MATHEMATICS OF COMPUTATION
VOLUME 64, NUMBER 210
APRIL 1995, PAGES 601–627

RUNGE-KUTTA APPROXIMATION OF QUASI-LINEAR PARABOLIC EQUATIONS

CHRISTIAN LUBICH AND ALEXANDER OSTERMANN

En l'honneur de Michel Crouzeix à l'occasion de son cinquantième anniversaire

ABSTRACT. We study the convergence properties of implicit Runge-Kutta methods applied to time discretization of parabolic equations with time- or solution-dependent operator. Error bounds are derived in the energy norm. The convergence analysis uses two different approaches. The first, technically simpler approach relies on energy estimates and requires algebraic stability of the Runge-Kutta method. The second one is based on estimates for linear time-invariant equations and uses Fourier and perturbation techniques. It applies to $A(\theta)$-stable Runge-Kutta methods and yields the precise temporal order of convergence. This order is noninteger in general and depends on the type of boundary conditions.

INTRODUCTION

In this paper we investigate the approximation properties of implicit Runge-Kutta methods applied to time discretization of parabolic equations with time- or solution-dependent operator. Apart from some results in Crouzeix's thesis [3], this appears not to have been studied previously. There are, however, a number of papers dealing with the backward Euler or Crank-Nicolson method, and a few papers studying multistep methods. These papers fall into two groups, depending on whether the results are obtained from[1]

(A) estimates for linear time-invariant equations coupled with perturbation techniques [3, 18, 21, 26], or

(B) energy estimates, e.g. [7, 27] (cf. also [14]).

Both approaches turn out to be useful also in the context of Runge-Kutta methods, and to offer different merits. They work with different assumptions about the equation (A: resolvent bounds, B: Gårding's inequality) and require different stability conditions on the part of the methods (A: $A(\theta)$-stability, B: B-stability or algebraic stability). When they apply, energy estimates provide far simpler stability and convergence proofs. It seems, however, that they do not yield the noninteger temporal convergence order which is actually observed in computations and can be explained via approach (A). When it comes to modified Runge-Kutta methods, in particular linearly implicit methods [24], there

Received by the editor October 7, 1993 and, in revised form, April 8, 1994.

1991 *Mathematics Subject Classification.* Primary 65M12, 65M15, 65M20, 65J10, 65J15.

Key words and phrases. Runge-Kutta time discretization, quasi-linear parabolic equations, error bounds, algebraic stability, $A(\theta)$-stability.

[1](A) relates to A-stability, (B) to B-stability.

is usually no alternative left to choosing (A). So we have found it worthwhile to present both approaches in this paper (also because (B) is quite short).

In §1 we use energy estimates to derive error bounds for algebraically stable Runge-Kutta methods applied to a class of quasi-linear parabolic equations, or to their spatial semidiscretizations in the method of lines. Our treatment here is certainly influenced by the classical paper of Douglas and Dupont [7], where the Crank-Nicolson method is studied. As algebraic stability has been tied to energy estimates since its introduction by Burrage and Butcher [1] and Crouzeix [4], a result like our Theorem 1.1 is possibly without surprise to the experts in the field. We note, however, that the somewhat related B-convergence theory of Frank et al. [8] does not apply to the equations studied here and, moreover, would only predict a smaller temporal convergence order than does Theorem 1.1 (q instead of $q+1$, where q is the stage order of the Runge-Kutta method).

An approach of type (A) is followed in the remaining §§2 to 5. It is different from Crouzeix's [3] approach to linear parabolic equations with time-dependent operator. Crouzeix uses a theorem of von Neumann and perturbation techniques to show step-by-step stability of A-stable methods and then obtains convergence by accumulating local errors similarly to the convergence proofs for nonstiff ordinary differential equations. Notwithstanding its merits, that result yields only suboptimal orders of convergence, it does not give error bounds in the energy norm, and it does not apply to $A(\theta)$-stable methods with $\theta < \pi/2$.

In §2 we derive some new stability estimates for strongly $A(\theta)$-stable Runge-Kutta methods applied to linear parabolic equations with constant operator. Generating functions, Parseval's formula, resolvent bounds, and techniques from [23] and [26] are the tools in this stability analysis.

In §3 we consider parabolic equations with time-dependent operator. The estimates of §2 are such that they extend to the time-dependent case in a very simple way (Lemma 3.1), by taking up an idea of Savaré [26], who recently studied multistep methods for such equations. Convergence results are then presented in Theorems 3.2 and 3.3. The temporal order of convergence in the energy norm is $\min(p, q+1+\beta)$, where p denotes the (nonstiff) order and q the stage order of the Runge-Kutta method, and β depends on the *spatial* smoothness of the solution and on the boundary conditions. In the case of time-dependent strongly elliptic second-order operators, we get the following values of β when the error is measured in temporally discrete versions of the $L_t^2(H_x^1) \cap L_t^\infty(L_x^2)$ norm: With homogeneous Dirichlet boundary conditions, we have $\beta = \frac{3}{4} - \varepsilon$ for arbitrary $\varepsilon > 0$, in the sense of an error bound $C(\varepsilon) \cdot h^{q+1+3/4-\varepsilon}$, which can probably be sharpened to $C \cdot h^{q+1+3/4} \cdot |\log h|^r$ with a small power of the logarithm. In the case of Neumann boundary conditions we get $\beta = \frac{5}{4} - \varepsilon$ in 1 space dimension, and $\beta = \frac{1}{4} - \varepsilon$ in 2 and more space dimensions. Periodic boundary conditions yield the full order p (always assuming sufficient temporal and spatial smoothness of the solution). We observed the sharpness of these convergence orders in our numerical experiments. Cf. also [25] and [23], where fractional convergence orders for linear and semilinear equations with constant operator are studied.

Section 4 extends these results to quasi-linear equations. For solution-dependent strongly elliptic second-order operators, we still get the results sketched above in 1 and 2 space dimensions, but our assumptions lead to some problems in 3 space dimensions.

Finally, §5 shows that the results of §§2 to 4 extend to variable stepsizes under mild restrictions on the time step sequence.

We conclude this section by recalling some terminology (cf. [2, 12]). A Runge-Kutta (RK) method applied to an initial value problem

$$u' = F(t, u), \qquad u(0) = u_0,$$

with a stepsize $h > 0$ yields at $t_n = nh$ an approximation u_n given recursively by

$$u_{n+1} = u_n + h\sum_{j=1}^{m} b_j U'_{nj}, \quad U_{ni} = u_n + h\sum_{j=1}^{m} a_{ij} U'_{nj},$$
$$U'_{ni} = F(t_n + c_i h, U_{ni}) \qquad (i = 1, \dots, m).$$

The Runge-Kutta method has *order* p, if the error of the method, when applied to ordinary differential equations with sufficiently differentiable right-hand side, satisfies $u_n - u(t_n) = O(h^p)$ as $h \to 0$, uniformly on bounded time intervals. We always assume $p \geq 1$. The method has *stage order* q, if $\sum_{j=1}^{m} a_{ij} c_j^{k-1} = c_i^k / k$ for $k = 1, \dots, q$ and all i. In the following we will use the notation

$$\mathscr{A} = (a_{ij})_{i,j=1}^{m}, \quad b^T = (b_1, \dots, b_m), \quad \mathbf{1} = (1, \dots, 1)^T.$$

A Runge-Kutta method is called $A(\theta)$-*stable*, if $I - z\mathscr{A}$ is nonsingular in the sector $|\arg(-z)| \leq \theta$, and if the absolute value of the *stability function* $R(z) = 1 + zb^T(I - z\mathscr{A})^{-1}\mathbf{1}$ is bounded by 1 for $|\arg(-z)| \leq \theta$. The method is called *strongly* $A(\theta)$-*stable*, if it is $A(\theta)$-stable and in addition has an invertible Runge-Kutta matrix $\mathscr{A}$, and the limit of the stability function at infinity, $R(\infty) = 1 - b^T\mathscr{A}^{-1}\mathbf{1}$, has absolute value strictly smaller than 1. The Runge-Kutta method is called *algebraically stable* if the matrix $(b_i a_{ij} + b_j a_{ji} - b_i b_j)_{i,j=1}^{m}$ is positive semidefinite and all weights b_i are positive.

Throughout the paper, C will denote a generic constant which takes on different values on different occurrences.

Energy Estimates

1. A convergence result for algebraically stable RK methods

In this section we use energy estimates to derive a convergence result for algebraically stable Runge-Kutta methods applied to the initial value problem

$$u' + A(u)u = f(t), \qquad u(0) = u_0. \tag{1.1}$$

The setting of this equation is as follows: Let H and V be (real, separable) Hilbert spaces with norms $|\cdot|$ and $\|\cdot\|$, respectively, such that V is embedded densely and continuously in H. The norm on the dual space V' is denoted by $\|\cdot\|_*$. We identify H and its dual H', so that $V \subset H = H' \subset V'$, and the duality $\langle \cdot, \cdot \rangle$ between V' and V coincides on $H \times V$ with the scalar product of H. We assume that, uniformly for all $u \in V$, the bilinear form associated with the linear operator $A(u) : V \to V'$ satisfies the Gårding inequality

$$\langle A(u)v, v \rangle \geq \alpha \cdot \|v\|^2 - c \cdot |v|^2, \quad \text{for } u, v \in V \tag{1.2}$$

with $\alpha > 0$ and $c \geq 0$, and is bounded by

$$|\langle A(u)v, w\rangle| \leq M \cdot \|v\| \cdot \|w\| \quad \text{for } u, v, w \in V. \tag{1.3}$$

Further we assume that there is a subset S of V such that the following Lipschitz condition is satisfied: For every $\delta > 0$, there exists $L = L(\delta, S)$ such that

$$\|(A(v) - A(w))u\|_* \leq \delta \cdot \|v - w\| + L \cdot |v - w| \quad \text{for } u \in S,\ v, w \in V. \tag{1.4}$$

Example (cf. [7]). On a smooth domain $\Omega \subset \mathbf{R}^d$, consider the quasi-linear parabolic equation

$$\frac{\partial u}{\partial t} = \sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}(u(x,t))\frac{\partial u}{\partial x_j}\right) + f(x,t), \qquad x \in \Omega,\ 0 < t \leq T$$

with Neumann boundary conditions

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

where $(n_1, \ldots, n_d)(x)$ denotes the normal vector. The coefficient functions $a_{ij} : \mathbf{R} \to \mathbf{R}$ are assumed to be bounded and Lipschitz bounded, and the matrices $(a_{ij}(\mu))$ $(\mu \in \mathbf{R})$ are uniformly positive definite. The variational formulation of this problem is of the form (1.1) on $H = L^2(\Omega)$ and $V = H^1(\Omega)$, with operators $A(u) : V \to V'$ defined by

$$\langle A(u)v, w\rangle = \int_\Omega \sum_{i,j=1}^{d} a_{ij}(u(x)) \cdot \frac{\partial v}{\partial x_j} \cdot \frac{\partial w}{\partial x_i} \cdot dx.$$

This satisfies (1.2) and (1.3). Condition (1.4) holds with

$$S = S(r) = \left\{u \in H^1(\Omega) : \sup_{x \in \Omega} |\nabla u(x)| \leq r\right\},$$

because for $u \in S(r)$ and $v, w, \varphi \in V$ we have

$$\left|\int_\Omega \sum_{i,j} (a_{ij}(v(x)) - a_{ij}(w(x))) \cdot \frac{\partial u}{\partial x_j} \cdot \frac{\partial \varphi}{\partial x_i} \cdot dx\right| \leq l \cdot |v - w|_{L^2} \cdot r \cdot \|\varphi\|_{H^1},$$

where l denotes a Lipschitz constant of the functions $a_{ij}(\cdot)$. This example is readily extended to include first-order terms, or to Dirichlet or mixed boundary conditions. □

We have the following convergence result.

Theorem 1.1. *Consider the initial value problem* (1.1)–(1.4). *Let* $\{u_n\} \subset V$ *be a Runge-Kutta solution obtained with an algebraically stable method of stage order* q *and order* $p \geq q + 1$ *having an invertible coefficient matrix* $\mathcal{A}$ *and* $|R(\infty)| < 1$. *If equation* (1.1) *has a solution* $u(t) \in S$ *for* $0 \leq t \leq T$, *with temporal derivatives* $u^{(q+1)} \in L^2(0, T; V)$ *and* $u^{(q+2)} \in L^2(0, T; V')$, *then for*

sufficiently small stepsizes h *(restricted only by the constants in* (1.2)–(1.4)*), the error is bounded for* $Nh \le T$ *by*

$$h\sum_{n=0}^{N}\|u_n - u(t_n)\|^2 + \max_{0\le n\le N}|u_n - u(t_n)|^2 \le C\cdot(h^{q+1})^2\cdot\left(\int_0^T \|u^{(q+1)}(t)\|^2\,dt + \int_0^T \|u^{(q+2)}(t)\|_*^2\,dt\right). \tag{1.5}$$

The constant C *depends on the Runge-Kutta method, on the constants in* (1.2)–(1.4)*, and on* T.

Remarks. (a) Theorem 1.1 generalizes to variable stepsizes. The proof makes no essential use of constant stepsizes, in contrast to the proofs in §§2 to 4.

(b) Theorem 1.1 has an obvious extension to the situation where the constants in (1.2) and (1.3) are allowed to depend on $\|u\|$ and to deteriorate with growing $\|u\|$. (The constant C in (1.5) then depends also on $\sup_{0\le t\le T}\|u(t)\|$.) For example, the incompressible Navier-Stokes equations in dimension 2 and 3 then fit into the framework. Moreover, $f(t)$ in equation (1.1) can be replaced by $f(t, u)$ satisfying a local Lipschitz condition $\|f(t, v) - f(t, w)\|_* \le \delta\cdot\|v - w\| + L(\delta, r)\cdot|v - w|$ for $\|v\| + \|w\| \le r$. This is often satisfied for first-order nonlinearities. Of course, the operator A may also depend on t.

(c) Equation (1.1) can also result from *space discretization* of a parabolic initial-boundary value problem, with conditions (1.2)–(1.4) holding uniformly in the meshwidth. In this situation, it is more interesting to compare the fully discrete solution to the solution of the PDE rather than that of the spatial semidiscretization. A projection $\hat{u}(t)$ of the PDE solution onto the finite-dimensional approximation space then satisfies a perturbed equation (1.1):

$$\hat{u}' + A(\hat{u})\hat{u} = f(t) + d(t), \qquad \hat{u}(0) = u_0 + e_0,$$

where $d(t)$ is the spatial truncation error. If $\hat{u}(t)$ is in S and sufficiently smooth, then the difference between the Runge-Kutta solution u_n of equation (1.1) and $\hat{u}(t)$ is bounded by

$$h\sum_{n=1}^{N}\|u_n - \hat{u}(t_n)\|^2 + \max_{1\le n\le N}|u_n - \hat{u}(t_n)|^2 \le C\cdot\left(\widehat{B} + |e_0|^2 + |R(\infty)|\cdot h\cdot\|e_0\|^2 + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|d(t_n + c_i h)\|_*^2\right),$$

where $\widehat{B}$ is the expression on the right-hand side of (1.5), with u replaced by $\hat{u}$. The proof of this estimate is a simple extension of the proof of Theorem 1.1. Errors resulting from the inexact solution of the nonlinear Runge-Kutta equations can be bounded similarly.

(d) In finite dimension, the *existence* of a numerical solution can be shown under the method assumption of [6, Thm. II.5.4]: There exists a positive diagonal matrix D such that $D\mathscr{A}+\mathscr{A}^T D$ is positive definite (cf. also [12, Ch. IV.14]). Using condition (1.2), one shows that the iteration $U_{ni}'^{(1)} + A(U_{ni}^{(0)})U_{ni}^{(1)} = f(t_n + c_i h)$ maps some ball into itself and one applies Brouwer's fixed-point theorem. *Uniqueness* is obtained from condition (1.4) only if there exists a numerical solution with internal stages $U_{ni} \in S$, which is not guaranteed in general.

Proof of Theorem 1.1. The proof combines arguments that are familiar from B-stability theory and from energy estimates for the time-continuous case.

(a) For brevity, we denote the solution values $\widetilde{U}_{ni} = u(t_n + c_i h)$, $\widetilde{U}'_{ni} = u'(t_n + c_i h)$, and $\tilde{u}_n = u(t_n)$. We then have

$$\widetilde{U}_{ni} = \tilde{u}_n + h\sum_{j=1}^{m} a_{ij}\widetilde{U}'_{nj} + D_{ni}, \quad \tilde{u}_{n+1} = \tilde{u}_n + h\sum_{j=1}^{m} b_j\widetilde{U}'_{nj} + d_{n+1},$$

where the defects are of the form

$$D_{ni} = h^q \int_{t_n}^{t_{n+1}} \kappa_i\left(\frac{t-t_n}{h}\right) u^{(q+1)}(t)\,dt,$$

$$d_{n+1} = h^{q+1}\int_{t_n}^{t_{n+1}} \kappa\left(\frac{t-t_n}{h}\right) u^{(q+2)}(t)\,dt$$

$$= -h^q \int_{t_n}^{t_{n+1}} \kappa'\left(\frac{t-t_n}{h}\right) u^{(q+1)}(t)\,dt$$

with bounded Peano kernels κ_i and κ (for simplicity we assume that all $c_i \in [0, 1]$). Here we have used Taylor expansion and the definition of the stage order q and the order $p \geq q+1$. We note for later use that, by the Cauchy-Schwarz inequality,

$$h\sum_{n=0}^{N}\sum_{j=1}^{m} \|D_{nj}\|^2 + h\sum_{n=0}^{N}(\|d_{n+1}\|^2 + \|d_{n+1}/h\|_*^2) \leq C\cdot B, \tag{1.6}$$

where B denotes the expression on the right-hand side of (1.5).

(b) The errors $E_{ni} = U_{ni} - \widetilde{U}_{ni}$, $E'_{ni} = U'_{ni} - \widetilde{U}'_{ni}$, and $e_n = u_n - \tilde{u}_n$ thus satisfy

$$E'_{ni} + A(U_{ni})E_{ni} = -(A(U_{ni}) - A(\widetilde{U}_{ni}))\widetilde{U}_{ni}, \tag{1.7a}$$

$$E_{ni} = e_n + h\sum_{j=1}^{m} a_{ij}E'_{nj} - D_{ni}, \tag{1.7b}$$

$$e_{n+1} = e_n + h\sum_{i=1}^{m} b_i E'_{ni} - d_{n+1}. \tag{1.7c}$$

We take the square of the H-norm in (1.7c) to obtain

$$|e_{n+1}|^2 = \left|e_n + h\sum_{i=1}^{m} b_i E'_{ni}\right|^2 - 2\left\langle d_{n+1}, e_n + h\sum_{i=1}^{m} b_i E'_{ni}\right\rangle + |d_{n+1}|^2. \tag{1.8}$$

We estimate the three terms on the right-hand side separately. Expressing e_n by equation (1.7b), we have

$$\left|e_n + h\sum_{i=1}^{m} b_i E'_{ni}\right|^2 = |e_n|^2 + 2h\sum_{i=1}^{m} b_i\langle E'_{ni}, E_{ni} + D_{ni}\rangle$$

$$+ h^2\sum_{i=1}^{m}\sum_{j=1}^{m}(b_i b_j - b_i a_{ij} - b_j a_{ji})\cdot\langle E'_{ni}, E'_{nj}\rangle.$$

As the method is algebraically stable, the last term is nonpositive. For the middle term we note that by (1.7a) we have (omitting all the subscripts n, i)

$$\langle E', E+D\rangle = -\langle A(U)E, E\rangle - \langle A(U)E, D\rangle - \langle (A(U)-A(\tilde{U}))\tilde{U}, E+D\rangle.$$

Using conditions (1.2)–(1.4) (note that $\tilde{U}_{ni} = u(t_n + c_i h) \in S$ by assumption), we can bound this by

$$\langle E', E+D\rangle \le -\alpha\cdot\|E\|^2 + c\cdot|E|^2 + M\cdot\|E\|\cdot\|D\| + (\delta\|E\| + L|E|)\cdot\|E+D\|$$

and hence, for sufficiently small δ,

$$\langle E', E+D\rangle \le -\frac{\alpha}{2}\cdot\|E\|^2 + C\cdot|E|^2 + C\cdot\|D\|^2.$$

The second term in (1.8) is estimated as

$$\left|\left\langle d_{n+1}, e_n + h\sum_{i=1}^m b_i E'_{ni}\right\rangle\right| \le \|d_{n+1}\|_*\cdot\|e_n\| + \|d_{n+1}\|\cdot h\sum_{i=1}^m b_i\|E'_{ni}\|_*$$

$$\le \frac{1}{2}h\delta\|e_n\|^2 + \frac{1}{2}h/\delta\cdot\|d_{n+1}/h\|_*^2 + Ch\delta\cdot\sum_{j=1}^m\|E_{nj}\|^2 + Ch/\delta\cdot\|d_{n+1}\|^2,$$

with a small δ. Here we have used the bound

$$\|E'_{ni}\|_* \le C\cdot\sum_{j=1}^m\|E_{nj}\|, \tag{1.9}$$

which is a consequence of equation (1.7a) and conditions (1.3) and (1.4). Finally, the last term in (1.8) is bounded by

$$|d_{n+1}|^2 \le \|d_{n+1}\|_*\cdot\|d_{n+1}\| \le \frac{1}{2}h\cdot\|d_{n+1}\|^2 + \frac{1}{2}h\cdot\|d_{n+1}/h\|_*^2.$$

Putting all these estimates together, we have shown (note that $b_i > 0$ for all i)

$$\begin{aligned}|e_{n+1}|^2 - |e_n|^2 + \alpha/3\cdot h\cdot\sum_{i=1}^m b_i\|E_{ni}\|^2 &\\ \le Ch\delta\cdot\|e_n\|^2 + Ch\sum_{i=1}^m|E_{ni}|^2 &\\ + Ch\sum_{i=1}^m\|D_{ni}\|^2 + Ch\cdot(\|d_{n+1}\|^2 + \|d_{n+1}/h\|_*^2).&\end{aligned} \tag{1.10}$$

(c) From equations (1.7b, c) we infer

$$e_{n+1} = R(\infty)\cdot e_n + b^T\mathscr{C}^{-1}(E_n + D_n) - d_{n+1},$$

and since $|R(\infty)| < 1$, this implies

$$h\sum_{n=0}^N\|e_{n+1}\|^2 \le Ch\sum_{n=0}^N\sum_{i=1}^m\|E_{ni}+D_{ni}\|^2 + Ch\sum_{n=0}^N\|d_{n+1}\|^2. \tag{1.11}$$

(d) Summing the inequalities (1.10) from $n = 0$ to N and inserting (1.6) and (1.11), we get

$$|e_{N+1}|^2 + h\sum_{n=0}^N\sum_{i=1}^m\|E_{ni}\|^2 \le Ch\sum_{n=0}^N\sum_{i=1}^m|E_{ni}|^2 + C\cdot B. \tag{1.12}$$

We next estimate

$$|E_{ni}|^2 \le \tfrac{1}{2}\delta \cdot \|E_{ni}\|^2 + \tfrac{1}{2}\delta^{-1} \cdot \|E_{ni}\|_*^2$$

and bound $\|E_{ni}\|_*$, using the triangle inequality in (1.7b), the continuity $\|e_n\|_* \le C \cdot |e_n|$ of the inclusion $H \subset V'$, and the estimate (1.9):

$$\|E_{ni}\|_* \le C \cdot |e_n| + Ch \sum_{j=1}^{m} \|E_{nj}\| + \|D_{ni}\| .$$

Hence,

$$(1.13)\quad h \sum_{n=0}^{N} \sum_{i=1}^{m} |E_{ni}|^2 \le (\delta/2 + Ch^2/\delta) \cdot h \sum_{n=0}^{N} \sum_{j=1}^{m} \|E_{nj}\|^2 + Ch \sum_{n=0}^{N} |e_n|^2 + C \cdot B .$$

We insert this bound into (1.12) to get (again for a suitable choice of δ)

$$|e_{N+1}|^2 \le Ch \sum_{n=0}^{N} |e_n|^2 + C \cdot B , \qquad 0 \le Nh \le T ,$$

and the discrete Gronwall inequality now gives us

$$|e_n|^2 \le C \cdot B , \qquad 0 \le nh \le T .$$

We insert this estimate back into (1.13), and (1.13) back into (1.12), and so obtain

$$h \sum_{n=0}^{N} \sum_{i=1}^{m} \|E_{ni}\|^2 \le C \cdot B .$$

This bound inserted into (1.11) finally gives us

$$h \sum_{n=0}^{N} \|e_{n+1}\|^2 \le C \cdot B ,$$

and the theorem is proved. □

Fourier and Perturbation Techniques

2. Stability estimates for linear time-invariant equations

In this section we derive some estimates for Runge-Kutta methods applied to equations with constant operator

$$(2.1)\qquad u' + Au = f(t) , \quad u(0) = u_0 \quad (t > 0) .$$

We study this equation in a Hilbert space framework of analytic semigroups (cf., e.g., Kato [15, Chs. VI and IX], Lions [19, Chs. IV and VI], and Lasiecka [16]), emphasizing the role of resolvent bounds. On a (complex, separable) Hilbert space H with scalar product $(\cdot\,,\,\cdot)$ and norm $|\cdot|$, let $-A$ be the generator of a bounded analytic semigroup that has 0 in its resolvent set. In other words, $A : D(A) \subset H \to H$ is a densely defined closed linear operator whose resolvent is bounded by

$$(2.2)\qquad |(\lambda + A)^{-1}|_{H \leftarrow H} \le \frac{M}{1 + |\lambda|} \quad \text{for } |\arg \lambda| \le \pi - \varphi \ \left(\varphi < \frac{\pi}{2}\right) .$$

We consider a second Hilbert space $V \subset H$ with norm $\|\cdot\|$ and assume that

$$V = D(A^{1/2}) = D(A^{*1/2}) \quad \text{with equivalent norms,} \tag{2.3}$$

where as usual the norm on $D(A^{1/2})$ is given by $\|v\|_{D(A^{1/2})} = |A^{1/2}v|$. In particular, $A^{1/2}$ and $A^{*1/2}$ are isomorphisms between V and H. It follows that the sesquilinear form defined by $(Au, v) = (A^{1/2}u, A^{*1/2}v)$ for $u \in D(A)$, $v \in V$ extends to a bounded sesquilinear form on $V \times V$, and consequently A extends to an isomorphism from V to its conjugate linear dual V' which we again denote by A:

$$A : V \to V' \text{ is bounded and invertible.} \tag{2.4}$$

The norm on V' will be denoted by $\|\cdot\|_*$. We always identify H and H', so that

$$V \subset H = H' \subset V', \text{ with duality } \langle v', v\rangle = (v', v) \text{ for } v' \in H, v \in V. \tag{2.5}$$

From (2.2) we get the bounds

$$\begin{aligned} \|(\lambda + A)^{-1}\|_{V \leftarrow V} &\le \frac{M_1}{1 + |\lambda|}, \\ \|(\lambda + A)^{-1}\|_{V \leftarrow V'} &\le M_2 \end{aligned} \tag{2.6}$$

for $|\arg\lambda| \le \pi - \varphi$, where M_1 and M_2 can be chosen to depend only on the constants in (2.2) and (2.3).

Remark. On finite time intervals, all our results remain valid if, for some $c > 0$, the operator $A + cI$ instead of A satisfies conditions (2.2) and (2.3).

Examples. For A a second-order strongly elliptic differential operator on a bounded domain Ω with Neumann boundary conditions (not necessarily self-adjoint), $A + cI$ satisfies for suitable $c \ge 0$ the above assumptions on $H = L^2(\Omega)$ and $V = H^1(\Omega)$. The bound (2.2) is well known, and condition (2.3) follows with the help of, e.g., Theorem 1.4.8 in Henry [13], or Lions [20]. The assumptions are equally met for Dirichlet boundary conditions (with $V = H_0^1(\Omega)$) and mixed boundary conditions. The conditions can also be verified for finite element discretizations of such operators, uniformly in the gridsize. □

We will now give stability bounds for Runge-Kutta solutions of equation (2.1). It is convenient to split the problem into the two special cases where either $u_0 = 0$ or $f(t) \equiv 0$. The general case then follows by superposition.

Lemma 2.1. *Consider equations* (2.1)–(2.3) *with* $u_0 = 0$. *Let the Runge-Kutta method be strongly* $A(\theta)$*-stable with* $\theta > \varphi$. *Then, the numerical solution* (u_n) *and the internal stages* (U_{ni}) *are bounded for* $h > 0$ *by*

$$h\sum_{n=0}^{N}\|u_{n+1}\|^2 + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|U_{ni}\|^2 \le C \cdot h\sum_{n=0}^{N}\sum_{i=1}^{m}\|f(t_n + c_i h)\|_*^2 \tag{2.7}$$

for every $N \ge 0$. *The constant* C *is independent of* h, N, *and* f.

Proof. (a) We consider the generating functions

$$u(\zeta) = \sum_{n=0}^{\infty} u_{n+1}\zeta^n, \quad U(\zeta) = \sum_{n=0}^{\infty} U_n\zeta^n, \quad F(\zeta) = \sum_{n=0}^{\infty} F_n\zeta^n,$$

with $U_n = (U_{ni})_{i=1}^m$, $F_n = (f(t_n + c_i h))_{i=1}^m$. A calculation yields (see Lemma 3.1 of [23]) that they are related by[2]

$$U(\zeta) = \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} F(\zeta), \tag{2.8a}$$

$$u(\zeta) = \frac{b^T \mathscr{C}^{-1}}{1 - R(\infty)\zeta} U(\zeta) \tag{2.8b}$$

with $\Delta(\zeta) = (\mathscr{C} + \frac{\zeta}{1-\zeta} \mathbf{1} b^T)^{-1}$. We will show in part (b) of the proof that

$$\left\| \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} \right\|_{V^m \leftarrow (V')^m} \leq C, \qquad |\zeta| \leq 1. \tag{2.9}$$

Then, using Parseval's formula in (2.8a) gives us the desired estimate for the internal stages U_{ni}. Since $|R(\infty)| < 1$, Parseval's formula used in (2.8b) yields

$$\sum_{n=0}^{N} \|u_{n+1}\|^2 \leq \text{Const} \sum_{n=0}^{N} \sum_{i=1}^{m} \|U_{ni}\|^2$$

and hence the bound (2.7).

(b) It remains to show (2.9). Let $|\zeta| \leq 1$, $\zeta \neq 1$. We have

$$(\Delta(\zeta)/h + A)^{-1} = \frac{1}{2\pi i} \int_\gamma (z/h + A)^{-1} \cdot (z - \Delta(\zeta))^{-1} \, dz, \tag{2.10}$$

where γ is a union of bounded contours that enclose the eigenvalues of $\Delta(\zeta)$. The formula of Lemma 2.4 in [23],

$$(\Delta(\zeta) - z)^{-1} = \mathscr{C}(I - z\mathscr{C})^{-1} + \frac{\zeta}{1 - R(z)\zeta} (I - z\mathscr{C})^{-1} \mathbf{1} b^T (I - z\mathscr{C})^{-1}, \tag{2.11}$$

shows that the eigenvalues of $\Delta(\zeta)$ are either eigenvalues of $\mathscr{C}^{-1}$ or satisfy $R(z) = 1/\zeta$. By $A(\theta)$-stability, they are all in the sector $|\arg z| \leq \pi - \theta < \pi - \varphi$. Moreover, for ζ bounded away from 1, all eigenvalues of $\Delta(\zeta)$ are bounded away from 0. These eigenvalues can be enclosed by bounded contours in $|\arg z| \leq \pi - \varphi$ that stay a fixed positive distance away from the eigenvalues, so that on these contours $(\Delta(\zeta) - z)^{-1}$ is uniformly bounded for $|\zeta| \leq 1$. By the bound (2.6), the contribution of these eigenvalues in equation (2.10) thus gives us uniformly bounded operators from $(V')^m$ to V^m for $|\zeta| \leq 1$.

It remains to study the contribution of the eigenvalue $z(\zeta)$ near 0 occurring for ζ near 1. Hence, let $z(\zeta)$ be defined by $R(z(\zeta)) = 1/\zeta$ for ζ near 1, with $z(1) = 0$. By $A(\theta)$-stability, $|\arg z(\zeta)| \leq \pi - \theta$ for $|\zeta| \leq 1$, ζ near 1. Formula (2.11) shows that

$$(\Delta(\zeta) - z)^{-1} = \frac{\text{Res}(z(\zeta))}{z(\zeta) - z} + O(1), \quad z \to z(\zeta), \ \zeta \text{ near } 1,$$

where the residue $\text{Res}(z) = (I - z\mathscr{C})^{-1} \mathbf{1} b^T (I - z\mathscr{C})^{-1} / R'(z)$ is bounded for z near 0 (since $R'(0) = 1$). Therefore, the contribution in (2.10) of the contour γ_0 encircling $z(\zeta)$ is equal to

$$\frac{1}{2\pi i} \int_{\gamma_0} (z/h + A)^{-1} \cdot (z - \Delta(\zeta))^{-1} \, dz = (z(\zeta)/h + A)^{-1} \cdot \text{Res}(z(\zeta)). \tag{2.12}$$

[2]As in (2.8a), we often write $\Delta(\zeta)$ instead of $\Delta(\zeta) \otimes I_V$, and A instead of $I_m \otimes A$.

Again by (2.6), this is uniformly bounded as an operator from $(V')^m$ to V^m for ζ near 1 with $|\zeta| \le 1$. □

The following lemma shows that, pointwise in time in the H-norm, the solution is again bounded as in Lemma 2.1. For different pointwise estimates, cf. Lemma 3.5 of [23].

Lemma 2.2. *Under the assumptions of Lemma* 2.1, *we have for all* $n \ge 0$

$$|u_{n+1}|^2 + \max_{i=1,\dots,m} |U_{ni}|^2 \le C \cdot h \sum_{\nu=0}^{n} \sum_{i=1}^{m} \|f(t_\nu + c_i h)\|_*^2 . \tag{2.13}$$

Here, C *is again independent of* n, h, *and* f.

Proof. (a) We start from the inequality

$$|u_{n+1}|^2 \le 2 \sum_{\nu=0}^{n} |(u_{\nu+1} - u_\nu , u_{\nu+1})| , \tag{2.14}$$

which holds for arbitrary sequences (u_n) with $u_0 = 0$. Inserting $u_{\nu+1} - u_\nu$ from the Runge-Kutta method and using the duality (2.5), we obtain

$$|u_{n+1}|^2 \le 2 \sum_{\nu=0}^{n} \left| \left\langle h \sum_{i=1}^{m} b_i(-AU_{\nu i} + f(t_\nu + c_i h)) , u_{\nu+1} \right\rangle \right| .$$

By the Cauchy-Schwarz inequality, this implies

$$|u_{n+1}|^2 \le 2 \left(h \sum_{\nu=0}^{n} \left\| \sum_{i=1}^{m} b_i(-AU_{\nu i} + f(t_\nu + c_i h)) \right\|_*^2 \right)^{1/2} \cdot \left(h \sum_{\nu=0}^{n} \|u_{\nu+1}\|^2 \right)^{1/2} .$$

By the bound (2.4) and Lemma 2.1, each of the sums on the right-hand side is bounded by a constant times the right-hand side of (2.7). This shows that $|u_{n+1}|^2$ is bounded as stated in (2.13).

(b) To prove the estimate for the internal stages, we note that $U_n = (U_{ni})_{i=1}^m$ is given from the Runge-Kutta formulas as

$$U_n = (I + h\mathscr{A} \otimes A)^{-1}(h\mathscr{A} F_n + \mathbf{1} u_n) .$$

Since the eigenvalues of $\mathscr{A}$ are all in the interior of the sector $|\arg \lambda| \le \pi - \varphi$, it follows from (2.2) that $(I + h\mathscr{A} \otimes A)^{-1}$ is uniformly bounded as an operator from H^m to H^m. We also have the resolvent bound

$$\|(\lambda + A)^{-1}\|_{H \leftarrow V'} \le C \cdot |\lambda|^{-1/2} , \quad |\arg \lambda| \le \pi - \varphi . \tag{2.15}$$

This follows from (2.6) via $|(\lambda + A)^{-1} w|^2 \le \|(\lambda + A)^{-1} w\|_* \cdot \|(\lambda + A)^{-1} w\| \le C \cdot |\lambda|^{-1} \cdot \|w\|_*^2$ for $w \in V'$. Now, (2.15) implies that $(I + h\mathscr{A} \otimes A)^{-1}$ is bounded by $O(h^{-1/2})$ as an operator from $(V')^m$ to H^m. Hence, we have

$$|U_{ni}|^2 \le C \cdot \left(|u_n|^2 + h \sum_{j=1}^{m} \|f(t_n + c_j h)\|_*^2 \right) .$$

As we know already that $|u_n|^2$ is bounded by (2.13), this gives the desired bound also for the internal stages U_{ni}. □

A dual version of Lemma 2.2 is the following.

Lemma 2.3. *Under the assumptions of Lemma* 2.2, *we have for all* $N \geq 0$

$$\left(h \sum_{n=0}^{N} \|u_{n+1}\|^2 + h \sum_{n=0}^{N} \sum_{i=1}^{m} \|U_{ni}\|^2 \right)^{1/2} \leq C \cdot h \sum_{n=0}^{N} \sum_{i=1}^{m} |f(t_n + c_i h)| \tag{2.16}$$

with a constant as in (2.13).

Proof. The result relies on a duality argument as in [26, Claim 2.8]. For the internal stages, we have to bound the mapping $l_N^1(H^m) \to l_N^2(V^m) : (F_n)_{n=0}^N \mapsto (U_n)_{n=0}^N$. This has the same operator norm as its adjoint $l_N^2(V'^m) \to l_N^\infty(H^m) : (\tilde{F}_n)_{n=0}^N \mapsto (\tilde{U}_n)_{n=0}^N$, which is given by the following scheme, derived by taking adjoints in (2.8a):

$$\begin{aligned} \tilde{U}_n &= b\tilde{u}_n - h(\mathscr{C}^T \otimes A^*)\tilde{U}_n + h\mathscr{C}^T \tilde{F}_n, \\ \tilde{u}_{n-1} &= \tilde{u}_n - h(\mathbf{1}^T \otimes A^*)\tilde{U}_n + h\mathbf{1}^T \tilde{F}_n, \quad \tilde{u}_N = 0. \end{aligned}$$

In the same way as in Lemma 2.2, one obtains the bound (2.13) for the dual variables:

$$\max_{0 \leq n \leq N} \max_{1 \leq i \leq m} |\tilde{U}_{ni}|^2 \leq C \cdot h \sum_{n=0}^{N} \sum_{i=1}^{m} \|\tilde{F}_{ni}\|_*^2,$$

i.e., the required bound for the adjoint mapping. We thus get the bound (2.16) for the internal stages (U_{ni}), and the result for $(u_{n+1})_{n=0}^N$ then follows from equation (2.8b) via Parseval's formula. □

Next we consider equation (2.1) with $f(t) \equiv 0$. Then the Runge-Kutta solution is just $u_n = R(-hA)^n u_0$. It is known from Le Roux [17] (cf. also [5]) that $R(-hA)$ is uniformly power-bounded if $R(z)$ is the stability function of a (strongly) $A(\theta)$-stable method and A satisfies (2.2). Hence,

$$|u_n| \leq C \cdot |u_0| \quad \text{for } n \geq 0.$$

For the l^2-norm in time/V-norm in space we have the following estimate:

Lemma 2.4. *Consider equations* (2.1)–(2.3) *with* $f(t) \equiv 0$. *Let the Runge-Kutta method be strongly* $A(\theta)$*-stable with* $\theta > \varphi$. *Then, the numerical solution and the internal stages are bounded for all* $N \geq 0$ *by*

$$h \sum_{n=0}^{N} \|u_{n+1} - R(\infty)^{n+1} u_0\|^2 + h \sum_{n=0}^{N} \sum_{i=1}^{m} \|U_{ni}\|^2 \leq C \cdot |u_0|^2. \tag{2.17}$$

The constant C *is independent of* h *and* N.

Remark. Note that $h \sum_{n=0}^{N} \|u_{n+1}\|^2 \leq C \cdot |u_0|^2$ if $R(\infty) = 0$. In this case one has also $\|u_n\| \leq C \cdot (nh)^{-1/2} \cdot |u_0|$ for $n \geq 1$, see [22, formula (3.31)].

Proof of Lemma 2.4. The generating functions satisfy (cf. (2.8))

$$U(\zeta) = \left(\frac{\Delta(\zeta)}{h} + A \right)^{-1} \frac{\Delta(\zeta)}{h} \frac{\mathbf{1}}{1-\zeta} u_0, \tag{2.18a}$$

$$u(\zeta) = \frac{1}{1 - R(\infty)\zeta} (b^T \mathscr{C}^{-1} U(\zeta) + R(\infty) u_0). \tag{2.18b}$$

Using the identity

$$\frac{\Delta(\zeta)\mathbf{1}}{1-\zeta} = \frac{\mathscr{C}^{-1}\mathbf{1}}{1-R(\infty)\zeta}, \tag{2.19}$$

which is verified by multiplying both sides with $\Delta(\zeta)^{-1} = \mathscr{C} + \frac{\zeta}{1-\zeta}\mathbf{1}b^T$, we rewrite (2.18a) in the form of (2.8a) with $F_n = (F_{ni})_{i=1}^m := \mathscr{C}^{-1}\mathbf{1}\cdot R(\infty)^n \cdot u_0/h$. For the internal stages we thus get the bound (2.17) by applying Lemma 2.3 with F_{ni} in the role of $f(t_n + c_i h)$, noting once more $|R(\infty)| < 1$. The result for the sequence $(u_{n+1} - R(\infty)^{n+1}u_0)$ then follows by using Parseval's formula in (2.18b). □

3. Linear equations with time-dependent operator

We consider time discretization of the initial value problem

$$u' + A(t)u = f(t), \quad u(0) = u_0 \quad (0 < t \le T). \tag{3.1}$$

Extending the setting of §2 to the time-dependent situation, we assume that the densely defined closed operators $A(t)\colon D(A(t)) \subset H \to H$ satisfy conditions (2.2) and (2.3) uniformly in $0 \le t \le T$:

$$|(\lambda + A(t))^{-1}|_{H\leftarrow H} \le \frac{M}{1+|\lambda|} \quad \text{for } |\arg\lambda| \le \pi - \varphi \ \left(\varphi < \frac{\pi}{2}\right), \tag{3.2}$$

with M independent of t, and

$$V = D(A(t)^{1/2}) = D(A(t)^{*1/2}) \quad \text{with equivalent norms, uniformly in } t, \tag{3.3}$$

where V is assumed not to depend on t. Of course, $A(t)$ then also satisfies (2.4) and (2.6) uniformly in t. In addition, we assume

$$\|A(t) - A(\tau)\|_{V'\leftarrow V} \le L\cdot|t-\tau|, \qquad 0 \le \tau \le t \le T. \tag{3.4}$$

Example. Consider the second-order strongly elliptic differential operator ($\partial_i := \partial/\partial x_i$) $A(t)u = \sum_{i,j}\partial_i(a_{ij}(x,t)\partial_j u) + \sum_i b_i(x,t)\partial_i u + c(x,t)u$ with smooth bounded coefficients on a smooth bounded domain Ω, equipped with Neumann boundary conditions. We take this as an unbounded operator on $H = L^2(\Omega)$. While $D(A(t))$ depends on t through the boundary conditions $\partial u/\partial n_{A(t)} = \sum_{i,j} n_i a_{ij}(x,t)\partial_j u = 0$, the space $D(A(t)^{1/2}) = V = H^1(\Omega)$ is independent of the problem coefficients (cf. Lions [19, Ch. VI.1]), and [20, §6].

Lemma 3.1. *On finite time intervals* $(0 \le nh \le Nh \le T < \infty)$, *the estimates of Lemmas* 2.1–2.4 *remain valid for the Runge-Kutta solutions of equations* (3.1)–(3.4) *with time-dependent operator.*

Proof. The proof is a discrete analogue of the following surprisingly simple proof of an estimate for the exact solution of equation (3.1). We learnt this from Savaré's paper [26], where time discretization of equation (3.1) by linear multistep methods is studied. Consider equation (3.1) rewritten in the form

$$u' + A(\bar t)u = \overline{f}(t) \quad \text{with } \overline{f}(t) = f(t) + (A(\bar t) - A(t))u(t) \tag{3.5}$$

for a fixed $\bar t \ge t$. For the time-invariant equation with operator $A(t)$, we have the estimate (cf. [16, §4])

$$\int_0^{\bar t} \|u(t)\|^2\,dt \le C\left(|u_0|^2 + \int_0^{\bar t}\|\overline{f}(t)\|_*^2\,dt\right) \tag{3.6}$$

with C independent of $\bar{t}$. To bound the term containing $\overline{f}$, one uses (3.4) and a partial integration:

$$\begin{aligned}\int_0^{\bar{t}} \|(A(\bar{t}) - A(t))u(t)\|_*^2 \, dt &\le \int_0^{\bar{t}} L^2(\bar{t}-t)^2 \|u(t)\|^2 \, dt \\ &= \int_0^{\bar{t}} 2L^2(\bar{t}-t) \left(\int_0^t \|u(\tau)\|^2 \, d\tau \right) dt .\end{aligned} \tag{3.7}$$

Hence, $w(t) = \int_0^t \|u(\tau)\|^2 \, d\tau$ satisfies an inequality of the form to which Gronwall's lemma applies. This provides an estimate (3.6) with the data f instead of $\overline{f}$.

The above arguments carry over to the discrete case without difficulty: Lemmas 2.1 and 2.4 establish the discrete version of (3.6), partial summation replaces the partial integration in (3.7), and a discrete Gronwall lemma then yields the desired estimates. We omit the details. □

If the solution of (3.1) is sufficiently smooth in time, one has the following convergence result.

Theorem 3.2. *For an initial value problem* (3.1)–(3.4) *consider a Runge-Kutta method of stage order* q *and order* $p \ge q+1$ *that is strongly* $A(\theta)$*-stable with* $\theta > \varphi$. *If* $u^{(q+1)} \in L^2(0, T; V)$ *and* $u^{(q+2)} \in L^2(0, T; V')$, *then the error is bounded for* $Nh \le T$ *by*

$$\begin{aligned}&h \sum_{n=0}^{N} \|u_n - u(t_n)\|^2 + \max_{0 \le n \le N} |u_n - u(t_n)|^2 \\ &\qquad \le C \cdot (h^{q+1})^2 \cdot \left(\int_0^T \|u^{(q+1)}(t)\|^2 \, dt + \int_0^T \|u^{(q+2)}(t)\|_*^2 \, dt \right) .\end{aligned} \tag{3.8}$$

The constant C *depends on the Runge-Kutta method, on the constants in* (3.2)–(3.4), *and on* T.

Proof. Let us first assume that the operator A is time-independent. The general case will be treated at the end of the proof.

(a) For the errors $e_{n+1} = u_{n+1} - u(t_{n+1})$ and $E_n = (E_{ni})_{i=1}^m$ with $E_{ni} = U_{ni} - u(t_n + c_i h)$ we have the recursion

$$\begin{aligned}(I + h\mathscr{A} \otimes A)E_n &= \mathbf{1} e_n - D_n , \\ e_{n+1} &= e_n - h(b^T \otimes A)E_n - d_{n+1} ,\end{aligned} \tag{3.9}$$

where d_{n+1} and $D_n = (D_{ni})_{i=1}^m$ are the defects obtained by inserting the exact solution values into the Runge-Kutta scheme. We recall that these defects satisfy the bound (1.6). The generating functions

$$\begin{aligned}e(\zeta) &= \sum_{n=0}^{\infty} e_{n+1} \zeta^n , \quad E(\zeta) = \sum_{n=0}^{\infty} E_n \zeta^n , \\ d(\zeta) &= \sum_{n=0}^{\infty} d_{n+1} \zeta^n , \quad D(\zeta) = \sum_{n=0}^{\infty} D_n \zeta^n\end{aligned} \tag{3.10}$$

are then related by (cf. (2.8))

$$E(\zeta) = -\left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} \frac{\Delta(\zeta)}{h}\left(D(\zeta) + \frac{\zeta}{1-\zeta}\mathbf{1}d(\zeta)\right), \tag{3.11a}$$

$$e(\zeta) = \frac{1}{1 - R(\infty)\zeta}(b^T\mathscr{C}^{-1}(E(\zeta) + D(\zeta)) - d(\zeta)). \tag{3.11b}$$

Using the identity (2.19), we rewrite (3.11a) as

$$E(\zeta) + D(\zeta) = \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} AD(\zeta) - \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} \frac{\mathscr{C}^{-1}\mathbf{1}}{1 - R(\infty)\zeta}\frac{\zeta d(\zeta)}{h}. \tag{3.12}$$

By (2.4) and (2.9), the operator $(\Delta(\zeta)/h + A)^{-1}A$ is uniformly bounded on V^m for $|\zeta| \le 1$. By (2.9) and $|R(\infty)| < 1$, the expression multiplying $\zeta d(\zeta)/h$ is a uniformly bounded operator from V' to V^m. Thus Parseval's formula applied to (3.12) gives

$$h\sum_{n=0}^{N}\sum_{i=1}^{m}\|E_{ni}\|^2 \le C\left(h\sum_{n=0}^{N}\sum_{i=1}^{m}\|D_{ni}\|^2 + h\sum_{n=0}^{N-1}\|d_{n+1}/h\|_*^2\right), \tag{3.13}$$

and by (1.6) this is bounded by the right-hand side of (3.8). Parseval's formula used once more in (3.11b) then yields the desired bound for $h\sum_0^N \|e_{n+1}\|^2$.

(b) The pointwise estimate in the H-norm follows as in the proof of Lemma 2.2 with u_ν, $U_{\nu i}$ replaced by e_ν, $E_{\nu i}$, using the second formula of (3.9) and the estimate (3.13) of the internal stages.

(c) For the time-dependent case, we use the ideas of Lemma 3.1. With fixed $A = A(\bar{t})$ for a suitable $\bar{t}$, we write the error recursion as

$$\begin{aligned}(I + h\mathscr{C} \otimes A)E_n &= \mathbf{1}e_n - D_n + h\mathscr{C}F_n,\\ e_{n+1} &= e_n - hb^T \otimes AE_n - d_{n+1} + hb^T F_n\end{aligned} \tag{3.14}$$

with $F_{ni} = (A - A(t_n + c_i h))E_{ni}$. This gives for the generating functions a combination of (2.8) and (3.11):

$$E(\zeta) = -\left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} \frac{\Delta(\zeta)}{h}\left(D(\zeta) + \frac{\zeta}{1-\zeta}\mathbf{1}d(\zeta)\right) + \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} F(\zeta). \tag{3.15}$$

Hence, Parseval's formula can again be applied, yielding

$$\begin{aligned}h\sum_{\nu=0}^{n}\sum_{i=1}^{m}\|E_{\nu i}\|^2 \le C\Bigg(&h\sum_{\nu=0}^{n}\sum_{i=1}^{m}\|D_{\nu i}\|^2 + h\sum_{\nu=0}^{n-1}\|d_{\nu+1}/h\|_*^2\\ &+h\sum_{\nu=0}^{n}\sum_{i=1}^{m}\|A - A(t_\nu + c_i h)\|_{V'\leftarrow V}^2\|E_{\nu i}\|^2\Bigg).\end{aligned}$$

Partial summation of the last term (with $A = A(t_{n+1})$) and the application of a discrete Gronwall lemma yield an l^2 bound of the form (3.13) for the internal stages. This leads to the desired bound for $h\sum_0^N \|e_{n+1}\|^2$ as in part (a). The pointwise bound in the H-norm is proved as in part (b), using the second formula of (3.14). □

Assuming more *spatial* regularity, we obtain an improved *temporal* order of approximation.

Theorem 3.3 (Refined error estimate). *In addition to the conditions of Theorem* 3.2 *let* $p \geq q+2$. *We further suppose that the regularity assumptions* $u^{(q+2)} \in L^2(0, T; V)$ *and* $u^{(q+3)} \in L^2(0, T; V')$ *hold. If* $\beta \in [0, 1]$ *is such that* $D(A(t)^{1/2+\beta})$ *is independent of* t *(with uniformly equivalent norms) and* $A^\beta u^{(q+1)} \in L^2(0, T; V)$, *then the error is bounded for* $Nh \leq T$ *by*

$$
\begin{aligned}
&h \sum_{n=0}^{N} \|u_n - u(t_n)\|^2 + \max_{0 \leq n \leq N} |u_n - u(t_n)|^2 \\
&\qquad \leq C \cdot (h^{q+1+\beta})^2 \cdot \int_0^T \|A^\beta u^{(q+1)}(t)\|^2 \, dt \\
&\qquad\quad + C \cdot (h^{q+2})^2 \cdot \left(\int_0^T \|u^{(q+2)}(t)\|^2 \, dt + \int_0^T \|u^{(q+3)}(t)\|_*^2 \, dt \right).
\end{aligned} \tag{3.16}
$$

Again, the constant C *depends only on the Runge-Kutta method, on the constants in* (3.2)–(3.4), *and on* T.

Remark. The restriction to $\beta \leq 1$ is not essential. If $A(t)$ depends smoothly on t as an operator from V to V' and if higher temporal derivatives of u satisfy some regularity assumptions as in Theorem 3.3, then the convergence order is $\min(p, q+1+\beta)$. We do not prove this extension of Theorem 3.3. The proof uses similar ideas but becomes very technical.

Example. We consider again a second-order strongly elliptic differential operator with time- (and space-) dependent smooth coefficients on a smooth bounded domain Ω, equipped with appropriate boundary conditions. We take it as an unbounded operator on $H = L^2(\Omega)$. The attainable value of β in Theorem 3.3 relies on the characterization of the domains of fractional powers of elliptic operators given by [9] and [10]:

(i) *Homogeneous Dirichlet boundary conditions.* For $\alpha < 5/4$ (and $\alpha \geq 1/2$) we have $D(A(t)^\alpha) = H^{2\alpha}(\Omega) \cap H_0^1(\Omega)$ with uniformly equivalent norms. However, for $\alpha > 5/4$ an element $v \in D(A(t)^\alpha)$ has to be such that $A(t)v$ vanishes on the boundary, and hence $D(A(t)^\alpha)$ depends in general on t for $\alpha > 5/4$. We next consider the condition $A^\beta u^{(q+1)}(t) \in V$ or equivalently $u^{(q+1)}(t) \in D(A(t)^{1/2+\beta})$. A smooth function over Ω that vanishes on the boundary is in $D(A(t)^{5/4-\varepsilon}) = H^{5/2-2\varepsilon}(\Omega) \cap H_0^1(\Omega)$ for arbitrary $\varepsilon > 0$. This is sharp unless further (unnatural) boundary conditions are satisfied. In the case of a temporally and spatially smooth solution, Theorem 3.3 is thus applicable with $1/2 + \beta = 5/4 - \varepsilon$, i.e., $\beta = 3/4 - \varepsilon$.

(ii) *Homogeneous Neumann boundary conditions* $\partial u / \partial n_{A(t)} \equiv \sum_{i,j} n_i a_{ij}(x, t) \partial_j u = 0$. In 2 and more space dimensions, $D(A(t)^\alpha) = H^{2\alpha}(\Omega)$ for $\alpha < 3/4$, but for larger values of α the domain depends on t through the boundary conditions. These do not depend on t in dimension 1, and then $D(A(t)^\alpha)$ is independent of t for $\alpha < 7/4$. On the other hand, a smooth function satisfying the boundary conditions is in $D(A(t)^{7/4-\varepsilon})$ for arbitrary $\varepsilon > 0$. The time derivatives of a smooth solution satisfy the boundary conditions only if they do not depend on t, hence in dimension 1. Otherwise, the solution derivatives are in $D(A(t)^{3/4-\varepsilon})$. We can thus use Theorem 3.3 (or the extension mentioned in the above remark) with $\beta = 5/4 - \varepsilon$ in 1 space dimension or when the coefficients a_{ij} do not depend on t on the boundary, and with

$\beta = 1/4 - \varepsilon$ otherwise in higher dimension. The latter value of β is also obtained with nonhomogeneous Neumann boundary conditions.

(iii) *Periodic boundary conditions.* Here we have for all α that $D(A(t)^\alpha)$ is independent of t, and a smooth solution is in $D(A(t)^\alpha)$ together with its time derivatives. Hence, in this case the above remark gives us the full convergence order p.

Remark. While the above values of β are sharp for the energy norm, they are not necessarily optimal for error estimates in the $L^2(\Omega)$-norm. In fact, we can also consider $A(t)$ as an unbounded operator on the space $H = H^{-1}(\Omega)$ with $D(A(t)) = H_0^1(\Omega)$, in the case of Dirichlet boundary conditions. We then have $V = L^2(\Omega)$ and $D(A(t)^{1/2+\beta}) = D(A_0(t)^\beta)$, where $A_0(t)$ is the same differential operator viewed as an unbounded operator on $L^2(\Omega)$. Hence we can choose $\beta = 5/4 - \varepsilon$ (larger by $1/2$) in this case, to obtain an estimate

$$\left(h \sum_{n=0}^{N} \|u_n - u(t_n)\|^2_{L^2(\Omega)} \right)^{1/2} = O(h^{q+1+5/4-\varepsilon})$$

for smooth solutions. In [23] such an estimate was shown pointwise in time for equations with time-independent operator. We do not know if this can be achieved in the time-dependent case.

Proof of Theorem 3.3. We concentrate on those aspects that go beyond the proof of Theorem 3.2.

(a) We first consider the estimates in the V-norm for the time-invariant situation. Since $p \geq q+2$, the defect in (3.9) can now be rewritten as

$$\begin{aligned} D_{ni} &= h^{q+1} \cdot \delta_i \cdot u^{(q+1)}(t_n) + h^{q+1} \int_{t_n}^{t_{n+1}} \tilde{\kappa}_i \left(\frac{t - t_n}{h} \right) u^{(q+2)}(t)\, dt, \\ d_{n+1} &= h^{q+2} \int_{t_n}^{t_{n+1}} \tilde{\kappa} \left(\frac{t - t_n}{h} \right) u^{(q+3)}(t)\, dt \end{aligned} \tag{3.17}$$

with bounded Peano kernels $\tilde{\kappa}$ and $\tilde{\kappa}_i$ and with

$$\delta_i = \frac{1}{(q+1)!} \left((q+1) \sum_{j=1}^{m} a_{ij} c_j^q - c_i^{q+1} \right) \text{ satisfying } \sum_{i=1}^{m} b_i \delta_i = 0. \tag{3.18}$$

We shall show that the estimate (3.16) also holds for $h \sum \|E_{ni} + D_{ni}\|^2$. The desired bound for $\sum \|e_n\|^2$ then follows from (3.11b). We start from identity (3.12). By (3.17), the term coming from $d(\zeta)$ is bounded as needed. It thus remains to consider the contribution of

$$G(\zeta) := \left(\frac{\Delta(\zeta)}{h} + A \right)^{-1} AD(\zeta) = \left(\frac{\Delta(\zeta)}{h} + A \right)^{-1} A^{1-\beta} \cdot A^\beta D(\zeta).$$

Using the estimate (see, e.g., [13, §1.4])

$$\|(\lambda + A)^{-1} A^{1-\beta}\|_{V \leftarrow V} \leq C \cdot |\lambda|^{-\beta}, \quad |\arg \lambda| \leq \pi - \varphi, \tag{3.19}$$

we obtain from the representation as a contour integral (2.10) that $\|G(\zeta)\| \leq C \cdot h^\beta \cdot \|A^\beta D(\zeta)\|$, uniformly for all ζ with $|\zeta| \leq 1$ that are bounded away from 1. For ζ near 1 (cf. proof of Lemma 2.1) one eigenvalue $z(\zeta)$ of $\Delta(\zeta)$

gets close to 0, and such an estimate can no longer be inferred. To overcome this problem, we use equation (2.12) and the identity

$$b^T(I - z\mathscr{Q})^{-1} = b^T + zb^T\mathscr{Q}(I - z\mathscr{Q})^{-1}.$$

The contribution of the eigenvalue $z(\zeta)$ to $G(\zeta)$ is thus (with $z = z(\zeta)$ for short)

$$\left(\frac{z}{h} + A\right)^{-1} A \cdot \frac{(I - z\mathscr{Q})^{-1}\mathbf{1}}{R'(z)} \cdot b^T D(\zeta)$$
$$+ \left(\frac{z}{h} + A\right)^{-1} A^{1-\beta} z^{\beta} \cdot \frac{z^{1-\beta}(I - z\mathscr{Q})^{-1}\mathbf{1}b^T\mathscr{Q}(I - z\mathscr{Q})^{-1}}{R'(z)} \cdot A^{\beta} D(\zeta).$$

For the first term, we use that $(z(\zeta)/h + A)^{-1}A$ is bounded as an operator on V^m for ζ near 1 with $|\zeta| \le 1$, whereas the second term can be bounded via (3.19) as before. This yields

$$\|G(\zeta)\| \le C \cdot h^{\beta} \cdot \|A^{\beta} D(\zeta)\| + C \cdot \|b^T D(\zeta)\|, \tag{3.20}$$

now valid uniformly in $|\zeta| \le 1$. We finally apply Parseval's formula and use (3.18) for $b^T D_n$.

(b) We next verify the pointwise estimates in the H-norm. As in the proof of Theorem 3.2, we start from

$$|e_{n+1}|^2 \le 2 \sum_{\nu=0}^{n} |\langle -h(b^T \otimes A)(I + h\mathscr{Q} \otimes A)^{-1}\mathbf{1}e_{\nu}, e_{\nu+1}\rangle$$
$$+ \langle h(b^T \otimes A)(I + h\mathscr{Q} \otimes A)^{-1} D_{\nu}, e_{\nu+1}\rangle - \langle d_{\nu+1}, e_{\nu+1}\rangle|$$

and use the Cauchy-Schwarz inequality. This yields immediately the bounds for the first and the third term. The second term is split as

$$(b^T \otimes A)(I + h\mathscr{Q} \otimes A)^{-1} D_{\nu}$$
$$= \sum_{i=1}^{m} b_i A D_{\nu i} - h^{\beta}(b^T \otimes A)(I + h\mathscr{Q} \otimes A)^{-1}\mathscr{Q} \otimes (hA)^{1-\beta} \cdot A^{\beta} D_{\nu},$$

and since (3.19) shows that $(I + h\mathscr{Q} \otimes A)^{-1}\mathscr{Q} \otimes (hA)^{1-\beta}$ is bounded as an operator on V^m, we obtain the desired bounds as in part (a).

(c) In the time-dependent case, we start from (3.12) and (3.15), which shows

$$E(\zeta)+D(\zeta) = -\left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} \frac{\mathscr{Q}^{-1}\mathbf{1}}{1 - R(\infty)\zeta} \frac{\zeta d(\zeta)}{h} + \left(\frac{\Delta(\zeta)}{h} + A\right)^{-1} (F(\zeta)+K(\zeta)),$$

with

$$F_{ni} = (A - A(t_n + c_i h))(E_{ni} + D_{ni})$$

and

$$K_{ni} = A(t_n)D_{ni} + (A(t_n + c_i h) - A(t_n))D_{ni}.$$

The term coming from $K(\zeta)$ can be bounded as in part (a), using the Lipschitz boundedness (3.4). The rest of the proof is then identical to part (c) of the

preceding proof. For the pointwise estimate in the H-norm, we use (2.14) with u_n replaced by e_n, further

$$
\begin{aligned}
e_{n+1} &= e_n - h\sum_{i=1}^{m} b_i A(t_n + c_i h)(E_{ni} + D_{ni}) \\
&\quad + h\sum_{i=1}^{m} b_i (A(t_n + c_i h) - A(t_n))D_{ni} + A(t_n)b^T D_n - d_{n+1}
\end{aligned} \tag{3.21}
$$

and the above estimate for $\sum \|E_{ni} + D_{ni}\|^2$. □

Remark. Concerning spatial discretization, the remark (c) after Theorem 1.1 applies *verbatim* to the present situation, with $\widehat{B}$ now given by the right-hand side of (3.8) or (3.16) for $\hat{u}$ instead of u. This follows by linearity and using Lemma 3.1 to treat the perturbations.

4. Quasi-linear equations

We now consider equations with solution-dependent operator,

$$
u' + A(u)u = f(t), \quad u(0) = u_0 \quad (0 < t \le T). \tag{4.1}
$$

We use the framework of §2 and consider again a Hilbert space triple $V \subset H = H' \subset V'$. For $v \in V$, let $A(v) : D(A(v)) \subset H \to H$ be a densely defined closed linear operator. Conditions (2.2) and (2.3) are assumed to hold *uniformly for v varying in bounded subsets of V*, viz.,

$$
|(\lambda + A(v))^{-1}|_{H \leftarrow H} \le \frac{M}{1 + |\lambda|} \quad \text{for } |\arg\lambda| \le \pi - \varphi \ \left(\varphi < \frac{\pi}{2}\right), \tag{4.2}
$$

$$
V = D(A(v)^{1/2}) = D(A(v)^{*1/2}) \quad \text{with equivalent norms.} \tag{4.3}
$$

Then $A(v)$ also satisfies the bounds of (2.4) and (2.6) uniformly for v in bounded subsets of V. We further assume that the following local Lipschitz condition is satisfied: For all $\delta > 0$ and all $r < \infty$, there exists $L = L(\delta, r)$ such that

$$
\|A(v) - A(w)\|_{V' \leftarrow V} \le \delta \cdot \|v - w\| + L \cdot |v - w| \quad \text{for } \|v\| \le r, \ \|w\| \le r. \tag{4.4}
$$

Example. Consider a second-order strongly elliptic differential operator $A(v)$ with smooth coefficients $a_{ij}(v(x))$ etc. over a smooth bounded domain $\Omega \subset \mathbf{R}^d$, equipped with Neumann boundary conditions. In 1 space dimension, taking $H = L^2(\Omega)$ and $V = H^1(\Omega) \subset C(\overline{\Omega})$ gives well-defined operators $A(v)$, $v \in V$, that satisfy the above conditions. In particular, condition (4.4) is obtained from the estimate

$$
\|A(v) - A(w)\|_{V' \leftarrow V} \le \sup_{x \in \Omega} |a(v(x)) - a(w(x))|,
$$

using the local Lipschitz boundedness of a and the imbedding $H^s(\Omega) \subset C(\overline{\Omega})$ for $s > \frac{1}{2}$ together with the bound [19, Prop. IV.4.1]

$$
\|v\|_{H^s} \le \delta \cdot \|v\|_{H^1} + C_s(\delta) \cdot |v|_{L^2} \quad \text{for } v \in H^1(\Omega) \text{ and } s < 1.
$$

This choice of H and V is no longer possible in 2 dimensions. Here, it can be shown that our conditions are met when the differential operator is

considered as an unbounded operator on the Sobolev space $H = H^s(\Omega)$ with $0 < s < 1/2$, with $V = H^{s+1}(\Omega)$. The lower bound on s originates from Sobolev's inequality in the requirement $V \subset C(\overline{\Omega})$. The upper bound comes from condition (4.3), because for values of $s \geq 1/2$ the Neumann boundary conditions (which depend on the coefficients $a_{ij}(v(x))$) enter into $D(A(v)^{1/2})$. Note that for $s < 3/2$ one has $H^s(\Omega) = D(A_0(v)^{s/2})$, where $A_0(v)$ is the same differential operator taken as an unbounded operator on $L^2(\Omega)$. In 3 space dimensions, there is a conflict between the Sobolev inequality for V and condition (4.3), so that the 3-dimensional quasi-linear Neumann problem falls outside our framework.

The situation is more favorable for the Dirichlet problem, for which condition (4.3) is less stringent. Here one can take $H = H_0^s(\Omega)$ and $V = H^{s+1}(\Omega) \cap H_0^1(\Omega)$ (and hence $V' = H^{s-1}(\Omega)$) with $0 < s \leq 1$ in 2 space dimensions, and still with $s = 1$ in 3 dimensions. □

We have the following convergence result for the case that the solution of (4.1) is sufficiently smooth in time.

Theorem 4.1 (Convergence of Runge-Kutta methods for quasi-linear parabolic equations). *For an initial value problem* (4.1)–(4.4) *consider a Runge-Kutta method of stage order* q *and order* $p \geq q+1$ *that is strongly* $A(\theta)$*-stable with* $\theta > \varphi$. *If* $u^{(q+1)} \in L^2(0, T; V)$ *and* $u^{(q+2)} \in L^2(0, T; V')$, *then for sufficiently small stepsizes* h *there exists a unique numerical solution* u_n $(0 \leq nh \leq T)$ *whose error is bounded by*

$$
\begin{aligned}
&h \sum_{n=0}^{N} \|u_n - u(t_n)\|^2 + \max_{0 \leq n \leq N} |u_n - u(t_n)|^2 \\
&\qquad \leq C \cdot (h^{q+1})^2 \cdot \left(\int_0^T \|u^{(q+1)}(t)\|^2 \, dt + \int_0^T \|u^{(q+2)}(t)\|_*^2 \, dt \right).
\end{aligned} \tag{4.5}
$$

The constant C *depends on the Runge-Kutta method, on the constants in* (4.2)–(4.4), *on* $\sup_{0 \leq t \leq T} \|u(t)\|$, *and on* T.

Proof. Let us assume for a moment that the numerical solution u_n and the internal stages U_{ni} exist for $0 \leq nh \leq T$ and that

$$
\|U_{ni}\| \leq r \tag{4.6}
$$

with $r = 2 \sup_{0 \leq t \leq T} \|u(t)\|$. For sufficiently small stepsizes h, this will be verified at the end of this proof.

(a) For a concise notation, we abbreviate the exact solution values by $\tilde{U}_{ni} = u(t_n + c_i h)$, $\tilde{u}_n = u(t_n)$ and we set

$$
\mathscr{A}_n = \operatorname{diag}(A(\tilde{U}_{n1}), \ldots, (\tilde{U}_{nm})). \tag{4.7}
$$

In this notation, the errors $E_{ni} = U_{ni} - \tilde{U}_{ni}$ and $e_n = u_n - \tilde{u}_n$ of the Runge-Kutta method applied to (4.1) are related by

$$
\begin{aligned}
(I + h\mathscr{Q}\mathscr{A}_n)E_n &= \mathbf{1}e_n + h\mathscr{Q}F_n - D_n, \\
e_{n+1} &= e_n - hb^T\mathscr{A}_n E_n + hb^T F_n - d_{n+1}
\end{aligned} \tag{4.8}
$$

with $F_{ni} = (A(\tilde{U}_{ni}) - A(U_{ni})) \cdot U_{ni}$ and with the defects D_{ni} and d_{n+1}. Recall that these defects satisfy the bound (1.6). Lemma 3.1 and Theorem 3.2 now give the following l^2-bound on the error (cf. (2.7) and (3.13)):

$$h\sum_{n=0}^{N}\|e_{n+1}\|^2 + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|E_{ni}\|^2 \le C\cdot\left(B_N + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|F_{ni}\|_*^2\right) \tag{4.9}$$

with

$$B_N = h\sum_{n=0}^{N}\sum_{i=1}^{m}\|D_{ni}\|^2 + h\sum_{n=0}^{N-1}\|d_{n+1}/h\|_*^2 + h\sum_{n=0}^{N}\|d_{n+1}\|^2 . \tag{4.10}$$

The rest of the proof is to show that the left-hand side of (4.9) can actually be bounded already by B_N.

(b) We thus have to estimate F_{ni}. This is the moment where condition (4.4) comes into play. Since the internal stages are bounded by assumption (4.6), we conclude that

$$\|F_{ni}\|_*^2 \le (\delta\|E_{ni}\| + L|E_{ni}|)^2 \cdot r^2$$

and further

$$\|F_{ni}\|_*^2 \le 2r^2\delta^2\|E_{ni}\|^2 + 4r^2L^2(|E_{ni} + D_{ni}|^2 + |D_{ni}|^2) . \tag{4.11}$$

(c) We next establish a bound for $|E_{ni} + D_{ni}|$. For technical reasons, we regroup (4.8) as

$$\begin{aligned}(I + h\mathscr{C}\otimes A(\tilde{u}_n))(E_n + D_n) &= \mathbf{1}e_n + h\mathscr{C}F_n + h\mathscr{C}\otimes A(\tilde{u}_n)D_n \\ &\quad + h(\mathscr{C}\otimes A(\tilde{u}_n) - \mathscr{C}\mathscr{A}_n)E_n .\end{aligned} \tag{4.12}$$

Since the operator $(I + h\mathscr{C}\otimes A(\tilde{u}_n))^{-1}$ is uniformly bounded on H^m and bounded by $O(h^{-1/2})$ from $(V')^m$ to H^m (see Proof of Lemma 2.2), we obtain

$$\begin{aligned}|E_{ni} + D_{ni}|^2 \le C\cdot\Bigg(&|e_n|^2 + h\sum_{j=1}^{m}\|F_{nj} + A(\tilde{u}_n)D_{nj}\|_*^2 \\ &+ h\|\mathscr{C}\otimes A(\tilde{u}_n) - \mathscr{C}\mathscr{A}_n\|_{V'^m\leftarrow V^m}^2\sum_{j=1}^{m}\|E_{nj}\|^2\Bigg) ,\end{aligned}$$

and by applying (4.4) to $\|A(\tilde{u}_n) - A(\tilde{U}_{ni})\|_{V'\leftarrow V}$ also

$$|E_{ni} + D_{ni}|^2 \le C\cdot\left(|e_n|^2 + h\sum_{j=1}^{m}\|F_{nj}\|_*^2 + h\sum_{j=1}^{m}\|D_{nj}\|^2 + h^3\sum_{j=1}^{m}\|E_{nj}\|^2\right) . \tag{4.13}$$

(d) Next we consider the second equation of (4.8). Recalling the techniques of the proof of Lemma 2.2 part (a) and substituting each occurrence of $\sum\sum\|E_{\nu i}\|^2$ or $\sum\|e_{\nu+1}\|^2$ by (4.9), one finds

$$|e_{n+1}|^2 \le C\cdot\left(B_n + h\sum_{\nu=0}^{n}\sum_{i=1}^{m}\|F_{\nu i}\|_*^2\right) , \tag{4.14}$$

which is the same bound as for the l^2-norm. We then insert (4.14) into (4.13) and use the estimate (4.11). This results in

$$|E_{ni}+D_{ni}|^2 \le C\cdot\left(B_n + h\sum_{\nu=0}^{n}\sum_{j=1}^{m}|E_{\nu j}+D_{\nu j}|^2 + h(h^2+\delta^2)\sum_{\nu=0}^{n}\sum_{j=1}^{m}\|E_{\nu j}\|^2\right)$$

for $i=1,\dots,m$. The application of a discrete Gronwall inequality now gives the bound

$$|E_{ni}+D_{ni}|^2 \le C\cdot\left(B_n + h(h^2+\delta^2)\sum_{\nu=0}^{n}\sum_{j=1}^{m}\|E_{\nu j}\|^2\right). \tag{4.15}$$

(e) We finally insert (4.15) into (4.11), and (4.11) into (4.9). Since δ can be chosen arbitrarily small, we obtain

$$h\sum_{n=0}^{N}\|e_{n+1}\|^2 + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|E_{ni}\|^2 \le C\cdot B_N,$$

and by (4.15), (4.14) also

$$\max_{0\le n\le N}|e_{n+1}|^2 + \max_{0\le n\le N}\max_{i=1,\dots,m}|E_{ni}|^2 \le C\cdot B_N.$$

These are the desired bounds that lead to the estimate (4.5) in the same way as in Theorem 3.2.

(f) It remains to prove the (local) uniqueness and existence of the numerical solution as well as the bound (4.6) for the internal stages. We use a fixed point argument.

First we will show that the iteration

$$U_{ni}^{(k+1)} = u_n - h\sum_{j=1}^{m}a_{ij}\left(A(U_{nj}^{(k)})U_{nj}^{(k+1)} - f(t_n+c_jh)\right), \qquad i=1,\dots,m, \tag{4.16}$$

is a contraction with respect to the weighted norm

$$|||U_n||| = \max_{i=1,\dots,m}\max(\|U_{ni}\|, h^{-1/2}|U_{ni}|)$$

in a ball (with respect to the $|||\cdot|||$ norm) around $\mathbf{1}\cdot u_n$ of a fixed, sufficiently large radius. The proof that the iteration maps the ball into itself for small h uses arguments similar to the proof of contractivity, and is therefore omitted. To show contractivity, we consider the difference of two sequences $U_n^{(k+1)}$ and $V_n^{(k+1)}$. We denote this difference by $E_n^{(k+1)} = U_n^{(k+1)} - V_n^{(k+1)}$, and like in (4.7), we write for short $\mathscr{A}(U_n) = \operatorname{diag}(A(U_{n1}),\dots,A(U_{nm}))$. We then have

$$\begin{aligned}(I+h\mathscr{C}\otimes A(u_n))E_n^{(k+1)} &= h\mathscr{C}(\mathscr{A}(V_n^{(k)}) - \mathscr{A}(U_n^{(k)}))V_n^{(k+1)} \\ &\quad + h(\mathscr{C}\otimes A(u_n) - \mathscr{C}\mathscr{A}(U_n^{(k)}))E_n^{(k+1)}.\end{aligned}$$

Recalling the uniform bounds on $(I+h\mathscr{C}\otimes A(u_n))^{-1}$ and the Lipschitz condition (4.4), and using that $V_n^{(k+1)}$ is again in the ball, we get

$$\begin{aligned}\|E_{ni}^{(k+1)}\| &\le C\cdot\left(\delta\cdot\max_{j=1,\dots,m}\|E_{nj}^{(k)}\| + Lh^{1/2}\cdot h^{-1/2}\max_{j=1,\dots,m}|E_{nj}^{(k)}|\right) \\ &\quad + C\cdot(\delta+Lh^{1/2})\cdot|||U_n^{(k)} - \mathbf{1}u_n|||\cdot\max_{j=1,\dots,m}\|E_{nj}^{(k+1)}\|\end{aligned}$$

and the same bound for $h^{-1/2} \cdot |E_{ni}^{(k+1)}|$. Thus, the recursion satisfies

$$|||E_n^{(k+1)}||| \le C \cdot (\delta + Lh^{1/2}) \cdot |||E_n^{(k)}|||,$$

and hence is a contraction for δ and h small enough. This proves the uniqueness of the numerical solution in the ball.

It remains to show the existence of a fixed point. For this we take $U_{ni}^{(0)} = u(t_n + c_i h)$ as starting value for the iteration and use $|||u_n - u(t_n)||| \le C \cdot h^{1/2}$, which follows from parts (a)–(e) of the proof (with $q = 0$). From the identity

$$\begin{aligned}(I + h\mathscr{C} \otimes A(u(t_n)))(U_n^{(1)} - U_n^{(0)}) \\ = \mathbf{1} \cdot (u_n - u(t_n)) - h(\mathscr{C}\mathscr{A}(U_n^{(0)}) - \mathscr{C} \otimes A(u(t_n)))(U_n^{(1)} - U_n^{(0)}) - D_n\end{aligned}$$

one deduces as above that

$$|||U_n^{(1)} - U_n^{(0)}||| \le C \cdot h^{1/2}.$$

Thus, if h is small enough, the iterates remain bounded by $2 \sup_{0 \le t \le T} \|u(t)\|$ in V-norm and converge to a fixed point U_n satisfying (4.6). □

Under slightly stronger assumptions, we obtain—as in the linear case—a higher temporal order of convergence. In addition to (4.2)–(4.4) we now assume that for all $r > 0$ there exists $\tilde{L} = \tilde{L}(r)$ such that

$$\left\| A(v) - A(w) - \frac{\partial A}{\partial u}(u)[v - w] \right\|_{V' \leftarrow V} \le \tilde{L} \cdot (\|u - v\| + \|u - w\|) \cdot \|v - w\| \tag{4.17}$$

for $\max(\|u\|, \|v\|, \|w\|) \le r$. This, together with (4.4), implies the estimate

$$\left\| \frac{\partial A}{\partial u}(v)[w] \right\|_{V' \leftarrow V} \le C \cdot \|w\| \tag{4.18}$$

uniformly for v and w in bounded subsets of V. Let further $\beta \in [0, 1]$ be such that $D(A^{1/2+\beta}(v))$ is independent of v (with uniformly equivalent norms) and

$$\left\| A^{\beta} \cdot \frac{\partial A}{\partial u}(u(t))[w]u(t) \right\|_{*} \le C \cdot \|A^{\beta} w\|, \qquad 0 \le t \le T, \tag{4.19}$$

uniformly for $w \in D(A^{1/2+\beta})$.

Theorem 4.2 (Refined error estimate for quasi-linear equations). *In addition to the conditions of Theorem* 4.1 *we assume* $p \ge q + 2$, (4.17) *and* (4.19). *We further suppose that the regularity assumptions* $u^{(q+2)} \in L^2(0, T; V)$ *and* $u^{(q+3)} \in L^2(0, T; V')$ *hold. If* $A^{\beta} u^{(q+1)} \in L^2(0, T; V)$, *then the error is bounded for* $Nh \le T$ *by*

$$\begin{aligned} h \sum_{n=0}^{N} \|u_n - u(t_n)\|^2 + \max_{0 \le n \le N} |u_n - u(t_n)|^2 \\ \le C \cdot (h^{q+1+\beta})^2 \cdot \int_0^T \|A^{\beta} u^{(q+1)}(t)\|^2 \, dt \\ + C \cdot (h^{q+2}) \cdot \left(\int_0^T \|u^{(q+2)}(t)\|^2 \, dt + \int_0^T \|u^{(q+3)}(t)\|_{*}^2 \, dt \right). \end{aligned} \tag{4.20}$$

Again, the constant C depends only on the Runge-Kutta method, on the constants in (4.2)–(4.4), (4.17), (4.19), on $\sup_{0\le t\le T}\|u(t)\|$, and on T.

Example. Consider again the example of solution-dependent strongly elliptic second-order operators from the beginning of this section. We compare the situation to that of the example after Theorem 3.3. In 1 space dimension, Theorem 4.2 gives us the same noninteger convergence order as Theorem 3.3. Condition (4.19) is not restrictive for spatially smooth solutions. In 2 space dimensions, we still get the same values of β by choosing $H = H^s(\Omega)$ with a small $s > 0$. For the Dirichlet problem in 3 dimensions, taking $H = H_0^1(\Omega)$ allows us to choose $\beta = 1/4 - \varepsilon$. For periodic boundary conditions, Theorem 4.2 gives us $\beta = 1$, but we expect that here again the full order p can be obtained under reasonable assumptions on the derivatives of A.

Proof of Theorem 4.2. We use the same notation as in the proof of Theorem 4.1. The main idea now is to show that under the present assumptions, the bounds (4.9), (4.14), and (4.15) also hold if $\|E_{ni}\|$ is replaced by $\|E_{ni} + D_{ni}\|$ and if $\sum_i \|D_{ni}\|^2$ is replaced by $\sum_i \|h^\beta A^\beta D_{ni}\|^2 + \|\sum_i b_i D_{ni}\|^2$. The rest is then identical to the preceding proof.

(a) We start again with (4.8), written as

$$\begin{aligned}(I + h\mathscr{C}\mathscr{A}_n)(E_n + D_n) &= \mathbf{1}e_n + h\mathscr{C}\tilde{F}_n + h\mathscr{C}K_n + h\mathscr{C}S_n\,,\\ e_{n+1} &= e_n - hb^T\mathscr{A}_n(E_n + D_n) + hb^T\tilde{F}_n + hb^TK_n + hb^TS_n - d_{n+1}\end{aligned} \tag{4.21}$$

with $\tilde{F}_{ni} = (A(\tilde{U}_{ni} - D_{ni}) - A(U_{ni}))\cdot U_{ni}$ and with $S_{ni} = A(\tilde{u}_n)D_{ni} + \frac{\partial A}{\partial u}(\tilde{u}_n)[D_{ni}]\tilde{u}_n$. The remaining term, namely

$$\begin{aligned}K_{ni} = &\left(A(\tilde{U}_{ni}) - A(\tilde{U}_{ni} - D_{ni}) - \frac{\partial A}{\partial u}(\tilde{u}_n)[D_{ni}]\right)\cdot U_{ni}\\ &+ \frac{\partial A}{\partial u}(\tilde{u}_n)[D_{ni}]\cdot(E_{ni} + D_{ni} + \tilde{U}_{ni} - \tilde{u}_n - D_{ni})\\ &+ (A(\tilde{U}_{ni}) - A(\tilde{u}_n))\cdot D_{ni}\end{aligned}$$

satisfies by (4.17), (4.18), and (4.4) the bound

$$\|K_{ni}\|_*^2 \le C\cdot h^2\cdot(\|D_{ni}\|^2 + \|E_{ni} + D_{ni}\|^2)\,. \tag{4.22}$$

The recursion (4.21) admits by Lemma 3.1 and Theorem 3.3 instead of (4.9)–(4.10) the (sharper) bound

$$h\sum_{n=0}^{N}\|e_{n+1}\|^2 + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|E_{ni} + D_{ni}\|^2 \le C\cdot\left(\tilde{B}_N + h\sum_{n=0}^{N}\sum_{i=1}^{m}\|\tilde{F}_{ni}\|_*^2\right), \tag{4.23}$$

where, owing to (4.22),

$$\begin{aligned}\tilde{B}_N = &\,h\sum_{n=0}^{N}\sum_{i=1}^{m}\|h^\beta A^\beta D_{ni}\|^2 + h\sum_{n=0}^{N}\left\|\sum_{i=1}^{m}b_iD_{ni}\right\|^2\\ &+ h^3\sum_{n=0}^{N}\sum_{i=1}^{m}\|E_{ni} + D_{ni}\|^2 + h\sum_{n=0}^{N-1}\|d_{n+1}/h\|_*^2 + h\sum_{n=0}^{N}\|d_{n+1}\|^2\,.\end{aligned} \tag{4.24}$$

Note that $\tilde{F}_{ni}$ is bounded by

$$\|\tilde{F}_{ni}\|_*^2 \le 2r^2\delta^2\|E_{ni} + D_{ni}\|^2 + 4r^2L^2|E_{ni} + D_{ni}|^2\,. \tag{4.25}$$

(b) To establish the bound for $|E_{ni} + D_{ni}|$, we use the identity

$$(I + h\mathscr{C} \otimes A(\tilde{u}_n))(E_n + D_n) = \mathbf{1}e_n + h\mathscr{C}\tilde{F}_n + h\mathscr{C}K_n + h\mathscr{C}S_n + h(\mathscr{C} \otimes A(\tilde{u}_n) - \mathscr{C}\mathscr{A}_n)(E_n + D_n),$$

further (4.23), the uniform boundedness of $(I + h\mathscr{C} \otimes A(\tilde{u}_n))^{-1}\mathscr{C} \otimes (hA(\tilde{u}_n))^{1-\beta}$ on V^m and (4.19). This gives

(4.26)

$$|E_{ni} + D_{ni}|^2 \leq C \cdot \left(|e_n|^2 + h\sum_{j=1}^{m} \|\tilde{F}_{nj}\|_*^2 + h\sum_{j=1}^{m} \|h^\beta A^\beta D_{nj}\|^2 + h^3 \sum_{j=1}^{m} \|E_{nj} + D_{nj}\|^2 \right).$$

(c) From the second line of (4.21) one obtains with the same techniques as in the preceding proof

$$|e_{n+1}|^2 \leq C \cdot \left(\tilde{B}_n + h\sum_{\nu=0}^{n}\sum_{i=1}^{m} \|\tilde{F}_{\nu i}\|_*^2 \right). \tag{4.27}$$

With the bounds (4.23), (4.25)–(4.27) available, one continues as in the proof of Theorem 4.1. □

Remark. The remark (c) after Theorem 1.1 about space discretization applies also to the present situation, with $\widehat{B}$ now given by the right-hand side of (4.5) or (4.20) with $\hat{u}$ instead of u. Also remark (b) after Theorem 1.1 about the generalization from $f(t)$ to $f(t, u)$ applies to Theorem 4.1, and to Theorem 4.2 with additional assumptions on $\partial f/\partial u$.

5. Variable time steps

The proofs of the results of the foregoing sections used in an essential way the assumption of a constant time step h. There is, however the following extension to variable stepsizes.

Theorem 5.1. *The lemmas and theorems of §§2–4 remain valid for Runge-Kutta solutions obtained with stepsize sequences $\{h_n\}$ satisfying*

$$\sum_{n=0}^{N} |h_{n+1}/h_n - 1| \leq C, \tag{5.1}$$

$$ch \leq h_n \leq h, \qquad 0 \leq n \leq N, \tag{5.2}$$

with a positive constant c.

Remark. Condition (5.1) is familiar from the convergence analysis of linear multistep methods for ODEs, see [11, Thm. III.5.7]. Condition (5.2) may appear rather restrictive. However, if there is a finite subdivision of the integration interval into subintervals on which stepsizes of different scales are used, then one can apply Theorem 5.1 separately on each of the subintervals.

We do not give a proof of Theorem 5.1, but only indicate how the variable stepsize version of Lemma 2.1 comes about. The basic idea is again that of Lemma 3.1. To simplify the presentation further, we consider here only the backward Euler method

$$\frac{u_{n+1} - u_n}{h_{n+1}} + Au_{n+1} = f_{n+1}, \quad u_0 = 0.$$

We rewrite this as

$$\frac{u_{n+1}-u_n}{h} + A_{n+1}u_{n+1} = g_{n+1}$$

with $A_{n+1} = (h_{n+1}/h)\cdot A$, $g_{n+1} = (h_{n+1}/h)\cdot f_{n+1}$, or again as (cf. (3.5))

$$\frac{u_{n+1}-u_n}{h} + A_N u_{n+1} = g_{n+1} + (A_N - A_{n+1})\cdot u_{n+1}.$$

By Lemma 2.1 and condition (5.2), we now have

$$h\sum_{n=1}^{N}\|u_n\|^2 \le Ch\sum_{n=1}^{N}\|f_n\|_*^2 + Ch\sum_{n=1}^{N}\|A_N - A_n\|_{V'\leftarrow V}^2\|u_n\|^2.$$

Using partial summation and the definition of A_n, and noting that

$$\left|\left(\frac{h_N - h_n}{h}\right)^2 - \left(\frac{h_N - h_{n+1}}{h}\right)^2\right| \le 2\frac{|h_{n+1}-h_n|}{h} =: a_n,$$

we get for all N

$$h\sum_{n=1}^{N}\|u_n\|^2 \le Ch\sum_{n=1}^{N}\|f_n\|_*^2 + C\sum_{n=1}^{N-1}a_n\cdot h\sum_{\nu=1}^{n}\|u_\nu\|^2.$$

By (5.1), (5.2), we have $\sum a_n \le C$, and hence a discrete Gronwall-type inequality gives

$$h\sum_{n=1}^{N}\|u_n\|^2 \le Ch\sum_{n=1}^{N}\|f_n\|_*^2,$$

which is the desired result.

Bibliography

1. K. Burrage and J. C. Butcher, *Stability criteria for implicit Runge-Kutta methods*, SIAM J. Numer. Anal. **16** (1979), 46–57.
2. J. C. Butcher, *The numerical analysis of ordinary differential equations*, Wiley, Chichester, 1987.
3. M. Crouzeix, *Sur l'approximation des équations différentielles opérationnelles linéaires par des méthodes de Runge-Kutta*, Thèse d'Etat, Univ. Paris 6, 1975.
4. ———, *Sur la B-stabilité des méthodes de Runge-Kutta*, Numer. Math. **32** (1979), 75–82.
5. M. Crouzeix, S. Larsson, S. Piskarev, and V. Thomée, *The stability of rational approximations of analytic semigroups*, BIT **33** (1993), 74–84.
6. M. Crouzeix and P.-A. Raviart, *Approximation des problèmes d'évolution*, Lecture Notes, Univ. Rennes, 1980.
7. J. Douglas, Jr. and T. Dupont, *Galerkin methods for parabolic equations*, SIAM J. Numer. Anal. **7** (1970), 575–626.
8. R. Frank, J. Schneid, and C. W. Ueberhuber, *Order results for implicit Runge-Kutta methods applied to stiff systems*, SIAM J. Numer. Anal. **22** (1985), 515–534.
9. 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.
10. P. Grisvard, *Caractérisation de quelques espaces d'interpolation*, Arch. Rational Mech. Anal. **25** (1967), 40–63.
11. E. Hairer, S. P. Nørsett, and G. Wanner, *Solving ordinary differential equations* I. *Nonstiff problems*, 2nd ed., Springer-Verlag, Berlin, 1993.

12. E. Hairer and G. Wanner, *Solving ordinary differential equations* II. *Stiff and differential-algebraic problems*, Springer-Verlag, Berlin, 1991.
13. D. Henry, *Geometric theory of semilinear parabolic equations*, Lecture Notes in Math., vol. 840, Springer-Verlag, Berlin, 1981.
14. C. Johnson, *Error estimates and adaptive time-step control for a class of one-step methods for stiff ordinary differential equations*, SIAM J. Numer. Anal. **25** (1988), 908–926.
15. T. Kato, *Perturbation theory for linear operators*, 2nd ed., Springer-Verlag, Berlin, 1976.
16. I. Lasiecka, *Unified theory for abstract parabolic boundary problems—a semigroup approach*, Appl. Math. Optim. **6** (1980), 287–333.
17. M. N. Le Roux, *Semidiscretization in time for parabolic problems*, Math. Comp. **33** (1979), 919–931.
18. ———, *Méthodes multipas pour des équations paraboliques non linéaires*, Numer. Math. **35** (1980), 143–162.
19. J. L. Lions, *Equations différentielles opérationnelles*, Springer-Verlag, Berlin, 1961.
20. ———, *Espaces d'interpolation et domaines de puissances fractionnaires d'opérateurs*, J. Math. Soc. Japan **14** (1962), 233–241.
21. Ch. Lubich, *On the convergence of multistep methods for nonlinear stiff differential equations*, Numer. Math. **58** (1991), 839–853; Erratum **61** (1992), 277–279.
22. Ch. Lubich and O. Nevanlinna, *On resolvent conditions and stability estimates*, BIT **31** (1991), 293–313.
23. Ch. Lubich and A. Ostermann, *Runge-Kutta methods for parabolic equations and convolution quadrature*, Math. Comp. **60** (1993), 105–131.
24. ———, *Linearly implicit time discretization of nonlinear parabolic equations*, submitted to IMA J. Numer. Anal., to appear (1995).
25. A. Ostermann and M. Roche, *Runge-Kutta methods for partial differential equations and fractional order of convergence*, Math. Comp. **59** (1992), 403–420.
26. G. Savaré, *$A(\theta)$-stable approximations of abstract Cauchy problems*, Numer. Math. **65** (1993), 319–336.
27. M. Zlámal, *Finite element methods for nonlinear parabolic equations*, RAIRO Anal. Numér. **11** (1977), 93–107.

INSTITUT FÜR ANGEWANDTE MATHEMATIK UND STATISTIK, UNIVERSITÄT WÜRZBURG, AM HUBLAND, D-97074 WÜRZBURG, GERMANY

Current address: Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10, D-72076 Tübingen, Germany

E-mail address: lubich@na.uni-tuebingen.de

INSTITUT FÜR MATHEMATIK UND GEOMETRIE, UNIVERSITÄT INNSBRUCK, TECHNIKERSTRASSE 13, A-6020 INNSBRUCK, AUSTRIA

E-mail address: alex@mat1.uibk.ac.at