac{h}{2}\right) + \left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma) g_n''\left(t_n + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \mathrm{d}\sigma,$$

where $0 < \tau < h$. We insert this expansion into (18) and express the terms involving g_n' with the help of the bounded linear operators

$$\varphi(hA_n) = \frac{1}{h}\int_0^h \mathrm{e}^{(h-\tau)A_n}\,\mathrm{d}\tau \quad \text{and} \quad \psi(hA_n) = \frac{1}{h^2}\int_0^h \mathrm{e}^{(h-\tau)A_n}\,\tau\,\mathrm{d}\tau. \tag{19}$$

Thus, we obtain the following representation of the defects

$$\begin{aligned}\delta_{n+1} = {} & h^2\left(\psi(hA_n) - \frac{1}{2}\varphi(hA_n)\right) g_n'\left(t_n + \frac{h}{2}\right) \\ & + \int_0^h \mathrm{e}^{(h-\tau)A_n}\left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma)\, g_n''\left(t_n + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \mathrm{d}\sigma\,\mathrm{d}\tau.\end{aligned}$$

For later it is also substantial that the equality

$$\psi(hA_n) - \tfrac{1}{2}\varphi(hA_n) = hA_n\chi(hA_n)$$

holds with some bounded linear operator $\chi(hA_n)$. Precisely, after possibly enlarging the constant $M_4 \geqslant 1$ in (10), we receive

$$\begin{aligned}&\|\varphi(hA_n)\|_{X\leftarrow X} + \|\varphi(hA_n)\|_{D\leftarrow D} + \|\psi(hA_n)\|_{X\leftarrow X} \\ &\quad + \|\psi(hA_n)\|_{D\leftarrow D} + \|\chi(hA_n)\|_{X\leftarrow X} + \|\chi(hA_n)\|_{D\leftarrow D} \leqslant M_4.\end{aligned} \tag{20}$$

The bounds for $\varphi(hA_n)$ and $\psi(hA_n)$ are a direct consequence of the defining relations (19) and (5), see also (10), whereas the boundedness of $\chi(hA_n)$ follows by means of the integral formula of Cauchy.

We first specify a convergence estimate under the assumption that the true solution of (1) possesses favourable regularity properties. Our main tool for the derivation of this error bound is the stability result stated in Section 3. In view of the proof of our convergence result, it is convenient to introduce several abbreviations. Accordingly to the above considerations, we split the defects $\delta_{j+1} = \delta_{j+1}^{(1)} + \delta_{j+1}^{(2)}$ where

$$\begin{aligned}\delta_{j+1}^{(1)} &= h^2\left(\psi(hA_j) - \frac{1}{2}\varphi(hA_j)\right) g_j'\left(t_j + \frac{h}{2}\right) = h^3 A_j\chi(hA_j) g_j'\left(t_j + \frac{h}{2}\right), \\ \delta_{j+1}^{(2)} &= \int_0^h \mathrm{e}^{(h-\tau)A_j}\left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma)\, g_j''\left(t_j + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \mathrm{d}\sigma\,\mathrm{d}\tau.\end{aligned} \tag{21a}$$

Analogously, the error is decomposed into $e_n = -e_n^{(1)} - e_n^{(2)}$ with

$$e_n^{(k)} = \sum_{j=0}^{n-1} \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i}\,\delta_{j+1}^{(k)}, \quad k = 1, 2. \tag{21b}$$

Henceforth, we denote by $\|g_n\|_{X,\infty} = \max\{\|g_n(t)\|_X : t_n \leqslant t \leqslant t_{n+1}\}$ the maximum value of the map $g_n = (A - A_n)u + b - b_n$ on the interval $[t_n, t_{n+1}]$. Recall the abbreviations $A_n = A(t_n + h/2)$ and $b_n = b(t_n + (h/2))$ introduced in (7) and (8). Further, we set

$$\|g\|_{X,\infty} = \max\{\|g_n\|_{X,\infty} : n \geqslant 0,\ t_{n+1} \leqslant T\}.$$

Theorem 2 (*Convergence*). *Under Hypotheses* 1–2 *with* $\alpha = 1$, *apply the Magnus integrator* (8) *to the initial value problem* (1). *Then, the convergence estimate*

$$\|u_n - u(t_n)\|_X \leqslant Ch^2(\|g'\|_{D,\infty} + \|g''\|_{X,\infty}),$$

is valid for $0 \leqslant t_n \leqslant T$, *provided that the quantities on the right-hand side are well-defined. The constant* $C > 0$ *does not depend on n and h.*

Proof. We successively consider the error terms $e_n^{(1)}$ and $e_n^{(2)}$ specified above. An application of Theorem 1 yields

$$\begin{aligned}
\|e_n^{(1)}\|_X &\leqslant \left\| \sum_{j=0}^{n-2} \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \delta_{j+1}^{(1)} \right\|_X + \|\delta_n^{(1)}\|_X \\
&\leqslant h^2 \cdot h \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \right\|_{X \leftarrow X} \|A_j\|_{X \leftarrow D} \|\chi(hA_j)\|_{D \leftarrow D} \left\| g_j'\left(t_j + \frac{h}{2}\right) \right\|_D \\
&\quad + h^2(\|\varphi(hA_{n-1})\|_{X \leftarrow X} + \|\psi(hA_{n-1})\|_{X \leftarrow X}) \left\| g_{n-1}'\left(t_{n-1} + \frac{h}{2}\right) \right\|_X \\
&\leqslant C\|g'\|_{D,\infty} h^2
\end{aligned}$$

with $C > 0$ depending on the constants M_5 and M_4 appearing in Theorem 1 and (20), on $\|A(t)\|_{X \leftarrow D}$, and on T. A direct estimation of $\delta_{j+1}^{(2)}$ with the help of (5) shows

$$\begin{aligned}
\|\delta_{j+1}^{(2)}\|_X &\leqslant \int_0^h \|\mathrm{e}^{(h-\tau)A_j}\|_{X \leftarrow X} \left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma) \left\| g_j''\left(t_j + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \right\|_X \mathrm{d}\sigma\, \mathrm{d}\tau \\
&\leqslant M_2 \|g''\|_{X,\infty} h^3.
\end{aligned}$$

Consequently, for the remaining term, we obtain by Theorem 1

$$\|e_n^{(2)}\|_X \leqslant \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \right\|_{X \leftarrow X} \|\delta_{j+1}^{(2)}\|_X + \|\delta_n^{(2)}\|_X \leqslant C\|g''\|_{X,\infty} h^2$$

with a constant $C > 0$ depending on M_2, M_5, and T. Altogether, the desired estimate follows. □

We remark that, in the situation of the theorem, Hypothesis 2 is always fulfilled with $\alpha = 1$. However, in view of applications, the condition on the derivative of g_n is often too restrictive. We next prove a convergence result under weaker assumptions on g_n'. For the proof of Theorem 3 an extension of our stability result is needed which we give at the end of this section.

Theorem 3 (*Convergence*). *Under Hypotheses* 1–2 *with* $\alpha = 1$, *the Magnus integrator* (8) *applied to* (1) *satisfies the bound*

$$\|u_n - u(t_n)\|_X \leqslant Ch^2((1 + |\log h|)\|g'\|_{X,\infty} + \|g''\|_{X,\infty})$$

for $0 \leqslant t_n \leqslant T$ *with some constant* $C > 0$ *not depending on* n *and* h.

Proof. Following the proof of Theorem 2, we show a refined error estimate for $e_n^{(1)}$. Due to Lemma 2 which is given at the end of this section, we have

$$\begin{aligned}\|e_n^{(1)}\|_X &\leqslant h^2 \cdot h \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} A_j \chi(hA_j) \right\|_{X \leftarrow X} \left\| g_j'\left(t_j + \frac{h}{2}\right) \right\|_X \\ &\quad + h^2(\|\varphi(hA_{n-1})\|_{X \leftarrow X} + \|\psi(hA_{n-1})\|_{X \leftarrow X}) \left\| g_{n-1}'\left(t_{n-1} + \frac{h}{2}\right) \right\|_X \\ &\leqslant C\|g'\|_{X,\infty} h^2(1 + |\log h|)\end{aligned}$$

which yields the result of the theorem. □

In the sequel, we analyse the convergence behaviour of (8) with respect to the norm in D. For that purpose, we introduce the notion of intermediate spaces, see also [10].

For some $0 < \vartheta < 1$ let $X_\vartheta = (X, D)_{\vartheta,p}$ denote the real interpolation space between X and D. Consequently, the norm in X_ϑ fulfills the relation

$$\|x\|_{X_\vartheta} \leqslant C_v \|x\|_D^{\vartheta} \|x\|_X^{1-\vartheta} \quad \text{for all } x \in D$$

with some constant $C_v > 0$. In particular, it follows

$$\|\mathrm{e}^{tA(s)}\|_{X_\vartheta \leftarrow X_\vartheta} + \|t^{1-\vartheta}\mathrm{e}^{tA(s)}\|_{D \leftarrow X_\vartheta} \leqslant M_2 \quad \text{for } 0 \leqslant t \leqslant T. \tag{22}$$

For the subsequent derivations, we choose ϑ in such a way that the interpolation space $X_{1+\vartheta} = (D, D(A(t)^2))_{\vartheta,p}$ between D and the domain of $A(t)^2$ is independent of t, and that the map A satisfies a Lipschitz-condition from $X_{1+\vartheta}$ to X_ϑ. In applications, this assumption is fulfilled for ϑ sufficiently small, see also Example 2.

Hypothesis 3. For some $0 < \vartheta < 1$, the interpolation space $X_{1+\vartheta}$ does not depend on t. Further, we suppose that the estimate

$$\|A(t) - A(s)\|_{X_\vartheta \leftarrow X_{1+\vartheta}} \leqslant M_3(t - s)$$

holds with some constant $M_3 > 0$ for all $0 \leqslant s \leqslant t \leqslant T$.

In this situation, following the proof of Theorem 1, we obtain

$$\left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} \right\|_{X_\vartheta \leftarrow X_\vartheta} \leqslant M_5 \quad \text{and} \quad \left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} \right\|_{D \leftarrow X_\vartheta} \leqslant M_5(t_{n+1} - t_m)^{-1+\vartheta}, \tag{23}$$

after a possible enlargement of $M_5 \geqslant 1$.

Example 2. In continuation of Example 1, we consider the second-order parabolic partial differential equation (11) subject to homogeneous Dirichlet boundary conditions and a certain initial condition. For this initial-boundary value problem, the admissible value of ϑ in Hypothesis 3 relies on the characterisation of the interpolation spaces between $D = W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$ and $D(A(t)^2)$. It follows from [4, Théorème 8.1′] that for $0 \leqslant \vartheta < 1/(2p)$ the interpolation space $X_{1+\vartheta}$ is isomorphic to $W^{2+2\vartheta,p}(\Omega) \cap W_0^{1,p}(\Omega)$ and thus independent of t. This is no longer true for $\vartheta > 1/(2p)$, since $X_{1+\vartheta}$, in general, depends on t through the boundary conditions $A(t)u = 0$ on $\partial\Omega$. Therefore, we may choose $0 \leqslant \vartheta < 1/(2p)$ in Hypothesis 3. Assuming that the spatial derivatives of the coefficients α_{ij}, β_i, and γ are Hölder continuous with respect to t, the required Hölder continuity of $A(t)$ on $X_{1+\vartheta}$ follows.

Under the requirement that the first derivative of g_n is bounded in D and that g_n'' belongs to the interpolation space X_γ for some $\gamma > 0$ arbitrarily small, the following result is valid. Note that for stepsizes $h > 0$ sufficiently small it follows $\gamma^{-1} h^\gamma \leqslant C|\log h|$.

Theorem 4. *Suppose that Hypotheses* 1–2 *with* $\alpha = 1$ *and Hypothesis* 3 *with* $\vartheta = \gamma$ *are fulfilled and apply the Magnus integrator* (8) *to the initial value problem* (1). *Then, the convergence estimate*

$$\|u_n - u(t_n)\|_D \leqslant Ch^2((1 + |\log h|)\|g'\|_{D,\infty} + (1 + \gamma^{-1}h^\gamma)\|g''\|_{X_\gamma,\infty})$$

holds true for $0 \leqslant t_n \leqslant T$. *The constant* $C > 0$ *is independent of* n *and* h.

Proof. Similarly as in the proof of Theorem 2, we successively analyse the error terms $e_n^{(1)}$ and $e_n^{(2)}$ defined in (21) by applying Theorem 1 and (20). On the one hand, we receive

$$\begin{aligned}
\|e_n^{(1)}\|_D &\leqslant \left\| \sum_{j=0}^{n-2} \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \delta_{j+1}^{(1)} \right\|_D + \|\delta_n^{(1)}\|_D \\
&\leqslant h^2 \cdot h \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \right\|_{D \leftarrow X} \|A_j\|_{X \leftarrow D} \|\chi(hA_j)\|_{D \leftarrow D} \left\| g_j'\left(t_j + \frac{h}{2}\right) \right\|_D \\
&\quad + h^2(\|\varphi(hA_{n-1})\|_{D \leftarrow D} + \|\psi(hA_{n-1})\|_{D \leftarrow D}) \left\| g_{n-1}'\left(t_{n-1} + \frac{h}{2}\right) \right\|_D \\
&\leqslant C\|g'\|_{D,\infty} h^2(1 + |\log h|).
\end{aligned}$$

A direct estimation of $\delta_{j+1}^{(2)}$ with the help of the relation (22) shows

$$\begin{aligned}
\|\delta_{n+1}^{(2)}\|_{X_\gamma} &\leqslant \int_0^h \|\mathrm{e}^{(h-\tau)A_n}\|_{X_\gamma \leftarrow X_\gamma} \left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma) \left\| g_n''\left(t_n + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \right\|_{X_\gamma} \mathrm{d}\sigma\,\mathrm{d}\tau \\
&\leqslant M_2 \|g''\|_{X_\gamma,\infty} h^3.
\end{aligned}$$

Besides, we receive

$$\begin{aligned}
\|\delta_{j+1}^{(2)}\|_D &\leqslant \int_0^h \|\mathrm{e}^{(h-\tau)A_j}\|_{D \leftarrow X_\gamma} \left(\tau - \frac{h}{2}\right)^2 \int_0^1 (1-\sigma) \left\| g_j''\left(t_j + \frac{h}{2} + \sigma\left(\tau - \frac{h}{2}\right)\right) \right\|_{X_\gamma} \mathrm{d}\sigma\,\mathrm{d}\tau \\
&\leqslant M_2 \|g''\|_{X_\gamma,\infty} \gamma^{-1} h^{2+\gamma}.
\end{aligned}$$

Consequently, together with (23) it follows

$$\begin{aligned}\|e_n^{(2)}\|_D &\leqslant \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} \right\|_{D\leftarrow X_\gamma} \|\delta_{j+1}^{(2)}\|_{X_\gamma} + \|\delta_n^{(2)}\|_D \\ &\leqslant C\|g''\|_{X_\gamma,\infty} h^2(1+\gamma^{-1}h^\gamma).\end{aligned}$$

This yields the given result. □

We next extend the above result to the situation where the first derivative of g belongs to the interpolation space $X_\beta = (X, D)_{\beta,p}$ for some $0<\beta<1$. If Hypothesis 3 holds with $\vartheta = \beta$, a proof similar to that of Lemma 2 below yields the auxiliary estimate

$$\left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} A_{m-1}\chi(hA_{m-1}) \right\|_{D\leftarrow X_\beta} \leqslant M_5 h^{-1+\beta}(t_{n+1}-t_m)^{-1}. \tag{24}$$

As before, we further suppose $g_n'' \in X_\gamma$ for some $\gamma>0$ arbitrarily small. Maximising the term $\gamma^{-1}h^\gamma$ with respect to γ yields $\gamma^{-1}h^\gamma \leqslant C|\log h|$ for $h>0$ sufficiently small.

Theorem 5. *Under Hypotheses* 1–2 *with* $\alpha = 1$ *and Hypothesis* 3 *with* $\vartheta = \beta$, *the Magnus integrator* (8) *for* (1) *satisfies the estimate*

$$\|u_n - u(t_n)\|_D \leqslant C(h^{1+\beta}(1+|\log h|)\|g'\|_{X_\beta,\infty} + h^2(1+\gamma^{-1}h^\gamma)\|g''\|_{X_\gamma,\infty})$$

for $0\leqslant t_n \leqslant T$ *with some constant* $C>0$ *independent of* n *and* h.

Proof. We follow the proof of Theorem 4 and modify the estimation of $e_n^{(1)}$. If $g' \in X_\beta$ the integral formula of Cauchy implies

$$\begin{aligned}\|\delta_n^{(1)}\|_D &\leqslant h^2 \left\| \psi(hA_{n-1}) - \frac{1}{2}\varphi(hA_{n-1}) \right\|_{D\leftarrow X_\beta} \left\| g_{n-1}'\left(t_{n-1}+\frac{h}{2}\right)\right\|_{X_\beta} \\ &\leqslant Ch^{1+\beta}\|g'\|_{X_\beta}.\end{aligned}$$

Together with (24) we thus receive

$$\begin{aligned}\|e_n^{(1)}\|_D &\leqslant \left\| \sum_{j=0}^{n-2} \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i}\delta_{j+1}^{(1)} \right\|_D + \|\delta_n^{(1)}\|_D \\ &\leqslant h^2 \cdot h \sum_{j=0}^{n-2} \left\| \prod_{i=j+1}^{n-1} \mathrm{e}^{hA_i} A_j\chi(hA_j) \right\|_{D\leftarrow X_\beta} \left\| g_j'\left(t_j+\frac{h}{2}\right)\right\|_{X_\beta} \\ &\quad + h^2 \left\| \psi(hA_{n-1}) - \frac{1}{2}\varphi(hA_{n-1}) \right\|_{D\leftarrow X_\beta} \left\| g_{n-1}'\left(t_{n-1}+\frac{h}{2}\right)\right\|_{X_\beta} \\ &\leqslant C\|g'\|_{X_\beta,\infty} h^{1+\beta}(1+|\log h|)\end{aligned}$$

which yields the given result. □

The following extension of Theorem 1 is needed in the proof of Theorem 3.

Lemma 2. *Assume that Hypotheses* 1–2 *with* $\alpha = 1$ *hold. Then, the bound*

$$\left\| \prod_{i=m}^{n} \mathrm{e}^{hA_i} A_{m-1} \chi(hA_{m-1}) \right\|_{X \leftarrow X} \leqslant M_5(1 + |\log h| + (t_{n+1} - t_m)^{-1}) \tag{25}$$

is valid for $0 \leqslant t_m < t_n \leqslant T$ *with some constant* $M_5 > 0$ *not depending on n and h.*

Proof. We note that by the integral formula of Cauchy, Theorem 1 and Hypotheses 1–2 it suffices to prove the desired bound (25) with A_{m-1} replaced by A_m. Thus, as in the proof of Theorem 1, we compare the discrete evolution operator with a frozen operator

$$\prod_{i=m}^{n} \mathrm{e}^{hA_i} A_m \chi(hA_m) = \Delta_m^n A_m \chi(hA_m) + A_m \mathrm{e}^{(t_{n+1}-t_m)A_m} \chi(hA_m).$$

Clearly, the second term is bounded by

$$\|A_m \mathrm{e}^{(t_{n+1}-t_m)A_m}\|_{X \leftarrow X} \|\chi(hA_m)\|_{X \leftarrow X} \leqslant C(t_{n+1} - t_m)^{-1},$$

see (5) and remark above as well as (20). For estimating the first term, we employ relation (13) for Δ_m^n given in the proof of Theorem 1 and receive

$$\begin{aligned} \Delta_m^n A_m \chi(hA_m) = &\sum_{j=m+1}^{n-1} \Delta_{j+1}^n (\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) A_m \mathrm{e}^{(t_j - t_m)A_m} \chi(hA_m) \\ &+ \sum_{j=m+1}^{n} \mathrm{e}^{(t_{n+1}-t_{j+1})A_m} (\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) A_m \mathrm{e}^{(t_j - t_m)A_m} \chi(hA_m). \end{aligned}$$

As a consequence of the integral formula of Cauchy, see also (15), we obtain

$$\|(\mathrm{e}^{hA_j} - \mathrm{e}^{hA_m}) A_m \mathrm{e}^{(t_j - t_m)A_m}\|_{X \leftarrow X} \leqslant Ch(t_j - t_m)^{-1}.$$

Together with (5), (14) and (20), it thus follows

$$\|\Delta_m^n A_m \chi(hA_m)\|_{X \leftarrow X} \leqslant Ch \sum_{j=m+1}^{n} (t_j - t_m)^{-1} \leqslant C(1 + |\log h|).$$

Altogether, this proves the desired result. □

5. Numerical examples

In order to illustrate the sharpness of the proven orders in our convergence bounds, we consider problem (11) in one space dimension for $x \in [0, 1]$ and $t \in [0, 1]$. We choose $\alpha(x, t) = 1 + \mathrm{e}^{-t}$ and $\beta(x, t) = \gamma(x, t) = 0$, and we determine $f(x, t)$ in such a way that the exact solution is given by $U(x, t) = x(1 - x)\mathrm{e}^{-t}$.

Table 1
Numerically observed temporal orders of convergence in different norms for discretisations with N spatial degrees of freedom and time stepsize $h = 1/128$

N	D_1	D_2	D_∞	L^1	L^2	L^∞
50	1.624	1.375	1.217	1.981	1.986	2.000
100	1.562	1.310	1.101	1.979	1.986	1.998
200	1.531	1.280	1.051	1.979	1.986	1.998
300	1.521	1.270	1.034	1.979	1.986	1.998
400	1.516	1.266	1.026	1.979	1.986	1.998

We discretise the partial differential equation in space by standard finite differences and in time by the Magnus integrator (8), respectively. Due to the particular form of the exact solution, the spatial discretisation error of our method is zero. The numerically observed temporal orders of convergence in various discrete norms are shown in Table 1. Recall that $X = L^p(\Omega)$ and $D_p = W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega)$.

The numerically observed order in the discrete L^2-norm is approximately 2, which is in accordance with Theorem 3. There is further a pronounced order reduction to approximately 1.25 in the discrete D_2-norm for sufficiently large N. This is explained as follows. The attainable value of β in Theorem 5 is restricted on the one hand by Hypothesis 3 and on the other hand by the domain of the function

$$g_n'(t) = A'(t)u(t) + (A(t) - A_n)u'(t) + b'(t), \quad t_n \leqslant t \leqslant t_{n+1},$$

see (16). In our example, g_n' is spatially smooth but does not satisfy the boundary conditions. For $X = L^2(\Omega)$ the optimal value is therefore $\beta = 1/4 - \varepsilon$ for arbitrarily small $\varepsilon > 0$, see [3,4] and the discussion in Example 2.

Similarly, for arbitrary $1 < p < \infty$, Theorem 3 predicts order 2 for the L^p-error, whereas an order reduction to approximately $1 + 1/(2p)$ in the discrete D_p-norm for large N is explained by Theorem 5. These numbers are in perfect agreement with Table 1, where we illustrated the limit cases $p = 1$ and $p = \infty$.

Acknowledgements

This work was supported by Acciones Integradas Austria-Spain 2002/03 under project 10/2002, Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under project H210-N13, and by DGI-MCYT under grants BFM2001-2138 and MTM 2004-02847.

References

[1] S. Blanes, P.C. Moan, Splitting methods for the time-dependent Schrödinger equation, Phys. Lett. A 265 (2000) 35–42.
[2] J. van den Eshof, M. Hochbruck, Preconditioning Lanczos approximations to the matrix exponential, SIAM J. Sci. Comput. (2004), to appear.
[3] D. Fujiwara, Concrete characterization of the domains of fractional powers of some elliptic differential operators of the second order, Proc. Japan Acad. 43 (1967) 82–86.
[4] P. Grisvard, Caractérisation de quelques espaces d'interpolation, Arch. Rational Mech. Anal. 25 (1967) 40–63.

[5] E. Hairer, Ch. Lubich, G. Wanner, Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations, Springer, Berlin, 2002.
[6] D. Henry, Geometric Theory of Semilinear Parabolic Equations, Lecture Notes in Mathematics, vol. 840, Springer, Berlin, 1981.
[7] M. Hochbruck, Ch. Lubich, On Krylov subspace approximations to the matrix exponential operator, SIAM J. Numer. Anal. 34 (1997) 1911–1925.
[8] M. Hochbruck, Ch. Lubich, On Magnus integrators for time-dependent Schrödinger equations, SIAM J. Numer. Anal. 41 (2003) 945–963.
[9] A. Iserles, S.P. Nørsett, On the solution of linear differential equations in Lie groups, Roy. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci. 357 (1999) 983–1019.
[10] A. Lunardi, Analytic Semigroups and Optimal Regularity in Parabolic Problems, Birkhäuser, Basel, 1995.
[11] W. Magnus, On the exponential solution of a differential equation for a linear operator, Comm. Pure Appl. Math. 7 (1954) 649–673.
[12] A. Ostermann, M. Thalhammer, Non-smooth data error estimates for linearly implicit Runge–Kutta methods, IMA J. Numer. Anal. 20 (2000) 167–184.
[13] A. Ostermann, M. Thalhammer, Convergence of Runge–Kutta methods for nonlinear parabolic equations, Appl. Numer. Math. 42 (2002) 367–380.
[14] A. Pazy, Semigroups of Linear Operators and Applications to Partial Differential Equations, Springer, New York, 1983.
[15] H. Tanabe, Equations of Evolution, Pitman, London, 1979.
[16] A. Zanna, Collocation and relaxed collocation for the Fer and the Magnus expansions, SIAM J. Numer. Anal. 36 (1999) 1145–1182.

2.2. A Magnus type integrator for quasilinear problems

A second-order Magnus type integrator for quasilinear parabolic problems

CÉSAREO GONZÁLEZ AND MECHTHILD THALHAMMER

To appear in Mathematics of Computation

MATHEMATICS OF COMPUTATION
S 0025-5718(XX)0000-0

A SECOND-ORDER MAGNUS TYPE INTEGRATOR FOR QUASILINEAR PARABOLIC PROBLEMS

C. GONZÁLEZ AND M. THALHAMMER

Abstract. In this paper, we consider an explicit exponential method of classical order two for the time discretisation of quasilinear parabolic problems. The numerical scheme is based on a Magnus integrator and requires the evaluation of two exponentials per step. Our convergence analysis includes parabolic partial differential equations under a Dirichlet boundary condition and provides error estimates in Sobolev-spaces. In an abstract formulation, the initial-boundary value problem is written as an initial value problem on a Banach space X

$$u'(t) = A\big(u(t)\big)u(t), \quad 0 < t \le T, \qquad u(0) \text{ given},$$

involving the sectorial operator $A(v) : D \to X$ with domain $D \subset X$ independent of $v \in V \subset X$. Under reasonable regularity requirements on the problem, we prove the stability of the numerical method and derive error estimates in the norm of certain intermediate spaces between X and D. Various applications and a numerical experiment illustrate the theoretical results.

1. Introduction

In this paper, we are concerned with the numerical solution of initial value problems of the form

$$u'(t) = A\big(u(t)\big)u(t), \quad 0 < t \le T, \qquad u(0) \text{ given}. \tag{1.1}$$

Our main interest is to study (1.1) in an abstract setting where $A(v) : D \subset X \to X$ is a family of sectorial operators on a Banach space X which is defined for elements $v \in V \subset X_\gamma$ in an open subset of some intermediate space $D \subset X_\gamma \subset X$. The scope of applications includes quasilinear parabolic partial differential equations under a boundary condition of Dirichlet type which arise in the modelling of diffusion processes with state-dependent diffusivity and in the study of fluids in porous media.

In the present work, we pursue our convergence and stability analysis of Magnus type integrators for the time discretisation of non-autonomous parabolic problems [11, 26, 27] and study an explicit exponential integration scheme for abstract quasilinear problems (1.1). The numerical method considered relies on a second-order Magnus integrator and requires the evaluation of two exponentials at each step.

In the last few years, due to the progress of the art and the increasing potentiality for the efficient calculation of the matrix exponential in *non*-dubious ways [22],

Received by the editor December 2004. Revised version September 2005.
2000 *Mathematics Subject Classification.* 35K55, 35K90, 65L20, 65M12.
Key words and phrases. Quasilinear parabolic problems, Magnus integrators, Stability, Convergence.

see [10, 16] and references cited therein, numerical methods based on the Magnus expansion have received a lot of attention. This is confirmed by a variety of recent works, as a small selection we mention [5, 7, 17, 18, 19, 29]. Following an approach studied by Magnus [21], for a linear system of non-autonomous ordinary differential equations

$$y'(t) = A(t)y(t), \qquad y(0) \text{ given}, \tag{1.2}$$

the solution is represented by the exponential of a time-dependent matrix Ω

$$y(t) = \mathrm{e}^{\Omega(t)} y(0), \qquad t \geq 0,$$

which is given by an infinite series of iterated integrals involving matrix commutators of A

$$\Omega(t) = \int_0^t A(\tau)\,\mathrm{d}\tau - \frac{1}{2}\int_0^t \Big[\int_0^\tau A(\sigma)\,\mathrm{d}\sigma, A(\tau)\Big]\,\mathrm{d}\tau + \ldots \tag{1.3}$$

In order to obtain a numerical approximation to the exact solution of (1.2), the Magnus expansion (1.3) is truncated and the integrals are determined by means of a quadrature formula. For instance, applying the midpoint rule to the first integral and omitting the remaining terms yields the second-order approximation

$$y_1 = \mathrm{e}^{hA(h/2)} y_0 \tag{1.4}$$

to the exact solution value at time $h > 0$. Here, the numerical starting value y_0 is a suitable approximation to the exact initial value $y(0)$. Such interpolatory Magnus integrators were considered in Iserles & Nørsett [18], e.g., in the context of geometric integration, and, as proven in Hochbruck & Lubich [17], this method class is also eminently suited for the time integration of spatial discretisations of time-dependent Schrödinger type equations. In [11, 26], the second-order Magnus integrator (1.4) was studied for abstract parabolic problems and further extended to linear and semilinear equations.

The above considerations motivate the following Magnus type integrator for differential equations of the form (1.1). For some initial value $u_0 \approx u(0)$ and a stepsize $h > 0$ the numerical solution u_1 is determined by the relation

$$u_1 = \mathrm{e}^{hA(U_{01})} u_0 \approx u(h). \tag{1.5a}$$

As auxiliary approximation to the exact solution value at the midpoint of the interval $[0, h]$, the additional internal stage U_{01} is calculated by means of a first-order integrator

$$U_{01} = \mathrm{e}^{h/2\, A(u_0)} u_0 \approx u\big(h/2\big). \tag{1.5b}$$

By Taylor series expansions it is straightforward to show that this scheme has classical order 2. It is notable that (1.5) can also be considered as a Runge-Kutta Munthe-Kaas method.

The objective of the present work is to analyse the stability and convergence behaviour of the numerical method (1.5) in the situation where (1.1) constitutes a quasilinear parabolic initial-boundary value problem written as an initial value problem on a Banach space.

Our paper is organised as follows. In Section 2, we state the fundamental hypotheses on the differential equation in (1.1) and further specify several applications that can be cast into our abstract setting. In Section 3, we introduce the Magnus type integrator whose favourable stability and convergence properties in connection

with parabolic problems are analysed in detail in the subsequent Sections 4 and 5. In particular, under reasonable regularity requirements on the data and the solution of the initial value problem (1.1), we state an error estimate in the norm of a certain intermediate space between the underlying Banach space X and the domain D. In Section 6, we finally comment on an extension of the Magnus type integrator to equations with an additional inhomogeneity and illustrate the theoretical result by a numerical example.

2. Problem class and applications

In this section, we state the fundamental assumptions on the problem class considered and illustrate the abstract framework by several applications. The hypotheses on the initial value problem (1.1) primarily rely on González & Palencia [13] where Runge-Kutta time discretisations for quasilinear parabolic problems were studied. However, in our notation we follow Lunardi [20] and the previous works [11, 26]. For an extensive treatise of quasilinear evolution equations, we refer to the works of Amann [1]-[4]. The theory of sectorial operators and analytic semigroups is described in detail in the monographs [15, 20, 25]. A comprehensive overview of interpolation theory is given in [20], see also [6, 28].

To simplify the notation, we henceforth do not distinguish the arising constants. Thus, the positive quantities $K, L, M > 0$ and $C > 0$ possibly have different values at different occurrences.

2.1. **Quasilinear equation.** We consider a complex Banach space $(X, \|\cdot\|_X)$ and a dense subspace $(D, \|\cdot\|_D)$ which we assume to be continuously embedded in X. For $0 < \mu < 1$ we denote by X_μ some intermediate space between X and D such that the norm in X_μ fulfills the relation

$$\|x\|_{X_\mu} \le K\|x\|_X^{1-\mu}\|x\|_D^{\mu}, \qquad x \in D,$$

with a constant $K > 0$. Specifically, we set $X_0 = X$ and $X_1 = D$.

The right-hand side of the differential equation in (1.1) is defined by the map $A : V \to L(D, X)$ where $V \subset X_\gamma$ is an open subset of some intermediate space X_γ with $0 \le \gamma < 1$. In view of applications, the requirement that the domain of the unbounded linear operator $A(v) : D \to X$ is independent of $v \in V$ implies that in general only initial-boundary value problems involving a boundary condition of Dirichlet type are covered by our analysis. The fundamental assumptions on A are as follows.

Hypothesis 2.1. (i) The closed linear operator $A(v) : D \to X$ is uniformly sectorial for $v \in V$. Thus, there exist constants $a \in \mathbb{R}$, $0 < \phi < \pi/2$, and $M > 0$ such that for every $v \in V$ and for any complex number $\lambda \in \mathbb{C}$ in the complement of the sector

$$S_\phi(a) = \big\{z \in \mathbb{C} : |\arg(a - z)| \le \phi\big\} \cup \{a\}$$

the resolvent $\big(\lambda I - A(v)\big)^{-1} : X \to X$ exists and further satisfies the estimate

$$\Big\|\big(\lambda I - A(v)\big)^{-1}\Big\|_{X \leftarrow X} \le \frac{M}{|\lambda - a|}, \qquad \lambda \in \mathbb{C} \setminus S_\phi(a). \tag{2.1}$$

(ii) The graph norm of $A(v)$ and the norm in D are equivalent, i.e., for every $v \in V$ the following relation holds with a constant $K > 0$

$$K^{-1}\|x\|_D \le \|x\|_X + \big\|A(v)x\big\|_X \le K\|x\|_D, \qquad x \in D. \tag{2.2}$$

(iii) For some $0 \le \vartheta < 1$ the intermediate space $X_{1+\vartheta}$ between D and the domain of $A(v)^2$ does not depend on $v \in V$. Moreover, the map $A : V \to L(X_{1+\vartheta}, X_\vartheta)$ is Lipschitz-continuous with respect to v, that is, the estimate

$$\big\|A(v) - A(w)\big\|_{X_\vartheta \leftarrow X_{1+\vartheta}} \le L\|v - w\|_{X_\gamma}, \qquad v, w \in V, \tag{2.3}$$

is valid with a constant $L > 0$.

By Hypothesis 2.1(ii), a suitable choice for the intermediate space X_γ is the real interpolation space or the intermediate Calderón space, whereas, due to the non-applicability of Heinz's theorem, the fractional power space may depend on $v \in V$.

Quasilinear parabolic initial-boundary value problems where the above assumptions hold true are specified below in Subsection 2.2.

Remark 2.2. In the situation of Hypothesis 2.1 with $\vartheta = 0$, the unique solvability of the abstract initial value problem (1.1) is ensured. Namely, it is shown in Amann [2] that the quasilinear differential equation defines a semiflow in $X_\beta \cap V$ for every $\gamma < \beta < 1$. However, the limiting case $\beta = \gamma$ is not covered by this result.

We note that for a linear operator $F : X \to D$ relation (2.2) implies the bounds

$$\big\|A(v)F\big\|_{X \leftarrow X} \le K\|F\|_{D \leftarrow X}, \qquad \|F\|_{D \leftarrow X} \le K\big(1 + \big\|A(v)F\big\|_{X \leftarrow X}\big).$$

Besides, after possibly enlarging the constant $M > 0$, the following extension of the resolvent estimate (2.1) is valid

$$\Big\|t^{\nu-\mu}\big(\lambda I - tA(v)\big)^{-1}\Big\|_{X_\nu \leftarrow X_\mu} \le \frac{M}{|\lambda - at|}, \qquad t > 0, \qquad 0 \le \mu \le \nu \le 1, \tag{2.4}$$

see also [12]. For any fixed $v \in V$ the sectorial operator $A(v) : D \to X$ is the infinitesimal generator of an analytic semigroup $\big(\mathrm{e}^{tA(v)}\big)_{t \ge 0}$ on X. Here, the linear operator

$$\mathrm{e}^{tA(v)} = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda \big(\lambda I - tA(v)\big)^{-1} \,\mathrm{d}\lambda, \qquad t > 0, \tag{2.5}$$

is defined through the integral formula of Cauchy, where Γ denotes a path that surrounds the spectrum of $A(v)$. If $t = 0$ let $\mathrm{e}^{tA(v)} = I$. Therefore, due to (2.4), the estimates

$$\begin{aligned} \big\|t^{\nu-\mu}\mathrm{e}^{tA(v)}\big\|_{X_\nu \leftarrow X_\mu} &\le M, \qquad 0 \le t \le T, \qquad 0 \le \mu \le \nu \le 1, \\ \big\|t^{1+\nu-\mu}A(v)\,\mathrm{e}^{tA(v)}\big\|_{X_\nu \leftarrow X_\mu} &\le M, \qquad 0 \le t \le T, \qquad 0 \le \mu, \nu \le 1, \end{aligned} \tag{2.6}$$

are valid, see also [20, Chapter 2]. Consequently, by means of the identity

$$\mathrm{e}^{tA(v)} - I = \int_0^t A(v)\mathrm{e}^{\tau A(v)} \,\mathrm{d}\tau,$$

we obtain the following bound

$$\big\|\mathrm{e}^{tA(v)} - I\big\|_{X_\nu \leftarrow X_\mu} \le Mt^{-\nu+\mu}, \qquad t > 0, \qquad 0 \le \mu, \nu \le 1. \tag{2.7}$$

For later use, we further introduce the bounded linear operators

$$\begin{aligned} \varphi\big(tA(v)\big) &= \frac{1}{t}\int_0^t \mathrm{e}^{(t-\tau)A(v)} \,\mathrm{d}\tau, \quad t > 0, \qquad \varphi\big(tA(v)\big) = I, \quad t = 0, \\ \psi\big(tA(v)\big) &= \frac{1}{t^2}\int_0^t \tau\mathrm{e}^{(t-\tau)A(v)} \,\mathrm{d}\tau, \quad t > 0, \qquad \psi\big(tA(v)\big) = \frac{1}{2}\,I, \quad t = 0, \end{aligned} \tag{2.8a}$$

which are related to the analytic semigroup. Moreover, with the help of the integral formula of Cauchy, the validity of the relation

$$\psi\big(tA(v)\big) - 1/2\,\varphi\big(tA(v)\big) = tA(v)\chi\big(tA(v)\big)$$

with a bounded linear operator $\chi\big(tA(v)\big)$ follows. More precisely, as a direct consequence of the defining relations and (2.6), we obtain the estimate

$$\begin{aligned}\big\|t^{\nu-\mu}\varphi\big(tA(v)\big)\big\|_{X_\nu\leftarrow X_\mu} + \big\|t^{\nu-\mu}\psi\big(tA(v)\big)\big\|_{X_\nu\leftarrow X_\mu} \\ + \big\|\chi\big(tA(v)\big)\big\|_{X_\mu\leftarrow X_\mu} \le M, \qquad 0\le t\le T, \qquad 0\le\mu\le\nu\le 1,\end{aligned} \tag{2.8b}$$

with a constant $M > 0$.

We close this subsection with some useful abbreviations. In the sequel, the closed ball in X_μ with radius $\varrho > 0$ and center $v^* \in X_\mu$ is denoted by

$$B_\mu(v^*, \varrho) = \big\{v \in X_\mu : \|v - v^*\|_{X_\mu} \le \varrho\big\} \subset X_\mu. \tag{2.9}$$

Further, for a family $f = (f_n)_{0\le n\le N}$ of bounded maps $f_n : I_n \subset \mathbb{R} \to X_\mu$ or for a sequence $g = (g_n)_{0\le n\le N}$ in X_μ we set

$$\begin{gathered}\big\|f\big\|_{X_\mu,\infty} = \max_{0\le n\le N}\big\|f_n\big\|_{X_\mu,\infty}, \qquad \big\|f_n\big\|_{X_\mu,\infty} = \max_{t\in I_n}\big\|f_n(t)\big\|_{X_\mu}, \\ \big\|g\big\|_{X_\mu,\infty} = \max_{0\le n\le N}\big\|g_n\big\|_{X_\mu}.\end{gathered} \tag{2.10}$$

2.2. **Applications.** The following initial-boundary value problem can be cast into the abstract setting of Subsection 2.1, see also [13].

Example 2.3. Let Ω be an open and bounded domain in $\mathbb{R}^d$ with regular boundary $\partial\Omega$. We consider the following partial differential equation for a real-valued function $U : \Omega \times [0,T] \to \mathbb{R} : (x,t) = (x_1, x_2, \dots, x_d, t) \to U(x,t)$

$$\partial_t U(x,t) = \mathcal{A}\big(U(x,t)\big)U(x,t), \qquad x\in\Omega, \quad 0 < t \le T, \tag{2.11a}$$

subject to a homogeneous Dirichlet boundary condition and an initial condition

$$U(x,t) = 0, \quad x\in\partial\Omega, \quad 0\le t\le T, \qquad U(x,0) = U_0(x), \quad x\in\Omega. \tag{2.11b}$$

Here, for $v \in C^1(\Omega)$ and $w \in C^2(\Omega)$ the second-order differential operator $\mathcal{A}$ is defined through

$$\mathcal{A}\big(v(x)\big)w(x) = \sum_{i,j=1}^{d} a_{ij}\big(x, v(x), \nabla v(x)\big)\partial_{x_i x_j} w(x), \qquad x\in\Omega. \tag{2.12}$$

We suppose that the real-valued coefficients a_{ij} which are defined on an open domain $\Omega\times\Lambda \subset \mathbb{R}^d\times\mathbb{R}\times\mathbb{R}^d$ satisfy suitable regularity and boundedness assumptions, and we further impose the ellipticity condition

$$\sum_{i,j=1}^{d} a_{ij}(x,p,q)\,\xi_i\xi_j \ge \kappa\sum_{i=1}^{d}\xi_i^2, \qquad (x,p,q)\in\Omega\times\Lambda, \quad \xi\in\mathbb{R}^d,$$

for some $\kappa > 0$.

By suppressing the spatial variable, the initial-boundary value problem (2.11) takes the form of an abstract initial value problem (1.1) on the Banach space

$$X = L^p(\Omega), \qquad d < p < \infty.$$

More precisely, we set $\big(u(t)\big)(x) = U(x,t)$ and define the linear operator $A(v)$ through $\big(A(v)w\big)(x) = \mathcal{A}\big(v(x)\big)w(x)$. Then, by choosing

$$D = W^{2,p}(\Omega) \cap W_0^{1,p}(\Omega), \qquad V = X_\gamma, \quad 1/2 + d\,(2p)^{-1} < \gamma < 1,$$

it follows that Hypothesis 2.1 is satisfied with $\vartheta = 0$. In particular, due to the imbedding $X_\gamma \subset C^1(\Omega)$, the linear operator $A(v) : D \to X$ is well-defined for elements $v \in X_\gamma$, see [15, Section 1.6]. If the coefficients of the differential operator do not depend on the derivative, that is, in (2.12) $a_{ij}\big(x, v(x), \partial_x v(x)\big)$ is replaced by $a_{ij}\big(x, v(x)\big)$, the less restrictive condition $d\,(2p)^{-1} < \gamma < 1$ follows.

The following illustration describes the movement of a fluid of variable density through a porous medium under the influence of gravity and hydrodynamic dispersion. It is shown in Clément et al. [9] that the specified system of elliptic-parabolic partial differential equations when reformulated as an abstract evolution equation on a suitably chosen Banach space leads to a quasilinear parabolic problem.

Example 2.4. Let $\Omega \subset \mathbb{R}^2$ be a rectangle or an open and bounded domain in $\mathbb{R}^2$ with regular boundary $\partial\Omega$. Elements $x = (x_1, x_2)^T \in \mathbb{R}^2$ are meanwhile interpreted as columns. We consider a system of elliptic-parabolic partial differential equations for functions $U, V : \Omega \times [0,T] \to \mathbb{R} : (x,t) \to U(x,t)$

$$\begin{cases} -\Delta V(x,t) = \partial_{x_1} U(x,t), \\ \partial_t U(x,t) + \operatorname{div} F(x,t) = 0, \end{cases} \qquad x \in \partial\Omega, \quad 0 < t \le T, \tag{2.13a}$$

with map $F = (F_1, F_2)^T : \Omega \times [0,T] \to \mathbb{R}^2$ defined by

$$F(x,t) = \operatorname{curl} V(x,t)\; U(x,t) - D\big(\operatorname{curl} V(x,t)\big)\nabla U(x,t). \tag{2.13b}$$

Here, we set $\operatorname{curl} V = (-\partial_{x_2} V, \partial_{x_1} V)^T$ and further employ the standard notations $\nabla U = (\partial_{x_1} U, \partial_{x_2} U)^T$, $\Delta V = \partial_{x_1}^2 V + \partial_{x_2}^2 V$, and $\operatorname{div} F = \partial_{x_1} F_1 + \partial_{x_2} F_2$. The system (2.13a) is subject to the boundary conditions

$$V(x,t) = 0, \quad \nu^T F(x,t) = 0, \qquad x \in \partial\Omega, \quad 0 < t \le T, \tag{2.13c}$$

where $\nu = (\nu_1, \nu_2)^T$ is the outward normal unit vector on $\partial\Omega$. Moreover, we impose a certain initial condition $U(x,0) = U_0(x)$ for $x \in \Omega$. Specifically, the real-valued functions $D_{ij} : \mathbb{R}^2 \to \mathbb{R} : q \to D(q)$ that define the hydrodynamic dispersion matrix $D(q) = \big(D_{ij}(q)\big)_{1 \le i,j \le 2}$ are given by

$$D_{ij}(q) = \begin{cases} \Big(c_1 + c_2\sqrt{q_1^2 + q_2^2}\Big)\delta_{ij} + c_3\,\frac{q_i q_j}{\sqrt{q_1^2+q_2^2}}, & \text{if } q \neq 0, \\ c_1 \delta_{ij}, & \text{if } q = 0. \end{cases} \tag{2.13d}$$

The positive constants c_1, c_2, and c_3 involve certain characteristic quantities such as the transversal and longitudinal dispersion length, the molecular diffusion coefficient as well as the tortuosity and porosity of the medium. As usual, δ_{ij} denotes the Kronecker symbol. In particular, the ellipticity condition

$$\sum_{i,j=1}^{2} D_{ij}(q)\,\xi_i \xi_j \ge \kappa\big(\xi_1^2 + \xi_2^2\big), \qquad q \in \mathbb{R}^2, \quad \xi \in \mathbb{R}^d,$$

holds for some $\kappa > 0$.

Using the well-known result that the differential operator $-\Delta$ subject to a homogeneous Dirichlet boundary condition is invertible in $L^p(\Omega)$, we express the

solution V of (2.13a) in terms of U. That is, denoting the inverse operator by $(-\Delta|\gamma)^{-1}$, we get the relation $V = (-\Delta|\gamma)^{-1}\partial_{x_1}U$. Furthermore, by introducing the solution-dependent coefficients a_i and $a_{ij}, 1 \le i,j \le 2$, such that

$$-\operatorname{curl} V(x,t) = -\operatorname{curl}\big((-\Delta|\gamma)^{-1}\partial_{x_1}U(x,t)\big) = \begin{pmatrix} a_1\big(U(x,t)\big) \\ a_2\big(U(x,t)\big) \end{pmatrix},$$

$$D\big(\operatorname{curl} V(x,t)\big) = \begin{pmatrix} a_{11}\big(U(x,t)\big) & a_{12}\big(U(x,t)\big) \\ a_{21}\big(U(x,t)\big) & a_{22}\big(U(x,t)\big) \end{pmatrix},$$

problem (2.13) takes the form

$$\partial_t U(x,t) = \mathcal{A}\big(U(x,t)\big)U(x,t), \qquad x \in \Omega, \quad 0 < t \le T, \tag{2.14a}$$

with differential operator $\mathcal{A}$ given by

$$\begin{aligned} \mathcal{A}\big(U(x,t)\big)U(x,t) &= \sum_{i=1}^{2} \partial_{x_i} a_i\big(U(x,t)\big)U(x,t) \\ &\quad + \sum_{i,j=1}^{2} \partial_{x_i} a_{ij}\big(U(x,t)\big)\partial_{x_j}U(x,t). \end{aligned} \tag{2.14b}$$

In addition, the solution U fulfills the boundary condition

$$\sum_{i=1}^{2} \nu_i a_i\big(U(x,t)\big)U(x,t) + \sum_{i,j=1}^{2} \nu_i a_{ij}\big(U(x,t)\big)\partial_{x_j}U(x,t) = 0 \tag{2.14c}$$

for $x \in \partial\Omega$ and $0 < t \le T$ as well as the initial condition $U(x,0) = U_0(x)$ for $x \in \Omega$.

In order to cast this parabolic initial-boundary value problem into our abstract framework, we set

$$Y = W^{1,p'}(\Omega), \qquad X = Y', \qquad D = W^{1,p}(\Omega),$$

for $2 < p < \infty$ and $1/p' = 1 - 1/p$. Besides, we define $A(u)u$ for $u \in D$ through

$$\big\langle A(u)u, v\big\rangle = \sum_{i=1}^{2} \big\langle \partial_{x_i} v, a_i(u)u\big\rangle + \sum_{i,j=1}^{2} \big\langle \partial_{x_i} v, a_{ij}(u)\partial_{x_j} u\big\rangle, \qquad v \in Y, \tag{2.15}$$

where we employ the standard notation

$$\big\langle f, g\big\rangle = \int_\Omega f(x)g(x)\,\mathrm{d}x, \qquad f \in L^p(\Omega), \quad g \in L^{p'}(\Omega).$$

In (2.15), due to the imbedding $W^{1,p}(\Omega) \subset C\big(\Omega\big)$, the coefficients $a_i(u)$ and $a_{ij}(u)$ are defined pointwise on the closure of Ω. The investigations in [9] imply that the operator family $A : V \to L(D,X)$ satisfies Hypothesis 2.1 with $V = X_\gamma$ for $1/2 + 1/p < \gamma < 1$ and $\vartheta = 0$.

3. Magnus type integrator

In the sequel, we specify the numerical scheme for the time discretisation of quasilinear parabolic problems.

Henceforth, for integers $n \ge 0$ let $t_n = nh$ be the grid points associated with a constant stepsize $h > 0$. The numerical approximation u_{n+1} to the value of the

exact solution of the abstract initial value problem (1.1) at time t_{n+1} is determined through the recurrence formula

$$\begin{aligned} U_{n1} &= \mathrm{e}^{h/2\,A_n} u_n, \qquad A_n = A(u_n), \\ u_{n+1} &= \mathrm{e}^{hA_{n1}} u_n, \qquad A_{n1} = A\big(U_{n1}\big), \qquad n \geq 0. \end{aligned} \tag{3.1}$$

Here, similarly as for Runge-Kutta methods, the numerical solution u_{n+1} is computed by means of an additional internal stage U_{n1} which is a first-order approximation to the exact solution value at the midpoint $t_{n1} = t_n + h/2$.

Provided that the exponential is available, the benefits of the Magnus type integrator (3.1) are its explicitness and favourable stability properties. Namely, the utilisation of exponentials instead of rational functions enhances the stability properties of the integrator. In this respect, we refer to González & Palencia [13] where the stability and convergence behaviour of Runge-Kutta time discretisations for quasilinear parabolic problems is studied. However, in [13] the requirement of strong $A(\theta)$-stability implies that the Runge-Kutta method is implicit.

In the non-stiff case, by employing Taylor series expansions, it is straightforward to prove that the numerical method (3.1) has classical order two. In the situation where (1.1) constitutes an abstract quasilinear parabolic problem on a Banach space its convergence behaviour is analysed in Section 5 below.

Remark 3.1. We note that the solution of (3.1) remains well-defined in $X_\beta \cap V$ for any $\gamma < \beta \leq 1$. Namely, whenever u_n lies in $X_\beta \cap V$ it follows from (2.6) that U_{n1} is bounded in X_β

$$\big\|U_{n1}\big\|_{X_\beta} \leq M \|u_n\|_{X_\beta}.$$

On the other hand, for $h > 0$ sufficiently small it holds

$$\big\|U_{n1} - u_n\big\|_{X_\gamma} \leq \big\|\mathrm{e}^{h/2\,A_n} - I\big\|_{X_\gamma \leftarrow X_\beta} \|u_n\|_{X_\beta} \leq M h^{\beta-\gamma} \|u_n\|_{X_\beta} \leq \varrho,$$

that is, the internal stage U_{n1} is contained in a ball $B_\gamma(u_n, \varrho) \subset X_\gamma$ and thus in V for suitably chosen $\varrho > 0$, see also (2.7) and (2.9). In particular, it follows $U_{n1} \in X_\beta \cap V$, and therefore the sectorial operator $A(U_{n1})$ is well-defined. Now, similar considerations to before show that also u_{n+1} belongs to $X_\beta \cap V$.

For a family $(F_i)_{i \geq 0}$ of non-commutative operators on a Banach space, we employ the product notation

$$\prod_{i=m}^{n} F_i = F_n F_{n-1} \cdots F_m, \quad n \geq m, \qquad \prod_{i=m}^{n} F_i = I, \quad n < m.$$

As a consequence, by solving the recursion for the numerical solution in (3.1), we get the relation

$$u_{n+1} = \prod_{i=0}^{n} \mathrm{e}^{hA_{i1}} u_0 = \mathrm{e}^{hA_{n1}} \mathrm{e}^{hA_{n-1,1}} \cdots \mathrm{e}^{hA_{01}} u_0, \qquad n \geq 0.$$

Our first objective is to study the stability behaviour of this numerical approximation. This is done in Section 4 below.

4. Stability

In this section, we analyse the stability behaviour of the numerical method (3.1), that is, we study the dependence of the numerical approximation on the initial value and the effect of additional perturbations. Several auxiliary estimates are collected in Subsection 4.2.

4.1. **Stability result.** For the following considerations, we employ the assumptions and notation introduced in the previous Sections 2 and 3. In particular, we denote by $0 \le \gamma < 1$ and $0 \le \vartheta < 1$ the constants specified in Hypothesis 2.1. Further, in view of Example 2.3 and the discussion in Subsection 6.2, it is senseful to suppose $\vartheta \le \gamma$.

Henceforth, we fix $\gamma < \beta \le 1$ and $u_0 \in X_\beta$. Accordingly to our numerical scheme (3.1), for initial values $v_0, w_0 \in X_\beta$ and additional perturbations $p_n, q_n \in X_\beta$ for $n \ge 1$ we consider the recursions

$$\begin{aligned} v_{n+1} &= \mathrm{e}^{hA(V_{n1})} v_n + h p_{n+1}, \qquad V_{n1} = \mathrm{e}^{h/2\, A(v_n)} v_n, \\ w_{n+1} &= \mathrm{e}^{hA(W_{n1})} w_n + h q_{n+1}, \qquad W_{n1} = \mathrm{e}^{h/2\, A(w_n)} w_n, \qquad n \ge 0. \end{aligned} \tag{4.1}$$

We note that similar considerations as in Remark 3.1 imply that V_{n1} and v_{n+1} belong to $X_\beta \cap V$ provided that $v_n \in X_\beta \cap V$, $p_{n+1} \in X_\beta$ bounded, and $h > 0$ sufficiently small. The analogue is valid for w_{n+1}.

The following result shows that furthermore these recurrence formulas remain bounded in X_β. Especially, it follows that the Magnus type integrator (3.1) starting from $u_0 \in X_\beta \cap V$ is applicable up to time T.

Theorem 4.1 (Stability)**.** *Suppose that Hypothesis 2.1 is fulfilled with $\vartheta > 0$. For $\gamma < \beta \le 1$ let $v_0 \in X_\beta \cap V$ and $w_0 \in X_\beta \cap V$ and assume that p_n and q_n are bounded in X_β for $n \ge 1$. Then, for $h > 0$ chosen sufficiently small the solutions of* (4.1) *satisfy the bound*

$$\|v_n - w_n\|_{X_\beta} \le C\Big(\|v_0 - w_0\|_{X_\beta} + \max_{1 \le j \le n} \|p_j - q_j\|_{X_\beta} \Big), \qquad 0 \le nh \le T,$$

with constant $C > 0$ not depending on n and h.

Proof. Our proof is based on a fixed-point iteration based on a global representation of the solutions in (4.1). For this purpose, we introduce several notations.

For the following, we choose $u_0 \in X_\beta$ and fix $\gamma < \zeta < \beta$ and $0 < \alpha < \beta - \zeta$. For constants $\varrho > 0$ and $\widetilde{L} > 0$ we set

$$\begin{aligned} \mathcal{Z} = \big\{ z = (z_n)_{0 \le nh \le T} : z_0 \in B_\beta(u_0, \varrho) \cap V, z_n \in X_\zeta \cap V \text{ for } n \ge 1 \\ \text{and } nh \le T, \|z_n - z_m\|_{X_\zeta} \le \widetilde{L}(t_n - t_m)^\alpha \text{ for } 0 \le mh \le nh \le T \big\}. \end{aligned} \tag{4.2}$$

In particular, for $z^* \in B_\beta(u_0, \varrho) \cap V$ we denote $\mathcal{Z}_{z^*} = \big\{ z \in \mathcal{Z} : z_0 = z^* \big\}$. Note that the sequence spaces $\mathcal{Z}$ and $\mathcal{Z}_{z^*}$ are complete metric spaces with the distance induced by the maximum norm

$$\|z\|_{X_\zeta, \infty} = \max_{0 \le nh \le T} \|z_n\|_{X_\zeta},$$

see also (2.10). Besides, for some $\sigma > 0$ we introduce the set

$$\mathcal{S} = \big\{ s = (s_n)_{h \le nh \le T} : s_n \in B_\beta(0, \sigma) \text{ for } n \ge 1 \text{ and } nh \le T \big\}.$$

For $z \in \mathcal{Z}$, accordingly to relation (4.1), we denote by $Z_1(z) = \big(Z_{n1}(z)\big)_{n\geq 0}$ the sequence defined through

$$Z_{n1}(z) = \mathrm{e}^{h/2\, A(z_n)} z_n, \qquad n \geq 0. \tag{4.3a}$$

Moreover, we introduce a family of linear operators $L(z) = \big(L_m^n(z)\big)_{n\geq m\geq 0}$ depending on the sequence Z_1 and thus on z

$$L_m^n(z) = \prod_{i=m}^{n} \mathrm{e}^{hA(Z_{i1}(z))}, \qquad 0 \leq m \leq n. \tag{4.3b}$$

For the following, we fix $z^* \in X_\beta \cap V$ and $s \in \mathcal{S}$ and let

$$\begin{aligned} &\mathcal{N} : \mathcal{Z}_{z^*} \longrightarrow \mathcal{Z}_{z^*} : z \longmapsto \mathcal{N}(z) = N(z,s) = \big(N_n(z,s)\big)_{0\leq nh\leq T}, \\ &N_0(z,s) = z^*, \qquad N_n(z,s) = L_0^{n-1}(z) z^* + h \sum_{j=0}^{n-1} L_{j+1}^{n-1}(z) s_{j+1}, \quad n \geq 1. \end{aligned} \tag{4.3c}$$

Clearly, a sequence $z \in \mathcal{Z}_{z^*}$ that is a fixed-point of the nonlinear operator $\mathcal{N}$, that is, z satisfies the relation $z = \mathcal{N}(z)$, also fulfills the recurrence formula

$$z_{n+1} = \mathrm{e}^{hA(Z_{n1})} z_n + h s_{n+1}, \qquad Z_{n1} = \mathrm{e}^{h/2\, A(z_n)} z_n, \qquad n \geq 0, \tag{4.4}$$

with initial value $z_0 = z^*$.

We next prove the unique solvability of the fixed-point equation $z = \mathcal{N}(z)$ and the continuous dependence of the fixed-point on the initial value and additional perturbations. Several auxiliary results needed for the following considerations are derived in the subsequent Subsection 4.2.

(i) Let $v, w \in \mathcal{Z}_{z^*}$ and $s \in \mathcal{S}$. Estimating the difference $N_n(v,s) - N_n(w,s)$ with the help of Lemma 4.6 and using that $\|s_{j+1}\| \leq \sigma$ for $0 \leq j \leq n-1$ gives

$$\begin{aligned} \big\|N_n(v,s) - N_n(w,s)\big\|_{X_\zeta} &\leq \big\|L_0^{n-1}(v) - L_0^{n-1}(w)\big\|_{X_\zeta \leftarrow X_\beta} \|z^*\|_{X_\beta} \\ &\quad + h \sum_{j=0}^{n-1} \big\|L_{j+1}^{n-1}(v) - L_{j+1}^{n-1}(w)\big\|_{X_\zeta \leftarrow X_\beta} \|s_{j+1}\|_{X_\beta} \\ &\leq C\Big(t_{n-1}^{\beta-\zeta} \|z^*\|_{X_\beta} + \sigma h \sum_{j=0}^{n-2} (t_{n-1} - t_{j+1})^{\beta-\zeta}\Big) \|v - w\|_{X_\zeta,\infty} \\ &\leq C t_n^{\beta-\zeta} \big(\|z^*\|_{X_\beta} + \sigma t_n\big) \|v - w\|_{X_\zeta,\infty}. \end{aligned}$$

If $\beta = 1$ an additional logarithmic term $(1 + |\log h|)$ arises. Thus, for $0 < t_n \leq T$ and $h > 0$ small enough the mapping $\mathcal{N}$ is contractive, that is, the estimate

$$\big\|\mathcal{N}(v) - \mathcal{N}(w)\big\|_{X_\zeta,\infty} \leq \kappa \|v - w\|_{X_\zeta,\infty}, \qquad v, w \in \mathcal{Z}_{z^*},$$

holds with constant $\kappa < 1$.

(ii) For any $z \in \mathcal{Z}_{z^*}$ and $s \in \mathcal{S}$ Lemma 4.4 and 4.5 imply

$$\begin{aligned}\big\|N_n(z,s) - N_m(z,s)\big\|_{X_\zeta} &\le \big\|L_0^{n-1}(z) - L_0^{m-1}(z)\big\|_{X_\zeta \leftarrow X_\beta} \|z^*\|_{X_\beta} \\ &\quad + h \sum_{j=0}^{m-1} \big\|L_{j+1}^{n-1}(z) - L_{j+1}^{m-1}(z)\big\|_{X_\zeta \leftarrow X_\beta} \|s_{j+1}\|_{X_\beta} \\ &\quad + h \sum_{j=m}^{n-1} \big\|L_{j+1}^{n-1}(z)\big\|_{X_\zeta \leftarrow X_\beta} \|s_{j+1}\|_{X_\beta} \\ &\le \widetilde{C}\big(1 + C t_{n-1}^\alpha\big)\big(\|z^*\|_{X_\beta}(t_n - t_m)^{\beta-\zeta} + \sigma t_{m-1}(t_n - t_m)^{\beta-\zeta} + \sigma(t_{n-1} - t_m)\big) \\ &\le \widetilde{C}\big(1 + C t_n^\alpha\big)(t_n - t_m)^{\beta-\zeta}\end{aligned}$$

with constant $\widetilde{C} > 0$ independent of the Hölder-constant $\widetilde{L}$. This relation shows that the constant $\widetilde{L} > 0$ can be chosen such that sequence $\mathcal{N}(z)$ belongs to $\mathcal{Z}$ for $T > 0$ sufficiently small. Again, if $\beta = 1$ an additional logarithmis factor appears in the estimate.

As a consequence, $\mathcal{N}$ is a contraction on $\mathcal{Z}_{z^*}$. Therefore, an application of the Banach contraction principle shows that $\mathcal{N}$ possesses a unique fixed-point $z \in \mathcal{Z}_{z^*}$. Consequently, for any $z^* \in X_\beta \cap V$ and $s \in \mathcal{S}$ the recursion (4.4) is solvable in the sequence space $\mathcal{Z}_{z^*}$.

As a further consequence, we obtain the stability estimate of the theorem. Assume that $v, w \in \mathcal{Z}$ and $p, q \in \mathcal{S}$ fulfill the identities

$$v = N(v,p), \qquad w = N(w,q).$$

The bound in (i) together with Lemma 4.4 shows

$$\begin{aligned}\|v - w\|_{X_\zeta,\infty} &= \big\|N(v,p) - N(w,q)\|_{X_\zeta,\infty} \\ &\le \big\|N(v,p) - N(w,p)\|_{X_\zeta,\infty} + \big\|N(w,p) - N(w,q)\|_{X_\zeta,\infty} \\ &\le \kappa \|v - w\|_{X_\zeta,\infty} + C\big(\|v_0 - w_0\|_{X_\beta} + \|p - q\|_{X_\beta,\infty}\big).\end{aligned}$$

Therefore, as $\kappa < 1$ we get the relation

$$\|v - w\|_{X_\zeta,\infty} \le C\big(\|v_0 - w_0\|_{X_\beta} + \|p - q\|_{X_\beta,\infty}\big).$$

Applying the above arguments together with the previous estimate finally proves the following bound in X_β

$$\begin{aligned}\|v - w\|_{X_\beta,\infty} &\le \big\|N(v,p) - N(w,p)\|_{X_\beta,\infty} + \big\|N(w,p) - N(w,q)\|_{X_\beta,\infty} \\ &\le C\|v - w\|_{X_\zeta,\infty} + C\big(\|v_0 - w_0\|_{X_\beta} + \|p - q\|_{X_\beta,\infty}\big) \\ &\le C\big(\|v_0 - w_0\|_{X_\beta} + \|p - q\|_{X_\beta,\infty}\big)\end{aligned}$$

which is the desired result.

We finally remark that the rather strong restrictions concerning the size of the end time $T > 0$ can be weakend by introducing exponential weights in the maximum norm. Alternatively, combining the stability and the convergence result given in Section 5 shows the validity of Theorem 4.1 on the whole interval of existence of the true solution $u : [0,T] \to X_\beta$ of (1.1). □

Remark 4.2. The analogue of Theorem 4.1 is valid for any Magnus type integrator of the form $u_{n+1} = \mathrm{e}^{hU_{n1}} u_n$ provided that the internal stages U_{n1} satisfy an estimate

of the form

$$\big\|U_{n1} - U_{m1}\big\|_{X_\gamma} \le C(t_n - t_m)^\alpha, \qquad 0 \le t_m \le t_n \le T,$$

see also Lemma 4.3.

4.2. **Auxiliary estimates.** In the sequel, we employ the assumptions and abbreviations introduced in the previous Subsection 4.1. In particular, as in the proof of Theorem 4.1, we choose $\gamma < \zeta < \beta$ and $0 < \alpha < \beta - \zeta$. In this subsection, we denote by $\widetilde{C} > 0$ a constant that only depends on the constants that appear in Hypothesis 2.1, but not on the Hölder-constant $\widetilde{L}$, see (4.2). Especially, the constants $\widetilde{C} > 0$ and $C > 0$ are independent of n and h.

At first, we show that for any sequence $z \in \mathcal{Z}$ the associated sequence $Z_1(z)$ reflects the Hölder-continuity of z, see also (4.2) and (4.3a). For the moment, as we consider a fixed sequence $z \in \mathcal{Z}$, we omit the dependence of Z_1 on z.

Lemma 4.3. *Assume that Hypothesis 2.1 holds with $\vartheta \ge 0$. Then, for any $z \in \mathcal{Z}$ the associated sequence $Z_1 = \big(Z_{n1}\big)_{n\ge 0}$ defined by* (4.3a) *satisfies the estimate*

$$\big\|Z_{n1} - Z_{m1}\big\|_{X_\gamma} \le C(t_n - t_m)^\alpha, \qquad 0 \le t_m \le t_n \le T,$$

with constant $C > 0$.

Proof. In order to estimate the difference $Z_{n1} - Z_{m1}$, we make use of the identity

$$Z_{n1} - Z_{m1} = \mathrm{e}^{h/2\,A(z_n)}(z_n - z_m) + \Big(\mathrm{e}^{h/2\,A(z_n)} - \mathrm{e}^{h/2\,A(z_m)}\Big) z_m.$$

Due to the fact that z lies in $\mathcal{Z}$, together with (2.6) it follows for $0 \le t_m \le t_n \le T$

$$\big\|\mathrm{e}^{h/2\,A(z_n)}(z_n - z_m)\big\|_{X_\gamma} \le \big\|\mathrm{e}^{h/2\,A(z_n)}\big\|_{X_\gamma \leftarrow X_\zeta} \|z_n - z_m\|_{X_\zeta} \le C(t_n - t_m)^\alpha.$$

On the other hand, let Γ be a path that surrounds the spectrum of the sectorial operators $A(z_n)$ and $A(z_m)$. Then, by means of the integral formula of Cauchy, we have the representation

$$\begin{aligned}\Big(\mathrm{e}^{h/2\,A(z_n)} - \mathrm{e}^{h/2\,A(z_m)}\Big) z_m = \frac{h}{\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda \big(\lambda I - h/2\,A(z_n)\big)^{-1} \\ \times \big(A(z_n) - A(z_m)\big)\big(\lambda I - h/2\,A(z_m)\big)^{-1} z_m \,\mathrm{d}\lambda,\end{aligned}$$

see also (2.5). We estimate this expression with the help of relation (2.3) and the resolvent bound (2.4). Note further that $\|z_n - z_m\|_{X_\gamma} \le K \|z_n - z_m\|_{X_\zeta}$ with some $K > 0$. As a consequence, we get the estimate

$$\begin{aligned}\Big\|\Big(\mathrm{e}^{h/2\,A(z_n)} - \mathrm{e}^{h/2\,A(z_m)}\Big) z_m\Big\|_{X_\gamma} &\le \frac{h}{\pi} \int_\Gamma |\mathrm{e}^\lambda| \Big\|\big(\lambda I - h/2\,A(z_n)\big)^{-1}\Big\|_{X_\gamma \leftarrow X} \\ &\quad \times \big\|A(z_n) - A(z_m)\big\|_{X \leftarrow D} \Big\|\big(\lambda I - h/2\,A(z_m)\big)^{-1}\Big\|_{D \leftarrow X_\zeta} \|z_m\|_{X_\zeta} \,|\mathrm{d}\lambda| \\ &\le C \|z_n - z_m\|_{X_\gamma} \|z_m\|_{X_\zeta} \le C(t_n - t_m)^\alpha.\end{aligned}$$

Altogether, this yields the desired result. □

As a direct consequence of (2.6) we obtain the following bound for the analytic semigroup generated by the sectorial operator $A(Z_{m1})$

$$\big\|\mathrm{e}^{(t_{n+1} - t_m) A(Z_{m1})}\big\|_{X_\nu \leftarrow X_\mu} \le M(t_{n+1} - t_m)^{-\nu + \mu}, \qquad 0 \le t_m \le t_n \le T, \tag{4.5}$$

whenever $0 \le \mu \le \nu \le 1$. Lemma 4.4 below shows that the corresponding estimate remains valid for $L = L(z)$, see (4.3b).

Lemma 4.4. *Suppose that Hypothesis 2.1 holds with $\vartheta > 0$. Then, for any $z \in \mathcal{Z}$ the associated linear operator family $L = \big(L_m^n\big)_{n \geq m \geq 0}$ defined by* (4.3b) *fulfills*

$$\big\|L_m^n\big\|_{X_\nu \leftarrow X_\mu} \leq \widetilde{C}\big(1 + C(t_{n+1} - t_m)^\alpha\big)(t_{n+1} - t_m)^{-\nu+\mu}, \qquad 0 \leq t_m \leq t_n \leq T,$$

for all $0 \leq \mu \leq \nu \leq 1$ provided that $\vartheta < \mu + \alpha$.

Proof. Our techniques for proving Lemma 4.4 are close to that applied in [11, 26]. The basic idea is to compare L_m^n with the frozen operator

$$\prod_{i=m}^{n} \mathrm{e}^{hA(Z_{m1})} = \mathrm{e}^{(t_{n+1}-t_m)A(Z_{m1})},$$

where the bound (4.5) is available. Thus, it remains to estimate the difference

$$\Delta_m^n = L_m^n - \mathrm{e}^{(t_{n+1}-t_m)A(Z_{m1})}, \qquad 0 \leq m < n. \tag{4.6a}$$

From a telescopic identity, we obtain the equality

$$\Delta_m^n = \sum_{j=m+1}^{n-1} \Delta_{j+1}^n \Xi_{jm} + \sum_{j=m+1}^{n} \mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\, \Xi_{jm} \tag{4.6b}$$

which involves the linear operator

$$\Xi_{jm} = \Big(\mathrm{e}^{hA(Z_{j1})} - \mathrm{e}^{hA(Z_{m1})}\Big)\mathrm{e}^{(t_j-t_m)A(Z_{m1})}, \qquad j > m.$$

The integral formula of Cauchy yields the following representation, where the path Γ is chosen in such a way that it surrounds the spectrum of the sectorial operators $A(Z_{j1})$ and $A(Z_{m1})$, see also (2.5) and the proof of Lemma 4.3

$$\begin{aligned}\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\, \Xi_{jm} = \frac{h}{2\pi\mathrm{i}} \int_\Gamma \mathrm{e}^\lambda\, \mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})} \big(\lambda I - hA(Z_{j1})\big)^{-1} \\ \times \big(A(Z_{j1}) - A(Z_{m1})\big)\big(\lambda I - hA(Z_{m1})\big)^{-1} \mathrm{e}^{(t_j-t_m)A(Z_{m1})}\, \mathrm{d}\lambda.\end{aligned}$$

We estimate this expression by applying the resolvent bound (2.4) and further (2.6). Due to relation (2.3) and Lemma 4.3, for $m < j < n$ we finally get

$$\begin{aligned}\big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\, \Xi_{jm}\big\|_{X_\nu \leftarrow X_\mu} &\leq Ch(t_{n+1} - t_{j+1})^{-\nu+\vartheta}(t_j - t_m)^{-1-\vartheta+\mu} \\ &\quad \times \big\|Z_{j1} - Z_{m1}\big\|_{X_\gamma} \\ &\leq Ch(t_{n+1} - t_{j+1})^{-\nu+\vartheta}(t_j - t_m)^{-1-\vartheta+\mu+\alpha}, \qquad m < j < n.\end{aligned}$$

Moreover, it follows

$$\big\|\Xi_{nm}\big\|_{X_\nu \leftarrow X_\mu} \leq Ch^{1-\nu+\vartheta}(t_n - t_m)^{-1-\vartheta+\mu+\alpha}.$$

Thus, by interpreting the last sum in (4.6b) as a Riemann-sum and estimating it by the associated integral, we have

$$\sum_{j=m+1}^{n-1} \big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\Xi_{jm}\big\|_{X_\nu \leftarrow X_\mu} \leq C(t_{n+1} - t_m)^{-\nu+\mu+\alpha}$$

provided that $\vartheta < \mu + \alpha$. Furthermore, we make use of the relation

$$\big\|\Xi_{jm}\big\|_{X_\mu \leftarrow X_\mu} \leq Ch^{1-\mu+\vartheta}(t_j - t_m)^{-1-\vartheta+\mu+\alpha}, \qquad j > m.$$

First, we estimate Δ_m^n as operator from X_ϑ to X_ν. With the help of the above relations, due to the fact that for $j > m$ and $n > m$ it holds

$$\big\|\Xi_{jm}\big\|_{X_\vartheta \leftarrow X_\vartheta} \le Ch(t_j - t_m)^{-1+\alpha}, \qquad \big\|\Xi_{nm}\big\|_{X_\nu \leftarrow X_\vartheta} \le Ch^{1-\nu+\vartheta}(t_n - t_m)^{-1+\alpha},$$
$$\sum_{j=m+1}^{n-1} \big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\Xi_{jm}\big\|_{X_\nu \leftarrow X_\vartheta} \le C(t_{n+1} - t_m)^{-\nu+\vartheta+\alpha},$$

we obtain the following bound

$$\begin{aligned}\big\|\Delta_m^n\big\|_{X_\nu \leftarrow X_\vartheta} &\le \sum_{j=m+1}^{n-1} \big\|\Delta_{j+1}^n\big\|_{X_\nu \leftarrow X_\vartheta}\big\|\Xi_{jm}\big\|_{X_\vartheta \leftarrow X_\vartheta} + \big\|\Xi_{nm}\big\|_{X_\nu \leftarrow X_\vartheta} \\ &\quad + \sum_{j=m+1}^{n-1} \big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\Xi_{jm}\big\|_{X_\nu \leftarrow X_\vartheta} \\ &\le Ch \sum_{j=m+1}^{n-1} \big\|\Delta_{j+1}^n\big\|_{X_\nu \leftarrow X_\vartheta}(t_j - t_m)^{-1+\alpha} + C(t_{n+1} - t_m)^{-\nu+\vartheta+\alpha}.\end{aligned}$$

Thus, from a Gronwall-type inequality with a weakly singular kernel, see [8, 24], e.g., it follows

$$\big\|\Delta_m^n\big\|_{X_\nu \leftarrow X_\vartheta} \le Ch(t_{n+1} - t_m)^{-\nu+\vartheta+\alpha}$$

with constant $C > 0$ possibly depending on T. Now, it is straightforward to estimate Δ_m^n as operator from X_μ to X_ν

$$\begin{aligned}\big\|\Delta_m^n\big\|_{X_\nu \leftarrow X_\mu} &\le \sum_{j=m+1}^{n-1} \big\|\Delta_{j+1}^n\big\|_{X_\nu \leftarrow X_\vartheta}\big\|\Xi_{jm}\big\|_{X_\vartheta \leftarrow X_\mu} + \big\|\Xi_{nm}\big\|_{X_\nu \leftarrow X_\mu} \\ &\quad + \sum_{j=m+1}^{n-1} \big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A(Z_{m1})}\Xi_{jm}\big\|_{X_\nu \leftarrow X_\mu} \\ &\le Ch \sum_{j=m+1}^{n-1} (t_{n+1} - t_j)^{-\nu+\vartheta+\alpha}(t_j - t_m)^{-1-\vartheta+\mu+\alpha} + C(t_{n+1} - t_m)^{-\nu+\mu+\alpha},\end{aligned}$$

wherefore we finally have

$$\big\|\Delta_m^n\big\|_{X_\nu \leftarrow X_\mu} \le C(t_{n+1} - t_m)^{-\nu+\mu+\alpha}. \tag{4.7}$$

Together with (4.5) this yields the desired result. □

Now, with the help of Lemma 4.4 we are in the position to show that L is Hölder-continuous.

Lemma 4.5. *In the situation of Lemma 4.4, for $z \in \mathcal{Z}$ the associated operator family $L = \big(L_m^n\big)_{n \ge m \ge 0}$ satisfies the following estimates*

$$\big\|L_j^n(z) - L_j^m\big(z\big)\|_{X_\nu \leftarrow X_\mu} \le \widetilde{C}\big(1 + Ct_n^\alpha\big)(t_n - t_m)^{-\nu+\mu}, \qquad \nu \ne 1,$$
$$\big\|L_j^n(z) - L_j^m\big(z\big)\|_{D \leftarrow X_\mu} \le \widetilde{C}\big(1 + Ct_n^\alpha\big)\big(1 + |\log h|\big)(t_n - t_m)^{-1+\mu},$$

where $0 \le \mu, \nu \le 1$ and $0 \le t_j \le t_m < t_n \le T$.

Proof. From a telescopic identity, for $j \le m < n$ we obtain

$$L_j^n - L_j^m = \big(L_{m+1}^n - I\big)L_j^m = \sum_{i=m+1}^{n} L_{i+1}^n \Big(\mathrm{e}^{hA(Z_{i1})} - I\Big) L_j^m, \qquad j \le m < n.$$

We note that the relations in (2.8) imply

$$\big\|\mathrm{e}^{hA(Z_{i1})} - I\big\|_{X \leftarrow D} = \big\|hA(Z_{i1})\varphi\big(hA(Z_{i1})\big)\big\|_{X \leftarrow D} \le Mh, \qquad 0 \le t_i \le T,$$

see also [25], e.g. Further, it holds

$$\big\|\mathrm{e}^{hA(Z_{i1})} - I\big\|_{X_\nu \leftarrow D} \le Mh^{1-\nu}, \qquad 0 \le t_i \le T.$$

Thus, by applying the bound from Lemma 4.4, for any $0 \le \mu, \nu \le 1$ we get

$$\begin{aligned}\big\|L_j^n - L_j^m\big\|_{X_\nu \leftarrow X_\mu} \le \sum_{i=m+1}^{n-1} &\big\|L_{i+1}^n\big\|_{X_\nu \leftarrow X} \big\|\mathrm{e}^{hA(Z_{i1})} - I\big\|_{X \leftarrow D} \big\|L_j^m\big\|_{D \leftarrow X_\mu} \\ &+ \big\|\mathrm{e}^{hA(Z_{i1})} - I\big\|_{X_\nu \leftarrow D} \big\|L_j^m\big\|_{D \leftarrow X_\mu} \\ &\le \widetilde{C}\big(1 + Ct_n^\alpha\big)\, h \sum_{i=m+1}^{n-1} (t_{n+1} - t_{i+1})^{-\nu} (t_{m+1} - t_j)^{-1+\mu}.\end{aligned}$$

Therefore, interpreting the sum as a Riemann-sum and estimating it by the associated integral yields for $\nu \ne 1$

$$\begin{aligned}\big\|L_j^n - L_j^m\big\|_{X_\nu \leftarrow X_\mu} &\le \widetilde{C}\big(1 + Ct_n^\alpha\big)(t_{n+1} - t_{m+1})^{1-\nu}(t_{m+1} - t_j)^{-1+\mu} \\ &\le \widetilde{C}\big(1 + Ct_n^\alpha\big)(t_n - t_m)^{-\nu+\mu}\Big(\frac{n-m}{m+1-j}\Big)^{1-\mu} \\ &\le \widetilde{C}\big(1 + Ct_n^\alpha\big)(t_n - t_m)^{-\nu+\mu}\end{aligned}$$

which proves the desired result. If $\nu = 1$ the additional term $\big(1 + |\log h|\big)$ arises in the estimate. □

In Lemma 4.6 we study the dependence of the operators $L_m^n(z)$ on z. For that purpose, for $v = (v_n)_{n\ge0}$ and $w = (w_n)_{n\ge0}$ in $\mathcal{Z}$ we denote by $\|v - w\|_{X_\varsigma,\infty}$ the maximum value of $\|v_n - w_n\|_{X_\varsigma}$ for $0 \le nh \le T$, see also (2.10).

Lemma 4.6. *Suppose that Hypothesis 2.1 is satisfied with $\vartheta > 0$. Then, for sequences $v = (v_n)_{n\ge0} \in \mathcal{Z}$ and $w = (w_n)_{n\ge0} \in \mathcal{Z}$ the following estimates are valid for arbitrary $0 \le \mu \le \nu \le 1$ and $0 \le t_m < t_n \le T$. If $\nu \ne 1$ and $\mu \ne 0$ it follows*

$$\big\|L_m^n(v) - L_m^n(w)\big\|_{X_\nu \leftarrow X_\mu} \le C(t_n - t_m)^{-\nu+\mu}\|v - w\|_{X_\varsigma,\infty},$$

else if $\nu = 1$ or $\mu = 0$ the bound

$$\big\|L_m^n(v) - L_m^n(w)\big\|_{X_\nu \leftarrow X_\mu} \le C(t_n - t_m)^{-\nu+\mu}\big(1 + |\log h|\big)\|v - w\|_{X_\varsigma,\infty}$$

holds.

Proof. For $v = (v_n)_{n\ge0} \in \mathcal{Z}$ and $w = (w_n)_{n\ge0} \in \mathcal{Z}$ we define the associated sequences $V_1 = \big(V_{n1}\big)_{n\ge0}$ and $W_1 = \big(W_{n1}\big)_{n\ge0}$ accordingly to (4.3a). An application of the telescopic identity yields

$$L_m^n(v) - L_m^n(w) = \sum_{j=m}^{n} L_{j+1}^n(v)\Big(\mathrm{e}^{hA(V_{j1})} - \mathrm{e}^{hA(W_{j1})}\Big) L_m^{j-1}(w),$$

see also the proof of the previous Lemma 4.5. We estimate $L_m^n(v) - L_m^n(w)$ as operator from X_μ to X_ν

$$\begin{aligned}\big\|L_m^n(v) - L_m^n(w)\big\|_{X_\nu \leftarrow X_\mu} &\le \big\|L_{m+1}^n(v)\big\|_{X_\nu \leftarrow X} \big\|\mathrm{e}^{hA(V_{m1})} - \mathrm{e}^{hA(W_{m1})}\big\|_{X \leftarrow X_\mu} \\ &+ \sum_{j=m+1}^{n-1} \big\|L_{j+1}^n(v)\big\|_{X_\nu \leftarrow X} \big\|\mathrm{e}^{hA(V_{j1})} - \mathrm{e}^{hA(W_{j1})}\big\|_{X \leftarrow D} \big\|L_m^{j-1}(w)\big\|_{D \leftarrow X_\mu} \\ &+ \big\|\mathrm{e}^{hA(V_{n1})} - \mathrm{e}^{hA(W_{n1})}\big\|_{X_\nu \leftarrow D} \big\|L_m^{n-1}(w)\big\|_{D \leftarrow X_\mu}.\end{aligned}$$

By the integral formula of Cauchy, we have the representation

$$\begin{aligned}\mathrm{e}^{hA(V_{j1})} - \mathrm{e}^{hA(W_{j1})} = \frac{h}{\pi \mathrm{i}} \int_\Gamma \mathrm{e}^\lambda \big(\lambda I - hA(V_{j1})\big)^{-1} \\ \times \big(A(V_{j1}) - A(W_{j1})\big) \big(\lambda I - h\,A(W_{j1})\big)^{-1} \,\mathrm{d}\lambda,\end{aligned}$$

see also (2.5). Consequently, with the help of (2.4) and (2.6) we have

$$\begin{aligned}\big\|\mathrm{e}^{hA(V_{j1})} - \mathrm{e}^{hA(W_{j1})}\big\|_{X \leftarrow X_\mu} &\le Ch^\mu \big\|A(V_{j1}) - A(W_{j1})\big\|_{X \leftarrow D}, \qquad 0 \le \mu \le 1, \\ \big\|\mathrm{e}^{hA(V_{j1})} - \mathrm{e}^{hA(W_{j1})}\big\|_{X_\nu \leftarrow D} &\le Ch^{1-\nu} \big\|A(V_{j1}) - A(W_{j1})\big\|_{X \leftarrow D}, \qquad 0 \le \nu \le 1.\end{aligned}$$

Hypothesis 2.1 and similar considerations as in the proof of Lemma 4.3 yield the bound

$$\begin{aligned}\big\|A(V_{n1}) - A(W_{n1})\big\|_{X \leftarrow D} &\le L \big\|V_{n1} - W_{n1}\big\|_{X_\gamma} \\ &= L \big\|\mathrm{e}^{h/2\,A(v_n)} v_n - \mathrm{e}^{h/2\,A(w_n)} w_n\big\|_{X_\gamma} \le C \|v_n - w_n\|_{X_\zeta}.\end{aligned}$$

As a consequence, by means of Lemma 4.4 we finally have

$$\big\|L_m^n(v) - L_m^n(w)\big\|_{X_\nu \leftarrow X_\mu} \le C(t_n - t_m)^{-\nu+\mu} \|v - w\|_{X_\zeta, \infty}.$$

Here, an additional logarithmic factor arises if $\nu = 1$ or $\mu = 0$. This proves the desired result. □

5. Convergence

In this section, we state a convergence result for the Magnus type integrator (3.1) applied to the quasilinear problem (1.1). Our proof relies on a favourable relation for the global error which we derive first.

5.1. **Relation for error.** For the subsequent considerations, we employ the abbreviations introduced before in Sections 2 and 3. In particular, for a constant stepsize $h > 0$ we let $t_n = nh$ and $t_{n1} = t_n + h/2$ and set $A_n = A(u_n)$ and $A_{n1} = A\big(U_{n1}\big)$ for $n \ge 0$. Furthermore, we define

$$\varphi_{n1} = \varphi\big(hA_{n1}\big), \qquad \psi_{n1} = \psi\big(hA_{n1}\big), \qquad \chi_{n1} = \chi\big(hA_{n1}\big), \qquad \psi_n = \big(h/2\,A_n\big),$$

see also (2.8a). Besides, it is convenient to denote the exact solution values by

$$\widehat{u}_{n+1} = u(t_{n+1}), \qquad \widehat{U}_{n1} = u(t_{n1}), \qquad \widehat{A}_n = A\big(\widehat{u}_n\big), \qquad \widehat{A}_{n1} = A\big(\widehat{U}_{n1}\big).$$

Then, the global error of the numerical approximation and the internal stage, respectively, equals

$$e_{n+1} = u_{n+1} - \widehat{u}_{n+1}, \qquad E_{n1} = U_{n1} - \widehat{U}_{n1}, \qquad n \ge 0.$$

Moreover, the discrete evolution operator is given by

$$\mathcal{E}_m^n = \prod_{i=m}^{n} \mathrm{e}^{hA_{i1}}, \qquad 0 \le m \le n. \tag{5.1}$$

In addition, we set $\mathcal{E}_m^n = I$ if $n < m$.

In order to represent the global error e_{n+1} in a suitable way, we consider the differential equation (1.1) on the subinterval $[t_n, t_{n+1}]$ and rewrite the right-hand side by adding and substracting A_{n1}

$$u'(t) = A_{n1}u(t) + g_n(t), \qquad g_n(t) = \big(A\big(u(t)\big) - A_{n1}\big)u(t).$$

Thus, with the help of the variation-of-constants formula, a relation similar to the second formula in (3.1) involving further the defect of the method follows

$$\widehat{u}_{n+1} = \mathrm{e}^{hA_{n1}}\widehat{u}_n + d_{n+1}, \qquad d_{n+1} = \int_0^h \mathrm{e}^{(h-\tau)A_{n1}} g_n(t_n+\tau)\,\mathrm{d}\tau. \tag{5.2}$$

By taking the difference of (3.1) and (5.2) and resolving the resulting recursion for e_{n+1}, we finally obtain

$$e_{n+1} = \mathcal{E}_0^n e_0 - \sum_{j=0}^{n} \mathcal{E}_{j+1}^n d_{j+1}. \tag{5.3}$$

For deriving a useful relation for the defects, we decompose g_n as follows

$$g_n(t) = f_n(t) + \big(\widehat{A}_{n1} - A_{n1}\big)u(t), \qquad f_n(t) = \big(A\big(u(t)\big) - \widehat{A}_{n1}\big)u(t). \tag{5.4}$$

Provided that the map A and the exact solution u satisfy suitable regularity assumptions, a Taylor series expansion of $f_n : [t_n, t_{n+1}] \to X$ yields

$$f_n(t_n+\tau) = \big(\tau - h/2\big) f_n'\big(t_{n1}\big) + \big(\tau - h/2\big)^2 \int_0^1 (1-\sigma)\, f_n''\big(t_{n1} + \sigma\big(\tau - h/2\big)\big)\,\mathrm{d}\sigma,$$

and, moreover, the following identity is valid

$$A_{n1} - \widehat{A}_{n1} = \mathcal{A}_{n1}E_{n1}, \qquad \mathcal{A}_{n1} = \int_0^1 A'\big(\sigma\, U_{n1} + (1-\sigma)\widehat{U}_{n1}\big)\,\mathrm{d}\sigma,$$

with $A'(v) : V \to L(D, X)$ denoting the Fréchet derivative of A at $v \in V$. Consequently, by integrating accordingly to (5.2) and applying (2.8a), the defects split up into $d_{n+1} = \delta_{n+1} + \theta_{n+1} = \delta_{n+1}^{(0)} + \delta_{n+1}^{(1)} + \theta_{n+1}$ where

$$\delta_{n+1}^{(0)} = h^2\big(\psi_{n1} - 1/2\,\varphi_{n1}\big) f_n'\big(t_{n1}\big) = h^3 A_{n1}\chi_{n1} f_n'\big(t_{n1}\big), \tag{5.5a}$$

$$\delta_{n+1}^{(1)} = \int_0^h \mathrm{e}^{(h-\tau)A_{n1}} \big(\tau - h/2\big)^2 \int_0^1 (1-\sigma)\, f_n''\big(t_{n1} + \sigma\big(\tau - h/2\big)\big)\,\mathrm{d}\sigma\,\mathrm{d}\tau, \tag{5.5b}$$

$$\theta_{n+1} = -\int_0^h \mathrm{e}^{(h-\tau)A_{n1}} \mathcal{A}_{n1}E_{n1}u(t_n+\tau)\mathrm{d}\tau. \tag{5.5c}$$

As the term θ_{n+1} involves the error of the internal stage, we next derive a suitable relation for E_{n1}. Rewriting again the right-hand side of (1.1)

$$u'(t) = A_n u(t) + G_n(t), \qquad G_n(t) = \big(A\big(u(t)\big) - A_n\big)u(t),$$

by the variation-of-constants formula, we obtain the representation

$$\widehat{U}_{n1} = \mathrm{e}^{h/2\,A_n}\widehat{u}_n + D_{n1}, \qquad D_{n1} = \int_0^{h/2} \mathrm{e}^{(h/2-\tau)A_n} G_n(t_n+\tau)\,\mathrm{d}\tau, \tag{5.6}$$

and, together with the first formula in (3.1) this implies

$$E_{n1} = \mathrm{e}^{h/2\,A_n} e_n - D_{n1}. \tag{5.7}$$

Similarly to before, we employ a decomposition of G_n

$$G_n(t) = F_n(t) + \big(\widehat{A}_n - A_n\big)u(t), \qquad F_n(t) = \big(A\big(u(t)\big) - \widehat{A}_n\big)u(t), \tag{5.8}$$

and thus obtain from a Taylor series expansion the identity

$$F_n(t_n + \tau) = \tau F_n'(t_n) + \tau^2 \int_0^1 (1-\sigma)\, F_n''(t_n + \sigma\tau)\,\mathrm{d}\sigma,$$

and, on the other hand, the relation

$$A_n - \widehat{A}_n = \mathcal{A}_n e_n, \qquad \mathcal{A}_n = \int_0^1 A'\big(\sigma\, u_n + (1-\sigma)\widehat{u}_n\big)\,\mathrm{d}\sigma,$$

follows. Consequently, determining the integral in (5.6) with the help of (2.8a), yields the splitting $D_{n1} = \Delta_{n1} + \Theta_{n1} = \Delta_{n1}^{(0)} + \Delta_{n1}^{(1)} + \Theta_{n1}$ for the defect of the internal stage where

$$\Delta_{n1}^{(0)} = h^2/4\,\psi_n F_n'(t_n), \tag{5.9a}$$

$$\Delta_{n1}^{(1)} = \int_0^{h/2} \mathrm{e}^{(h/2-\widetilde{\tau})A_n} \widetilde{\tau}^2 \int_0^1 (1-\sigma)\, F_n''\big(t_n + \sigma\widetilde{\tau}\big)\,\mathrm{d}\sigma\,\mathrm{d}\widetilde{\tau}, \tag{5.9b}$$

$$\Theta_{n1} = -\int_0^{h/2} \mathrm{e}^{(h/2-\widetilde{\tau})A_n} \mathcal{A}_n e_n u\big(t_n + \widetilde{\tau}\big)\,\mathrm{d}\widetilde{\tau}. \tag{5.9c}$$

Finally, we expand relation (5.3) by inserting successively formula (5.5c) for the defect $d_{n+1} = \delta_{n+1} + \theta_{n+1}$, formula (5.7), and further (5.9c) for $D_{n1} = \Delta_{n1} + \Theta_{n1}$. Altogether, we have the following representation for the global error

$$\begin{aligned} e_n = \mathcal{E}_0^{n-1} e_0 &+ \sum_{j=0}^{n-1} \mathcal{E}_{j+1}^{n-1} \int_0^h \mathrm{e}^{(h-\tau)A_{j1}} \mathcal{A}_{j1} \\ &\times \int_0^{h/2} \mathrm{e}^{(h/2-\widetilde{\tau})A_j} \mathcal{A}_j\, e_j\, u\big(t_j + \widetilde{\tau}\big)\,\mathrm{d}\widetilde{\tau}\, u\big(t_j + \widetilde{\tau}\big)\,\mathrm{d}\widetilde{\tau}\, u(t_j + \tau)\,\mathrm{d}\tau \\ &+ \sum_{j=0}^{n-1} \mathcal{E}_{j+1}^{n-1} \int_0^h \mathrm{e}^{(h-\tau)A_{j1}} \mathcal{A}_{j1} \mathrm{e}^{h/2\,A_j} e_j\, u(t_j + \tau)\,\mathrm{d}\tau \\ &- \sum_{j=0}^{n-1} \mathcal{E}_{j+1}^{n-1} \delta_{j+1} - \sum_{j=0}^{n-1} \mathcal{E}_{j+1}^{n-1} \int_0^h \mathrm{e}^{(h-\tau)A_{j1}} \mathcal{A}_{j1} \Delta_{j1} u(t_j + \tau)\,\mathrm{d}\tau, \end{aligned} \tag{5.10}$$

where the defects $\delta_{j+1} = \delta_{j+1}^{(0)} + \delta_{j+1}^{(1)}$ and $\Delta_{j1} = \Delta_{j1}^{(0)} + \Delta_{j1}^{(1)}$ are defined through the formulas (5.5a)-(5.5b) and (5.9a)-(5.9b).

5.2. **Error estimate.** We next analyse the error behaviour of the Magnus type integrator (3.1) for the quasilinear parabolic problem (1.1) and state a convergence estimate with respect to the norm of the intermediate space X_β where $\gamma < \beta < 1$.

For the derivation of Theorem 5.1, our main tools are the global representation (5.10) as well as the stability estimate of Theorem 4.1. In order to obtain the optimal convergence order, we further make use of a refined stability bound specified in Lemma 5.2 at the end of this subsection. Regarding the error estimate it is notable that the differentiability of the functions f_n and F_n introduced

in (5.4) and (5.8) is governed by the smoothness of the exact solution u and the operator family A, that is, the requirement that the first derivatives of f_n and F_n are bounded in X_ϑ for a certain $\vartheta > 0$ is satisfied in various applications, see also Subsection 6.2. We finally note that the restriction $\beta < 1$ is senseful in view of Remark 3.1, however, the statement of Theorem 5.1 remains valid for the limiting case $\beta = 1$.

In the sequel, for maps $g : [0,T] \to X$ and $G_j : [t_j, t_{j+1}] \to X$ defined for integers $j \geq 0$ we employ the abbreviations

$$\big\|g\big\|_{X,\infty} = \max_{0\leq t\leq t_n} \big\|g(t)\big\|_X, \qquad \big\|G\big\|_{X,\infty} = \max_{0\leq j\leq n-1} \big\|G_j\big\|_{X,\infty},$$

where $\big\|G_j\big\|_{X,\infty} = \max\big\{\big\|G_j(t)\big\|_X : t_j \leq t \leq t_{j+1}\big\}$, see also (2.10).

Theorem 5.1 (Convergence)**.** *Suppose that Hypothesis 2.1 is fulfilled for constants $0 < \vartheta \leq \gamma < 1$ and choose $\gamma < \beta < 1$. Assume further that the exact solution of (1.1) is bounded in $X_{1+\vartheta}$ and that $A'(v) : V \to L(X_{1+\vartheta}, X_\vartheta)$ is bounded for every $v \in V$. Besides, we require $u : [0,T] \to X_\beta$ to be Lipschitz-continuous with respect to t. Then, for $h > 0$ chosen sufficiently small the numerical method (3.1) applied to the abstract initial value problem (1.1) satisfies the convergence estimate*

$$\begin{aligned}\big\|u_n - u(t_n)\big\|_{X_\beta} \leq{}& C\|u_0 - u(0)\|_{X_\beta},\\ &+ Ch^{2-\beta+\vartheta}\Big(\big(1+|\log h\,|\big)\big\|f'\big\|_{X_\vartheta,\infty} + \big\|F'\big\|_{X_\vartheta,\infty}\Big)\\ &+ Ch^2\Big(\big\|f''\big\|_{X,\infty} + h^{1-\beta}\big\|F''\big\|_{X,\infty}\Big), \qquad 0 \leq t_n \leq T,\end{aligned}$$

provided that the quantities on the right-hand side are well-defined. The constant $C > 0$ is independent of n and h.

Proof. We note that the existence of the numerical solution in X_β is ensured by Theorem 4.1. Thus, it remains to derive the desired convergence bound. For this purpose, we consider relation (5.10) for the global error e_n and estimate it in X_β. On the one hand, for the error terms involving the initial values and e_j, $0 \leq j \leq n-1$, we thus obtain the bound

$$\begin{aligned}\big\|e_n^{(0)}\big\|_{X_\beta} \leq{}& \big\|\mathcal{E}_0^{n-1}\big\|_{X_\beta\leftarrow X_\beta}\|e_0\|_{X_\beta}\\ &+ \sum_{j=0}^{n-1}\int_0^h\int_0^{h/2} \big\|\mathcal{E}_{j+1}^{n-1}\mathrm{e}^{(h-\tau)A_{j1}}\big\|_{X_\beta\leftarrow X_\vartheta}\big\|\mathcal{A}_{j1}\big\|_{L\left(X_{1+\vartheta},X_\vartheta\right)\leftarrow X_\gamma}\\ &\qquad\times\big\|\mathrm{e}^{(h/2-\widetilde{\tau})A_j}\big\|_{X_\gamma\leftarrow X_\vartheta}\big\|\mathcal{A}_j\big\|_{L\left(X_{1+\vartheta},X_\vartheta\right)\leftarrow X_\gamma}\big\|e_j\big\|_{X_\beta}\\ &\qquad\quad\times\big\|u\big(t_j+\widetilde{\tau}\big)\big\|_{X_{1+\vartheta}}\big\|u(t_j+\tau)\big\|_{X_{1+\vartheta}}\mathrm{d}\widetilde{\tau}\,\mathrm{d}\tau\\ &+ \sum_{j=0}^{n-1}\int_0^h \big\|\mathcal{E}_{j+1}^{n-1}\mathrm{e}^{(h-\tau)A_{j1}}\big\|_{X_\beta\leftarrow X_\vartheta}\big\|\mathcal{A}_{j1}\big\|_{L\left(X_{1+\vartheta},X_\vartheta\right)\leftarrow X_\gamma}\\ &\qquad\times\big\|\mathrm{e}^{h/2\,A_j}\big\|_{X_\gamma\leftarrow X_\beta}\big\|e_j\big\|_{X_\beta}\big\|u(t_j+\tau)\big\|_{X_{1+\vartheta}}\mathrm{d}\tau.\end{aligned}$$

On the other hand, inserting the representation (5.5a) for the defects $\delta_{n+1}^{(0)}$ yields the following estimate for the remaining terms

$$\begin{aligned}\big\|e_n^{(1)}\big\|_{X_\beta} &\le h^3 \sum_{j=0}^{n-2} \big\|\mathcal{E}_{j+1}^{n-1} A_{j1}\chi_{j1}\big\|_{X_\beta \leftarrow X_\vartheta} \big\|f_j'\big(t_{j1}\big)\big\|_{X_\vartheta} \\ &\quad + h^2\big(\big\|\psi_{n-1,1}\big\|_{X_\beta \leftarrow X_\vartheta} + 1/2\,\big\|\varphi_{n-1,1}\big\|_{X_\beta \leftarrow X_\vartheta}\big)\big\|f_{n-1}'\big(t_{n-1,1}\big)\big\|_{X_\vartheta} \\ &\qquad + \sum_{j=0}^{n-2} \big\|\mathcal{E}_{j+1}^{n-1}\big\|_{X_\beta \leftarrow X} \big\|\delta_{j+1}^{(1)}\big\|_X + \big\|\delta_n^{(1)}\big\|_{X_\beta} \\ &\qquad\quad + \sum_{j=0}^{n-1} \int_0^h \big\|\mathcal{E}_{j+1}^{n-1} \mathrm{e}^{(h-\tau)A_{j1}}\big\|_{X_\beta \leftarrow X_\vartheta} \big\|\mathcal{A}_{j1}\big\|_{L\left(X_{1+\vartheta}, X_\vartheta\right) \leftarrow X_\gamma} \\ &\qquad\qquad \times \big\|\Delta_{j1}\big\|_{X_\beta} \big\|u(t_j+\tau)\big\|_{X_{1+\vartheta}} \mathrm{d}\tau.\end{aligned}$$

We next apply the bounds for the analytic semigroup and the related operators, see (2.6) and (2.8b), as well as the stability bounds of Lemma 5.2. Note further that for any $0 \le \mu \le 1$ the relation

$$\begin{aligned}\big\|\delta_{n+1}^{(1)}\big\|_{X_\mu} &\le \int_0^h \int_0^1 \big|\tau - h/2\big|^2 \big\|\mathrm{e}^{(h-\tau)A_{n1}}\big\|_{X_\mu \leftarrow X} \\ &\qquad \times \big\|f_n''\big(t_{n1} + \sigma\big(\tau - h/2\big)\big)\big\|_X \,\mathrm{d}\sigma\,\mathrm{d}\tau \le Ch^{3-\mu}\big\|f''\big\|_{X,\infty}\end{aligned}$$

holds, and that we moreover have

$$\begin{aligned}\big\|\Delta_{n1}\big\|_{X_\beta} &\le h^2 \big\|\psi_n\big\|_{X_\beta \leftarrow X_\vartheta} \big\|F_n'(t_n)\big\|_{X_\vartheta} \\ &\qquad + \int_0^{h/2} \int_0^1 \widetilde{\tau}^2 \big\|\mathrm{e}^{(h/2-\widetilde{\tau})A_n}\big\|_{X_\beta \leftarrow X} \big\|F_n''\big(t_n + \sigma\widetilde{\tau}\big)\big\|_{X,\infty} \mathrm{d}\sigma\,\mathrm{d}\widetilde{\tau} \\ &\le Ch^{2-\beta+\vartheta}\big\|F'\big\|_{X_\vartheta,\infty} + Ch^{3-\beta}\big\|F''\big\|_{X,\infty}.\end{aligned}$$

Therefore, under the assumptions of the theorem it follows

$$\begin{aligned}\big\|e_n\big\|_{X_\beta} &\le \big\|e_n^{(0)}\big\|_{X_\beta} + \big\|e_n^{(1)}\big\|_{X_\beta} \\ &\le C\|e_0\|_{X_\beta} + Ch \sum_{j=0}^{n-1} (t_n - t_j)^{-\beta+\vartheta} \big\|e_j\big\|_{X_\beta} \\ &\qquad + C\Big(h^{1+\alpha}\big(1 + |\log h|\big)\big\|f'\big\|_{X_\vartheta,\infty} + h^{2-\beta+\vartheta}\,\big\|F'\big\|_{X_\vartheta,\infty} \\ &\qquad\qquad + h^2\,\big\|f''\big\|_{X,\infty} + h^{3-\beta}\,\big\|F''\big\|_{X,\infty}\Big)\, h \sum_{j=0}^{n-2} (t_n - t_{j+1})^{-\beta+\vartheta} \\ &\qquad + Ch^{2-\beta+\vartheta}\, h \sum_{j=0}^{n-2} (t_n - t_{j+1})^{-1} \big\|f'\big\|_{X_\vartheta,\infty}.\end{aligned}$$

As a consequence, by interpreting the sums as Riemann-sums and estimating them by the corresponding integrals we get

$$\begin{aligned}\big\|e_n\big\|_{X_\beta} &\le C\|e_0\|_{X_\beta} + Ch\sum_{j=0}^{n-1}(t_n-t_j)^{-\beta+\vartheta}\big\|e_j\big\|_{X_\beta}\\ &\quad + C\min\big\{h^{1+\alpha}, h^{2-\beta+\vartheta}\big\}\big(1+|\log h|\big)\big\|f'\big\|_{X_\vartheta,\infty}\\ &\quad + Ch^{2-\beta+\vartheta}\big\|F'\big\|_{X_\vartheta,\infty} + Ch^2\big\|f''\big\|_{X,\infty} + Ch^{3-\beta}\big\|F''\big\|_{X,\infty}.\end{aligned}$$

Finally, the application of a Gronwall lemma shows

$$\begin{aligned}\big\|e_n\big\|_{X_\beta} &\le C\|e_0\|_{X_\beta} + C\min\big\{h^{1+\alpha}, h^{2-\beta+\vartheta}\big\}\big(1+|\log h|\big)\big\|f'\big\|_{X_\vartheta,\infty}\\ &\quad + Ch^{2-\beta+\vartheta}\big\|F'\big\|_{X_\vartheta,\infty} + Ch^2\big\|f''\big\|_{X,\infty} + Ch^{3-\beta}\big\|F''\big\|_{X,\infty},\end{aligned}\tag{5.11}$$

see also the proof of Lemma 4.4.

We note that the exponent α in the bound (5.11) as it is restricted by the condition $0<\alpha<\beta-\zeta$ with $\gamma<\zeta<\beta$ possibly is close to 0. However, regarding the numerical experiments of Section 6 it is essential to raise the size of α. For that purpose, let u denote the exact solution of (1.1) started at the numerical initial value $u_0\in X_\beta$ and assume that it is Lipschitz-continuous on X_β, i.e.

$$\big\|u(t_n)-u(t_m)\big\|_{X_\beta} \le C(t_n-t_m).$$

In particular, this relation holds true if the first derivative u' is bounded in X_β. Consequently, due to the convergence estimate (5.11) which implies that the order of the numerical scheme in X_β is at least one, we have

$$\begin{aligned}\big\|u_n-u_m\big\|_{X_\beta} &\le \big\|u_n-u(t_n)\big\|_{X_\beta} + \big\|u_m-u(t_m)\big\|_{X_\beta} + \big\|u(t_n)-u(t_m)\big\|_{X_\beta}\\ &\le Ch + C(t_n-t_m) \le C(t_n-t_m), \qquad 0\le t_m\le t_n\le T.\end{aligned}$$

Altogether, these considerations show that we may set $\alpha=1$ in (5.11) which proves the desired result. □

For the proof of the above convergence estimate, the following stability result is needed. Recall the abbreviation $\chi_{m1}=\chi\big(hA_{m1}\big)$.

Lemma 5.2. *Assume that Hypothesis 2.1 is valid with $\vartheta>0$. Then, the discrete evolution operator $\mathcal{E}_m^n$ defined in* (5.1) *fulfills the estimates*

$$\begin{aligned}\big\|\mathcal{E}_m^n\big\|_{X_\beta\leftarrow X_\beta} + \big\|(t_{n+1}-t_m)^{\beta-\vartheta}\mathcal{E}_m^n\big\|_{X_\beta\leftarrow X_\vartheta} &\le C,\\ \big\|\mathcal{E}_m^n A_{m1}\chi_{m1}\big\|_{X_\beta\leftarrow X_\vartheta} &\le Ch^{-1+\alpha}\big(1+|\log h|\big)(t_{n+1}-t_m)^{-\beta+\vartheta}\\ &\quad + Ch^{-\beta+\vartheta}(t_{n+1}-t_m)^{-1}, \qquad 0\le t_m\le t_n\le T,\end{aligned}$$

with constant $C>0$ not depending on n and h.

Proof. The first estimate of Lemma 5.2 is a direct consequence of Lemma 4.4. For proving the second bound, we correlate the discrete evolution operator with the analytic semigroup generated by A_{m1}. That is, similarly as in the proof of Lemma 4.4, we make use of the identity

$$\begin{aligned}\mathcal{E}_m^n A_{m1}\chi_{m1} &= \Delta_m^n A_{m1}\,\chi_{m1} + A_{m1}\,\mathrm{e}^{(t_{n+1}-t_m)A_{m1}}\,\chi_{m1}\\ &= \sum_{j=m+1}^{n-1}\Delta_{j+1}^n\,\widetilde{\Xi}_{jm} + \sum_{j=m+1}^{n}\mathrm{e}^{(t_{n+1}-t_{j+1})A_{m1}}\,\widetilde{\Xi}_{jm} + A_{m1}\,\mathrm{e}^{(t_{n+1}-t_m)A_{m1}}\,\chi_{m1},\end{aligned}$$

where $\Delta_m^n = \mathcal{E}_m^n - \mathrm{e}^{(t_{n+1}-t_m)A_{m1}}$ and

$$\widetilde{\Xi}_{jm} = \big(\mathrm{e}^{hA_{j1}} - \mathrm{e}^{hA_{m1}}\big) A_{m1}\, \mathrm{e}^{(t_j-t_m)A_{m1}} \chi_{m1}, \qquad j > m.$$

By means of the integral formula of Cauchy, we obtain the relations

$$\big\|\widetilde{\Xi}_{jm}\big\|_{X_\vartheta \leftarrow X_\vartheta} \le Ch(t_j - t_m)^{-2+\alpha}, \qquad \big\|\widetilde{\Xi}_{nm}\big\|_{X_\beta \leftarrow X_\vartheta} \le Ch^{1-\beta+\vartheta}(t_n - t_m)^{-2+\alpha}.$$

Consequently, with the help of the estimate

$$\big\|A_{m1}\, \mathrm{e}^{(t_{n+1}-t_m)A_{m1}} \chi_{m1}\big\|_{X_\beta \leftarrow X_\vartheta} \le M(t_{n+1} - t_m)^{-1-\beta+\vartheta},$$

see also (4.5) and (2.8b), and (4.7) we obtain

$$\begin{aligned}
\big\|\mathcal{E}_m^n A_{m1}\chi_{m1}\big\|_{X_\beta \leftarrow X_\vartheta} &\le \sum_{j=m+1}^{n-1} \big\|\Delta_{j+1}^n\big\|_{X_\beta \leftarrow X_\vartheta} \big\|\widetilde{\Xi}_{jm}\big\|_{X_\vartheta \leftarrow X_\vartheta} \\
&\quad + \sum_{j=m+1}^{n-1} \big\|\mathrm{e}^{(t_{n+1}-t_{j+1})A_{m1}}\big\|_{X_\beta \leftarrow X_\vartheta} \big\|\widetilde{\Xi}_{jm}\big\|_{X_\vartheta \leftarrow X_\vartheta} + \big\|\widetilde{\Xi}_{nm}\big\|_{X_\beta \leftarrow X_\vartheta} \\
&\quad + \big\|A_{m1}\, \mathrm{e}^{(t_{n+1}-t_m)A_{m1}}\, \chi_{m1}\big\|_{X_\beta \leftarrow X_\vartheta} \\
&\le Ch \sum_{j=m+1}^{n-1} (t_{n+1} - t_{j+1})^{-\beta+\vartheta+\alpha}(t_j - t_m)^{-2+\alpha} + C(t_{n+1} - t_m)^{-1-\beta+\vartheta} \\
&\le Ch^{-1+\alpha}\big(1 + |\log h|\big)(t_{n+1} - t_m)^{-\beta+\vartheta} + C(t_{n+1} - t_m)^{-1-\beta+\vartheta}
\end{aligned}$$

which yields the specified estimate. □

6. Extension and numerical example

In this section, we discuss a possible extension of the Magnus type integrator (3.1) to quasilinear equations with an additional inhomogeneity and illustrate the theoretical convergence result by a numerical example. Throughout, we employ the hypotheses and notation introduced in Sections 2-5.

6.1. **Extension to inhomogeneous quasilinear problems.** The convergence analysis of the previous Section 5 easily generalises to problems with an additional inhomogeneous part. In view of our numerical example, we consider an abstract initial value problem of the form

$$u'(t) = A\big(u(t)\big)u(t) + b(t), \quad 0 < t \le T, \qquad u(0) \text{ given}, \tag{6.1}$$

involving a time-dependent map $b : [0,T] \to X$. In this case, the numerical method (3.1) for the quasilinear parabolic equation (1.1) is modified as follows

$$\begin{aligned}
U_{n1} &= \mathrm{e}^{h/2\, A_n} u_n + h/2\, \varphi\big(h/2\, A_n\big) b_n, \qquad b_n = b(t_n), \\
u_{n+1} &= \mathrm{e}^{hA_{n1}} u_n + h\varphi\big(hA_{n1}\big) b_{n1}, \qquad b_{n1} = b(t_{n1}), \qquad n \ge 0,
\end{aligned} \tag{6.2}$$

see (2.8a). Similar considerations as in Section 5 show that the following convergence result is valid with maps $\widetilde{f}_n$ and $\widetilde{F}_n$ defined by

$$\widetilde{f}_n(t) = f_n(t) + b(t) - b_{n1}, \qquad \widetilde{F}_n(t) = F_n(t) + b(t) - b_n, \qquad n \ge 0, \tag{6.3}$$

provided that first and second derivatives of b are bounded in certain intermediate spaces, see also (5.4) and (5.8).

Theorem 6.1 (Convergence)**.** *Assume that the requirements of Theorem 5.1 are satisfied. Then, for $h > 0$ chosen sufficiently small the numerical method* (6.2) *applied to the abstract initial value problem* (6.1) *fulfills the convergence bound*

$$\begin{aligned}\big\|u_n - u(t_n)\big\|_{X_\beta} &\le C\|u_0 - u(0)\|_{X_\beta} \\ &\quad + Ch^{2-\beta+\vartheta}\Big(\big(1+|\log h\,|\big)\big\|\widetilde{f}'\big\|_{X_\vartheta,\infty} + \big\|\widetilde{F}'\big\|_{X_\vartheta,\infty}\Big) \\ &\quad + Ch^2\Big(\big\|\widetilde{f}''\big\|_{X,\infty} + h^{1-\beta}\big\|\widetilde{F}''\big\|_{X,\infty}\Big), \qquad 0 \le t_n \le T,\end{aligned}$$

with constant $C > 0$ not depending on n and h.

6.2. **Numerical example.** The following application illustrates the above convergence result. In order to keep the realisation simple, we restrict ourselves to a parabolic initial-boundary value problem in one space dimension.

Example 6.2. We consider a one-dimensional initial-boundary value problem for a function $U : [0,1]^2 \to \mathbb{R} : (x,t) \to U(x,t)$ comprising a quasilinear partial differential equation with additional inhogoneneous part

$$\partial_t U(x,t) = \mathcal{A}\big(U(x,t)\big)U(x,t) + B(x,t), \qquad 0 < x \le 1, \quad 0 < t \le 1, \tag{6.4a}$$

subject to a homogeneous Dirichlet boundary condition and an initial condition

$$U(0,t) = 0 = U(1,t), \quad 0 \le t \le 1, \qquad U(x,0) = U_0(x), \quad 0 \le x \le 1. \tag{6.4b}$$

For functions $v \in C^1(0,1)$ and $w \in C^2(0,1)$ the differential operator $\mathcal{A}$ is given by

$$\mathcal{A}\big(v(x)\big)w(x) = a\big(x, v(x), \partial_x v(x)\big)\partial_x^2 w(x), \qquad 0 < x \le 1, \tag{6.4c}$$

with coefficient $a : \mathbb{R}^3 \to \mathbb{R}$ satisfying suitable regularity and boundedness assumptions. Specifically, for the numerical example we set

$$a(x,p,q) = 1 + p^2 + c\,q^2, \qquad c = 0, 1, \tag{6.4d}$$

and determine the function B and the initial condition U_0 such that the exact solution of (6.4) is given by $U(x,t) = \mathrm{e}^{-t}x(1-x)$. Note that U fulfills the homogeneous Dirichlet boundary condition.

We let $\big(u(t)\big)(x) = U(x,t)$, $\big(A(v)w\big)(x) = \mathcal{A}\big(v(x)\big)w(x)$, and $\big(b(t)\big)(x) = B(x,t)$. With this notation, the initial-boundary value problem (6.4) takes the form of an initial value problem (6.1) on the Banach space $X = L^p(\Omega)$ for $1 < p < \infty$ with domain of $A(v)$ given by the function space $D = W^{2,p}(0,1) \cap W_0^{1,p}(0,1)$. From the previous Example 2.3 we thus conclude that the linear operator family $A : X_\gamma \to L(X_{1+\vartheta}, X_\vartheta)$ satisfies Hypothesis 2.1 with $\vartheta = 0$ and constant γ restricted by the condition $c/2 + (2p)^{-1} < \gamma < 1$. Furthermore, due to the fact that the domain of $A(v)^2$ equals

$$D^2 = D\big(A(v)^2\big) = \Big\{w \in W^{4,p}(0,1) \cap W_0^{1,p}(0,1) : \partial_x^2 w(x)\big|_{x=0,1} = 0\Big\}$$

and therefore does not depend on $v \in V$, the same holds true for any intermediate space $D \subset X_{1+\vartheta} \subset D\big(A(v)^2\big)$. Besides, for $A : X_\gamma \to L\big(D^2, D\big)$ is Lipschitz-continuous with respect to v. As a consequence, Hypothesis 2.1 remains valid for every $0 \le \vartheta \le 1$.

In the present situation, all requirements of Theorem 5.1 are fulfilled. Namely, the exact solution $U(x,t)$ and the data $a(x,p,q)$ and $B(x,t)$ are sufficiently regular. Therefore, the maps $\widetilde{f}_n$ and $\widetilde{F}_n$ defined in (6.3) are twice differentiable in X, and, besides, the Fréchet derivative $A'(v) : X_\gamma \to L(X_{1+\vartheta}, X_\vartheta)$ is bounded. A result

h\M	50	100	150
2^{-2}	1.8988	1.8987	1.8986
2^{-3}	1.9021	1.9018	1.9017
2^{-4}	1.8965	1.8959	1.8957
2^{-5}	1.9078	1.9067	1.9064
2^{-6}	1.9184	1.9163	1.9159
2^{-7}	1.9291	1.9252	1.9244
2^{-8}	1.9409	1.9333	1.9319
2^{-9}	1.9553	1.9415	1.9388
2^{-10}	1.9728	1.9508	1.9457

h\M	50	100	150
2^{-2}	1.3462	1.3335	1.3293
2^{-3}	1.2770	1.2621	1.2572
2^{-4}	1.2987	1.2760	1.2686
2^{-5}	1.3185	1.2847	1.2738
2^{-6}	1.3480	1.2977	1.2817
2^{-7}	1.3947	1.3181	1.2946
2^{-8}	1.4679	1.3495	1.3141
2^{-9}	1.5817	1.3977	1.3437
2^{-10}	1.7389	1.4730	1.3889

TABLE 1. Numerically observed temporal convergence order in the discrete X_β-norm for $c = 0$, $p = 2$, $\beta = (2p)^{-1} = 1/4$ (left), $\beta = 1$ (right). Expected values $\kappa_{1/4} \approx 2$, $\kappa_1 \approx 1 + 1/4$, see (6.5b).

in Grisvard [14] which characterises the intermediate spaces $X \subset X_\vartheta \subset D$ implies that any function which is spatially smooth but does not satisfy further boundary conditions belongs to X_ϑ as long as $\vartheta < (2p)^{-1}$, see also the discussion in [11]. That is, the first derivatives of $\widetilde{f}_n$ and $\widetilde{F}_n$ are bounded in X_ϑ for $\vartheta < (2p)^{-1}$. Moreover, the exact solution of (6.4) lies in the intermediate space $X_{1+\vartheta}$ if $\vartheta < (2p)^{-1}$ and its first time derivative $\partial_t U(x,t) = -U(x,t)$ remains bounded in X_β for arbitrary $0 \le \beta \le 1$. As a consequence, accordingly to Theorem 5.1, the expected convergence order with respect to the norm of the Sobolev-space X_β is

$$\kappa_\beta = 2 - \beta + \vartheta, \qquad c/2 + (2p)^{-1} < \gamma < 1, \qquad \vartheta < (2p)^{-1}, \tag{6.5a}$$

where $\gamma < \beta < 1$.

For the numerical example, the partial differential equation is discretised in space by symmetric finite differences of grid length $\Delta x = (M+2)^{-1}$, and, for the time integration, we apply the numerical method (6.2) with stepsize $h > 0$. The numerical temporal order of convergence measured in the discrete X_β-norm is determined

h\M	50	100	150
2^{-2}	2.0180	2.0180	2.0180
2^{-3}	2.0465	2.0464	2.0463
2^{-4}	1.9818	1.9813	1.9812
2^{-5}	1.9827	1.9819	1.9817
2^{-6}	1.9859	1.9843	1.9840
2^{-7}	1.9910	1.9880	1.9874
2^{-8}	1.9968	1.9920	1.9909
2^{-9}	2.0001	1.9965	1.9943
2^{-10}	2.0137	2.0012	1.9978

h\M	50	100	150
2^{-2}	1.0854	1.0661	1.0601
2^{-3}	1.0752	1.0492	1.0408
2^{-4}	1.0895	1.0504	1.0375
2^{-5}	1.1184	1.0616	1.0429
2^{-6}	1.1662	1.0831	1.0560
2^{-7}	1.2396	1.1169	1.0775
2^{-8}	1.3500	1.1678	1.1101
2^{-9}	1.5106	1.2439	1.1584
2^{-10}	1.7118	1.3572	1.2302

TABLE 2. Numerically observed temporal convergence order in the discrete X_β-norm for $c = 0$, $p = 100$, $\beta = (2p)^{-1} = 1/200$ (left), $\beta = 1$ (right). Expected values $\kappa_{1/200} \approx 2$, $\kappa_1 \approx 1 + 1/200$, see (6.5b).

h\M	50	100	150
2^{-2}	1.5985	1.5955	1.5948
2^{-3}	1.4579	1.4533	1.4523
2^{-4}	1.4644	1.4568	1.4550
2^{-5}	1.4922	1.4788	1.4756
2^{-6}	1.5154	1.4920	1.4863
2^{-7}	1.5474	1.5067	1.4968
2^{-8}	1.5963	1.5263	1.5090
2^{-9}	1.6737	1.5560	1.5261
2^{-10}	1.7854	1.6040	1.5528

h\M	50	100	150
2^{-2}	1.2614	1.2482	1.2438
2^{-3}	1.2056	1.1915	1.1868
2^{-4}	1.2529	1.2315	1.2244
2^{-5}	1.2864	1.2546	1.2443
2^{-6}	1.3222	1.2748	1.2599
2^{-7}	1.3712	1.2995	1.2775
2^{-8}	1.4432	1.3326	1.2997
2^{-9}	1.5524	1.3802	1.3301
2^{-10}	1.7069	1.4520	1.3741

TABLE 3. Numerically observed temporal convergence order in the discrete X_β-norm for $c = 1$, $p = 2$, $\beta = 1/2 + (2p)^{-1} = 3/4$ (left), $\beta = 1$ (right). Expected values $\kappa_{3/4} \approx 1 + 1/2$, $\kappa_1 \approx 1 + 1/4$, see (6.5c).

from the numerical and exact solution values. In particular, if the differential operator involves no first derivative, i.e., $c = 0$ in (6.4d), for the limiting cases $\beta = (2p)^{-1}$ and $\beta = 1$ we expect a numerical convergence order of approximately

$$\kappa_{(2p)^{-1}} = 2 - (2p)^{-1} + \vartheta \approx 2, \qquad \kappa_1 = 1 + \vartheta \approx 1 + (2p)^{-1}, \tag{6.5b}$$

see (6.5a). On the other hand, for the case where $c = 1$ we have

$$\kappa_{1/2+(2p)^{-1}} = 1 + 1/2 - (2p)^{-1} + \vartheta \approx 1 + 1/2, \qquad \kappa_1 = 1 + \vartheta \approx 1 + (2p)^{-1}. \tag{6.5c}$$

The results of the numerical experiment for $p = 2$ and $p = 100$ are displayed in Tables 1-4. The observed numbers are in good agreement with the expected values. We remark that for the chosen values of M and h the problem becomes non-stiff as the temporal stepsize h tends to 2^{-10}, wherefore the numerical order approaches the classical convergence order 2.

h\M	50	100	150
2^{-2}	1.6447	1.6441	1.6440
2^{-3}	1.4681	1.4669	1.4667
2^{-4}	1.4697	1.4677	1.4673
2^{-5}	1.4835	1.4791	1.4784
2^{-6}	1.4858	1.4865	1.4849
2^{-7}	1.5559	1.4946	1.4904
2^{-8}	1.4690	1.4935	1.4951
2^{-9}	1.5942	1.5602	1.4983
2^{-10}	1.7540	1.4754	1.5582

h\M	50	100	150
2^{-2}	0.9801	0.9598	0.9535
2^{-3}	1.0091	0.9838	0.9757
2^{-4}	1.0474	1.0102	0.9979
2^{-5}	1.0875	1.0338	1.0161
2^{-6}	1.1397	1.0613	1.0358
2^{-7}	1.2132	1.0979	1.0609
2^{-8}	1.3199	1.1488	1.0947
2^{-9}	1.4742	1.2222	1.1423
2^{-10}	1.6743	1.3298	1.2109

TABLE 4. Numerically observed temporal convergence order in the discrete X_β-norm for $c = 1$, $p = 100$, $\beta = 1/2 + (2p)^{-1} = 1/2 + 1/200$ (left), $\beta = 1$ (right). Expected values $\kappa_{1/2+1/200} \approx 1 + 1/2$, $\kappa_1 \approx 1 + 1/200$, see (6.5c).

Acknowledgement

The authors are grateful to Alexander Ostermann for several discussions related to this work and valuable comments on the manuscript. The work of Mechthild Thalhammer was supported by Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under project H210-N13. The research stays of Mechthild Thalhammer in Valladolid in June 2004 and of Césareo González in Innsbruck in October 2004 were supported by DGI-MCYT under grant BFM2001-2013/2138 co-financed by FEDER funds and by JCYL under grant VA112/02.

References

1. H. Amann, *Quasilinear evolution equations and parabolic systems.* Trans. Amer. Math. Soc. 293 (1986) 191-227.
2. H. Amann, *Dynamic theory of quasilinear parabolic equations - I. Abstract evolution equations.* Nonlin. Anal. Th. Meth. and Appl. 12 (1988) 895-919.
3. H. Amann, *Highly degenerate quasilinear parabolic systems.* Ann. Scuola Norm. Sup. Pisa Ser. IV 18 (1991) 135-166.
4. H. Amann, *Linear and Quasilinear Parabolic Problems. Vol. I Abstract Theory.* In: Monographs in Mathematics Vol. 89, Birkhäuser, Basel, 1995.
5. N. Aparicio, S. Malham, and M. Oliver, *Numerical evaluation of the Evans function by Magnus integration* (2004). To appear in BIT.
6. J. Bergh and J. Löfström, *Interpolation Spaces. An Introduction.* Springer, Berlin, 1976.
7. S. Blanes, F. Casas, and J. Ros, *Improved high order integrators based on the Magnus expansion.* BIT 40 (2000) 434-450.
8. H. Brunnner and P.J. van der Houwen, *The numerical solution of Volterra equations.* CWI Monographs 3. North-Holland, Amsterdam, 1986.
9. PH. Clément, C. J. van Duijn, and Shuanhu Li, *On a nonlinear elliptic-parabolic partial differential equation system in a two-dimensional groundwater flow problem.* SIAM J. Math. Anal. 23/4 (1992) 836-851.
10. J. Van den Eshof and M. Hochbruck, *Preconditioning Lanczos approximations to the matrix exponential* (2004). To appear in SIAM J. Sci. Comp.
11. C. González, A. Ostermann, and M. Thalhammer, *A second-order Magnus integrator for non-autonomous parabolic problems* (2004). J. Comp. Appl. Math. (in press).
12. C. González and C. Palencia, *Stability of time-stepping methods for time-dependent parabolic problems: the Hölder case.* Math. Comp. 68 (1999) 73-89.
13. C. González and C. Palencia, *Stability of Runge-Kutta methods for quasilinear parabolic problems.* Math. Comp. 69 (2000) 609-628.
14. P. Grisvard, *Caractérisation de quelques espaces d'interpolation.* Arch. Rat. Mech. Anal. 25 (1967) 40-63.
15. D. Henry, *Geometric Theory of Semilinear Parabolic Equations.* Lecture Notes in Mathematics 840, Springer, Berlin, 1981.
16. M. Hochbruck and Ch. Lubich, *On Krylov subspace approximations to the matrix exponential operator.* SIAM J. Numer. Anal. 34 (1997) 1911-1925.
17. M. Hochbruck and Ch. Lubich, *On Magnus integrators for time-dependent Schrödinger equations.* SIAM J. Numer. Anal. 41 (2003) 945-963.
18. A. Iserles and S.P. Nørsett, *On the solution of linear differential equations in Lie groups.* Phil. Trans. R. Soc. Lond. A 357 (1999) 983-1019.
19. Ch. Lubich, *Integrators for Quantum Dynamics: A Numerical Analyst's Brief Review.* In: Quantum Simulations of Complex Many-Body Systems: From Theory to Algorithms, Lecture Notes, J. Grotendorst, D. Mary, A. Muramatsu (Eds.), John von Neumann Institute for Computing, Jülich, NIC Series 10 (2002) 459-466.
20. A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems.* Birkhäuser, Basel, 1995.
21. W. Magnus, *On the exponential solution of a differential equation for a linear operator.* Comm. Pure Appl. Math. 7 (1954) 649-673.

22. C. Moler and Ch. Van Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*. SIAM Rev. 45 no. 1 (2003) 3-49.
23. H. Munthe-Kaas and B. Owren, *Computations in a free Lie algebra*. Phil. Trans. R. Soc. Lond. A 357 (1999) 957-981.
24. A. Ostermann and M. Thalhammer, *Non-smooth data error estimates for linearly implicit Runge-Kutta methods*. IMA J. Numer. Anal. 20 (2000) 167-184.
25. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York, 1983.
26. M. Thalhammer, *A second-order Magnus type integrator for non-autonomous semilinear parabolic problems* (2004). Submitted to IMA J. Numer. Anal.
27. M. Thalhammer, *A fourth-order commutator-free exponential integrator for non-autonomous differential equations* (2005). Submitted to SIAM J. Numer. Anal.
28. H. Triebel, *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam, 1978.
29. J. Wensch, M. Däne, W. Hergert, and A. Ernst, *The solution of stationary ODE problems in quantum mechanics by Magnus methods with stepsize control*. Comp. Phys. Comm. 160 (2004) 129-139.

Mailing address: Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, E-47011 Valladolid, Spain.
E-mail address: cesareo@mac.cie.uva.es

Mailing address: Institut für Mathematik, Fakultät für Mathematik, Informatik und Physik, Universität Innsbruck, Technikerstrasse 25/7, A-6020 Innsbruck, Austria.
E-mail address: Mechthild.Thalhammer@uibk.ac.at

2.3. Commutator-free integrators for non-autonomous problems

A fourth-order commutator-free exponential integrator for non-autonomous differential equations

MECHTHILD THALHAMMER

To appear in SIAM Journal on Numerical Analysis

A FOURTH-ORDER COMMUTATOR-FREE EXPONENTIAL INTEGRATOR FOR NON-AUTONOMOUS DIFFERENTIAL EQUATIONS

MECHTHILD THALHAMMER*

Abstract. In the present work, we study the convergence behaviour of commutator-free exponential integrators for abstract non-autonomous evolution equations

$$u'(t) = A(t)u(t), \qquad 0 < t \leq T.$$

In particular, we focus on a fourth-order scheme that relies on the composition of two exponentials involving the values of the linear operator family A at the Gaussian nodes

$$u_1 = \mathrm{e}^{h(a_2A_1+a_1A_2)}\,\mathrm{e}^{h(a_1A_1+a_2A_2)}\,u_0, \quad a_i = \tfrac{1}{4} \pm \tfrac{\sqrt{3}}{6}, \quad c_i = \tfrac{1}{2} \mp \tfrac{\sqrt{3}}{6}, \quad A_i = A(c_ih), \quad i = 1, 2.$$

We prove that the numerical scheme is stable and derive an error estimate with respect to the norm of the underlying Banach space. The theoretically expected order reduction is illustrated by a numerical example for a parabolic initial-boundary value problem subject to a homogeneous Dirichlet boundary condition.

Key words. Exponential integrators, commutator-free methods, non-autonomous differential equations, parabolic evolution equations, stability, convergence

AMS subject classifications. 65L05, 65M12, 65J10

1. Introduction. In the present paper, we consider a non-autonomous differential equation involving a time-dependent linear operator A

$$u'(t) = A(t)u(t), \quad 0 < t \leq T, \qquad u(0) \text{ given.} \tag{1.1}$$

Our setting includes parabolic initial-boundary value problems that take the form (1.1) when written as an abstract initial value problem on a Banach space. The objective of this work is to analyse the error behaviour of the fourth-order commutator-free exponential integrator

$$\begin{aligned} u_1 &= \mathrm{e}^{h(a_2A_1+a_1A_2)}\,\mathrm{e}^{h(a_1A_1+a_2A_2)}\,u_0, \\ a_i = \tfrac{1}{4} \pm \tfrac{\sqrt{3}}{6}, &\quad c_i = \tfrac{1}{2} \mp \tfrac{\sqrt{3}}{6}, \quad A_i = A(c_ih), \qquad i = 1, 2, \end{aligned} \tag{1.2}$$

to explain the substantial order reduction for problems of parabolic type. For that purpose, we derive a representation for the defect of (1.2) which remains valid within the framework of sectorial operators and analytic semigroups. In situations, where $A(t)$ is a bounded linear operator, the Campbell-Baker-Hausdorff formula is a powerful tool for the error analysis of (1.2) and higher order schemes, respectively. However, it is problematic to justify its validity in the context of parabolic evolution equations. Therefore, in this paper, we follow a different approach based on the variation-of-constants formula.

Numerical schemes that involve the evaluation of the exponential and related functions were proposed in the middle of the past century already. For a historical review, see [24]. At present, a variety of works confirms the renewed interest in such exponential integrators. As a small selection, we mention the recent works [5, 8, 14, 16, 19, 20] and refer to the references given therein. A reason for these research

*INSTITUT FÜR MATHEMATIK, UNIVERSITÄT INNSBRUCK, TECHNIKERSTRASSE 13, 6020 INNSBRUCK, AUSTRIA. MECHTHILD.THALHAMMER@UIBK.AC.AT

activities are advances in the computation of the product of a matrix exponential with a vector, see for instance [10, 15, 25]. As a consequence, numerical integrators based on the Magnus expansion [23] and related method classes [2, 3, 6, 7, 17, 21] are practicable in the numerical solution of non-autonomous stiff differential equations, see also [11, 12, 30] and references cited therein.

The excellent error behaviour of interpolatory Magnus integrators for time-dependent Schrödinger equations is explained in Hochbruck and Lubich [14]. There, it is proven that the exponential midpoint rule applied to ordinary differential equations (1.1)

$$u_1 = \mathrm{e}^{hA_1}\, u_0, \qquad A_1 = A\big(\tfrac{h}{2}\big), \tag{1.3}$$

is convergent of order 2 without any restriction on the size of $h\big\|A(t)\big\|$. Moreover, under a mild stepsize restriction, a fourth-order error bound is valid for the Magnus integrator

$$u_1 = \mathrm{e}^{ha_1(A_1+A_2)+h^2a_2[A_2,A_1]}\, u_0,$$
$$a_1 = \tfrac{1}{2}, \quad a_2 = \tfrac{\sqrt{3}}{12}, \quad c_i = \tfrac{1}{2} \mp \tfrac{\sqrt{3}}{6}, \quad A_i = A(c_i h), \qquad i = 1, 2,$$

where $\big[A_1, A_2\big] = A_1A_2 - A_2A_1$ denotes the matrix commutator. In [11], we considered the numerical scheme (1.3) in the context of parabolic evolution equations and showed that the full convergence order 2 is obtained when the error is measured in the norm of the underlying Banach space, provided that the data and the exact solution of (1.1) are sufficiently smooth in time.

The purpose of the present work is to investigate the convergence properties of higher-order methods for linear non-autonomous parabolic problems (1.1). Provided that the time-dependent sectorial operator $A(t)$ is Hölder-continuous with respect to t, it is ensured that any linear operator defined through $B = \alpha A(\xi_1) + (1-\alpha)A(\xi_2)$ with $\alpha, \xi_1, \xi_2 \in \mathbb{R}$ generates an analytic semigroup $\big(\mathrm{e}^{tB}\big)_{t\geq 0}$, that is, numerical schemes such as (1.2) remain well-defined for abstract evolution equations (1.1). For that reason, we focus on commutator-free exponential integrators that rely on the composition of exponentials involving linear combinations of values of A. We show that the fourth-order scheme (1.2) is stable, however, unless the operator familiy A fulfills unnatural requirements, a substantial order reduction is encountered. For instance, for one-dimensional initial-boundary value problems subject to a homogeneous Dirichlet boundary condition, the order of convergence with respect to a discrete L^p-norm is $2 + \kappa$ where $0 \leq \kappa < (2p)^{-1}$, in general.

The present work is organised as follows. In Section 2, we first state the fundamental hypotheses on the non-autonomous evolution equation (1.1). The considered commutator-free exponential integration scheme is then introduced in Section 3. The numerical approximation is based on the composition of two exponentials that involve the values of A at certain nodal points. Sections 4 and 5 are concerned with a stability and convergence analysis for parabolic problems. In Section 5.1, we derive an expansion of the numerical solution defect which remains well-defined for abstract differential equations (1.1) involving an unbounded linear operator $A(t)$, provided that the data and the exact solution of the problem are sufficiently differentiable with respect to time. The main result, a convergence estimate for the fourth-order scheme (1.2) is given in Section 5.2. Important tools for its proof are the stability bound and the representation of the defect derived before. Section 6 is finally devoted to a numerical example that illustrates the expected order reduction.

2. Parabolic problems. Henceforth, we denote by $(X, \|\cdot\|_X)$ the underlying Banach space. Our basic requirements on the unbounded linear operator family A defining the right-hand side of the differential equation in (1.1) are that of [11, 30]. For a detailed treatise of time-dependent evolution equations we refer to [22, 29]. The monographs [13, 27] delve into the theory of sectorial operators and analytic semigroups.

HYPOTHESIS 1. *We assume that the densly defined and closed linear operator $A(t) : D \subset X \to X$ is uniformly sectorial for $0 \le t \le T$. Thus, there exist constants $a \in \mathbb{R}$, $0 < \phi < \frac{\pi}{2}$, and $M > 0$ such that for all $0 \le t \le T$ the resolvent of $A(t)$ satisfies the condition*

$$\Big\| \big(\lambda I - A(t)\big)^{-1} \Big\|_{X \leftarrow X} \le \frac{M}{|\lambda - a|} \tag{2.1}$$

for any complex number $\lambda \notin S_\phi(a) = \big\{\lambda \in \mathbb{C} : |\arg(a - \lambda)| \le \phi\big\} \cup \{a\}$. The graph norm of $A(t)$ and the norm in D fulfill the following relation with a constant $K > 0$

$$K^{-1}\|x\|_D \le \|x\|_X + \big\|A(t)x\big\|_X \le K\|x\|_D, \qquad x \in D, \qquad 0 \le t \le T.$$

Moreover, it holds $A \in \mathscr{C}^\vartheta\big([0,T], L(D,X)\big)$ for some $0 < \vartheta \le 1$, i.e., the bound

$$\big\|A(t) - A(s)\big\|_{X \leftarrow D} \le L(t-s)^\vartheta, \qquad 0 \le s \le t \le T, \tag{2.2}$$

is valid with a constant $L > 0$.

For any $0 \le s \le T$ the sectorial operator $\Omega = A(s)$ generates an analytic semigroup $\big(\mathrm{e}^{t\Omega}\big)_{t \ge 0}$ on X which is defined by means of the integral formula of Cauchy

$$\mathrm{e}^{t\Omega} = \frac{1}{2\pi\mathrm{i}} \int_\Gamma \mathrm{e}^\lambda \big(\lambda I - t\Omega\big)^{-1} \mathrm{d}\lambda, \quad t > 0, \qquad \mathrm{e}^{t\Omega} = I, \quad t = 0. \tag{2.3}$$

Here, Γ denotes a path that surrounds the spectrum of Ω.

Henceforth, for $0 < \mu < 1$, we denote by X_μ some intermediate space between the Banach spaces $D = X_1$ and $X = X_0$ such that the norm in X_μ satisfies the bound $\|x\|_{X_\mu} \le K\|x\|_D^\mu \|x\|_X^{1-\mu}$ with a constant $K > 0$ for all elements $x \in D$. Examples for intermediate spaces are real interpolation spaces, see Lunardi [22], or fractional power spaces, see Henry [13]. Then, for all $0 \le \mu \le \nu \le 1$ and integers $k \ge 0$ the following bound is valid

$$\big\|t^{\nu-\mu}\,\mathrm{e}^{t\Omega}\big\|_{X_\nu \leftarrow X_\mu} + \big\|t^{k+\nu-\mu}\,\Omega^k\,\mathrm{e}^{t\Omega}\big\|_{X_\nu \leftarrow X_\mu} \le M, \qquad 0 \le t \le T, \tag{2.4}$$

with a constant $M > 0$. As a consequence, the linear operator φ_m which is given by

$$\varphi_m(t\Omega) = \frac{1}{(m-1)!\,t^m} \int_0^t \mathrm{e}^{(t-\tau)\Omega}\,\tau^{m-1}\mathrm{d}\tau, \quad t > 0, \qquad \varphi_m(0) = \frac{1}{(m-1)!}\,I, \tag{2.5a}$$

for integers $m \ge 1$, remains bounded on X_μ for any $0 \le t \le T$ and $0 \le \mu \le 1$. In the subsequent sections, we make use of the identities

$$\mathrm{e}^{t\Omega} = I + t\Omega\,\varphi_1(t\Omega), \qquad \varphi_{m-1}(t\Omega) = \frac{1}{(m-1)!}\,I + t\Omega\,\varphi_m(t\Omega), \quad m \ge 1. \tag{2.5b}$$

Furthermore, it is substantial that the relation

$$\varphi_m(t\Omega) - \varphi_m(0) = t\Omega\,\chi(t\Omega) \tag{2.5c}$$

holds with a linear operator $\chi(t\Omega)$ that is bounded on X_μ, see [13, 22] and also [11, 30].

3. Commutator-free exponential integrator. In this section, we introduce an integration method for linear non-autonomous parabolic problems (1.1) which relies on the composition of two exponentials. We note that the considered scheme is an example of a *Crouch-Grossman method* [9].

For a constant stepsize $h > 0$ the associated grid points are denoted by $t_j = jh$ for $j \geq 0$. The numerical approximation $u_{n+1} \approx u(t_{n+1})$ to the true solution of (1.1) is given by the recurrence formula

$$u_{n+1} = \mathrm{e}^{\widetilde{\zeta} h C_n} \, \mathrm{e}^{\zeta h B_n} \, u_n, \qquad n \geq 0. \tag{3.1a}$$

Here, we employ the following abbreviations

$$\begin{aligned} A_{ni} &= A(t_n + c_i h), \qquad i = 1, 2, \\ B_n = \alpha A_{n1} + \beta A_{n2}, &\qquad C_n = \gamma A_{n1} + \delta A_{n2}. \end{aligned} \tag{3.1b}$$

Throughout, we assume that the nodal points $\zeta, \widetilde{\zeta}, c_1, c_2 \in \mathbb{R}$ and the coefficients $\alpha, \beta, \gamma, \delta \in \mathbb{R}$ satisfy

$$0 < \zeta < 1, \qquad \widetilde{\zeta} = 1 - \zeta, \qquad 0 \leq c_1 \leq c_2 \leq 1, \qquad \alpha + \beta = 1, \qquad \gamma + \delta = 1. \tag{3.1c}$$

The following remark shows that relation (3.1a) remains senseful within the analytical framework of Section 2.

Remark 1. Under the assumptions of Hypothesis 1, the linear operator

$$\alpha A_{n1} + (1 - \alpha) A_{n2} = A_{n2} + \alpha\big(A_{n1} - A_{n2}\big), \qquad \alpha \in \mathbb{R},$$

is sectorial, see also [13, Theorem 1.3.2]. Therefore, the commutator-free exponential integrator (3.1) is well-defined for abstract evolution equations (1.1).

4. Stability. The stability properties of the commutator-free exponential integrator (3.1) are determined by the behaviour of the evolution operator

$$\prod_{i=m}^{n} \mathrm{e}^{\widetilde{\zeta} h C_i} \, \mathrm{e}^{\zeta h B_i} = \mathrm{e}^{\widetilde{\zeta} h C_n} \, \mathrm{e}^{\zeta h B_n} \, \mathrm{e}^{\widetilde{\zeta} h C_{n-1}} \, \mathrm{e}^{\zeta h B_{n-1}} \, \cdots \, \mathrm{e}^{\widetilde{\zeta} h C_m} \, \mathrm{e}^{\zeta h B_m}, \tag{4.1}$$

where $n \geq m \geq 0$. The following result implies that the numerical solution u_n remains bounded for arbitrarily chosen stepsizes $h > 0$ as long as $nh \leq T$.

Theorem 1 (Stability). *Under the requirements of Hypothesis 1 on A, the discrete evolution operator* (4.1) *fulfills the bound*

$$\Big\| \prod_{i=m}^{n} \mathrm{e}^{\widetilde{\zeta} h C_i} \, \mathrm{e}^{\zeta h B_i} \Big\|_{X \leftarrow X} \leq M, \qquad 0 \leq mh \leq nh \leq T,$$

with a constant $M > 0$ that does not depend on n and h.

Proof. As in our preceeding works [11, 30], the proof the above stability result relies on the telescopic identity and the integral formula of Cauchy. In the present situation, it is useful to compare the discrete evolution operator (4.1) with the linear operator

$$\prod_{i=m}^{n} \mathrm{e}^{\widetilde{\zeta} h A_{m2}} \, \mathrm{e}^{\zeta h A_{m2}} = \prod_{i=m}^{n} \mathrm{e}^{h A_{m2}} = \mathrm{e}^{(t_{n+1} - t_m) A_{m2}},$$

which satisfies the well-known bound

$$\left\|\mathrm{e}^{(t_{n+1}-t_m)A_{m2}}\right\|_{X\leftarrow X} + \left\|(t_{n+1}-t_m)\,A_{m2}\,\mathrm{e}^{(t_{n+1}-t_m)A_{m2}}\right\|_{X\leftarrow X} \leq C$$

for $0 \leq t_m \leq t_n \leq T$. Therefore, it suffices to estimate the difference

$$\begin{aligned}
\Delta_m^n &= \prod_{i=m}^{n} \mathrm{e}^{\widetilde{\zeta}hC_i}\,\mathrm{e}^{\zeta hB_i} - \mathrm{e}^{(t_{n+1}-t_m)A_{m2}} \\
&= \sum_{j=m}^{n-1} \Delta_{j+1}^n \Big(\mathrm{e}^{\widetilde{\zeta}hC_j}\,\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}}\Big)\mathrm{e}^{(t_j-t_m)A_{j2}} \\
&\qquad + \sum_{j=m}^{n} \mathrm{e}^{(t_n-t_j)A_{j2}} \Big(\mathrm{e}^{\widetilde{\zeta}hC_j}\,\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}}\Big)\mathrm{e}^{(t_j-t_m)A_{j2}}.
\end{aligned}$$

For this purpose, it is notable that the following relation holds true

$$\mathrm{e}^{\widetilde{\zeta}hC_j}\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}} = \Big(\mathrm{e}^{\widetilde{\zeta}hC_j} - \mathrm{e}^{\widetilde{\zeta}hA_{m2}}\Big)\,\mathrm{e}^{\zeta hB_j} + \mathrm{e}^{\widetilde{\zeta}hA_{m2}}\Big(\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{\zeta hA_{m2}}\Big).$$

By means of the integral formula of Cauchy, the resolvent identity

$$(\lambda I - \Omega_1)^{-1} - (\lambda I - \Omega_2)^{-1} = (\lambda I - \Omega_1)^{-1}(\Omega_1 - \Omega_2)(\lambda I - \Omega_2)^{-1},$$

and the relations given in (3.1), we receive

$$\begin{aligned}
&\Big(\mathrm{e}^{\widetilde{\zeta}hC_j}\,\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}}\Big)\,\mathrm{e}^{(t_j-t_m)A_{j2}} \\
&\quad = \frac{\widetilde{\zeta}h}{2\pi\mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda}\,(\lambda - \widetilde{\zeta}hC_j)^{-1}\big(\gamma(A_{j1}-A_{j2}) + A_{j2} - A_{m2}\big) \\
&\qquad\qquad \times (\lambda - \widetilde{\zeta}hA_{m2})^{-1}\,\mathrm{e}^{\zeta hB_j}\,\mathrm{e}^{(t_j-t_m)A_{j2}}\,\mathrm{d}\lambda \\
&\qquad + \frac{\zeta h}{2\pi\mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda}\,\mathrm{e}^{\widetilde{\zeta}hA_{m2}}\,(\lambda - \zeta hB_j)^{-1}\big(\alpha(A_{j1}-A_{j2}) + A_{j2} - A_{m2}\big) \\
&\qquad\qquad \times (\lambda - \zeta hA_{m2})^{-1}\,\mathrm{e}^{(t_j-t_m)A_{j2}}\,\mathrm{d}\lambda.
\end{aligned}$$

With the help of the resolvent bound (2.1), the Hölder estimate (2.2) for A, and (2.4) it thus follows

$$\begin{aligned}
\left\|\Big(\mathrm{e}^{\widetilde{\zeta}hC_j}\,\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}}\Big)\,\mathrm{e}^{(t_j-t_m)A_{j2}}\right\|_{X\leftarrow X} &\leq Mh^{\vartheta}, \qquad j = m, \\
\left\|\Big(\mathrm{e}^{\widetilde{\zeta}hC_j}\,\mathrm{e}^{\zeta hB_j} - \mathrm{e}^{hA_{m2}}\Big)\,\mathrm{e}^{(t_j-t_m)A_{j2}}\right\|_{X\leftarrow X} &\leq Mh(t_j-t_m)^{-1+\vartheta}, \qquad j > m.
\end{aligned}$$

Consequently, a further application of (2.4) together with a Gronwall-type inequality with a weakly singular kernel, see also [4, 26], yields the desired stability bound. □

5. Convergence. In this section, we analyse the convergence behaviour of the considered commutator-free exponential integrator for parabolic problems (1.1). As a first step, we next derive a useful relation for the defect of (3.1) by means of a suitable linearisation of the differential equation and an application of the variation-of-constants formula. Similar techniques have been used in the study of exponential splitting methods, see [1, 18, 28] and references therein. The following considerations also explain the definition of the numerical method.

5.1. Expansion of the defect. Replacing in (3.1) the numerical by the exact solution values defines the defect of the method

$$u(t_{n+1}) = \mathrm{e}^{\widetilde{\zeta} h C_n}\, \mathrm{e}^{\zeta h B_n}\, u(t_n) + d_{n+1}, \qquad n \geq 0. \tag{5.1}$$

Our basic approach is to consider the initial value problem (1.1) on the subinterval $[t_n, t_{n+1}]$ and to derive an analogous relation to (3.1a) for the exact solution values. For that purpose, we set

$$G_n(t) = \big(A(t) - B_n\big)u(t), \qquad H_n(t) = \big(A(t) - C_n\big)u(t). \tag{5.2}$$

On the one hand, rewriting the right-hand side of the differential equation in (1.1) as $u'(t) = B_n u(t) + G_n(t)$ and applying the variation-of-constants formula, see [22], yields the following relation for the solution value at time $t_n + \zeta h$

$$u(t_n + \zeta h) = \mathrm{e}^{\zeta h B_n}\, u(t_n) + \int_0^{\zeta h} \mathrm{e}^{(\zeta h - \tau) B_n}\, G_n(t_n + \tau)\, \mathrm{d}\tau.$$

On the other hand, by linearising (1.1) around C_n and inserting the above representation for $u(t_n + \zeta h)$, we further obtain

$$\begin{aligned} u(t_{n+1}) = \mathrm{e}^{\widetilde{\zeta} h C_n}\, \mathrm{e}^{\zeta h B_n}\, u(t_n) + \mathrm{e}^{\widetilde{\zeta} h C_n} \int_0^{\zeta h} \mathrm{e}^{(\zeta h - \tau) B_n}\, G_n(t_n + \tau)\, \mathrm{d}\tau \\ + \int_0^{\widetilde{\zeta} h} \mathrm{e}^{(\widetilde{\zeta} h - \tau) C_n}\, H_n(t_n + \zeta h + \tau)\, \mathrm{d}\tau. \end{aligned}$$

Consequently, the defect of the numerical method (3.1) equals

$$d_{n+1} = \mathrm{e}^{\widetilde{\zeta} h C_n} \int_0^{\zeta h} \mathrm{e}^{(\zeta h - \tau) B_n}\, G_n(t_n + \tau)\, \mathrm{d}\tau + \int_0^{\widetilde{\zeta} h} \mathrm{e}^{(\widetilde{\zeta} h - \tau) C_n}\, H_n(t_n + \zeta h + \tau)\, \mathrm{d}\tau. \tag{5.3}$$

In order to derive a suitable expansion of d_{n+1}, it is useful to introduce some additional notation.

The time-derivatives of the linear operator A and the exact solution u of (1.1) at time t_n are denoted by

$$A_n^{(i)} = A^{(i)}(t_n), \quad i \geq 0, \qquad \widehat{u}_n^{(j)} = u^{(j)}(t_n), \quad j \geq 0. \tag{5.4a}$$

For the coefficients of the numerical scheme, we define

$$\mu_i = \alpha c_1^i + \beta c_2^i, \qquad \nu_i = \gamma c_1^i + \delta c_2^i, \qquad i = 1, 2, 3, \tag{5.4b}$$

see (3.1). We note that for a sufficiently differentiable function $f : [t_n, t_{n+1}] \to X$ a Taylor series expansion yields

$$\begin{aligned} f(t_n + \tau) &= \sum_{i=0}^{m} \frac{\tau^i}{i!}\, f_n^{(i)} + R\big(\tau^{m+1}, f^{(m+1)}\big), \qquad 0 \leq \tau \leq h, \\ R\big(\tau^{m+1}, f^{(m+1)}\big) &= \frac{1}{m!}\, \tau^{m+1} \int_0^1 (1 - \sigma)^m f^{(m+1)}(t_n + \sigma\tau)\, \mathrm{d}\sigma, \end{aligned} \tag{5.5}$$

where $f_n^{(i)} = f^{(i)}(t_n)$. Thus, provided that the quantity

$$\big\|f^{(m+1)}\big\|_{X,\infty} = \max_{t_n \leq t \leq t_{n+1}} \big\|f^{(m+1)}(t)\big\|_X$$

is well-defined, the remainder fulfills

$$\big\|R\big(\tau^{m+1}, f^{(m+1)}\big)\big\|_X \le Mh^{m+1}\big\|f^{(m+1)}\big\|_{X,\infty}, \qquad 0 \le \tau \le h,$$

with some constant $M > 0$. Terms that satisfy an estimate of this form are henceforth denoted by $\mathscr{R}\big(h^{m+1}, f^{(m+1)}\big)$. In particular, the abbrevitation $\mathscr{R}\big(h^k, A^{(i)}\,u^{(j)}\big)$ signifies that the bound

$$\big\|\mathscr{R}\big(h^k, A^{(i)}\,u^{(j)}\big)\big\|_X \le Mh^k \max_{t_n\le s,t\le t_{n+1}} \big\|A^{(i)}(s)\,u^{(j)}(t)\big\|_{X,\infty}$$

holds true.

Provided that the involved derivatives of A and u are well-defined, the following representation is valid for the defect d_{n+1} given by (5.1). We recall formula (2.5a) for the linear operator φ_m.

LEMMA 1. *The numerical solution defect of* (3.1) *fulfills the relation*

$$\begin{aligned} d_{n+1} = \sum_{(i,j)\in\mathscr{J}} h^{i+j+1}\,\Phi_{ij}\,A_n^{(i)}\widehat{u}_n^{(j)} + \mathscr{R}\big(h^5, A^{(4)}u\big) \\ + \mathscr{R}\big(h^5, A'''u'\big) + \mathscr{R}\big(h^5, A''u''\big) + \mathscr{R}\big(h^5, A'u'''\big), \end{aligned}$$

where $\Phi_{ij} = \Phi_{ij}(hB_n, hC_n)$ *is defined through*

$$\begin{aligned} \Phi_{ij} = \frac{1}{i!j!}\Big\{ &\zeta^{j+1}\mathrm{e}^{\widetilde{\zeta}hC_n}\Big((i+j)!\,\zeta^i\varphi_{i+j+1}(\zeta hB_n) - j!\,\mu_i\varphi_{j+1}(\zeta hB_n)\Big) \\ &+ \sum_{\ell=j+1}^{i+j} \ell!\,\tbinom{i+j}{\ell}\zeta^{i+j-\ell}\widetilde{\zeta}^{\ell+1}\varphi_{\ell+1}(\widetilde{\zeta}hC_n) \\ &+ \sum_{\ell=0}^{j} \ell!\,\zeta^{j-\ell}\widetilde{\zeta}^{\ell+1}\Big(\tbinom{i+j}{\ell}\zeta^i - \nu_i\tbinom{j}{\ell}\Big)\varphi_{\ell+1}(\widetilde{\zeta}hC_n)\Big\} \end{aligned}$$

and $\mathscr{J} = \big\{(1,0),(2,0),(1,1),(3,0),(2,1),(1,2)\big\}$.

PROOF. We first derive a useful relation for the maps G_n and H_n defined in (5.2). With the help of (5.5), by combining the expansions

$$\begin{aligned} A(t_n+\tau) - B_n &= \sum_{i=0}^{3}\frac{1}{i!}\big(\tau^i - \mu_i h^i\big)A_n^{(i)} + \mathscr{R}\big(h^4, A^{(4)}\big), \\ u(t_n+\tau) &= \sum_{j=0}^{2}\frac{1}{j!}\,\tau^j\widehat{u}_n^{(j)} + R\big(\tau^3, u^{(3)}\big), \end{aligned}$$

we receive the following representation

$$\begin{aligned} G_n(t_n+\tau) &= \sum_{(i,j)\in\mathscr{J}}\frac{1}{i!j!}\big(\tau^i - \mu_i h^i\big)\tau^j A_n^{(i)}\widehat{u}_n^{(j)} + \mathscr{R}\big(h^4\big), \\ \mathscr{R}\big(h^4\big) &= \mathscr{R}\big(h^4, A^{(4)}u\big) + \mathscr{R}\big(h^4, A'''u'\big) + \mathscr{R}\big(h^4, A''u''\big) + \mathscr{R}\big(h^4, A'u'''\big), \end{aligned} \tag{5.6a}$$

see also (3.1b)-(3.1c) and (5.4). Similarly, it follows

$$H_n(t_n+\zeta h+\tau) = \sum_{(i,j)\in\mathscr{J}}\frac{1}{i!j!}\big((\zeta h+\tau)^i - \nu_i h^i\big)(\zeta h+\tau)^j A_n^{(i)}\widehat{u}_n^{(j)} + \mathscr{R}\big(h^4\big). \tag{5.6b}$$

We next insert the above expansions (5.6) into (5.3) and express the resulting integrals by means of (2.5a). Altogether, this yields the given result. □

In the situation of Section 2, a reasonable smoothness assumption on (1.1) is that the linear operator A and the exact solution u are sufficiently differentiable with respect to the variable t. Precisely, we suppose $A^{(4)}(t)$ and $u^{(4)}(t)$ to be bounded in the underlying Banach space X for all $0 \le t \le T$. The following remark states that then the expansion of Lemma 1 is well-defined. However, unless the exact solution satisfies additional (unnatural) requirements such as $A'(t)u(t) \in D$ for $0 \le t \le T$, in general, it is not possible to further expand the defect.

REMARK 2. Provided that $u'(t) \in X$ it follows from the differential equation in (1.1) that $A(t)u(t) \in X$ and therefore $u(t) \in D$ for $0 \le t \le T$. Differentiating (1.1) with respect to the variable t implies $A(t)u'(t) = u''(t) - A'(t)u(t) \in X$, and, as a consequence, $u'(t) \in D$ for any $0 \le t \le T$. Similarly, it follows $u^{(j-1)}(t) \in D$ if $u^{(j)}(t) \in X$ for $0 \le t \le T$ and $j = 3, 4$. Thus, under the regularity requirements $A \in \mathscr{C}^4\big([0,T], L(D,X)\big)$ and $u \in \mathscr{C}^4\big([0,T], X\big)$, the representation of the defect given in Lemma 1 is well-defined.

5.2. Error estimate. With the help of the stability estimate and the expansion of the defect given in Sections 4 and 5.1, we are able to prove the following convergence result.

THEOREM 2 (Convergence). *Assume that the requirements of Hypothesis 1 are fulfilled and that further $A \in \mathscr{C}^4\big([0,T], L(D,X)\big)$ and $u \in \mathscr{C}^4\big([0,T], X\big)$. Then, provided that $A^{(i)}(t)\, u^{(j)}(t)$ belongs to the intermediate space X_κ with $0 \le \kappa < 1$ for $0 \le t \le T$ and $(i,j) \in \big\{(1,0),(2,0),(1,1)\big\}$, the fourth-order commutator-free exponential integrator* (1.2) *satisfies the error estimate*

$$\|u_n - u(t_n)\|_X \le C\Big(\big\|u_0 - u(0)\big\|_X + h^{2+\kappa}\big(1 + |\log h|\big)\Big), \qquad 0 \le t_n \le T,$$

with some constant $C > 0$ independent of n and h.

PROOF. In order to obtain a suitable relation for the global error $e_n = u_n - u(t_n)$, we first resolve the recurrence formula (3.1a) for the numerical approximation

$$u_n = \prod_{i=0}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, u_0, \qquad n \ge 0.$$

Furthermore, by using (5.1), we receive $e_n = e_n^{(1)} + e_n^{(2)}$ with

$$e_n^{(1)} = \prod_{i=0}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\big(u_0 - u(0)\big), \qquad e_n^{(2)} = -\sum_{j=0}^{n-1} \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, d_{j+1}. \tag{5.7}$$

We next estimate the terms in (5.7) with respect to the norm of the underlying Banach space X. An application of Theorem 1 shows that the first term is bounded by a constant times the error of the initial value

$$\big\|e_n^{(1)}\big\|_X \le \Big\| \prod_{i=0}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i} \Big\|_{X \leftarrow X} \|u_0 - u(0)\|_X \le C \|u_0 - u(0)\|_X.$$

For estimating the second term $e_n^{(2)}$, we employ the representation of the defect given in Lemma 1. Making use of the fact that the sums involving the remainder and the

terms where $i+j \geq 3$ are bounded by constant times h^3, we receive

$$\begin{aligned}\big\|e_n^{(2)}\big\|_X \leq h^2 \sum_{j=0}^{n-1} \Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{10}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} \big\|A_j' \widehat{u}_j\big\|_{X_\kappa} \\ + h^3 \sum_{j=0}^{n-1} \Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{20}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} \big\|A_j'' \widehat{u}_j\big\|_{X_\kappa} \\ + h^3 \sum_{j=0}^{n-1} \Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{11}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} \big\|A_j' \widehat{u}_j'\big\|_{X_\kappa} \\ + Ch^3.\end{aligned} \tag{5.8}$$

We note that the coefficients of the fourth-order scheme (1.2) satisfy the conditions

$$\begin{aligned}\Phi_{20}(0,0) &= \tfrac{1}{2}\Big\{\zeta\big(\tfrac{1}{3}\,\zeta^2 - \mu_2\big) + \widetilde{\zeta}\Big(\tfrac{1}{3}\,\widetilde{\zeta}^2 + \zeta\widetilde{\zeta} + \zeta^2 - \nu_2\Big)\Big\} = 0, \\ \Phi_{11}(0,0) &= \zeta^2\big(\tfrac{1}{3}\,\zeta - \tfrac{1}{2}\,\mu_1\big) + \widetilde{\zeta}\Big(\tfrac{1}{3}\,\widetilde{\zeta}^2 + \tfrac{1}{2}\,\widetilde{\zeta}(2\zeta - \nu_1) + \zeta(\zeta - \nu_1)\Big) = 0.\end{aligned} \tag{5.9a}$$

Therefore, similar arguments as in the proof of Theorem 1 show the refined bounds

$$\begin{aligned}\Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{20}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} &\leq Mh(t_n - t_j)^{-1+\kappa}, \\ \Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{11}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} &\leq Mh(t_n - t_j)^{-1+\kappa},\end{aligned}$$

see also (2.4) and (2.5c). In relation (5.8), it remains to estimate the sum involving Φ_{10}. For that purpose, we apply (2.5b) together with suitable Taylor expansions of B_j and C_j. Moreover, the coefficients of (1.2) fulfill

$$\begin{aligned}\Phi_{10}(0,0) &= \zeta\big(\tfrac{1}{2}\,\zeta - \mu_1\big) + \tfrac{1}{2}\,\widetilde{\zeta}^2 + \widetilde{\zeta}(\zeta - \nu_1) = 0, \\ \Psi_{10}(0,0) &= \zeta^2\big(\tfrac{1}{6}\,\zeta - \tfrac{1}{2}\,\mu_1\big) + \widetilde{\zeta}\Big(\tfrac{1}{6}\,\widetilde{\zeta}^2 + \tfrac{1}{2}\,\widetilde{\zeta}(\zeta - \nu_1) + \zeta\big(\tfrac{1}{2}\,\zeta - \mu_1\big)\Big) = 0.\end{aligned} \tag{5.9b}$$

As a consequence, we finally obtain the refined estimate

$$\Big\| \prod_{i=j+1}^{n-1} \mathrm{e}^{\widetilde{\zeta} hC_i}\, \mathrm{e}^{\zeta hB_i}\, \Phi_{10}(hB_j, hC_j)\Big\|_{X \leftarrow X_\kappa} \leq Mh^{1+\kappa}\big(1 + |\log h| + (t_{n+1} - t_m)^{-1}\big).$$

Altogether, this implies

$$\begin{aligned}\big\|e_n^{(2)}\big\|_X \leq Ch^{3+\kappa} \sum_{j=0}^{n-1} \big(1 + |\log h| + (t_n - t_j)^{-1}\big) \\ + Ch^4 \sum_{j=0}^{n-1} (t_n - t_j)^{-1+\kappa} + Ch^3 \leq Ch^{2+\kappa}\big(1 + |\log h|\big)\end{aligned}$$

which proves the given error estimate. □

REMARK 3. Going over the proof of Theorem 2 shows that the essential conditions for a fractional convergence order of $2 + \kappa$ are (5.9). We note that the conditions for

Stepsize h	1/2	1/4	1/8	1/16	1/32
Method 1 (M = 50)	2.0076	1.9632	1.9597	1.9699	1.9822
Method 1 (M = 100)	2.0075	1.9631	1.9595	1.9696	1.9818
Method 2 (M = 50)	1.0924	1.9634	2.2295	2.3162	2.4248
Method 2 (M = 100)	1.0949	1.9604	2.2267	2.3153	2.4181
Method 3 (M = 50)	2.2597	2.1983	2.3386	2.4337	2.4999
Method 3 (M = 100)	2.2591	2.1960	2.3348	2.4227	2.4782
Method 4 (M = 50)	3.3250	3.5115	3.3419	3.0490	2.8486
Method 4 (M = 100)	3.0426	3.4011	3.4838	3.2384	2.9488

TABLE 6.1
Numerical temporal convergence orders in a discrete L^1-norm for spatial discretisations of grid length $\Delta x = (M+1)^{-1}$.

a classical convergence order 3 are equivalent to the relations in (5.9). However, it is not possible to construct a commutator-free exponential integrator of classical order 3 that is based on the evaluation of one exponential only, that is, the validity of relation (5.9) implies $0 < \zeta < 1$ in (3.1).

6. Numerical example. In this section, we illustrate the error estimate of Theorem 2 by a numerical example for a parabolic initial-boundary value problem subject to a homogeneous Dirichlet boundary condition. We start with a brief discussion of the considered time integration schemes. For notational simplicity, we only give the first step and denote $A_i = A(c_i h)$.

METHOD 1. For parabolic problems (1.1), it follows from the error estimate given in our previous work [11] that the exponential midpoint rule

$$u_1 = \mathrm{e}^{hA_1}\, u_0, \qquad c_1 = \tfrac{1}{2},$$

is convergent of order 2 with respect to the norm of the underlying Banach space.

METHOD 2. The commutator-free exponential integration scheme

$$u_1 = \mathrm{e}^{(1-\zeta)h(a_1 A_1 + (1-a_1)A_2)}\, \mathrm{e}^{\zeta h A_1}\, u_0,$$
$$\zeta = \tfrac{\sqrt{3}}{3}\,, \qquad a_1 = \tfrac{1}{4} - \tfrac{\sqrt{3}}{4}\,, \qquad c_i = \tfrac{1}{2} \mp \tfrac{\sqrt{3}}{6}\,, \quad i = 1, 2,$$

has a classical convergence order 3.

METHOD 3. The numerical method

$$u_1 = \mathrm{e}^{h(a_2 A_1 + a_1 A_2)}\, \mathrm{e}^{h(a_1 A_1 + a_2 A_2)}\, u_0, \qquad a_i = \tfrac{1}{4} \pm \tfrac{\sqrt{3}}{6}\,, \qquad c_i = \tfrac{1}{2} \mp \tfrac{\sqrt{3}}{6}\,, \quad i = 1, 2,$$

is the unique scheme of the form (3.1) that satisfies the conditions for a classical convergence order 4, see also (1.2).

In the numerical example, as an illustration, the fourth-order commutator-free exponential integrator given before is compared with a fourth-order interpolatory Magnus integrator. To explain the stability and error behaviour of this method for parabolic problems is beyond the purpose of the present work.

METHOD 4. The fourth-order interpolatory Magnus integrator

$$u_1 = \mathrm{e}^{ha_1(A_1+A_2) + h^2 a_2 [A_2, A_1]}\, u_0, \qquad a_1 = \tfrac{1}{2}\,, \qquad a_2 = \tfrac{\sqrt{3}}{12}\,,$$

requires the evaluation of the linear operator $\big[A_2, A_1\big] = A_2 A_1 - A_1 A_2$.

Stepsize h	1/2	1/4	1/8	1/16	1/32
Method 1 (M = 50)	2.0120	1.9740	1.9723	1.9786	1.9879
Method 1 (M = 100)	2.0120	1.9739	1.9722	1.9785	1.9878
Method 2 (M = 50)	1.1979	1.9223	2.0992	2.1336	2.1732
Method 2 (M = 100)	1.1985	1.9208	2.0977	2.1303	2.1666
Method 3 (M = 50)	2.0197	2.0409	2.1271	2.1917	2.2331
Method 3 (M = 100)	2.0194	2.0397	2.1244	2.1859	2.2210
Method 4 (M = 50)	3.3204	3.5217	2.9654	2.4024	2.3609
Method 4 (M = 100)	3.0341	3.4204	3.3656	2.6010	2.3197

TABLE 6.2
Numerical temporal convergence orders in a discrete L^2-norm for spatial discretisations of grid length $\Delta x = (M+1)^{-1}$.

We consider a one-dimensional initial-boundary value problem for a real-valued function $U : [0,1] \times [0,T] \to \mathbb{R} : (x,t) \mapsto U(x,t)$ comprising the partial differential equation

$$\partial_t U(x,t) = \mathscr{A}(x,t)\, U(x,t), \qquad 0 < x < 1, \quad 0 < t \leq T, \tag{6.1a}$$

subject to a homogeneous Dirichlet boundary condition and an initial condition

$$U(0,t) = 0 = U(1,t), \quad 0 \leq t \leq T, \qquad U(x,0) = U_0(x), \quad 0 \leq x \leq 1. \tag{6.1b}$$

The differential equation involves a second-order differential operator

$$\mathscr{A}(x,t) = \alpha(x,t)\, \partial_x^2 + \beta(x,t)\, \partial_x + \gamma(x,t) \tag{6.1c}$$

which we assume to satisfy the condition of strong ellipticity. We further suppose that the space and time-dependent coefficients α, β and γ fulfill suitable regularity and boundedness requirements. For $v \in \mathscr{C}_0^\infty(0,1)$ we define $u(t)$ and $A(t)$ through $\big(u(t)\big)(x) = U(x,t)$ and $\big(A(t)v\big)(x) = \mathscr{A}(x,t)v(x)$. Then, problem (6.1) can be cast into the abstract framework of Section 2 for

$$X = L^p(0,1), \qquad D = W^{p,2}(0,1) \cap W_0^{p,1}(0,1), \qquad 1 < p < \infty,$$

see [11] and references therein. In view of the numerical experiment, we choose

$$\alpha(x,t) = \mathrm{e}^{x-t}, \qquad \beta(x,t) = xt, \qquad \gamma(x,t) = x^2\big(1 + \mathrm{e}^t\big).$$

The admissible values of κ in Theorem 2 are $0 \leq \kappa < (2p)^{-1}$. Thus, the expected fractional convergence order in $X = L^p(0,1)$ is $2 + \kappa$ where $\kappa < (2p)^{-1}$.

In the numerical experiment, we discretise the problem in space by symmetric finite differences of grid length $\Delta x = (M+1)^{-1}$. In time, we apply the exponential integrators given above with stepsize $h = 2^{-i}$ for $1 \leq i \leq 5$ and integrate the problem up to time $T = 1$. A reference solution is determined for a temporal stepsize $h = 2^{-10}$. The numerical temporal order of convergence with respect to a discrete L^p-norm is determined in a standard way from the numerical solution values. The obtained numbers for $p = 2$ and the limiting cases $p = 1$ and $p = \infty$ are displayed in Tables 6.1-6.3. The convergence order 2 for the exponential midpoint rule (Method 1) is explained by a convergence result proven in [11]. For the commutator-free exponential integrators of classical order 3 (Method 2) and classical order 4 (Method 3), respectively, the values of approximately $2 + (2p)^{-1}$ are in accordance with the convergence orders predicted by Theorem 2.

Stepsize h	1/2	1/4	1/8	1/16	1/32
Method 1 (M = 50)	2.0250	2.0065	2.0208	2.0226	2.0149
Method 1 (M = 100)	2.0250	2.0063	2.0207	2.0222	2.0129
Method 2 (M = 50)	1.2328	1.7318	1.8169	1.8604	1.9092
Method 2 (M = 100)	1.2341	1.7313	1.8135	1.8559	1.9072
Method 3 (M = 50)	1.7384	1.8369	1.9113	1.9649	1.9851
Method 3 (M = 100)	1.7391	1.8347	1.9103	1.9604	1.9736
Method 4 (M = 50)	3.3042	3.0169	1.9200	2.0864	2.1880
Method 4 (M = 100)	3.0257	3.4434	2.0132	1.9839	2.0752

TABLE 6.3
Numerical temporal convergence orders in a discrete L^∞-norm for spatial discretisations of grid length $\Delta x = (M+1)^{-1}$.

7. Conclusions. In the present work, we studied the convergence properties of a commutator-free exponential integrator that relies on the composition of two exponentials for parabolic initial value problems of the form (1.1). In particular, we focused on the fourth-order scheme (1.2) which is based on the Gaussian nodes. We showed that the exponential integration scheme remains stable for arbitrarily large stepsizes. But, it is seen from the theoretical investigations and as well in a numerical experiment that a substantial order reduction occurs, in general. For instance, for one-dimensional parabolic initial-boundary value problems under a homogeneous Dirichlet boundary condition a fractional convergence order of at most $2 + (2p)^{-1}$ can be expected in the norm of the function space L^p. The order reduction is explained by the fact that even if the exact solution of the initial-boundary value problem belongs to the domain of the differential operator and further is temporally smooth, it in general does not fulfill additional boundary conditions, that is, combinations of the form $A(s)A(t)u(t)$ are not well-defined for all $0 \leq s, t \leq T$.

For that reason, concerning the derivation of high-order exponential integrators for non-autonomous parabolic problems, it seems more promising to employ a suitable linearisation and to base the numerical schemes on explicit exponential methods of Runge–Kutta or multistep type. Also, the error analysis for non-autonomous parabolic equations is of theoretical value as it gives insight how to construct and study numerical methods for quasilinear equations which are of particular interest in view of practical applications. For example, quasilinear parabolic problems are used in the modelling of diffusion processes with state-dependent diffusitivity and arise in the study of fluids in porous media, see [12]. It is intended to investigate this approach in a future work.

Acknowledgement. I am grateful to Jitse Niesen. His talk and interest at MaGIC 2005 motivated me to finish this work. I thank Alexander Ostermann for inspiring discussions on exponential integrators. This work was supported by Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under project H210-N13.

REFERENCES

[1] Ch. Besse, B. Bidegaray, and S. Descombes, *Order estimates in time of splitting methods for the nonlinear Schrödinger equation.* SIAM J. Numer. Anal. 40 (2002) 26-40.

[2] S. Blanes, F. Casas, J.A. Oteo, and J. Ros, *Magnus and Fer expansions for matrix differential equations: the convergence problem.* J. Phys. A Math. Gen. 31 (1998) 259-268.

[3] S. Blanes and P.C. Moan, *Splitting methods for the time-dependent Schrödinger equation.* Phys. Lett. A 265 (2000) 35-42.
[4] H. Brunnner and P.J. van der Houwen, *The numerical solution of Volterra equations.* CWI Monographs 3. North-Holland, Amsterdam, 1986.
[5] M.P. Calvo and C. Palencia, *A class of explicit multistep exponential integrators for semilinear problems* (2005). To appear in Numer. Math.
[6] E. Celledoni, *Eulerian and Semi-Lagrangian schemes based on commutator free exponential integrators* (2004). To appear in CRM Proceedings Series.
[7] E. Celledoni, A. Martinsen, and B. Owren, *Commutator-free Lie group methods.* FGCS 19/3 (2003) 341-352.
[8] S.M. Cox and P.C. Matthews, *Exponential time differencing for stiff systems.* J. Comp. Phys. 176 (2002) 430-455.
[9] P.E. Crouch and R. Grossman, *Numerical integration of ordinary differential equations on manifolds.* J. Nonlinear Sci. 3 (1993) 1-33.
[10] J. Van den Eshof and M. Hochbruck, *Preconditioning Lanczos approximations to the matrix exponential* (2004). To appear in SIAM J. Sci. Comput.
[11] C. González, A. Ostermann, and M. Thalhammer, *A second-order Magnus integrator for non-autonomous parabolic problems* (2004). To appear in J. Comp. Appl. Math.
[12] C. González and M. Thalhammer, *A second-order Magnus type integrator for quasilinear parabolic problems* (2004). Submitted to Math. Comp.
[13] D. Henry, *Geometric Theory of Semilinear Parabolic Equations.* Lecture Notes in Mathematics 840, Springer, Berlin, 1981.
[14] M. Hochbruck and Ch. Lubich, *On Magnus integrators for time-dependent Schrödinger equations.* SIAM J. Numer. Anal. 41 (2003) 945-963.
[15] M. Hochbruck and Ch. Lubich, *On Krylov subspace approximations to the matrix exponential operator.* SIAM J. Numer. Anal. 34 (1997) 1911-1925.
[16] M. Hochbruck and A. Ostermann, *Explicit exponential Runge-Kutta methods for semilinear parabolic problems* (2004). To appear in SIAM J. Numer. Anal.
[17] A. Iserles, *On the global error of discretization methods for highly-oscillatory ordinary differential equations.* BIT 42 (2002) 561-599.
[18] T. Jahnke and Ch. Lubich, *Error bounds for exponential operator splittings.* BIT 40 (2000) 735-744.
[19] A.K. Kassam and L.N. Trefethen, *Fourth-order time stepping for stiff PDEs.* SIAM J. Sci. Comput. 26/4 (2005) 1214-1233.
[20] S. Krogstad, *Generalized integrating factor methods for stiff PDEs.* J. Comp. Phys. 203 (2005) 72-88.
[21] Y.Y. Lu, *A fourth order Magnus scheme for Helmholtz equation.* J. Comp. Appl. Math. 173/2 (2005) 247-258.
[22] A. Lunardi, *Analytic Semigroups and Optimal Regularity in Parabolic Problems.* Birkhäuser, Basel, 1995.
[23] W. Magnus, *On the exponential solution of a differential equation for a linear operator.* Comm. Pure Appl. Math. 7 (1954) 649-673.
[24] B.V. Minchev and W.M. Wright, *A review of exponential integrators for first order semi-linear problems.* Tech. report 2/05, Department of Mathematics, NTNU, April 2005.
[25] C. Moler and Ch. Van Loan, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later.* SIAM Rev. 45/1 (2003) 3-49.
[26] A. Ostermann and M. Thalhammer, *Non-smooth data error estimates for linearly implicit Runge-Kutta methods.* IMA J. Numer. Anal. 20 (2000) 167-184.
[27] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations.* Springer, New York, 1983.
[28] Q. Sheng, *Global error estimates for exponential splitting.* IMA J. Numer. Anal. 14 (1993) 27-56.
[29] H. Tanabe, *Equations of Evolution.* Pitman, London, 1979.
[30] M. Thalhammer, *A second-order Magnus type integrator for non-autonomous semilinear parabolic problems* (2004). Submitted to IMA J. Numer. Anal.

2.4. A class of explicit exponential general linear methods

A class of explicit exponential general linear methods

ALEXANDER OSTERMANN, MECHTHILD THALHAMMER, AND WILL WRIGHT

To appear in BIT Numerical Mathematics

BIT Numerical Mathematics (2006)46:000-000
DOI:10.1007/s10543-000-0000-x

A CLASS OF EXPLICIT EXPONENTIAL GENERAL LINEAR METHODS*

ALEXANDER OSTERMANN[1], MECHTHILD THALHAMMER[1], and WILL WRIGHT[2]

[1] *Institut für Mathematik, Universität Innsbruck, A-6020 Innsbruck, Austria. email:* {*alexander.ostermann, mechthild.thalhammer*} *@uibk.ac.at*

[2] *Department of Mathematical and Statistical Sciences, La Trobe University, Melbourne, Victoria 3086, Australia. email: w.wright@latrobe.edu.au*

Abstract.

In this paper, we consider a class of explicit exponential integrators that includes as special cases the explicit exponential Runge–Kutta and exponential Adams–Bashforth methods. The additional freedom in the choice of the numerical schemes allows, in an easy manner, the construction of methods of arbitrarily high order with good stability properties.

We provide a convergence analysis for abstract evolution equations in Banach spaces including semilinear parabolic initial-boundary value problems and spatial discretizations thereof. From this analysis, we deduce order conditions which in turn form the basis for the construction of new schemes. Our convergence results are illustrated by numerical examples.

AMS subject classification (2000): 65L05, 65L06, 65M12, 65J10

Key words: Exponential integrators, general linear methods, explicit schemes, abstract evolution equations, semilinear parabolic problems, convergence, high-order methods.

1 Introduction

In the past few years, exponential time-integrators for semilinear problems

$$y'(t) = L\,y(t) + N\big(t, y(t)\big), \quad 0 \le t \le T, \qquad y(0) \text{ given}, \tag{1.1}$$

have attracted a lot of interest. They are particularly appealing in situations where this differential equation comes from the spatial discretization of a partial differential equation. Exponential integrators were for the first time considered in the sixties and seventies of the last century. For a historical survey, we refer to Minchev and Wright [15].

For exponential Runge–Kutta methods, a convergence analysis for parabolic problems has recently been given by Hochbruck and Ostermann [10, 11]. The stage order for explicit schemes, however, is one at most. For that reason,

*Submitted version, August 2005. Revised version, February 2006.

the construction of high-order methods is rather complicated, due to the large number of additional conditions required for stiff problems. On the other hand, the convergence of exponential Adams-type methods has been studied in Calvo and Palencia [5]. This class easily enables the construction of high-order schemes, although the resulting methods are only weakly stable in the sense that all parasitic roots for $y' = 0$ lie on the unit circle.

In the present paper, we are considering a class of explicit exponential integrators that combines the benefits of exponential Runge–Kutta and exponential Adams–Bashforth methods. There, it is possible to achieve high stage order which facilitates the construction of high-order methods with favorable stability properties for stiff problems. In addition, all methods included in our class are zero-stable with parasitic roots equal to zero.

An outline of the paper is as follows: In Section 2, we introduce a class of explicit exponential general linear methods based on the Adams–Bashforth schemes and further give the stage order and quadrature order conditions. These conditions form the basis for the construction of schemes of arbitrarily high order for stiff problems. In Section 3, we state our hypotheses on the problem class (1.1) employing the theory of sectorial operators in Banach spaces. In particular, parabolic initial-boundary value problems are included in our framework. The core of Section 3 is devoted to convergence estimates. Our main result is Theorem 3.4 proving that, for sufficiently smooth solutions of (1.1), the order of convergence is essentially $\min\{P, Q+1\}$. Here, P and Q denote the quadrature order and the stage order of the method, respectively. In Section 4, we exploit the order conditions in a systematic way to construct new schemes. In particular, we show that the class of two-stage methods of order p involving $p-1$ steps is uniquely determined up to a free parameter. Moreover, we derive a three-stage two-step method of order 4. The favorable convergence properties of our methods are illustrated in Section 5. In Section 6, we finally indicate how the convergence analysis given extends to exponential integrators with variable stepsizes.

The functions introduced below are commonly associated with *exponential time differencing methods* where the method coefficients are (linear) combinations of these functions. As we will see in Section 4, they also naturally arise in the construction of exponential general linear methods.

1.1 Exponential and related functions

For integers $j \geq 0$ and complex numbers $z \in \mathbb{C}$, we define $\varphi_j(z)$ through

$$\varphi_j(z) = \int_0^1 \mathrm{e}^{(1-\tau)z} \frac{\tau^{j-1}}{(j-1)!}\,\mathrm{d}\tau, \quad j \geq 1, \qquad \varphi_0(z) = \mathrm{e}^z. \tag{1.2a}$$

Consequently, the recurrence relation

$$\varphi_j(z) = \frac{1}{j!} + z\,\varphi_{j+1}(z), \qquad z \in \mathbb{C}, \quad j \geq 0, \tag{1.2b}$$

is valid.

The following result provides an expansion of the solution of a linear differential equation which is needed in the convergence analysis of Section 3.3.

LEMMA 1.1. *The exact solution of the initial value problem*

$$y'(t) = L\,y(t) + f(t), \quad t \geq t_n, \qquad y(t_n) \text{ given},$$

has the following representation

$$y(t_n+\tau) = \mathrm{e}^{\tau L}\,y(t_n) + \sum_{\ell=0}^{m-1} \tau^{\ell+1}\varphi_{\ell+1}(\tau L)\,f^{(\ell)}(t_n) + R_n(m,\tau),$$
$$R_n(m,\tau) = \int_0^\tau \mathrm{e}^{(\tau-\sigma)L} \int_0^\sigma \frac{(\sigma-\xi)^{m-1}}{(m-1)!}\,f^{(m)}(t_n+\xi)\,\mathrm{d}\xi\,\mathrm{d}\sigma, \qquad \tau \geq 0,$$

provided that the function f is sufficiently many times differentiable.

PROOF. Substituting the Taylor series expansion of f

$$f(t_n+\sigma) = \sum_{\ell=0}^{m-1} \frac{\sigma^\ell}{\ell!}\,f^{(\ell)}(t_n) + S_n(m,\sigma),$$
$$S_n(m,\sigma) = \int_0^\sigma \frac{(\sigma-\xi)^{m-1}}{(m-1)!}\,f^{(m)}(t_n+\xi)\,\mathrm{d}\xi, \tag{1.3}$$

into the variation-of-constants formula

$$y(t_n+\tau) = \mathrm{e}^{\tau L}\,y(t_n) + \int_0^\tau \mathrm{e}^{(\tau-\sigma)L}\,f(t_n+\sigma)\,\mathrm{d}\sigma, \qquad \tau \geq 0, \tag{1.4}$$

and applying the definition (1.2a) of the φ-functions yields the desired result. □

2 Exponential general linear methods

In this paper, we study a class of explicit exponential general linear methods that in particular contains the exponential Runge–Kutta methods and the exponential Adams-type methods considered recently in the literature, see [3, 6, 11, 12, 13] and further [5]. As will be seen from the theoretical results and the illustrations that follow in Sections 3-5, the extra freedom in the choice of the numerical method allows the construction of high-order schemes that possess favorable stability properties and exhibit no order reduction when applied to parabolic problems.

2.1 Method class

We study explicit exponential general linear methods for the autonomous problem

$$y'(t) = L\,y(t) + N\big(y(t)\big), \quad 0 \leq t \leq T, \qquad y(0) \text{ given}. \tag{2.1}$$

For given starting values $y_0, y_1, \ldots, y_{q-1}$, the numerical approximation y_{n+1} at time t_{n+1}, $n \geq q-1$, is given by the recurrence formula

$$y_{n+1} = \mathrm{e}^{hL}\, y_n + h \sum_{i=1}^{s} B_i(hL)\, N(Y_{ni}) + h \sum_{k=1}^{q-1} V_k(hL)\, N(y_{n-k}). \tag{2.2a}$$

The internal stages Y_{ni}, $1 \leq i \leq s$, are defined through

$$Y_{ni} = \mathrm{e}^{c_i hL}\, y_n + h \sum_{j=1}^{i-1} A_{ij}(hL)\, N(Y_{nj}) + h \sum_{k=1}^{q-1} U_{ik}(hL)\, N(y_{n-k}). \tag{2.2b}$$

The method coefficient functions $A_{ij}(hL)$, $U_{ik}(hL)$, $B_i(hL)$, and $V_k(hL)$ are linear combinations of the exponential and related φ-functions, see Section 1.1. The numerical scheme extends in an obvious way to non-autonomous problems (1.1) by replacing $N(Y_{ni})$ with $N(t_n + c_i h, Y_{ni})$ and $N(y_{n-k})$ with $N(t_{n-k}, y_{n-k})$, respectively.

The preservation of equilibria of (2.1) is guaranteed under the following conditions

$$\begin{aligned} &\sum_{i=1}^{s} B_i(hL) + \sum_{k=1}^{q-1} V_k(hL) = \varphi_1(hL), \\ &\sum_{j=1}^{i-1} A_{ij}(hL) + \sum_{k=1}^{q-1} U_{ik}(hL) = c_i\, \varphi_1(c_i hL), \qquad 1 \leq i \leq s. \end{aligned} \tag{2.3}$$

Moreover, these conditions also ensure the equivalence of our numerical methods for autonomous and non-autonomous problems. Throughout the paper, we tacitly assume (2.3) to be satisfied. We further suppose $U_{1k}(hL) = 0$ which implies $c_1 = 0$ and thus $Y_{n1} = y_n$.

c_2	$A_{21}(hL)$				$U_{21}(hL)$	$\ldots$	$U_{2,q-1}(hL)$
$\vdots$	$\vdots$	$\ddots$			$\vdots$		$\vdots$
c_s	$A_{s1}(hL)$	$\ldots$	$A_{s,s-1}(hL)$		$U_{s1}(hL)$	$\ldots$	$U_{s,q-1}(hL)$
	$B_1(hL)$	$\ldots$	$B_{s-1}(hL)$	$B_s(hL)$	$V_1(hL)$	$\ldots$	$V_{q-1}(hL)$

Table 2.1: The exponential general linear method (2.2) in tableau form.

The explicit exponential Runge–Kutta methods considered in Hochbruck and Ostermann [11], see also [6, 7, 12, 13, 19], are contained in our method class (2.2) when setting $q = 1$. The exponential Adams–Bashforth methods [3, 6, 16, 20] result from (2.2) for the special case of a single stage $s = 1$.

2.2 Order conditions

For deriving the order conditions for the method class (2.2), we assume the data in (2.1) to be sufficiently regular. In particular, we require that the nonlinearity

evaluated at the exact solution $f(t) = N\big(y(t)\big)$ is sufficiently often differentiable with respect to t for $0 < t < T$.

Substituting the exact solution values

$$\widehat{y}_n = y(t_n), \qquad \widehat{Y}_{ni} = y(t_n + c_i h), \qquad 1 \le i \le s, \quad n \ge 0, \tag{2.4}$$

into the numerical scheme (2.2) defines the defects of the internal stages

$$\begin{aligned} D_{ni} = \widehat{Y}_{ni} - \mathrm{e}^{c_i hL}\, \widehat{y}_n - h \sum_{j=1}^{i-1} A_{ij}(hL)\, f(t_n + c_j h) \\ - h \sum_{k=1}^{q-1} U_{ik}(hL)\, f(t_{n-k}), \qquad 1 \le i \le s, \end{aligned} \tag{2.5a}$$

and the defect of the numerical solution

$$\begin{aligned} d_{n+1} = \widehat{y}_{n+1} - \mathrm{e}^{hL}\, \widehat{y}_n - h \sum_{i=1}^{s} B_i(hL)\, f(t_n + c_i h) \\ - h \sum_{k=1}^{q-1} V_k(hL)\, f(t_{n-k}), \qquad n \ge q-1. \end{aligned} \tag{2.5b}$$

We next make use of the representation for the exact solution values given in Lemma 1.1 and further expand the nonlinear term in a Taylor series, see (1.3). This leads to the following expansions for the defects of the internal stages

$$\begin{aligned} D_{ni} &= \sum_{\ell=1}^{Q} h^\ell\, \Theta_{\ell i}(hL)\, f^{(\ell-1)}(t_n) + R_{ni}^{(Q)}, \\ \Theta_{\ell i}(hL) &= c_i^\ell\, \varphi_\ell(c_i hL) - \sum_{j=1}^{i-1} \frac{c_j^{\ell-1}}{(\ell-1)!}\, A_{ij}(hL) - \sum_{k=1}^{q-1} \frac{(-k)^{\ell-1}}{(\ell-1)!}\, U_{ik}(hL). \end{aligned} \tag{2.6a}$$

Likewise, the numerical solution defect equals

$$\begin{aligned} d_{n+1} &= \sum_{\ell=1}^{P} h^\ell\, \vartheta_\ell(hL)\, f^{(\ell-1)}(t_n) + r_{n+1}^{(P)}, \\ \vartheta_\ell(hL) &= \varphi_\ell(hL) - \sum_{i=1}^{s} \frac{c_i^{\ell-1}}{(\ell-1)!}\, B_i(hL) - \sum_{k=1}^{q-1} \frac{(-k)^{\ell-1}}{(\ell-1)!}\, V_k(hL). \end{aligned} \tag{2.6b}$$

The remainders are defined through

$$\begin{aligned} R_{ni}^{(Q)} &= R_n(Q, c_i h) - h \sum_{j=1}^{i-1} A_{ij}(hL)\, S_n(Q, c_j h) \\ &\quad - h \sum_{k=1}^{q-1} U_{ik}(hL)\; S_n(Q, -kh), \end{aligned} \tag{2.6c}$$

$$r_{n+1}^{(P)} = R_n(P, h) - h \sum_{i=1}^{s} B_i(hL)\, S_n(P, c_i h) - h \sum_{k=1}^{q-1} V_k(hL)\, S_n(P, -kh),$$

see Lemma 1.1 for the definition of R_n and S_n.

The numerical scheme (2.2) is said to be of *stage order* Q and *quadrature order* P if $D_{ni} = \mathcal{O}(h^{Q+1})$ for $1 \le i \le s$ and $d_{n+1} = \mathcal{O}(h^{P+1})$. That is, requiring $\Theta_{\ell i}(hL) = 0$ for $1 \le i \le s$ and $1 \le \ell \le Q$ as well as $\vartheta_\ell(hL) = 0$ for $1 \le \ell \le P$, we obtain the order conditions

$$c_i^\ell \, \varphi_\ell(c_i hL) = \sum_{j=1}^{i-1} \frac{c_j^{\ell-1}}{(\ell-1)!} \, A_{ij}(hL) + \sum_{k=1}^{q-1} \frac{(-k)^{\ell-1}}{(\ell-1)!} \, U_{ik}(hL), \qquad 1 \le i \le s, \quad 1 \le \ell \le Q, \tag{2.7a}$$

$$\varphi_\ell(hL) = \sum_{i=1}^{s} \frac{c_i^{\ell-1}}{(\ell-1)!} \, B_i(hL) + \sum_{k=1}^{q-1} \frac{(-k)^{\ell-1}}{(\ell-1)!} \, V_k(hL), \qquad 1 \le \ell \le P. \tag{2.7b}$$

Here, by definition $c_i^0 = 1$ for all $1 \le i \le s$.

In Section 3, we will show that the convergence order of explicit exponential general linear methods (2.2) when applied to parabolic problems (2.1) is essentially $p = \min\{P, Q+1\}$. Therefore, it is desirable to construct numerical schemes of high stage order.

3 Parabolic evolution equations

In this section, we provide a convergence analysis for explicit exponential general linear methods within the framework of abstract semilinear parabolic evolution equations. For a thorough treatment of the theory of sectorial operators and analytic semigroups, we refer to the monographs [8, 14, 18].

3.1 Analytical framework

Let X be a complex Banach space endowed with the norm $\|\cdot\|_X$ and $D \subset X$ another densely embedded Banach space. For any $0 < \vartheta < 1$ we denote by X_ϑ some intermediate space between $D = X_1$ and $X = X_0$ such that the norm in X_ϑ fulfills the relation

$$\|x\|_{X_\vartheta} \le C \|x\|_D^\vartheta \|x\|_X^{1-\vartheta}, \qquad x \in D, \qquad 0 < \vartheta < 1,$$

with a constant $C > 0$. Examples are real interpolation spaces, see Lunardi [14], or fractional power spaces, see Henry [8].

We consider initial value problems of the form (2.1) where the right-hand side of the differential equation is defined by a linear operator $L : D \to X$ and a sufficiently regular nonlinear map

$$N : X_\alpha \to X : v \mapsto N(v), \qquad D \subset X_\alpha \subset X, \qquad 0 \le \alpha < 1. \tag{3.1}$$

This requirement together with Hypothesis 3.1 renders (2.1) a semilinear parabolic problem.

HYPOTHESIS 3.1. *We assume that the closed and densely defined linear operator $L : D \to X$ is sectorial. Thus, there exist constants $a \in \mathbb{R}$, $0 < \phi < \pi/2$, and $M \geq 1$ such that L satisfies the resolvent condition*

$$\left\| \left(\lambda I - L\right)^{-1} \right\|_{X \leftarrow X} \leq \frac{M}{|\lambda - a|}, \qquad \lambda \in \mathbb{C} \setminus S_\phi(a), \tag{3.2}$$

on the complement of the sector $S_\phi(a) = \left\{\lambda \in \mathbb{C} : \left|\arg(a - \lambda)\right| \leq \phi\right\} \cup \{a\}$. Moreover, we suppose that the graph norm of L and the norm in D are equivalent, that is, the estimate

$$C^{-1}\|x\|_D \leq \|x\|_X + \|Lx\|_X \leq C\|x\|_D, \qquad x \in D, \tag{3.3}$$

is valid for a constant $C > 0$.

From the results in [8, 14, 18] it is well-known that the sectorial operator L is the infinitesimal generator of an analytic semigroup $\left(\mathrm{e}^{tL}\right)_{t \geq 0}$ on the underlying Banach space X. Precisely, for $L : D \to X$ sectorial and any positive t the linear operator $\mathrm{e}^{tL} : X \to X$ is given by Cauchy's integral formula

$$\mathrm{e}^{tL} = \frac{1}{2\pi \mathrm{i}} \int_\Gamma \mathrm{e}^{\lambda} \left(\lambda I - tL\right)^{-1} \mathrm{d}\lambda, \qquad t > 0, \tag{3.4}$$

with Γ denoting a path that surrounds the spectrum of L. Especially, if $t = 0$ one defines $\mathrm{e}^{tL} = I$. By means of (3.2), it is shown that the estimate

$$\left\| t^{\nu - \mu} \mathrm{e}^{tL} \right\|_{X_\nu \leftarrow X_\mu} \leq C, \qquad 0 \leq t \leq T, \qquad 0 \leq \mu \leq \nu \leq 1, \tag{3.5}$$

is valid with a constant $C > 0$, see [14, Prop. 2.3.1]. Furthermore, for the functions defined in (1.2) the same type of bound

$$\left\| t^{\nu - \mu} \varphi_\ell(tL) \right\|_{X_\nu \leftarrow X_\mu} \leq C, \qquad 0 \leq t \leq T, \qquad 0 \leq \mu \leq \nu \leq 1, \tag{3.6}$$

follows for every $\ell \geq 1$.

REMARK 3.2. Under the assumption that the nonlinear map N in (3.1) is locally Lipschitz-continuous

$$\left\| N(v) - N(w) \right\|_X \leq C(\varrho) \, \|v - w\|_{X_\alpha}, \qquad \|v\|_{X_\alpha} + \|w\|_{X_\alpha} \leq \varrho, \tag{3.7}$$

the existence and uniqueness of a local solution of the semilinear parabolic problem (2.1), with initial value $y(0) \in X_\alpha$, is guaranteed. Moreover, the solution is represented by the variation-of-constants formula (1.4), see [8, Sect. 3.3].

The following example can be cast into our abstract framework of semilinear parabolic problems. For simplicity and in view of our numerical experiments, we restrict ourselves to one space dimension. Accordingly to Henry [8], X_ϑ denotes a fractional power space.

EXAMPLE 3.3. We consider the following initial-boundary value problem for a real-valued function $Y : [0, 1] \times [0, T] \to \mathbb{R}$ comprising a semilinear partial

differential equation subject to a homogeneous Dirichlet boundary condition and an additional initial condition

$$\begin{aligned} &\partial_t Y(x,t) = \mathscr{L}(x)\, Y(x,t) + f\big(x, Y(x,t), \partial_x Y(x,t)\big), \\ &Y(0,t) = 0 = Y(1,t), \quad Y(x,0) = Y_0(x), \qquad 0 \le x \le 1, \quad 0 \le t \le T. \end{aligned} \tag{3.8a}$$

Here, the second-order strongly elliptic differential operator

$$\mathscr{L}(x) = \alpha(x)\, \partial_{xx} + \beta(x)\, \partial_x + \gamma(x) \tag{3.8b}$$

involves the coefficients $\alpha, \beta, \gamma : [0,1] \to \mathbb{R}$ which we require to be sufficiently smooth, and, in particular, $\alpha(x)$ has to be positive and bounded away from 0. Besides, we suppose the function f to be regular in all variables and to satisfy a certain growth condition in the third argument, see Henry [8, Example 3.6].

By defining a linear operator L and a map N through

$$\big(L\, v\big)(x) = \mathscr{L}(x)\, v(x), \quad \big(N(v)\big)(x) = f\big(x, v(x), \partial_x v(x)\big), \qquad v \in \mathscr{C}_0^\infty(0,1),$$

the above initial-boundary value problem takes the form of an initial value problem (2.1) for $\big(y(t)\big)(x) = Y(x,t)$. The results in [8] imply that L, when considered as an unbounded operator on the Hilbert space $X = L^2(0,1)$, satisfies Hypothesis 3.1 with $D = H^2(0,1) \cap H_0^1(0,1)$. Further, a suitable choice for the domain of the nonlinearity is the Sobolev space $X_\alpha = X_{1/2} = H_0^1(0,1)$.

3.2 Global error relation

Under the requirements of Section 3.1 on the initial value problem (2.1), we analyze the convergence behavior of the method class (2.2). We start with deriving a useful relation for the global error.

The errors of the numerical solution values and the internal stages, respectively, are defined through

$$e_n = \widehat{y}_n - y_n, \qquad E_{ni} = \widehat{Y}_{ni} - Y_{ni}, \qquad 1 \le i \le s,$$

see (2.4). Moreover, we introduce the abbreviations

$$\Delta N_n = N(\widehat{y}_n) - N(y_n), \qquad \Delta N_{ni} = N(\widehat{Y}_{ni}) - N(Y_{ni}), \qquad 1 \le i \le s.$$

Comparing formulas (2.2) and (2.5), we receive for $n \ge q-1$

$$E_{ni} = \mathrm{e}^{c_i hL} e_n + h \sum_{j=1}^{i-1} A_{ij}(hL)\, \Delta N_{nj} + h \sum_{k=1}^{q-1} U_{ik}(hL)\, \Delta N_{n-k} + D_{ni}, \tag{3.9a}$$

$$e_{n+1} = \mathrm{e}^{hL} e_n + h \sum_{i=1}^{s} B_i(hL)\, \Delta N_{ni} + h \sum_{k=1}^{q-1} V_k(hL)\, \Delta N_{n-k} + d_{n+1}. \tag{3.9b}$$

Resolving the recurrence formula for e_n leads to

$$\begin{aligned} e_n = {} & \mathrm{e}^{(t_n - t_{q-1})L} e_{q-1} + \sum_{\ell=q}^{n} \mathrm{e}^{(t_n - t_\ell)L} d_\ell + h \sum_{\ell=q-1}^{n-1} \mathrm{e}^{(t_n - t_{\ell+1})L} \\ & \times \Bigg(\sum_{i=1}^{s} B_i(hL)\, \Delta N_{\ell i} + \sum_{k=1}^{q-1} V_k(hL)\, \Delta N_{\ell-k} \Bigg), \qquad n \geq q-1. \end{aligned} \tag{3.10}$$

In Section 3.3, we exploit the above error relation under certain requirements on the order of the method and the smoothness properties of the nonlinearity.

3.3 Convergence estimates

Throughout, we employ the assumption that the starting values $y_0, y_1, \ldots, y_{q-1}$ have been computed using some starting procedure and that they belong to X_α. Further, we suppose that the method coefficients are sufficiently regular and satisfy

$$\begin{aligned} & \big\|A_{ij}(hL)\big\|_{X_\nu \leftarrow X_\mu} + \big\|B_i(hL)\big\|_{X_\nu \leftarrow X_\mu} + \big\|U_{ik}(hL)\big\|_{X_\nu \leftarrow X_\mu} \\ & \qquad + \big\|V_k(hL)\big\|_{X_\nu \leftarrow X_\mu} \leq C h^{-\nu+\mu}, \qquad h > 0, \qquad 0 \leq \mu \leq \nu \leq 1. \end{aligned} \tag{3.11}$$

In particular, the exponential general linear methods considered in Section 4 fulfill these requirements, see (1.2) and (3.6). As before, we set $f(t) = N\big(y(t)\big)$ and denote $\big\|f\big\|_{X_\vartheta,\infty} = \max\big\{\big\|f(t)\big\|_{X_\vartheta} : 0 \leq t \leq T\big\}$ for $0 \leq \vartheta \leq 1$.

It is straightforward to deduce the following convergence result from the global error relation (3.10).

THEOREM 3.4. *Under the requirements of Hypothesis 3.1, assume that the explicit exponential general linear method* (2.2) *applied to the initial value problem* (2.1) *satisfies* (3.11) *and further fulfills the order conditions* (2.7). *Suppose that* $f^{(Q)}(t) \in X_\beta$ *for some* $0 \leq \beta \leq \alpha$ *and* $f^{(P)}(t) \in X$. *Then, for stepsizes* $h > 0$ *the estimate*

$$\begin{aligned} \big\|y(t_n) - y_n\big\|_{X_\alpha} \leq {} & C \sum_{\ell=0}^{q-1} \big\|y(t_\ell) - y_\ell\big\|_{X_\alpha} + C h^{Q+1-\alpha+\beta} \sup_{0 \leq t \leq t_n} \big\|f^{(Q)}(t)\big\|_{X_\beta} \\ & + C h^P \sup_{0 \leq t \leq t_n} \big\|f^{(P)}(t)\big\|_X, \qquad t_q \leq t_n \leq T, \end{aligned}$$

holds with a constant $C > 0$ *independent of* n *and* h.

PROOF. We estimate (3.10) in the domain of the nonlinear term and obtain

$$\begin{aligned} \big\|e_n\big\|_{X_\alpha} \leq {} & \big\|\mathrm{e}^{(t_n - t_{q-1})L}\big\|_{X_\alpha \leftarrow X_\alpha} \big\|e_{q-1}\big\|_{X_\alpha} + \Big\| \sum_{\ell=q}^{n} \mathrm{e}^{(t_n - t_\ell)L} d_\ell \Big\|_{X_\alpha} \\ & + h \sum_{\ell=q-1}^{n-1} \sum_{i=1}^{s} \big\|\mathrm{e}^{(t_n - t_{\ell+1})L}\, B_i(hL)\big\|_{X_\alpha \leftarrow X} \big\|\Delta N_{\ell i}\big\|_X \\ & + h \sum_{\ell=q-1}^{n-1} \sum_{k=1}^{q-1} \big\|\mathrm{e}^{(t_n - t_{\ell+1})L}\, V_k(hL)\big\|_{X_\alpha \leftarrow X} \big\|\Delta N_{\ell-k}\big\|_X. \end{aligned}$$

Consequently, using the bound (3.5) for the analytic semigroup, relation (3.11) and the Lipschitz-property (3.7), we receive

$$\begin{aligned}\|e_n\|_{X_\alpha} &\le C\|e_{q-1}\|_{X_\alpha} + \Big\|\sum_{\ell=q}^{n} \mathrm{e}^{(t_n-t_\ell)L} d_\ell\Big\|_{X_\alpha} \\ &\quad + Ch\sum_{\ell=q-1}^{n-1}(t_n-t_\ell)^{-\alpha}\Big(\sum_{i=1}^{s}\|E_{\ell i}\|_{X_\alpha} + \sum_{k=1}^{q-1}\|e_{\ell-k}\|_{X_\alpha}\Big).\end{aligned} \tag{3.12}$$

For the error of the internal stages (3.9a), measured in the norm of X_α, we have

$$\begin{aligned}\|E_{\ell i}\|_{X_\alpha} &\le \|\mathrm{e}^{c_i hL}\|_{X_\alpha\leftarrow X_\alpha}\|e_\ell\|_{X_\alpha} + h\sum_{j=1}^{i-1}\|A_{ij}(hL)\|_{X_\alpha\leftarrow X}\|\Delta N_{\ell j}\|_X \\ &\quad + h\sum_{k=1}^{q-1}\|U_{ik}(hL)\|_{X_\alpha\leftarrow X}\|\Delta N_{\ell-k}\|_X + \|D_{\ell i}\|_{X_\alpha}.\end{aligned}$$

Using again (3.5), (3.7), and (3.11), the bound

$$\|E_{\ell i}\|_{X_\alpha} \le C\|e_\ell\|_{X_\alpha} + Ch^{1-\alpha}\sum_{j=1}^{i-1}\|E_{\ell j}\|_{X_\alpha} + Ch^{1-\alpha}\sum_{k=1}^{q-1}\|e_{\ell-k}\|_{X_\alpha} + \|D_{\ell i}\|_{X_\alpha}$$

and therefore the estimate

$$\|E_{\ell i}\|_{X_\alpha} \le C\|e_\ell\|_{X_\alpha} + Ch^{1-\alpha}\sum_{k=1}^{q-1}\|e_{\ell-k}\|_{X_\alpha} + C\sum_{j=1}^{i}\|D_{\ell j}\|_{X_\alpha}$$

follows. The constant $C > 0$ in particular depends on T, but is independent of h. Inserting this relation into (3.12), leads to

$$\begin{aligned}\|e_n\|_{X_\alpha} &\le C\|e_{q-1}\|_{X_\alpha} + Ch\sum_{\ell=0}^{n-1}(t_n-t_\ell)^{-\alpha}\|e_\ell\|_{X_\alpha} \\ &\quad + Ch\sum_{\ell=q-1}^{n-1}\sum_{i=1}^{s}(t_n-t_\ell)^{-\alpha}\|D_{\ell i}\|_{X_\alpha} + \Big\|\sum_{\ell=q}^{n}\mathrm{e}^{(t_n-t_\ell)L}d_\ell\Big\|_{X_\alpha}.\end{aligned} \tag{3.13}$$

It remains to estimate the terms involving the defects (2.6). From the assumption that the stage order conditions (2.7a) are fulfilled, it follows $D_{\ell i} = R_{\ell i}^{(Q)}$ for $1 \le \mathrm{i} \le s$. Therefore, provided that the Q-th order derivative of the map f is bounded in X_β, by (3.5) and (3.11), we obtain

$$\begin{aligned}\|R_{\ell i}^{(Q)}\|_{X_\alpha} &\le \|R_\ell(Q, c_i h)\|_{X_\alpha} + h\sum_{j=1}^{i-1}\|A_{ij}(hL)\|_{X_\alpha\leftarrow X_\beta}\|S_\ell(Q, c_j h)\|_{X_\beta} \\ &\quad + h\sum_{k=1}^{q-1}\|U_{ik}(hL)\|_{X_\alpha\leftarrow X_\beta}\|S_\ell(Q,-kh)\|_{X_\beta} \\ &\le Ch^{Q+1-\alpha+\beta}\|f^{(Q)}\|_{X_\beta,\infty}, \qquad 1 \le i \le s,\end{aligned}$$

see also (2.6c) and Lemma 1.1. Moreover, the validity of the order conditions (2.7b) implies $d_\ell = r_\ell^{(P)}$. It then holds

$$\begin{aligned}\big\|r_\ell^{(P)}\big\|_X &\le \big\|R_{\ell-1}(P,h)\big\|_X + h\sum_{i=1}^{s}\big\|B_i(hL)\big\|_{X\leftarrow X}\big\|S_{\ell-1}(P,c_ih)\big\|_X \\ &\quad + h\sum_{k=1}^{q-1}\big\|V_k(hL)\big\|_{X\leftarrow X}\big\|S_{\ell-1}(P,-kh)\big\|_X \\ &\le Ch^{P+1}\big\|f^{(P)}\big\|_{X,\infty}.\end{aligned}$$

Similarly, we obtain $\big\|r_n^{(P)}\big\|_{X_\alpha} \le Ch^{P+1-\alpha}\big\|f^{(P)}\big\|_{X,\infty}$. Thus, a direct estimation of the last sum in (3.13) gives

$$\begin{aligned}&\sum_{\ell=q}^{n-1}\big\|\mathrm{e}^{(t_n-t_\ell)L}\big\|_{X_\alpha\leftarrow X}\big\|r_\ell^{(P)}\big\|_X + \big\|r_n^{(P)}\big\|_{X_\alpha} \\ &\qquad\le Ch^{P+1}\sum_{\ell=q}^{n-1}(t_n-t_\ell)^{-\alpha}\,\big\|f^{(P)}\big\|_{X,\infty}.\end{aligned} \tag{3.14}$$

We insert the above estimates in (3.13) and interprete the arising sums as Riemann-sums and bound it by the corresponding integrals. From a Gronwall-type inequality with a weakly singular kernel, see [4, 17], the result follows. □

The example methods given in Section 4 comprise explicit exponential general linear methods with high stage order $Q = P-1$. In many practical examples, the exact solution of (2.1) and the map N defining the nonlinearity are sufficiently often differentiable. That is, the assumptions $f^{(P)}(t) \in X$ and $f^{(P-1)}(t) \in X_\alpha$ are fulfilled for all $0 \le t \le T$. Therefore, the convergence order predicted by Theorem 3.4 is $p = P$. This result is also confirmed by the numerical examples presented in Section 5.

Remark 3.5. Let $\alpha = 0$ and L be the generator of a $\mathscr{C}_0$-semigroup, see [18]. Then, the bound $\big\|\varphi_j(tL)\big\|_{X\leftarrow X} \le C$ is valid for finite times $0 \le t \le T$ and any $j \ge 0$. Returning to the above proof shows that the convergence estimate of Theorem 3.4 remains valid for $\mathscr{C}_0$-semigroups with the choice $\beta = 0$.

The following result shows that for parabolic problems it suffices to satisfy, instead of (2.7b), the weakened quadrature order conditions

$$\begin{aligned}\varphi_\ell(hL) &= \sum_{i=1}^{s}\frac{c_i^{\ell-1}}{(\ell-1)!}\,B_i(hL) + \sum_{k=1}^{q-1}\frac{(-k)^{\ell-1}}{(\ell-1)!}\,V_k(hL), \quad 1 \le \ell \le P-1, \\ \frac{1}{P} &= \sum_{i=1}^{s}c_i^{P-1}\,B_i(0) + \sum_{k=1}^{q-1}(-k)^{P-1}\,V_k(0),\end{aligned} \tag{3.15}$$

to obtain the full convergence order $p = P$. That is, the condition where $\ell = P$ is fulfilled for $L = 0$, but not necessarily for arbitrary arguments, see also (1.2).

THEOREM 3.6. *Assume that the requirements of Hypothesis 3.1 are valid and that the explicit exponential general linear method* (2.2) *fulfills* (3.11). *Further, suppose that the stage order conditions* (2.7a) *and the weak quadrature order conditions* (3.15) *are valid for* $Q = P - 1$ *and that* $0 \le \beta \le \alpha$. *Then, for stepsizes* $h > 0$ *the estimate*

$$\begin{aligned}\big\|y(t_n) - y_n\big\|_{X_\alpha} \le C \sum_{\ell=0}^{q-1} \big\|y(t_\ell) - y_\ell\big\|_{X_\alpha} + Ch^{P-\alpha+\beta} \sup_{0\le t\le t_n} \big\|f^{(P-1)}(t)\big\|_{X_\beta} \\ + Ch^P \sup_{0\le t\le t_n} \big\|f^{(P)}(t)\big\|_X, \qquad t_q \le t_n \le T,\end{aligned}$$

holds with a constant $C > 0$ *independent of* n *and* h, *provided that the quantities on the right-hand side are well-defined.*

PROOF. The derivation of the above result follows the lines of the proof of Theorem 3.4. For simplicity, we assume $\beta = \alpha$. It suffices to derive a refined bound for the last sum in (3.13) involving the numerical solution defects d_ℓ. Under the weak order conditions (3.15), the representation

$$d_\ell = h^P\, s_\ell^{(P)} + r_\ell^{(P)}, \qquad s_\ell^{(P)} = \vartheta_P(hL)\, f^{(P-1)}(t_{\ell-1}),$$

is valid, see (2.6). The remainder is estimated in the same way as before and yields a contribution of $Ch^P \big\|f^{(P)}\big\|_{X,\infty}$ in the convergence bound, see (3.14). We need to show that the sum

$$S = \sum_{\ell=q}^{n} \mathrm{e}^{(t_n - t_\ell)L}\, s_\ell^{(P)},$$

when measured in the norm of X_α, is bounded by a constant. For that purpose, we employ Abel's partial summation formula to obtain the identity

$$S = \mathscr{E}_n\, s_n^{(P)} - \mathscr{E}_{q-1}\, s_q^{(P)} - \sum_{\ell=q}^{n-1} \mathscr{E}_\ell \Big(s_{\ell+1}^{(P)} - s_\ell^{(P)} \Big), \qquad \mathscr{E}_\ell = \sum_{j=0}^{\ell} \mathrm{e}^{(t_n - t_j)L}.$$

We notice that the second condition in (3.15) implies $\vartheta_P(0) = 0$, and, by Cauchy's integral formula (3.4), we further receive $\vartheta_P(hL) = hL\,\psi(hL)$. In particular, if the method coefficients are (linear) combinations of the φ-functions, the linear operator ψ is bounded on X. The bound

$$\begin{aligned}\big\|hL\,\mathscr{E}_\ell\,\psi(hL)\big\|_{X_\nu \leftarrow X_\mu} &\le \big\|\mathrm{e}^{(t_n - t_\ell)L}\big\|_{X_\nu \leftarrow X_\mu} \Big\|hL \sum_{j=0}^{\ell} \mathrm{e}^{(t_\ell - t_j)L}\,\psi(hL)\Big\|_{X_\mu \leftarrow X_\mu} \\ &\le C(t_n - t_\ell)^{-\nu+\mu}, \qquad q-1 \le \ell < n, \qquad 0 \le \mu \le \nu < 1,\end{aligned}$$

follows by Cauchy's integral formula, see also [10, Lemma1.1]. Further, it holds

$$\big\|hL\,\mathscr{E}_n\,\psi(hL)\big\|_{X_\nu \leftarrow X_\mu} \le Ch^{-\nu+\mu}, \qquad 0 \le \mu \le \nu < 1.$$

As a consequence, we obtain the estimate

$$\begin{aligned}\|S\|_{X_\alpha} &\le \|hL\,\mathscr{E}_n\,\psi(hL)\|_{X_\alpha\leftarrow X_\alpha}\|f^{(P-1)}(t_{n-1})\|_{X_\alpha}\\ &\quad+\|hL\,\mathscr{E}_{q-1}\,\psi(hL)\|_{X_\alpha\leftarrow X_\alpha}\|f^{(P-1)}(t_{q-1})\|_{X_\alpha}\\ &\quad+\sum_{\ell=q}^{n-1}\|hL\,\mathscr{E}_\ell\,\psi(hL)\|_{X_\alpha\leftarrow X}\int_0^h\|f^{(P)}(t_{\ell-1}+\xi)\|_X\,\mathrm{d}\xi\\ &\le C\|f^{(P-1)}\|_{X_\alpha,\infty}+C\|f^{(P)}\|_{X,\infty}\end{aligned}$$

which yields the desired result. □

4 Example methods

In this section, we construct explicit exponential general linear methods (2.2) which have a favorable convergence behavior. We mainly focus on two-stage schemes with stage order $Q = P - 1$ where by Theorem 3.4 the full convergence order $p = P$ is ensured for abstract evolution equations. In the subsequent Section 5, the schemes are tested numerically on a semilinear parabolic initial-boundary value problem. Further, a table comparing the computational effort of various exponential integrators is included there, see Table 5.1. Among others, we count the number of evaluations of the nonlinear map N required at each step. However, by making use of the previous steps, the number of function evaluations can be reduced considerably.

Henceforth, for notational simplicity, we set $z = hL$ and $\varphi_{ij} = \varphi_i(c_j z)$. As well, we occasionally omit the argument in the method coefficient functions and write $A_{ij} = A_{ij}(z)$ etc.

4.1 Two-stage schemes

We first discuss explicit exponential general linear methods (2.2) with $s = 2$. Requiring the quadrature order and stage order conditions (2.7) to be fulfilled for $q + 1 = p = P = Q + 1$ determines the coefficients of the numerical scheme up to a free parameter, as the following result shows.

THEOREM 4.1. *For any $0 < c_2 \le 1$ there exists a unique explicit exponential general linear method of the form* (2.2) *with two stages and q steps that is convergent of order $p = q + 1$ for abstract parabolic problems* (2.1).

PROOF. The stage order conditions (2.7a) yield the following linear equations in the unknowns A_{21} and U_{2k}

$$A_{21}+\sum_{k=1}^{p-2}U_{2k}=c_2\,\varphi_{12},\quad \sum_{k=1}^{p-2}\frac{(-k)^{\ell-1}}{(\ell-1)!}\,U_{2k}=c_2^\ell\,\varphi_{\ell 2},\quad 2\le\ell\le p-1.$$

As this system is of Vandermonde form, it possesses a unique solution which depends on $0 < c_2 \le 1$. Similarly, the order conditions (2.7b)

$$B_1+B_2+\sum_{k=1}^{p-2}V_k=\varphi_1,\qquad \frac{c_2^{\ell-1}}{(\ell-1)!}\,B_2+\sum_{k=1}^{p-2}\frac{(-k)^{\ell-1}}{(\ell-1)!}\,V_k=\varphi_\ell,\quad 2\le\ell\le p,$$

uniquely determine the coefficient functions B_i and V_k. By Theorem 3.4, the order of convergence for abstract evolution equations (2.1) equals p, provided that the nonlinear term f satisfies suitable regularity assumptions. □

To minimize the number of φ-function evaluations, we choose to set the parameter $c_2 = 1$. The resulting methods EGLM*psq* can be considered as generalizations of the PEC schemes using a generalized Adams–Bashforth predictor of order $p-1$ and a generalized Adams–Moulton corrector of order p.

1	φ_1	
	$\varphi_1 - \varphi_2$	φ_2

1	$\varphi_1 + \varphi_2$		$-\varphi_2$
	$\varphi_1 - 2\,\varphi_3$	$\frac{1}{2}\,\varphi_2 + \varphi_3$	$-\frac{1}{2}\,\varphi_2 + \varphi_3$

Table 4.1: Coefficients of EGLM221 (left) and EGLM322 (right) for $c_2 = 1$.

Order 2. To achieve order two, one has to satisfy the order conditions given in the proof of Theorem 4.1 with $p = 2$. To this set of equations, the uniquely determined solution is an exponential Runge–Kutta method with coefficients given in Table 4.1. The scheme EGLM221 requires three φ-function evaluations, four matrix-vector products, and two function evaluations of the nonlinear map N, provided that the values of the previous step are available.

Order 3. For convergence of order $p = 3$, the resulting two-stage two-step method EGLM322 requires four φ-function evaluations, six matrix-vector products, and two new function evaluations, see Table 4.1.

1	$\varphi_1 + \frac{3}{2}\,\varphi_2 + \varphi_3$		$-2\,\varphi_2 - 2\,\varphi_3$	$\frac{1}{2}\,\varphi_2 + \varphi_3$
	$\varphi_1 + \frac{1}{2}\,\varphi_2 - 2\,\varphi_3 - 3\,\varphi_4$	$\frac{1}{3}\,\varphi_2 + \varphi_3 + \varphi_4$	$-\varphi_2 + \varphi_3 + 3\,\varphi_4$	$\frac{1}{6}\,\varphi_2 - \varphi_4$

Table 4.2: Coefficients of EGLM423 for $c_2 = 1$.

Order 4. The two-stage three-step method EGLM423 with coefficients given in Table 4.2 is convergent of order $p = 4$ and requires five φ-function evaluations, eight matrix-vector products, and two function evaluations.

A MAPLE code for generating the coefficients of the schemes EGLM*psq* involving $s = 2$ stages and $q = p - 1$ steps is downloadable from the webpage `http://www.math.ntnu.no/num/expint/`.

4.2 Schemes involving $s \geq 3$ stages

For explicit exponential general linear methods (2.2) involving $s \geq 3$ stages, contrary to two-stage schemes, there is some freedom available in the choice of the method. This makes it feasible to suitably weight desirable properties of the numerical scheme such as stability, small error coefficients, the number of φ-function evaluations, matrix-vector products, or function evaluations. Another possibility would be to use the extra freedom available to increase the convergence order of the scheme. This involves a thorough investigation of the global error (3.10) in the lines of Hochbruck and Ostermann [11] which is beyond the scope of the present work.

For the purpose of illustration, however, we briefly describe the construction of a scheme involving $s = 3$ stages and $q = 2$ steps. For simplicity, we now assume that the nonlinear map N defining the right-hand side of the differential equation in (2.1) is defined on $X_\alpha = X$, see (3.1). We require the weak quadrature order conditions (3.15) and the stage order conditions (2.7a) to be fulfilled for $P = 4$ and $Q = 2$. This implies that the defects of the internal stages are of the form

$$D_{ni} = h^3\,\Theta_{3i}(hL)\,f''(t_n) + R_{ni}^{(4)},$$
$$\Theta_{3i}(hL) = c_i^3\,\varphi_3(c_i hL) - \sum_{j=1}^{i-1} \frac{c_j^2}{2}\,A_{ij}(hL) - \frac{1}{2}\,U_{i1}(hL),$$

see (2.6). A suitable relation for the error of the internal stages (3.9a) together with Taylor series expansions of the nonlinearity finally shows that the error (3.10), when measured in X, is bounded by Ch^4 provided that the term

$$\sum_{i=1}^{s} B_i(hL)\,N'(y_n)\,\Theta_{3i}(hL)$$

vanishes. Altogether, the conditions for the order of convergence $p = 4$ with respect to the norm in X are

$$\sum_{j=1}^{i-1} \frac{c_j^{\ell-1}}{(\ell-1)!}\,A_{ij}(hL) + \frac{(-1)^{\ell-1}}{(\ell-1)!}\,U_{i1}(hL) = c_i^\ell\,\varphi_\ell(c_i hL), \qquad 1 \le \ell \le 2, \tag{4.1a}$$

$$\sum_{i=1}^{3} \frac{c_i^{\ell-1}}{(\ell-1)!}\,B_i(hL) + \frac{(-1)^{\ell-1}}{(\ell-1)!}\,V_1(hL) = \varphi_\ell(hL), \qquad 1 \le \ell \le 3, \tag{4.1b}$$

$$\sum_{i=2}^{3} B_i(hL)\,J\,\Theta_{3i}(hL) = 0, \tag{4.1c}$$

$$\sum_{i=1}^{3} c_i^3\,B_i(0) - V_1(0) = \frac{1}{4}, \tag{4.1d}$$

where J is an arbitrary and bounded linear operator on X.

We note that it is not possible to achieve $\Theta_{32} = 0$. Therefore, in order to satisfy condition (4.1c), we set $B_2 = \kappa B_3$ for a scalar κ. Inserting this ansatz into (4.1c) results in $B_3\,J\left(\kappa\,\Theta_{32} + \Theta_{33}\right) = 0$. This condition can be satisfied either by setting $\kappa = 0$ or by choosing $c_2 = c_3$. In view of the computational effort required, we fulfill both which gives $B_2 = 0$ and $c_2 = c_3 = 7/10$. We refer to the resulting scheme as EGLM432. It requires eight φ-function evaluations, eight matrix-vector products, and three function evaluations.

We conclude this subsection with a brief remark on exponential *generalized Runge–Kutta–Lawson methods* which provided the initial motivation for considering exponential general linear methods of the form (2.2), see also [13, 15]. The basic idea for constructing these schemes is to replace the nonlinear part by

an interpolation polynomial and to perform the Lawson transformation involving the exponential function. Then, a classical explicit Runge–Kutta method is used on the transformed problem and the obtained numerical solution is finally transformed back into the original variable. However, the resulting numerical schemes are inferior to methods constructed directly from the order conditions.

As an example, we mention the four-stage three-step scheme GLRK34 which satisfies the order conditions (2.7) with $Q = P = 3$ and the weakened quadrature order conditions (3.15) for $P = 4$. Thus, the order of convergence is $p = 4$. A MAPLE code to generate the coefficients of the generalized Lawson methods is downloadable from the website `http://www.math.ntnu.no/num/expint/`. The numerical experiments in Section 5 show that GLRK34 is not competitive with EGLM432 when comparing the computational effort and the size of the error.

4.3 Multistep schemes

We conclude this section on example methods with a remark on explicit exponential multistep methods that are contained in our method class (2.2) by setting $s = 1$. Under the requirements $q = p = P = Q + 1$, the order conditions (2.7) simplify as follows

$$B_1 + \sum_{k=1}^{p-1} V_k = \varphi_1, \qquad \sum_{k=1}^{p-1} (-k)^{\ell-1} V_k = (\ell-1)!\,\varphi_\ell, \quad 2 \le \ell \le p,$$

and uniquely define the coefficients of the method. An alternative way for deriving these schemes is to represent the exact solution of (2.1) by means of the variation-of-constants formula

$$y(t_{n+1}) = \mathrm{e}^{hL}\, y(t_n) + \int_0^h \mathrm{e}^{(h-\tau)L}\, N\big(y(t_n+\tau)\big)\,\mathrm{d}\tau$$

and to replace the nonlinear map N with the interpolation polynomial through the points $\big(t_{n-i}, N(y_{n-i})\big)$ for $0 \le i \le q-1$. Such exponential Adams–Bashforth methods were considered in [3, 6, 16, 20]. As an illustration, we include the four-step method EGLM414

$$\begin{aligned} y_{n+1} = \mathrm{e}^{hL}\, y_n &+ h\, B_1(hL)\, N(y_n) + h\, V_1(hL)\, N(y_{n-1}) \\ &+ h\, V_2(hL)\, N(y_{n-2}) + h\, V_3(hL)\, N(y_{n-3}), \end{aligned}$$

with coefficient functions

$$\begin{aligned} B_1 &= \varphi_1 + \tfrac{11}{6}\,\varphi_2 + 2\,\varphi_3 + \varphi_4, & V_1 &= -3\,\varphi_2 - 5\,\varphi_3 - 3\,\varphi_4, \\ V_2 &= \tfrac{3}{2}\,\varphi_2 + 4\,\varphi_3 + 3\,\varphi_4, & V_3 &= -\tfrac{1}{3}\,\varphi_2 - \varphi_3 - \varphi_4, \end{aligned}$$

that is convergent of order $p = 4$ for abstract evolution equations and requires five φ-function evaluations, five matrix-vector products, and one function evaluation.

The exponential multistep schemes studied in Calvo and Palencia [5] are instead based on the following representation of the exact solution

$$y(t_{n+1}) = \mathrm{e}^{qhL}\, y(t_{n-q+1}) + \int_0^{qh} \mathrm{e}^{(qh-\tau)L}\, N\big(y(t_{n-q+1}+\tau)\big)\,\mathrm{d}\tau.$$

method	p	P	Q	s	q	$\#\varphi$	#vec	#fun
EGLM221	2	2	1	2	1	3	4	2
EGLM322	3	3	2	2	2	4	6	2
EMAM4	4	4	3	1	4	5	5	1
EGLM414	4	4	3	1	4	5	5	1
EGLM423	4	4	3	2	3	5	8	2
EGLM432	$4-\gamma$	4 (weak)	2	3	2	8	8	3
ERKM4	$4-\gamma$	4 (weak)	1	5	1	8	13	5
GLRK34	4	4 (weak)	3	4	3	8	16	4

Table 5.1: Computational effort of various exponential integrators (order of convergence p for Problem 5.1, quadrature order P, stage order Q, number of stages s, number of steps q, number of distinct φ-functions needed to be evaluated, number of required matrix-vector products, number of (new) function evaluations of the nonlinear map N per step).

For instance, the four-step method which we refer to as EMAM4

$$\begin{aligned} y_{n+1} = \mathrm{e}^{4hL}\, y_{n-3} &+ h\, B_1(4hL)\, N(y_n) + h\, V_1(4hL)\, N(y_{n-1}) \\ &+ h\, V_2(4hL)\, N(y_{n-2}) + h\, V_3(4hL)\, N(y_{n-3}), \end{aligned}$$

with coefficient functions

$$\begin{aligned} B_1 &= \tfrac{16}{3}\,\varphi_2 - 64\,\varphi_3 + 256\,\varphi_4, & V_1 &= -24\,\varphi_2 + 256\,\varphi_3 - 768\,\varphi_4, \\ V_2 &= 48\,\varphi_2 - 320\,\varphi_3 + 768\,\varphi_4, & V_3 &= 4\,\varphi_1 - \tfrac{88}{3}\,\varphi_2 + 128\,\varphi_3 - 256\,\varphi_4, \end{aligned}$$

retains the full convergence order $p = 4$ for semilinear parabolic problems (2.1) and requires the same computational effort as the fourth-order scheme EGLM414.

We note that the exponential Adams–Bashforth methods are zero-stable with parasitic roots equal to zero. They have thus superior stability properties compared to the methods considered in [5] which are only weakly stable in the sense that all parasitic roots for $y' = 0$ lie on the unit circle.

5 Numerical experiments

In this section, we illustrate the theoretical results given in Section 3.3 on the convergence behavior of explicit exponential general linear methods for abstract evolution equations. As test problem, we choose a one-dimensional semilinear parabolic initial-boundary value problem.

Problem 5.1 (Parabolic problem). We consider the following parabolic differential equation under a homogeneous Dirichlet boundary condition

$$\begin{aligned} \partial_t Y(x,t) &= \partial_{xx}\, Y(x,t) - \, Y(x,t)\, \partial_x Y(x,t) + \Phi(x,t), \\ Y(0,t) = Y(1,t) = 0, \quad Y(x,0) &= x(1-x), \qquad 0 \le x \le 1, \quad 0 \le t \le T, \end{aligned}$$

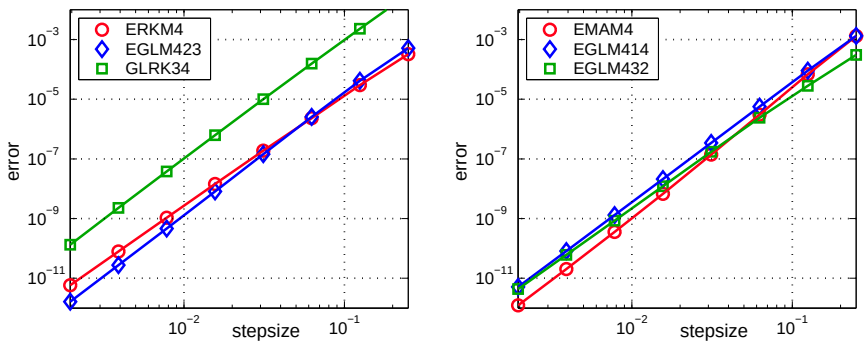


Figure 5.1: The numerically observed convergence orders of various explicit exponential integrators when applied to Problem 5.1. The error measured in a discrete H_0^1-norm is plotted versus the time stepsize.

where Φ is chosen such that the exact solution is $Y(x,t) = x(1-x)\,\mathrm{e}^t$.

As in Example 3.3, the above initial-boundary value problem is written as an abstract initial value problem of the form (1.1) for $\big(y(t)\big)(x) = Y(x,t)$ with linear operator L and nonlinearity N defined by

$$\big(L\,v\big)(x) = \partial_{xx}\,v(x), \qquad \big(N(t,v)\big)(x) = -\,v(x)\,\partial_x v(x) + \Phi(x,t),$$

for $v \in \mathscr{C}_0^\infty(0,1)$. A suitable choice for the underlying Banach space is the Hilbert space $X = L^2(0,1)$. Then, it holds $D = H^2(0,1) \cap H_0^1(0,1)$ and the domain of the nonlinearity N is equal to $[0,T]\times X_\alpha$ where $X_\alpha = X_{1/2} = H_0^1(0,1)$, see also Henry [8, Sect. 3.3].

We note that, accordingly to Lunardi [14, Sect. 7.3], an alternative choice is $X = \mathscr{C}(0,1)$, $D = \mathscr{C}_0^2(0,1)$, and $X_\alpha = X_{1/2} = \mathscr{C}_0^1(0,1)$. Here, we denote

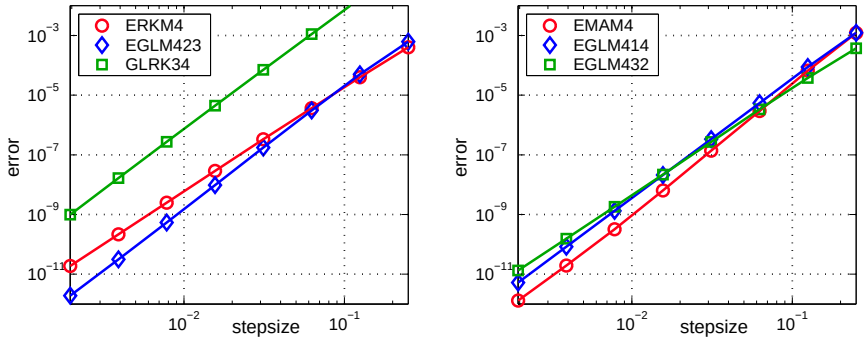


Figure 5.2: The numerically observed convergence orders of various explicit exponential integrators when applied to Problem 5.1. The error measured in a discrete $\mathscr{C}_0^1$-norm is plotted versus the time stepsize.

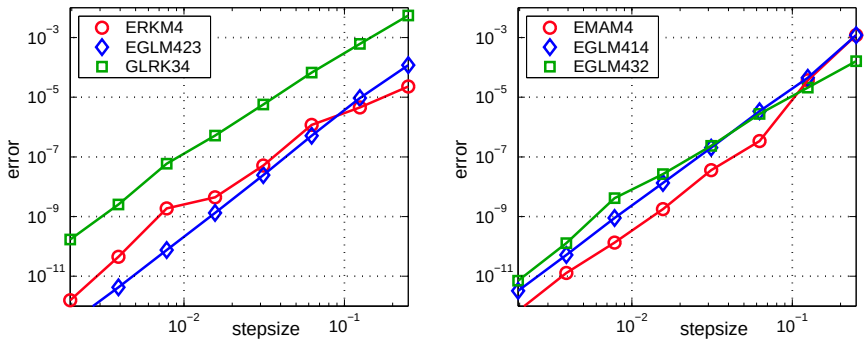


Figure 5.3: The numerically observed convergence orders of various explicit exponential integrators when applied to Problem 5.2. The error measured in a discrete L^2-norm is plotted versus the time stepsize.

$\mathscr{C}_0^k(0,1) = \{v \in \mathscr{C}^k(0,1) : v(0) = 0 = v(1)\}$ for $k = 1, 2$.

In order to solve Problem 5.1 numerically, we use a spatial discretization by standard finite differences of grid length $\Delta x = (M+1)^{-1}$ with $M = 200$. For various explicit exponential general linear methods discussed in Section 4, the resulting system of ordinary differential equations is integrated up to time $T = 1$. The numerical convergence orders with respect to a discrete X_α-norm are determined from the exact and numerical solution values.

The numerically observed convergence orders for the explicit exponential general linear methods are in exact agreement with the values expected from the theoretical results given in Section 3.3. For example, the schemes EGLM423, EGLM414, and GLRK34 show full order $p = 4$, see Figures 5.1-5.2. We point out that the scheme EGLM432 discussed in Section 4.2 and as well the exponential Runge–Kutta method ERKM4 considered in [11, Eq. (5.19)] suffer from a slight order reduction. The convergence order with respect to a discrete H_0^1-norm is approximately $p = 4 - \gamma$ with $\gamma = 1/4$. When the error is measured in a discrete $\mathscr{C}_0^1$-norm, an additional order reduction down to approximately $p = 4 - \gamma$ with $\gamma = 1/2$ is encountered. These fractional orders can be explained using arguments as in [11, Sect. 6].

In the following example, we illustrate the convergence behavior of our method class for an evolution equation which is governed by a $\mathscr{C}_0$-semigroup.

Problem 5.2 (Hyperbolic problem). We consider the hyperbolic initial-boundary value problem

$$\mathrm{i}\,\partial_t Y(x,t) = \partial_{xx}\, Y(x,t) + \frac{1}{1+Y(x,t)^2} + \Phi(x,t),$$
$$Y(0,t) = Y(1,t) = 0, \quad Y(x,0) = x(1-x), \qquad 0 \le x \le 1, \quad 0 \le t \le 1,$$

as abstract initial value problem on $X = L^2(0,1)$. The function Φ is determined such that the exact solution equals $Y(x,t) = x(1-x)\,\mathrm{e}^t$.

As before, we use standard finite differences of grid length $\Delta x = (M+1)^{-1}$

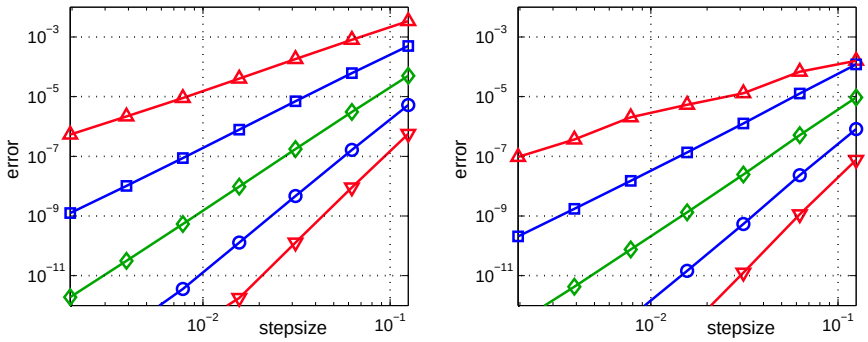


Figure 5.4: The numerically observed convergence orders of the two-stage schemes EGLM$p2q$ with $q = p - 1$ steps of orders $2 \le p \le 6$ when applied to Problem 5.1 (left) and Problem 5.2 (right). The error measured in a discrete $\mathscr{C}_0^1$-norm and L^2-norm, respectively, is plotted versus the time stepsize.

with $M = 200$ to discretize the problem in space. The obtained values for the error between the numerical and exact solution, measured in a discrete L^2-norm, are displayed in Figure 5.3.

We note that the exact solution has bounded time derivatives of moderate size. We are therefore in the situation of Remark 3.5, and, in particular, the error bound of Theorem 3.4 applies with $\alpha = \beta = 0$. The observed convergence orders confirm the theoretically predicted values.

We conclude this section with an additional numerical experiment where we illustrate the error behavior of the exponential methods EGLMpsq with $s = 2$ stages and $q = p - 1$ steps of order p for the above test problems, see Figure 5.4.

Due to the special structures of the above test problems, Fourier techniques are applicable for the numerical implementation of the φ-functions. We therefore used this approach in our numerical experiments. In more general situations where spectral techniques do not apply, matrix functions can be computed by subspace methods such as Krylov subspace techniques, see [9] and references cited therein. If the dimension of the involved matrices is moderate, an alternative implementation of the φ-functions is provided by the MATLAB package [2], downloadable from the website `http://www.math.ntnu.no/num/expint/`.

6 Extension to variable stepsizes

In this section, we briefly indicate how the techniques employed in this paper extend to variable stepsizes.

We let $(h_j)_{j \ge 0}$ be a sequence of positive stepsizes and define the associated grid points through $t_{j+1} = t_j + h_j$ for $j \ge 0$, where $t_0 = 0$. The stepsize ratios $(\omega_j)_{j \ge 1}$ are given by $h_j = \omega_j h_{j-1}$. As described in Section 4.3, a generic tool for the construction of numerical methods for (2.1) is the variation-of-constants

formula

$$y(t_{n+1}) = \mathrm{e}^{h_n L}\, y(t_n) + \int_0^{h_n} \mathrm{e}^{(h_n-\tau)L}\, N\big(y(t_n+\tau)\big)\,\mathrm{d}\tau \tag{6.1}$$

together with a replacement of the nonlinear term by some interpolation polynomial.

To keep the presentation simple, we illustrate the basic ideas by an explicit exponential integrator involving two stages and two steps. This generalizes the scheme EGLM322 of Section 4.1 to variable stepsizes. In order to determine the internal stage Y_{n2}, we replace N in (6.1) with the polynomial through $\big(t_{n-1}, N(y_{n-1})\big)$ and $\big(t_n, N(y_n)\big)$. Integration yields

$$\begin{aligned} Y_{n2} &= \mathrm{e}^{h_n L}\, y_n + h_n\, A_{21}^{(n)}(h_n L)\, N(y_n) + h_n\, U_{21}^{(n)}(h_n L)\, N(y_{n-1}), \\ A_{21}^{(n)} &= \varphi_1 + \omega_n\, \varphi_2, \qquad U_{21}^{(n)} = -\,\omega_n\, \varphi_2. \end{aligned}$$

Similarly, we obtain the numerical solution value

$$\begin{aligned} y_{n+1} &= \mathrm{e}^{h_n L}\, y_n + h_n\, B_1^{(n)}(h_n L)\, N(y_n) + h_n\, B_2^{(n)}(h_n L)\, N(Y_{n2}) \\ &\quad + h_n\, V_1^{(n)}(h_n L)\, N(y_{n-1}), \\ B_1^{(n)} &= A_{21}^{(n)} - (\varphi_2 + 2\,\omega_n\,\varphi_3), \qquad B_2^{(n)} = \frac{1}{1+\omega_n}\,(\varphi_2 + 2\,\omega_n\,\varphi_3), \\ V_1^{(n)} &= U_{21}^{(n)} + \frac{\omega_n}{1+\omega_n}\,(\varphi_2 + 2\,\omega_n\,\varphi_3)\,, \end{aligned}$$

by interpolating through the above points and $\big(t_{n+1}, N(Y_{n2})\big)$.

More generally, we allow explicit exponential general linear methods with coefficients depending on several subsequent stepsize ratios

$$\begin{aligned} y_{n+1} &= \mathrm{e}^{h_n L}\, y_n + h_n \sum_{i=1}^{s} B_i^{(n)}(h_n L)\, N(Y_{ni}) + h_n \sum_{k=1}^{q-1} V_k^{(n)}(h_n L)\, N(y_{n-k}), \\ Y_{ni} &= \mathrm{e}^{c_i h_n L}\, y_n + h_n \sum_{j=1}^{i} A_{ij}^{(n)}(h_n L)\, N(Y_{nj}) + h_n \sum_{k=1}^{q-1} U_{ik}^{(n)}(h_n L)\, N(y_{n-k}), \end{aligned}$$

see also (2.2). Provided that the stepsize ratios are bounded from above and below, that is, it holds

$$C_1 \le \omega_j \le C_2, \qquad j \ge 1, \tag{6.2}$$

with (moderate) constants $C_1, C_2 > 0$, the coefficient operators satisfy an estimate of the form (3.11) with h replaced by h_n. We emphasize that assumption (6.2) is always fulfilled in practical implementations. Due to the special form of the considered method class, no further requirements on the stepsize sequence are needed. As a consequence, it is straightforward to generalize the convergence analysis of Section 3.3. More precisely, by means of a Gronwall-type inequality derived in Bakaev [1, Lemma 4.4], the proof of Theorem 3.4 extends literally to variable stepsizes. We do not elaborate the details here.

7 Conclusions

The present work shows that the considered class of exponential integrators has the following benefits. It allows, in an easy manner, the construction of methods with high stage order and excellent convergence properties for stiff problems. Further, the combination of exponential Runge–Kutta and exponential Adams–Bashforth methods results in schemes with favorable stability properties.

It is beyond the scope of this paper to identify methods which are competitive with established schemes. To reach this aim, it is indispensable to implement the method with variable stepsizes based on an error control. In particular, an efficient implementation of the φ-functions plays a crucial role here. Besides, it remains to look into the computation of the starting values. These investigations are part of future work.

Acknowledgement

Mechthild Thalhammer was supported by Fonds zur Förderung der wissenschaftlichen Forschung (FWF) under project H210-N13.

REFERENCES

1. Bakaev, N.Yu. *On variable stepsize Runge–Kutta approximations of a Cauchy problem for the evolution equation.* BIT 38: 462–485 (1998).
2. Berland, H., Skaflestad, B., and Wright, W.M. *Expint – A Matlab package for exponential integrators.* Tech. report 4/05, Department of Mathematics, NTNU, 2005.
3. Beylkin, G., Keiser, J.M., and Vozovoi, L. *A new class of time discretization schemes for the solution of nonlinear PDEs.* J. Comput. Phys. 147: 362–387 (1998).
4. Brunner, H. and van der Houwen, P.J. *The Numerical Solution of Volterra Equations.* CWI Monographs 3, North-Holland, Amsterdam, 1986.
5. Calvo, M.P. and Palencia, C. *A class of explicit multistep exponential integrators for semilinear problems.* Numer. Math. 102: 367-381 (2006).
6. Cox, P.M. and Matthews, P.C. *Exponential time differencing for stiff systems.* J. Comput. Phys. 176: 430–455 (2002).
7. Friedli, A. *Verallgemeinerte Runge–Kutta Verfahren zur Lösung steifer Differentialgleichungssysteme.* In: Numerical Treatment of Differential Equations. Bulirsch, R., Grigorieff, R., and Schröder, J., eds., Lecture Notes in Mathematics 631, pp. 35–50, Springer, Berlin, 1978.
8. Henry, D. *Geometric Theory of Semilinear Parabolic Equations.* Lecture Notes in Mathematics 840, Springer, Berlin, 1981.
9. Hochbruck, M. and Hochstenbach, M.E. *Subspace extraction for matrix functions.* Preprint, Department of Mathematics, Case Western Reserve University, 2005.

10. Hochbruck, M. and Ostermann, A. *Exponential Runge–Kutta methods for parabolic problems*. Appl. Numer. Math. 53: 323–339 (2005).

11. Hochbruck, M. and Ostermann, A. *Explicit exponential Runge–Kutta methods for semilinear parabolic problems*. SIAM J. Numer. Anal. 43: 1069-1090 (2005).

12. Kassam, A.K. and Trefethen, L.N. *Fourth-order time stepping for stiff PDEs*. SIAM J. Sci. Comput. 26: 1214–1233 (2005).

13. Krogstad, S. *Generalized integrating factor methods for stiff PDEs*. J. Comput. Phys. 203: 72–88 (2005).

14. Lunardi A. *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Basel, 1995.

15. Minchev, B.V. and Wright, W.M. *A review of exponential integrators for first order semi-linear problems*. Tech. report 2/05, Department of Mathematics, NTNU, April 2005.

16. Nørsett, S.P. *An A-stable modification of the Adams–Bashforth methods*. In: Conference on the Numerical Solution of Differential Equations, J. Morris, ed., Lecture Notes in Mathematics 109: 214–219, Springer, Berlin, 1969.

17. Ostermann, A. and Thalhammer, M. *Non-smooth data error estimates for linearly implicit Runge–Kutta methods*. IMA J. Numer. Anal. 20: 167–184 (2000).

18. Pazy, A. *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York, 1983.

19. Strehmel, K. and Weiner, R. *B-convergence results for linearly-implicit one step methods*. BIT 27: 264–281 (1987).

20. Verwer, J.G. *On generalized linear multistep methods with zero-parasitic roots and an adaptive principal root*. Numer. Math. 27: 143–155 (1977).

A. Appendix

A.1. Positivity of exponential multistep methods

Positivity of exponential multistep methods

ALEXANDER OSTERMANN AND MECHTHILD THALHAMMER

Submitted to Proceedings of ENUMATH 2005

Positivity of exponential multistep methods

Alexander Ostermann and Mechthild Thalhammer

Institut für Mathematik, Leopold-Franzens-Universität Innsbruck
Mailing address: Technikerstraße 13, A-6020 Innsbruck, Austria
Email: {alexander.ostermann, mechthild.thalhammer}@uibk.ac.at

Summary. In this paper, we consider exponential integrators that are based on linear multistep methods and study their positivity properties for abstract evolution equations. We prove that the order of a positive exponential multistep method is two at most and further show that there exist second-order methods preserving positivity.

1 Introduction

Integration schemes that involve the evaluation of the exponential were first proposed in the 1960s for the numerical approximation of stiff ordinary differential equations. Nowadays, due to advances in the computation of the product of a matrix exponential with a vector, such methods are considered as practicable also for high-dimensional systems of differential equations. The renewed interest in exponential integrators is further enhanced by recent investigations which showed that they have excellent stability and convergence properties. In particular, they perform well for differential equations that result from a spatial discretisation of nonlinear parabolic and hyperbolic initial-boundary value problems, see [4, 9] and references therein.

However, aside from a favourable convergence behaviour, the usability of a numerical method for practical applications is substantially affected by its qualitative behaviour, and, in many cases, it is inevitable to ensure that certain geometric properties of the underlying problem are well preserved by the discretisation. In particular, it is desirable that the positivity of the true solution is retained by the numerical approximation. More precisely, if the solution of a linear abstract evolution equation

$$u'(t) = A\,u(t) + f(t), \quad 0 < t \le T, \qquad u(0) \text{ given}, \tag{1}$$

remains positive, the numerical solution should retain this property. Unfortunately, as proven by Bolley and Crouzeix [3], the order of positive rational one-step and linear multistep methods, respectively, is restricted by one.

The objective of the present paper is to investigate exponential multistep methods where the coefficients are combinations of the exponential and closely related functions. The general form of the considered schemes is introduced below in Section 3. Examples include Adams-type methods that were studied recently in [4, 9] for parabolic problems, see also the earlier works [8, 12].

The main result, which we deduce in Section 4, states that positive exponential multistep methods are of order two at most. Further, we show that there exist second-order methods which preserve positivity. Thus, the order barrier of [3] is raised by one. For exponential Runge–Kutta methods, a similar result has recently been obtained in [10].

Our analysis of exponential multistep methods for abstract evolution equations is based on an operator calculus which allows to define the Laplace-Stieltjes transform involving the generator of a positive C_0-semigroup. We refer to the subsequent Section 2, where the basic hypotheses on the differential equation and some fundamental tools of the employed analytical framework are recapitulated.

2 Analytical framework

In this section, we state the basic assumptions on the abstract initial value problem (1).

Throughout, we let $\bigl(V,\|\cdot\|\bigr)$ denote the underlying Banach space. Further, we suppose $A : D \subset V \to V$ to be a densely defined and closed linear operator on V that generates a *strongly continuous* semigroup $\bigl(\mathrm{e}^{tA}\bigr)_{t\geq 0}$ of type (M,ω), that is, there exist constants $M \geq 1$ and $\omega \in \mathbb{R}$ such that the bound

$$\bigl\|\mathrm{e}^{tA}\bigr\| \leq M\mathrm{e}^{\omega t}, \qquad t \geq 0, \tag{2}$$

is valid. For a detailed treatment of C_0-semigroups, we refer to the monographs [6, 11].

The notion of positivity requires the Banach space V to be endowed with an additional order structure. In the present paper, to keep the analytical framework simple, we restrict ourselves to the consideration of the Lebesgue spaces and subspaces thereof, respectively, as it is then straightforward to define the positivity of an element pointwise.[1] In general, an appropriate setting is provided by the theory of Banach lattices treated in Yosida [13, Chap. XII]. Our results remain valid within this framework.

We recall that a bounded linear operator $B : V \to V$ is said to be *positive* if for any element $v \in V$ satisfying $v \geq 0$ it follows $Bv \geq 0$.

[1] A function $v : \Omega \subset \mathbb{R}^d \to \mathbb{R}$ in $\mathrm{L}^p(\Omega)$, $1 \leq p \leq \infty$, is said to be *positive* if it is pointwise positive, i.e., $v(x) \geq 0$ for almost all $x \in \Omega$. In that case, we write $v \geq 0$ for short. We employ here the standard terminology, although the term *non-negative* would be more appropriate.

Example 1. We consider the differential operator ∂_{xx} subject to a mixed boundary condition on the Banach space of continuous functions, that is, for some $c_1, c_2 \in \mathbb{R}$ we let $A : D \to V : v \mapsto \partial_{xx} v$ where $V = \mathrm{C}([0,1])$ and $D = \big\{v \in \mathrm{C}^2([0,1]) : v'(0) + c_1 v(0) = 0 = v'(1) + c_2 v(1)\big\}$. It is shown in Arendt et al. [1, p. 134] that the associated semigroup $\big(\mathrm{e}^{tA}\big)_{t\geq 0}$ is positive.

Henceforth, we assume that the linear operator $A : D \to V$ is the generator of a positive semigroup $\big(\mathrm{e}^{tA}\big)_{t\geq 0}$ of type (M, ω), see (2). Then, from the formulation of the linear evolution equation (1) as a Volterra integral equation

$$u(t) = \mathrm{e}^{tA}\, u(0) + \int_0^t \mathrm{e}^{(t-\tau)A} f(\tau)\, \mathrm{d}\tau, \qquad 0 \leq t \leq T, \tag{3}$$

it is seen that the solution u remains positive, provided that the initial value $u(0)$ and the function f are positive.

Let $a \in \mathrm{BV}$ denote a function of bounded variation that is normalised at its discontinuities and satisfies $a(0) = 0$. The associated *Laplace-Stieltjes transform* is defined through

$$G(z) = \int_0^\infty \mathrm{e}^{tz}\, \mathrm{d}a(t), \tag{4}$$

see Hille and Phillips [6, Sect. 6.2]. We recall that a real-valued function G is said to be *absolutely monotonic* on an interval $I \subset \mathbb{R}$ if

$$G^{(j)}(x) \geq 0, \qquad x \in I, \quad j \geq 0.$$

The following result by Bernstein [2], which characterises absolutely monotonic functions of the form (4), is the basis of our analysis in Section 4.

Theorem 2 (Bernstein). *A function G is absolutely monotonic on the half line $(-\infty, \omega]$ iff it is the Laplace-Stieltjes transform of a non-decreasing function $a \in \mathrm{BV}$ such that*

$$\int_0^\infty \mathrm{e}^{\omega t}\, |\mathrm{d}a(t)| < \infty.$$

A well-known operational calculus described in Hille and Phillips [6, Chap. XV] allows to extend (4) to unbounded linear operators. More precisely, for A being the generator of a strongly continuous semigroup $\big(\mathrm{e}^{tA}\big)_{t\geq 0}$ on V, it holds

$$G(hA)\, v = \int_0^\infty \mathrm{e}^{tA}\, v\, \mathrm{d}a(t), \qquad h \geq 0, \quad v \in V, \tag{5}$$

where the integral is defined in the sense of Bochner. It is thus straightforward to deduce the following corollary from Theorem 2, see also Kovács [7].

Corollary 3. *Suppose that the linear operator A generates a positive and strongly continuous semigroup of type (M, ω). Assume further that the function G is absolutely monotonic on $(-\infty, h\omega]$ for some $h \geq 0$. Then, the linear operator $G(hA)$ defined by (5) is positive.*

Remark 4. We note that the converse of the above corollary is true as well. Namely, if $G(hA)$ is positive for any generator A of a positive and strongly continuous semigroup, then the function G is absolutely monotonic. The proof of this statement is in the lines of Bolley and Crouzeix [3, Proof of Lemma 1].

The construction of exponential integrators often relies on the variation-of-constants formula (3) and a replacement of the integrand f by an interpolation polynomial. As a consequence, the linear operators $\varphi_j(hA)$ defined through

$$\varphi_j(z) = \int_0^1 \mathrm{e}^{tz} \, \frac{(1-t)^{j-1}}{(j-1)!} \, \mathrm{d}t, \qquad j \geq 1, \qquad z \in \mathbb{C}, \tag{6}$$

naturally arise in the numerical schemes. By the above Theorem 2, these functions are absolutely monotonic, and thus the positivity of the associated operators $\varphi_j(hA)$ follows from Corollary 3.

3 Exponential multistep methods

In this section, we introduce the considered exponential multistep methods for the time integration of the linear evolution equation (1) and state the order conditions. The positivity properties of the numerical schemes are then studied in Section 4.

We let $t_j = jh$ denote the grid points associated with a constant stepsize $h > 0$. Besides, we suppose that the starting values $u_0, u_1, \ldots, u_{k-1} \in V$ are approximations the exact solution values of (1). Then, for integers $j \geq k$, the numerical solution values $u_j \approx u(t_j)$ are given by the k-step recursion

$$\sum_{\ell=0}^{k} \alpha_\ell(hA) \, u_{n+\ell} = h \sum_{\ell=0}^{k} \beta_\ell(hA) \, f(t_{n+\ell}), \qquad n \geq 0. \tag{7a}$$

Throughout, we choose $\alpha_k = 1$. Furthermore, we assume that the coefficient functions α_ℓ and β_ℓ are given as Laplace-Stieltjes transforms of certain functions a_ℓ and b_ℓ. Thus, it holds

$$\alpha_\ell(z) = \int_0^\infty \mathrm{e}^{tz} \, \mathrm{d}a_\ell(t), \quad \beta_\ell(z) = \int_0^\infty \mathrm{e}^{tz} \, \mathrm{d}b_\ell(t), \qquad z \in (-\infty, \omega]. \tag{7b}$$

For simplicity, we require b_ℓ to be piecewise differentiable such that the left-sided limit of $b_\ell'(t)$ exist at $t = j$ for all integers $j \geq 0$. In particular, these assumptions are satisfied if the coefficients functions are (linear) combinations of the exponential and the related φ-functions (6). We therefore refer to (7) as an *exponential linear k-step method*. Due to (7b), the operators $\alpha_\ell(hA)$ and $\beta_\ell(hA)$ are bounded on V.

Examples that have recently been studied in literature for the time integration of semilinear evolution equations are exponential Adams-type methods. For the choice $\alpha_1 = \ldots = \alpha_{k-1} = 0$ and $\beta_k = 0$, the resulting methods are discussed in Calvo and Palencia [4]. On the other hand, the case

$\alpha_0 = \ldots = \alpha_{k-2} = 0$ and $\beta_k = 0$ generalising the classical Adams–Bashforth methods is covered by the analysis given in [9].

In the following, we derive the order conditions for the exponential k-step method. We note that the arguments given below extend to semilinear problems $u'(t) = A\,u(t) + F\big(t, u(t)\big)$ by setting $f(t) = F\big(t, u(t)\big)$. As usual, the numerical method (7) is said to be *consistent* of order p, if the local error

$$d(t,h) = \sum_{\ell=0}^{k} \alpha_\ell(hA)\, u(t+\ell h) - h \sum_{i=0}^{k} \beta_\ell(hA)\, f(t+\ell h) \tag{8}$$

is of the form $d(t,h) = \mathscr{O}\big(h^{p+1}\big)$ for $h \to 0$, provided that the function f is sufficiently smooth, see Hairer, Nørsett, and Wanner [5, Chap. III.2].

In order to determine the leading h-term in $d(t,h)$, we make use of the variation-of-constants formula

$$u(t+\ell h) = \mathrm{e}^{\ell h A}\, u(t) + \int_0^{\ell h} \mathrm{e}^{(\ell h - \tau)A}\, f(t+\tau)\, \mathrm{d}\tau,$$

see also (3). We expand all occurrences of f in Taylor series at t and apply the definition of the φ-functions (6). A comparison in powers of h finally yields the following result.

Lemma 5. *The order conditions for exponential multistep methods* (7) *are*

$$\sum_{\ell=0}^{k} \alpha_\ell(hA)\, \mathrm{e}^{\ell h A} = 0, \tag{9a}$$

$$\sum_{\ell=1}^{k} \alpha_\ell(hA)\, \ell^q\, \varphi_q(\ell h A) \;=\; \sum_{\ell=0}^{k} \beta_\ell(hA)\, \frac{\ell^{q-1}}{(q-1)!}, \qquad 1 \le q \le p, \tag{9b}$$

where by definition $\ell^0 = 1$ *for* $\ell = 0$.

The first condition corresponds to the requirement that the exponential multistep method (7) is exact for the homogeneous equation $u'(t) = A\,u(t)$. By setting $A = 0$ in (9), the usual order conditions

$$\sum_{\ell=0}^{k} \alpha_\ell(0) = 0, \qquad \sum_{\ell=1}^{k} \alpha_\ell(0)\, \ell^q = q \sum_{\ell=0}^{k} \beta_\ell(0)\, \ell^{q-1}, \quad 1 \le q \le p$$

for a linear multistep method with coefficients $\alpha_\ell(0)$ and $\beta_\ell(0)$ follow, see also [5, Chap. III.2].

4 Positivity and order barrier

In this section, we derive an order barrier for positive exponential multistep methods. According to Bolley and Crouzeix [3], the numerical method (7) is

said to be *positive*, if the numerical solution values u_n remain positive for all $n \geq k$, provided that the semigroup $\left(\mathrm{e}^{tA}\right)_{t\geq 0}$, the function f, and further the starting values $u_0, u_1, \ldots, u_{k-1}$ are positive. We note that the requirement of positivity implies that the coefficients operators $\alpha_\ell(hA)$ satisfy

$$-\alpha_\ell(hA) \geq 0, \qquad 0 \leq \ell \leq k-1. \tag{10}$$

We next give the main result of the paper.

Theorem 6. *The order of a positive exponential k-step method is two at most.*

Proof. Our main tools for the proof of Theorem 6 are the representation (7b) of the coefficient functions as Laplace-Stieltjes transforms and further the characterisation of positivity given in Section 2. We note that due to Corollary 3, it is justified to work with the complex variable z instead of the linear operator hA. For the characteristic function of the interval $[r, s)$, we henceforth employ the abbreviation

$$Y_{[r,s)}(t) = \begin{cases} 1 & \text{if } \ r \leq t < s, \\ 0 & \text{else.} \end{cases}$$

(i) We first show that the validity of the first order condition (9a) together with the requirement (10) imply that the coefficient functions α_ℓ are of the form

$$\alpha_\ell(z) = -\mu_{k-\ell}\,\mathrm{e}^{(k-\ell)z}, \qquad \mu_{k-\ell} \geq 0, \qquad 0 \leq \ell \leq k-1, \tag{11}$$

or, equivalently, that the associated functions a_ℓ are given by

$$a_\ell(t) = -\mu_{k-\ell}\,Y_{[k-\ell,\infty)}(t), \qquad \mu_{k-\ell} \geq 0, \qquad 0 \leq \ell \leq k-1. \tag{12}$$

Inserting (7b) into (9a) and applying $\alpha_k(z) = 1$, we get

$$\mathrm{e}^{kz} = -\sum_{\ell=0}^{k-1} \alpha_\ell(z)\,\mathrm{e}^{\ell z} = -\sum_{\ell=0}^{k-1} \int_0^\infty \mathrm{e}^{tz}\,Y_{[\ell,\infty)}(t)\,\mathrm{d}a_\ell(t-\ell)$$

and furthermore conclude

$$Y_{[k,\infty)}(t) = -\sum_{\ell=0}^{k-1} a_\ell(t-\ell)\,Y_{[\ell,\infty)}(t). \tag{13}$$

From (10) and Remark 4 we deduce that the function $-\alpha_\ell$ is absolutely monotonic and thus Theorem 2 shows that $-a_\ell$ is non-decreasing. Due to the fact that $a_\ell(0) = 0$, we finally obtain (12). For the following considerations, accordingly to our choice $\alpha_k(z) = 1$, it is useful to define $\mu_0 = -1$. As a consequence, inserting (11) into (9a) we have

$$\sum_{\ell=1}^{k} \mu_\ell = -\mu_0 = 1. \tag{14}$$

(ii) We next reformulate the order conditions in terms of the functions a_ℓ and b_ℓ given by (7b). Inserting (11) into (9b), we have

$$-\sum_{\ell=1}^{k} \mu_{k-\ell}\, \ell^q\, \mathrm{e}^{(k-\ell)z}\, \varphi_q(\ell z) = \sum_{\ell=0}^{k} \beta_\ell(z) \frac{\ell^{q-1}}{(q-1)!}, \qquad 1 \le q \le p.$$

Moreover, making use of the fact that

$$\mathrm{e}^{(k-\ell)z}\, \varphi_q(\ell z) = \frac{1}{\ell^q} \int_0^\infty \mathrm{e}^{tz}\, \frac{(k-t)^{q-1}}{(q-1)!}\, Y_{[k-\ell,k)}(t)\, \mathrm{d}t,$$

see (6) for the definition of φ_q, we obtain

$$-(k-t)^{q-1} \sum_{\ell=1}^{k} \mu_{k-\ell}\, Y_{[k-\ell,k)}(t) = \sum_{\ell=0}^{k} \ell^{q-1} b_\ell(t), \qquad 1 \le q \le p. \tag{15}$$

For the following considerations, it is convenient to employ the abbreviation

$$\chi_j(t) = Y_{[0,j)}(t) - \sum_{\ell=1}^{j-1} \mu_\ell\, Y_{[\ell,j)}(t). \tag{16}$$

Obviously, the support of the function χ_j is contained in the interval $[0, j)$.

(iii) Exploiting the relations given above, we now show that the assumption $p \ge 3$ and the requirement of positivity, that is, the assumptions $\mu_\ell \ge 0$ for $1 \le \ell \le k$ and $b_\ell(t) \ge 0$ for any $t \in \mathbb{R}$ and $0 \le \ell \le k$, lead to a contradiction. Regarding the order conditions (15), we introduce the following relations for the functions b_ℓ

$$\sum_{\ell=0}^{k} b_\ell(t) = \chi_j(t), \tag{17a}$$

$$\sum_{\ell=0}^{k} (\ell + j - k)\, b_\ell(t) = (j-t)\, \chi_j(t), \tag{17b}$$

$$\sum_{\ell=0}^{k} (\ell + j - k)^2\, b_\ell(t) = (j-t)^2\, \chi_j(t), \tag{17c}$$

see also (16). Clearly, when setting $j = k$, we retain (15) with $p = 3$. Using that the functions b_ℓ are positive, we infer from (17c) that the values[2] at $t = j$ fulfil $b_\ell(j) = 0$ for $\ell \ne k-j$. Consequently, the derivatives satisfy $b'_\ell(j) \le 0$ for $\ell \ne k-j$. Taking the derivative of (17c) implies $b'_\ell(j) = 0$ for $\ell \ne k-j$. Further, differentiating (17b) yields $\mu_j = 0$ and thus $\chi_j = \chi_{j-1}$. Finally, taking suitable linear combinations of (17) shows that the order conditions (17) also hold for $j-1$ in place of j. By induction, we therefore conclude $\mu_j = 0$ for $1 \le j \le k$ which contradicts (14). □

[2] Here and henceforth, all function evaluations are understood as left-sided limits.

Remark 7. The order two barrier of Theorem 6 is sharp in the sense that there exist positive second-order schemes. A simple example is given by the exponential trapezoidal rule where $k = 1$, $\alpha_0(z) = -\mathrm{e}^z$, $\alpha_1 = 1$, $\beta_0 = \varphi_1 - \varphi_2$, and $\beta_1 = \varphi_2$.

For *analytic* semigroups it is well-known that the order conditions (9b) can be weakened, see e.g. [9]. Following the lines of [10] it can be shown that an order two barrier holds in this case, too. For instance, the exponential midpoint rule with $k = 2$, $\alpha_0(z) = -\mathrm{e}^{2z}$, $\alpha_1 = 0$, $\alpha_2 = 1$, $\beta_1(z) = 2\,\varphi_1(2z)$, and $\beta_0 = \beta_2 = 0$ has weak order two and preserves positivity.

References

[1] W. Arendt, A. Grabosch, G. Greiner, U. Groh, H.P. Lotz, U. Moustakas, R. Nagel, F. Neubrander, and U. Schlotterbeck, *One-parameter Semigroups of Positive Operators*. Springer, Berlin (1980)

[2] S. Bernstein, *Sur les fonctions absolument monotones*. Acta Mathematica **51**, 1–66 (1928)

[3] C. Bolley and M. Crouzeix, *Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques*. R.A.I.R.O. Anal. Numér. **12**, 237–245 (1978)

[4] M.P. Calvo and C. Palencia, *A class of explicit multistep exponential integrators for semilinear problems* (2005). To appear in Numer. Math.

[5] E. Hairer, S.P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, Berlin (1993)

[6] E. Hille and R.S. Phillips, *Functional Analysis and Semi-Groups*. American Mathematical Society, Providence (1957)

[7] M. Kovács, *On positivity, shape, and norm-bound preservation of time-stepping methods for semigroups*. J. Math. Anal. Appl. **304**, 115–136 (2005)

[8] S.P. Nørsett, *An A-stable modification of the Adams–Bashforth methods*. In: Conference on the Numerical Solution of Differential Equations, J. Morris, ed., Lecture Notes in Mathematics **109**, 214–219, Springer, Berlin (1969)

[9] A. Ostermann, M. Thalhammer, and W. Wright, *A class of explicit exponential general linear methods*. Preprint, University of Innsbruck (2005)

[10] A. Ostermann and M. Van Daele, *Positivity of exponential Runge–Kutta methods*. Preprint, University of Innsbruck (2005)

[11] A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*. Springer, New York (1983)

[12] J.G. Verwer, *On generalized linear multistep methods with zero-parasitic roots and an adaptive principal root*. Numer. Math. **27**, 143–155 (1977)

[13] K. Yosida, *Functional Analysis*. Springer, Berlin (1965)

Bibliography

[1] J.C. BUTCHER, *Numerical Methods for Ordinary Differential Equations*. John Wiley & Sons, Chichester, 2003.

[2] C. GONZÁLEZ, A. OSTERMANN, C. PALENCIA, AND M. THALHAMMER, *Backward Euler discretization of fully nonlinear parabolic problems*. Math. Comp. (2001) 71, 125-145.

[3] C. GONZÁLEZ, A. OSTERMANN, AND M. THALHAMMER, *A second-order Magnus integrator for nonautonomous parabolic problems*. J. Comp. Appl. Math. (2006) 189, 142-156.

[4] C. GONZÁLEZ AND M. THALHAMMER, *A second-order Magnus type integrator for quasilinear parabolic problems*. To appear in Math. Comp.

[5] E. HAIRER, S.P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations I. Nonstiff Problems*. Springer, Berlin, 1993.

[6] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems*. Springer, Berlin, 1996.

[7] E. HAIRER, CH. LUBICH, AND G. WANNER, *Geometric Numerical Integration. Structure-Preserving Algorithms for Ordinary Differential Equations*. Springer, Berlin, 2002.

[8] D. HENRY, *Geometric Theory of Semilinear Parabolic Equations*. Lecture Notes in Mathematics 840, Springer, Berlin, 1981.

[9] M. HOCHBRUCK AND M.E. HOCHSTENBACH, *Subspace extraction for matrix functions*. Preprint, Department of Mathematics, Case Western Reserve University, 2005.

[10] W. HUNDSDORFER AND J.G. VERWER, *Numerical Solution of Time-Dependent Advection-Diffusion-Reaction Equations*. Springer, Berlin, 2003.

[11] A. LUNARDI, *Analytic Semigroups and Optimal Regularity in Parabolic Problems*. Birkhäuser, Basel, 1995.

[12] C. MOLER AND CH. VAN LOAN, *Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later*. SIAM Rev. (2003) 45, 3-49.

[13] A. OSTERMANN AND M. THALHAMMER, *Non-smooth data error estimates for linearly implicit Runge-Kutta methods*. IMA J. Numer. Anal. (2000) 20, 167-184.

[14] A. OSTERMANN AND M. THALHAMMER, *Convergence of Runge-Kutta methods for nonlinear parabolic equations.* Applied Numerical Math. (2002) 42, 367-380.

[15] A. OSTERMANN AND M. THALHAMMER, *Positivity of exponential multistep methods.* Submitted to Proceedings of ENUMATH 2005.

[16] A. OSTERMANN, M. THALHAMMER, AND G. KIRLINGER, *Stability of linear multistep methods and applications to nonlinear parabolic problems.* Applied Numerical Math. (2004) 48, 389-407.

[17] A. OSTERMANN, M. THALHAMMER, AND W. WRIGHT, *A class of explicit exponential general linear methods.* To appear in BIT.

[18] M. THALHAMMER, *Runge-Kutta Time Discretization of Fully Nonlinear Parabolic Problems.* Doctoral thesis, Preprint, Institut für Technische Mathematik, Geometrie und Bauinformatik, Universität Innsbruck, 2000.

[19] M. THALHAMMER, *On the convergence behaviour of variable stepsize multistep methods for singularly perturbed problems.* BIT (2004) 44, 343-361.

[20] M. THALHAMMER, *A second-order Magnus type integrator for non-autonomous semilinear parabolic problems.* Preprint, Institut für Mathematik, Universität Innsbruck, 2005.

[21] M. THALHAMMER, *A fourth-order commutator-free exponential integrator for non-autonomous differential equations.* To appear in SIAM J. Numer. Anal.