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.

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

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

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*

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

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

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

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

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

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

Here a denotes a positive constant.

Choosing $X = C([0,1])$ and $D = \{v \in C^2([0,1]):\ v'(0) = v'(1) = 0\}$ 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

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

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

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

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*

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

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*

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

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

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

we arrive at the formally semilinear problem

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

Due to our assumptions and Lemma 3, we know that

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

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*

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

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

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

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

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

This recursion differs from (13) by the defects

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

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

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

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

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

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

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

which implies the bound

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

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

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

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

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

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)

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

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

$$\bigl\| A_0(I - hA_0)^{-n} \bigr\|_{X \leftarrow X} \leqslant C t_n^{-1}, \quad 0 < nh \leqslant T, \tag{22}$$

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

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

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

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

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

$$\mathcal{Q} = (a_{ij})_{i,j=1}^s, \qquad \mathbb{1} = (1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^s, \qquad b = (b_1, \ldots, b_s)^{\mathrm{T}}.$$

Then the *stability function* of (24) is defined through

$$R(z) = 1 + zb^{\mathrm{T}}(I - z\mathcal{Q})^{-1}\mathbb{1}.$$

The Runge–Kutta method is $A(\vartheta)$*-stable* if $I - z\mathcal{Q}$ 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 $\mathcal{Q}$ is invertible and the module of R at infinity, $R(\infty) = 1 - b^{\mathrm{T}}\mathcal{Q}^{-1}\mathbb{1}$, 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

$$\left|\big(\lambda - A(t)\big)^{-1}\right|_{X \leftarrow X} \leqslant \frac{M}{|\lambda - a|} \quad \text{for } \left|\arg(\lambda - a)\right| \leqslant \pi - \varphi \tag{25}$$

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*

$$\left\|u_n - u(t_n)\right\| + \max_{1 \leqslant i \leqslant s}\left\|U_{ni} - u(t_n + c_i h)\right\| \leqslant Ch^r\big(1 + |\log h|\big), \quad 0 \leqslant nh \leqslant T.$$

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

$$U_n' = f_{n+1}(U_n), \qquad U_n = \mathbb{1}u_n + h\mathcal{Q}U_n', \qquad u_{n+1} = u_n + hb^{\mathrm{T}}U_n'. \tag{26}$$

Here, the matrix $\mathcal{Q}$ is considered as a linear operator on X^s and the ith component of $\mathcal{Q}U_n'$ 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

$$A_{n+1} = \operatorname{diag}\bigl(A(t_n + c_1 h), \dots, A(t_n + c_s h)\bigr).$$

Due to the resolvent condition (25) and the $A(\vartheta)$-stability of the method, the operators

$$J_{n+1} = (I - h\mathcal{Q}A_{n+1})^{-1} \quad \text{and} \quad K_{n+1} = (I - hA_{n+1}\mathcal{Q})^{-1}$$

are well-defined and bounded for h sufficiently small.

In this notation, the stages are given by

$$U_n = J_{n+1}\mathbb{1}u_n + hJ_{n+1}\mathcal{Q}g_{n+1}(U_n), \tag{27a}$$

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

$$R(hA_{n+1}) = 1 + hb^{\mathrm{T}}A_{n+1}(I - h\mathcal{Q}A_{n+1})^{-1}\mathbb{1}.$$

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

$$\widehat{U}_n' = f_{n+1}\bigl(\widehat{U}_n\bigr), \quad \widehat{U}_n = \mathbb{1}\hat{u}_n + h\mathcal{Q}\widehat{U}_n' + \Delta_n, \quad \hat{u}_{n+1} = \hat{u}_n + hb^{\mathrm{T}}\widehat{U}_n' + \delta_{n+1}, \tag{29}$$

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

$$\begin{aligned}\bigl(\Psi(V)\bigr)_n = {} & J_{n+1}\mathbb{1}R(t_n,0)u_0 + hJ_{n+1}\mathcal{Q}g_{n+1}(V_n)\\ & + h\sum_{j=0}^{n-1} J_{n+1}\mathbb{1}R(t_n,t_{j+1})b^{\mathrm{T}}K_{j+1}g_{j+1}(V_j).\end{aligned} \tag{31}$$

For some $c_0 > 0$ and $0 < \gamma < 1$ we choose the set

$$\mathcal{D}_h = \bigl\{V = (V_n)_{n=0}^N \in \mathcal{D}^{(N+1)s}\colon\ \bigl\|V - \widehat{U}\bigr\|_\infty \leqslant c_0 h^\gamma\bigr\}$$

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)

$$\bigl|g_{j+1}(V_j) - g(t_{j+1},W_j)\bigr| \leqslant c_0 L h^\gamma \|V_j - W_j\|. \tag{32}$$

With the help of the stability result (28) and (32), we thus receive

$$\bigl\|\bigl(\Psi(V)\bigr)_n - \bigl(\Psi(W)\bigr)_n\bigr\| \leqslant c_0 c_1 L\bigl(1 + |\log h|\bigr)h^\gamma\|V - W\|_\infty,$$

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

$$\begin{aligned}\bigl\|\widehat{U}_n - \bigl(\Psi(\widehat{U})\bigr)_n\bigr\| \leqslant {} & \sum_{j=0}^{n-1}\bigl\|J_{n+1}\mathbb{1}R(t_n,t_{j+1})\bigr\|_{D\leftarrow X}|\delta_{j+1}| + \|J_{n+1}\|_{D\leftarrow D}\|\Delta_n\|\\ & + h\sum_{j=0}^{n-1}\bigl\|J_{n+1}\mathbb{1}R(t_n,t_{j+1})b^{\mathrm{T}}K_{j+1}\bigr\|_{D\leftarrow X}|A_{j+1}\Delta_j|.\end{aligned}$$

Applying the bounds (28) and (30) yields

$$\bigl\|\widehat{U} - \Psi(\widehat{U})\bigr\|_\infty \leqslant Ch^r\bigl(1 + |\log h|\bigr). \tag{33}$$

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)

$$\bigl\|U_n - \widehat{U}_n\bigr\| \leqslant \bigl\|U - \widehat{U}\bigr\|_\infty \leqslant \frac{1}{1-\kappa}\bigl\|\widehat{U} - \Psi(\widehat{U})\bigr\|_\infty \leqslant Ch^r\bigl(1 + |\log h|\bigr). \tag{34}$$

In order to estimate the error between the numerical and the exact solution, we use the relation

$$u_{n+1} - \hat{u}_{n+1} = \bigl(1 - b^{\mathrm{T}}\mathcal{Q}^{-1}\mathbb{1}\bigr)\bigl(u_n - \hat{u}_n\bigr) + b^{\mathrm{T}}\mathcal{Q}^{-1}\bigl(U_n - \widehat{U}_n + \Delta_n\bigr) - \delta_{n+1}.$$

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.