Electronic Transactions on Numerical Analysis.
Volume 27, pp. 78-93, 2007.

ISSN 1068-9613.

ETNA
Kent State University
etna@mcs.kent.edu

ON DIFFERENCE SCHEMES FOR QUASILINEAR EVOLUTION PROBLEMS*

RITA MEYER-SPASCHE†

Abstract. We review several methods leading to variable-coefficient schemes and/or to exact difference schemes for ordinary differential equations (error elimination; functional fitting; Principle of Coherence). Necessary and suffient conditions are given for t-independence of fitted RK coefficients. Conditions for τ-independence are investigated, τ the time-step. The theory is illustrated by examples. In particular, examples are given for non-uniqueness of exact schemes and for efficient difference schemes based on exact schemes and well suited for highly oscillatory ordinary differential systems or for parabolic equations with blow-up solutions.

Key words. difference schemes, time stepping, nonstandard schemes, exact schemes, exponential fitting, functional fitting, Runge-Kutta, collocation methods, review

AMS subject classifications. 65L05, 65M06, 65P99

1. Introduction. Time discretization for the numerical solution of initial value problems means that we approximate a continuous dynamical system by a family of discrete dynamical systems. We introduce the additional parameter $\Delta t =: \tau$ and require $\mathcal{O}(\tau^s)$-convergence for $\tau \to 0$, $s \geq 1$. If the dynamics of the discrete and continuous systems are very different for larger τ, then the step-size τ must be small for satisfactory results. If the dynamics of the systems are very similar, τ may be larger, computations are more efficient. In the ideal case, the step-size of the computations is determined by the solution to be computed: by its structure and by the accuracy required. In many applications, for instance in equilibrium computations and turbulence computations in plasma physics, the bounds for the step-size have to be determined by properties of the numerical method instead.

Let us look at a very simple example:

$$\dot{u} = u^2, \qquad u(0) = u_0 > 0, \tag{1.1}$$

with solution

$$u(t) = \frac{u_0}{1 - tu_0}. \tag{1.2}$$

This solution ceases to exist when the denominator vanishes, i.e. at its blow-up time $T = 1/u_0$. We take $u_0 = U_0$ and compute discrete solutions $\{U_n\}_{n=1}^N$.

If we discretize eq. (1.1) with the explicit forward Euler scheme, we obtain

$$\frac{U_{n+1} - U_n}{\tau} = U_n^2, \quad \text{or} \quad U_{n+1} = U_n + \tau U_n^2,$$

and the iterates exist for all times, independent of the value of U_0. Moreover, the step-size τ must be small enough to prohibit instability of the scheme.

If we discretize eq. (1.1) with the implicit backward Euler scheme, we obtain

$$\frac{U_{n+1} - U_n}{\tau} = U_{n+1}^2, \quad \text{or} \quad U_{n+1}^{\pm} = \frac{1}{2\tau}(1 \pm \sqrt{1 - 4\tau U_n^-}).$$

A choice in favor of the value U_{n+1}^- has to be made in each time step: this ensures that we get convergence to a continuous function in the limit $\tau \to 0$, and it enforces uniqueness of

*Received December 22, 2003. Accepted for publication May 28, 2004. Recommended by A. Ruffing.
†Max-Planck-Institut für Plasmaphysik, EURATOM-Association, D-85748 Garching, Germany (meyer-spasche@ipp-garching.mpg.de).

the solution. It does not prohibit though that the iterates on the two branches meet and turn complex at a time $\tilde{T}(\tau) < T(u_0)$. Real iterates thus cease to exist, but the non-existence happens in a way which is different from blow-up. Moreover, implicit difference schemes have a tendency to turn superstable and thus to produce qualitatively wrong solutions for step-sizes τ which are not small enough.

If we discretize eq. (1.1) with the 'nonstandard scheme'

$$\frac{U_{n+1}-U_n}{\tau} = U_n U_{n+1}, \quad \text{or} \quad U_{n+1} = \frac{U_n}{1-\tau U_n}, \tag{1.3}$$

we find that this scheme is exact, i.e. for any step-size τ it reproduces the solution (1.2) without discretization error, as long as $n \cdot \tau < T(u_0) = 1/u_0$.

Given any individual differential equation, how to find an optimal scheme for it? How should nonlinear terms in differential equations be discretized? Attempts are made for developing a theory of nonstandard schemes 'optimal for individual differential equations' [16]. In the case of eq. (1.1) and $f(u) = u^2$, however, we notice that

$$\frac{U_{n+1}-U_n}{\tau} = U_n U_{n+1} = f(U_n) + f'(U_n)\frac{U_{n+1}-U_n}{2},$$

and this is a linearly implicit *standard* scheme, a so-called Rosenbrock-Wanner scheme. Linearly implicit schemes were introduced by Rosenbrock in 1963. Today they are standard in the numerical treatment of stiff differential equations and of differential-algebraic equations [7]. Also other 'nonstandard' schemes found in the literature turned out to be standard [15].

In the following we shall have a 'nonstandard' view on standard schemes. We shall discuss on which functions given, well-known schemes are exact (on which $f(u)$, on which $u(t)$?) and we shall discuss several methods for finding schemes which are exact on given s-dimensional function spaces. Depending on the chosen function space, the schemes have constant coefficients, or coefficients depending on time t and/or time-step τ. Necessary and sufficient conditions for t-independence are given, τ-independence is discussed.

These investigations mostly lead to results for simple quasilinear equations. Exact schemes for simple equations have been used successfully for designing efficient schemes relevant to applications. As examples, Denk-Bulirsch schemes and LeRoux schemes are discussed in this text. Kojouharov-Chen schemes for advection-diffusion-convection equations [8] and structure-preserving schemes for canonical and non-canonical Hamiltonian systems [4] must at least be mentioned. This is an updated, enlarged version of [13].

2. Exact schemes.

We start by considering non-autonomous systems

$$\dot{u} = f(t,u), \qquad u(0) = u_0, \tag{2.1}$$

with smooth functions $f : (t_1,t_2) \times I\!R^q \to I\!R^q$, $q \geq 1$. We assume $[0,T] \subset (t_1,t_2) \subset I\!R$ and consider one-step schemes

$$U_{n+1} = \mathcal{A}(f)(U_n,\tau), \qquad U_0 = u_0, \tag{2.2}$$

for the numerical solution of system (2.1) on the interval $[0,T]$. Here $\tau = \Delta t$ is the time step, U_n is an approximation to the exact solution $u(t_n)$ at time $t_n = n\tau$, $n = 0,1,2,\ldots,$ and $\mathcal{A}(f)$ denotes the evolution map given by the numerical scheme. Note that for implicit methods, the evolution map $\mathcal{A}(f)$ requires a non-linear solve. In this text we assume that the explicit form (2.2) can always be obtained uniquely in exact arithmetic and we neglect the presence of rounding errors. We also allow numerical methods for which the evolution map $\mathcal{A}(f)$ involves derivatives of f with respect to u.

We define the *truncation error* $T(u_0, \tau, f)$ of scheme (2.2) by

$$T(u_0, \tau, f) := \frac{1}{\tau}(u(\tau) - U_1) = \frac{1}{\tau}(u(\tau) - \mathcal{A}(f)(u_0, \tau)).$$

Scheme (2.2) is
* *of order* m *on eq. (2.1)*, if m is the largest integer such that

$$\lim_{\tau \to 0} \frac{||T(u_0, \tau, f)||}{\tau^m} < \infty$$

for all smooth f and arbitrary u_0;
* *exact on the solution* $u(t; u_0)$ *of eq. (2.1) for given* f, if $T(u_0, \tau, f)$ vanishes for arbitrary step-size $\tau \leq \tau_0 \in [0, T]$, $\tau_0 > 0$ small enough;
* *exact on eq. (2.1) for given* f, if $T(u_0, \tau, f)$ vanishes for arbitrary initial value $u_0 \in \mathbb{R}^q$ and arbitrary step-size $\tau \leq \tau_0 \in [0, T]$.

2.1. Error expansions for constant-coefficient schemes. We now confine to autonomous scalar initial value problems

$$\dot{u} = f(u), \qquad u(0) = u_0 \in \mathbb{R}. \tag{2.3}$$

For the analysis we expand $T(u_0, \tau, f)$ in a Taylor series in τ,

$$T(u_0, \tau, f) = \sum_{j=0}^{\infty} B_j(f) \tau^j. \tag{2.4}$$

Scheme (2.2) is *of order* m on eq. (2.3), if $B_j(f) = 0$ for all $j < m$, arbitrary u_0 and all smooth functions f. Scheme (2.2) is *exact* on eq. (2.3) *for a given function* f, if $B_j(f) = 0$ for arbitrary $u_0 \in \mathbb{R}$ and for all $j \geq 0$.

LEMMA 2.1. *The trapezoidal rule*

$$\frac{U_{n+1} - U_n}{\tau} = \frac{f(U_{n+1}) + f(U_n)}{2} \tag{2.5}$$

is exact on equation (2.3) for those functions $f : \mathbb{R} \to \mathbb{R}$ *satisfying*

$$f'' f + (f')^2 = 0, \tag{2.6}$$

i.e. for $f(u) = \pm\sqrt{a_1 u + a_2}$*,* a_1, a_2 *constant,* $a_1 u + a_2 \geq 0$*. It follows that it is exact for solutions* $u(t)$ *satisfying* $u(t) \in$ *span*$\{1, t, t^2\}$.

The proof was already given in [14] and [4]. We thus only sketch it here. For the truncation error we obtain the expansion (2.4) with

$$B_j(f) = \begin{cases} 0 & j < 2, \\ \left(\dfrac{1}{(j+1)!} - \dfrac{1}{2j!}\right) \left.\dfrac{d^j f(u(t))}{dt^j}\right|_{t=0} & j \geq 2. \end{cases} \tag{2.7}$$

For general smooth f we thus obtain $B_0 = B_1 = 0$ and

$$B_2(f) = -\frac{1}{12}[f''(u) f^2(u) + (f'(u))^2 f(u)]|_{u=u_0} \neq 0.$$

The method is second order in general. For those f which satisfy eq. (2.6), we obtain $B_2(f) = 0$, and moreover $B_j(f) = 0$ for all j. Thus the trapezoidal rule is exact for those

ETNA
Kent State University
etna@mcs.kent.edu

f satisfying eq. (2.6). We then get by integration $f(u) = \pm(a_1u + a_2)^{1/2}$, a_1, a_2 constants. Special solutions of eq. (2.3) with this f are $u(t) = \tilde{a}_2 t + u_0$ and $u(t) = \left(\tilde{a}_1 t/2 + u_0^{1/2}\right)^2$. More general we obtain from the differential equation that

$$0 = \frac{d^2 f(u(t))}{dt^2} = \frac{d^3 u(t)}{dt^3}\ .$$

Thus the solution space is $U = \mathrm{span}\{1, t, t^2\}$. $\diamond$

In a similar way the following lemma was proved as well:

LEMMA 2.2. *The implicit midpoint rule*

$$\frac{U_{n+1} - U_n}{\tau} = f(\frac{U_{n+1} + U_n}{2})$$

is exact on equation (2.3) for those functions $f : I\!R \to I\!R$ satisfying

$$f'' f - 2(f')^2 = 0, \tag{2.8}$$

i.e. for $f(u) = a_2(a_3 - a_1u)^{-1}$, a_1, a_2, a_3 constants, $a_3 - a_1u(t) \neq 0$ for all t. Solutions u of eq. (2.3) then satisfy

$$u(t) = a_2 t + u_0 \qquad \text{for } a_1 = 0,\ a_3 = 1,$$

or

$$u(t) = \frac{a_3 \pm \sqrt{(a_3 - a_1u_0)^2 - 2a_1t}}{a_1} \qquad \text{for } a_1 \neq 0,\ a_3 - a_1u_0 \neq 0,\ a_2 = 1. \tag{2.9}$$

Note that this time the solutions form a nonlinear manifold, not a linear space like in the previous case. In particular, $u(t) = \sqrt{t}$ does not belong to the solution manifold for $t_0 = 0$.

In general, the solutions of eq. (2.3) should not be expected to form a linear space for nonlinear f. If $u(t)$ solves $\dot{u} = f(u),\ u(0) = u_0$, then $w(t) := a_1u(t) + a_2,\ a_1 \neq 0$, solves $\dot{w} = a_1 f([w - a_2]/a_1),\ w(0) = a_1u_0 + a_2$. That's all we can say for general f.

Also, uniqueness of exact schemes should not be expected: already in [4] it was shown that both the second-order Taylor method

$$U_{n+1} = U_n + \tau f(U_n) + \frac{\tau^2}{2} f'(U_n) f(U_n), \tag{2.10}$$

and the trapezoidal rule (2.5) are exact on the same set of differential equations (2.3), i.e. on those with a r.h.s.-function f satisfying (2.6). They are clearly different difference schemes, one explicit, one implicit, and also their error expansions are different: expansion (2.4) for scheme (2.10) has the coefficients

$$B_j(f) = \begin{cases} 0 & j < 2, \\ \dfrac{1}{(j+1)!} \dfrac{d^j f(u(t))}{dt^j}\Big|_{t=0} & j \geq 2. \end{cases}$$

So all $B_j,\ j \geq 2$, are different from those in (2.7), but vanish for the same fs.

This non-uniqueness should not surprise: difference schemes are equations to be satisfied by the approximate solutions of the differential equations under consideration. So there is not *the* exact difference scheme, there might be many of them, differing by terms which vanish for those differential equations on which they are exact. What we have to require, of course, is the unique solvability of the difference scheme for given initial value and sufficiently small step-size τ.

The trapezoidal rule and the implicit midpoint rule both are Runge-Kutta methods. We thus look at exact schemes within the framework of RK methods now.

ETNA
Kent State University
etna@mcs.kent.edu

2.2. Constant coefficient RK methods as exact schemes. In this subsection we collect some facts on RK methods to be used lateron. We consider here non-autonomous differential equations (2.1) again, and we will always assume $U_0 = u_0$ for the discrete iteration.

2.2.1. Constant coefficient RK methods. Let $b_i, a_{ij} (i, j = 1, \dots, s)$ be real numbers and let

$$c_i = \sum_{j=1}^{s} a_{ij}, \qquad i = 1, \dots, s. \tag{2.11}$$

The method defined by

$$k_i = f(t_n + c_i\tau, U_n + \tau \sum_{j=1}^{s} a_{ij}k_j), \qquad i = 1, \dots, s$$
$$U_{n+1} = U_n + \tau \sum_{i=1}^{s} b_i k_i,$$

is called an *s-stage Runge-Kutta scheme (RK scheme)*. An alternative definition is

$$Y_i = U_n + \tau \sum_{j=1}^{s} a_{ij} f(t_n + c_j\tau, Y_j), \qquad i = 1, \dots, s, \tag{2.12}$$

$$U_{n+1} = U_n + \tau \sum_{i=1}^{s} b_i f(t_n + c_i\tau, Y_i). \tag{2.13}$$

The connection between both is given by

$$k_i = f(t_n + c_i\tau, Y_i),$$
$$Y_i = U_n + \tau \sum_{j=1}^{s} a_{ij}k_j, \qquad i = 1, \dots, s.$$

There is a certain redundancy: formally different schemes can define the same numerical integration method, even when they have different stage numbers s_1 and s_2. In the following we consider *Runge-Kutta methods*, assuming that the resulting numerical integration method is at least of first order, and that the scheme representing it has minimal stage number s and satisfies $c_i \neq c_j$ for $i \neq j$.

2.2.2. Collocation methods. Remember the following facts [3, p.58], [6, p.211f]:
* If an RK method satisfies the simplifying conditions

$$B(\xi): \qquad \sum_{i=1}^{s} b_i c_i^{k-1} = 1/k, \quad 1 \le k \le \xi, \tag{2.14}$$

and is used for integrating

$$\dot{y} = f(t), \qquad y(t_n) = 0, \tag{2.15}$$

on the interval (t_n, t_{n+1}), then eq. (2.13) is an integration method of order ξ. Equation (2.13) is then exact on $f \in \text{span}\{1, t, \dots, t^{\xi-1}\}$. Trivial consequence: all consistent RK schemes are exact on $\dot{u} = \text{const}$.

* If an RK method satisfies the simplifying conditions

$$C(\xi): \qquad \sum_{j=1}^{s} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k, \quad 1 \le k \le \xi, \quad 1 \le i \le s, \tag{2.16}$$

then the stage equations (2.12) define integration methods of order ξ for eq. (2.15) on the intervals $(t_n, t_n + c_i\tau)$. Thus they are exact there for $f \in \text{span}\{1, t, \dots, t^{\xi-1}\}$. Note that the validity of $C(1)$ is ensured by eq. (2.11).

* If an s-stage RK method satisfies $B(s)$ and $C(s)$, it is called a *collocation method* (after Burrage 1978).

* Apply a collocation method (2.12), (2.13) to the differential equation (2.1). The *collocation polynomial*, i.e. the polynomial $p(t)$ interpolating the numerical solution of (2.12), (2.13) in the points t_n and $t_{ni} := t_n + c_i\tau$, $i = 1, \dots, s$, then has degree s. Its derivative $\dot{p}$ has degree $s-1$ and is integrated exactly by the stage equations (2.12) on the intervals $[t_n, t_{ni}]$. We thus obtain

$$p(t_n + c_i\tau) = p(t_n) + \tau \sum_{j=1}^{s} a_{ij}\, \dot{p}(t_n + c_j\tau), \qquad i = 1, \dots, s,$$

$$p(t_n + \tau) = p(t_n) + \tau \sum_{i=1}^{s} b_i\, \dot{p}(t_n + c_i\tau).$$

Put $p(t_n) = u_n$, where $u_n = u(t_n)$ and $u(t)$ is the solution of eq. (2.1) to be approximated. Then the collocation polynomial satisfies eq. (2.1) at the internal abscissas $t_{ni} = t_n + c_i\tau$, $i = 1, \dots, s$,

$$\dot{p}(t_n + c_i\tau) = f(t_n + c_i\tau, p(t_n + c_i\tau)), \quad i = 1, \dots, s.$$

* Given a positive integer s and numbers $c_1, \dots, c_s \in I\!R$, $0 \le c_i \le 1$, $i = 1, \dots, s$, $c_i \ne c_j$ for $i \ne j$. The collocation method satisfying

$$\begin{aligned} p(t_n) &= U_n \\ \dot{p}(t_n + c_i\tau) &= f(t_n + c_i\tau, p(t_n + c_i\tau)), \quad i = 1, \dots, s, \\ p(t_n + \tau) &=: U_{n+1} \end{aligned}$$

is equivalent to the s-stage RK method (2.12), (2.13) with coefficients

$$a_{ij} := \int_0^{c_i} \ell_j(t)dt, \qquad b_j := \int_0^1 \ell_j(t)dt, \qquad i, j = 1, \dots, s, \tag{2.17}$$

where the $\ell_j(t)$ are the Lagrange polynomials

$$\ell_j(t) = \prod_{m \ne j} \frac{(t - c_m)}{(c_j - c_m)}.$$

Note that $c_i \ne c_j$ for $i \ne j$ is essential here.

2.2.3. RK methods as exact schemes. The trapezoidal rule can be written as a Runge-Kutta method,

$$\begin{aligned} Y_1 &= U_n, \\ Y_2 &= U_n + \tau(f(t_n, Y_1) + f(t_n + \tau, Y_2))/2, \\ U_{n+1} &= U_n + \tau(f(t_n, Y_1) + f(t_n + \tau, Y_2))/2. \end{aligned}$$

It is a second-order 2-stage method and it satisfies the simplifying conditions $B(k)$ and $C(k)$ for $k \leq 2$, but not for $k = 3$. It thus is a collocation method and integrates eq. (2.15) exactly for $f \in \text{span}\{1, t\}$. As we have seen earlier, it is also exact on autonomous eqs. (2.3) if f satisfies eq. (2.6), which implies that $u(t) \in \text{span}\{1, t, t^2\}$.

The implicit midpoint rule can also be written as a Runge-Kutta method,

$$\begin{aligned} Y_1 &= U_n + \tau f(t_n + \tau/2, Y_1)/2, \\ U_{n+1} &= U_n + \tau f(t_n + \tau/2, Y_1). \end{aligned}$$

It is a second-order 1-stage method satisfying $B(2)$ and $C(1)$. It thus is a collocation method and integrates eq. (2.15) exactly for $f \in \text{span}\{1, t\}$. As we have seen earlier, it is also exact on autonomous eqs. (2.3) if f satisfies eq. (2.8).

The question thus arises if every RK method is exact on some nontrivial differential equation. As the following example shows, the answer is *no*. Consider the μ-dependent family of second-order 2-stage schemes for autonomous f,

$$\begin{aligned} Y_1 &= U_n, \\ Y_2 &= U_n + \tau(\mu f(Y_1) + (1-\mu) f(Y_2)), \\ U_{n+1} &= U_n + \tau(f(Y_1) + f(Y_2))/2. \end{aligned} \tag{2.18}$$

For $\mu = 1/2$ this is the trapezoidal rule. For the counterexample we choose $\mu = 2/3$. For $\mu = 2/3$ we obtain $B_0 = B_1 = 0$ and

$$B_2(f) = (\frac{1}{2} \cdot 0 \cdot 1 + \frac{1}{2} \cdot \frac{1}{3} \cdot 1 - \frac{1}{6}) f f'^2 + (\frac{1}{2} \cdot \frac{1}{2} \cdot 1 - \frac{1}{6}) f^2 f'' = \frac{1}{12} f^2 f''.$$

Hence the scheme is second order in general. $B_2(f)$ vanishes if either $f = 0$ or $f'' = 0$. This means $f(u) = a_1 u + a_2$ for arbitrary constants a_1 and a_2. Thus for the differential equation

$$\dot{u} = a_1 u + a_2 \tag{2.19}$$

the scheme is at least third order. With $f'' = 0$ we find

$$B_3(f) = \frac{1}{72} f f'^3 + \frac{1}{24} f''' f^3 + \frac{1}{12} f^2 f' f'' = \frac{1}{72} f f'^3.$$

Hence B_3 does not vanish when B_2 does. Thus the 2nd order scheme (2.18) with $\mu = 2/3$ is only third order for eq. (2.19) and not an exact scheme.

Thus some RK schemes are exact for larger classes of autonomous differential equations, others only for the trivial case where f is constant. The vanishing of the first non-zero term in the error expansion by particular choice of the r.h.s. function f does not guarantee exactness as one might have hoped for from the analysis of the classical schemes at the beginning of this section.

We now turn to different approaches which do allow to find exact schemes for equation (2.19) and for linear systems of type (2.19). It turns out, however, that the coefficients of the schemes must be allowed to depend on the step-size. Examples are given in eqs. (2.26) and (3.2).

2.3. Functional fitting: variable-coefficient RK schemes. In recent years much research has been performed for finding efficient numerical methods for system (2.1) with oscillatory solutions. If a good estimate of the frequency is known in advance, exact integration of the linear part of system (2.1) leads to very useful integration methods which are called 'exponentially-fitted' integration methods. For a list of references we refer to [17] and [18] and the references therein. Here we take a closer look at two different methods in this family: the application of the Principle of Coherence by Denk [1, 2] and the 'functional-fitting RK methods' as introduced by Ozawa [17].

2.3.1. Invariant spaces. Functional fitting methods approach exactness from the solution side. They do not look at the r.h.s. function f to find conditions under which a given scheme becomes exact, but they construct schemes which allow given functions $u(t)$ to be represented exactly. To formulate the following existence theorem without inconvenient restrictions, we consider non-autonomous systems (2.1) again and allow variable coefficients in the RK schemes, i.e. we consider schemes whose coefficients $a_{ij},\ b_i$ depend on the independent variable t and on the step-size τ. Ozawa [17] proved the following results, which we repeat here in the formulation of [4]:

LEMMA 2.3. *Let $\{c_i\}_{i=1}^s \in I\!R$ be given, $c_i \neq c_j$ for $i \neq j$. Let $\{v_m(t)\}_{m=1}^s \in C^s[0,T]$ be linearly independent functions, sufficiently smooth such that each of them satisfies*

$$\dot{v}_m(t + c_i\tau) = \sum_{j=1}^{s} \frac{(c_i\tau)^{j-1}}{(j-1)!} v_m^{(j)}(t) + \mathcal{O}(\tau^s), \tag{2.20}$$

and suppose that they solve in $[0,T]$ a homogeneous linear differential equation

$$\sum_{m=0}^{s} p_m(t) v^{(m)}(t) = 0 \qquad \text{with } \ p_s(t) \equiv 1,\ \ p_0(t) \neq 0, \tag{2.21}$$

with continuous coefficients $p_m \in C[0,T]$. Then the linear system

$$\begin{array}{rcl} v_m(t + c_i\tau) & = & v_m(t) + \tau \sum_{j=1}^{s} a_{ij}(t,\tau)\dot{v}_m(t + c_j\tau) \\ v_m(t + \tau) & = & v_m(t) + \tau \sum_{i=1}^{s} b_i(t,\tau)\dot{v}_m(t + c_i\tau) \end{array} \tag{2.22}$$

is uniquely solvable for $a_{ij}(t,\tau)$ and $b_i(t,\tau)$, with $t \in [0,T]$ and $0 < \tau < \tau_0$ for τ_0 small enough.

Note that the assumptions on the functions $v_m(t)$ exclude the constant $u(t) \equiv const$ from the set $\{v_m\}_{m=1}^s$. Nevertheless, the constant function will be contained in any linear space spanned by functions satisfying system (2.22): it satisfies system (2.22) for any set of coefficients $a_{ij}(t,\tau)$ and $b_i(t,\tau)$ and thus for any set of functions $v_1(t), \ldots, v_s(t)$.

The idea of the *proof* given by Ozawa is the following: For fixed t and τ, system (2.22) is a collection of $s+1$ linear systems of order s with matrix

$$(\dot{v}_m(t + c_j\tau))_{m,j=1,\ldots s} =: \bar{U}(t,\tau), \tag{2.23}$$

inhomogeneities

$$\left((\frac{v_m(t + c_i\tau) - v_m(t)}{\tau})_m \right)_i \ \text{and}\ (\frac{v_m(t + \tau) - v_m(t)}{\tau})_m,$$

and $s^2 + s$ unknowns $a_{ij}(t,\tau),\ b_i(t,\tau)$. These systems are uniquely solvable if the matrix $\bar{U}(t,\tau)$ is nonsingular. To prove that $\bar{U}(t,\tau)$ is nonsingular for all $t \in [0,T]$ and small enough

ETNA
Kent State University
etna@mcs.kent.edu

τ, conditions (2.20) and (2.21) are used. Condition (2.21) ensures that the Wronskian matrix of the linearly independent functions $\{v_1, \ldots, v_s\}$ is nonsingular [6, p. 64ff].$\diamond$

Now we consider the performance of a scheme obtained via Lemma 2.3. Assume that a function $u \in U := \text{span}\{1, v_1, \ldots. v_s\}$ satisfies the non-autonomous system (2.1). Then we expect that the RK scheme with coefficients attained according to Lemma 2.3 will be exact on $u(t)$. We get more: from this exactness on the linear space U it follows that the scheme has order s:

LEMMA 2.4. *Let the coefficients of the variable-coefficient s-stage RK scheme*

$$\begin{aligned} Y_i &= U_n + \tau \textstyle\sum_{j=1}^{s} a_{ij}(t_n, \tau) f(t_n + c_j\tau, Y_j) \\ U_{n+1} &= U_n + \tau \textstyle\sum_{i=1}^{s} b_i(t_n, \tau) f(t_n + c_i\tau, Y_i) \\ & \quad i = 1, \ldots, s \qquad t_n = t_0 + n\tau, \qquad U_0 = u_0, \end{aligned}$$

be obtained according to Lemma 2.3. Then the order of the scheme is at least s. If the abscissae c_i, $i = 1, \ldots, s$, are taken to satisfy

$$\int_0^1 t^{q-1} \prod_{i=1}^{s} (t - c_i)dt = 0, \qquad q = 1, \ldots, \nu, \qquad 1 \le \nu \le s, \tag{2.24}$$

then the order of accuracy is $s + \nu$. The maximum attainable order is $2s$.

The *proof* of these statements uses results on RK collocation methods with constant coefficients and can be found in Ozawa [17]. It is shown there that the first terms $a_{ij}^{(0)}$ and $b_i^{(0)}$ in the power expansions of $a_{ij}(t, \tau)$ and $b_i(t, \tau)$ with respect to τ satisfy the simplifying conditions $B(s)$ and $C(s)$ and depend only on $\{c_i\}_{i=1}^s$, but not on the generating functions $\{v_m\}_{m=1}^s$. They thus agree with the coefficients of the corresponding collocation scheme. If the abscissae c_i satisfy the additional condition (2.24), both the RK collocation scheme and the scheme obtained according to (2.22) have order $p = s + \nu \le 2s$. $\diamond$

The s-stage RK scheme obtained with the linearly independent functions $\{v_m\}_{i=1}^s$ is exact on the solution $u(t)$ whenever $u(t) \in U = \text{span}\{1, v_1, \ldots. v_s\}$. If all solutions of (2.1) happen to belong to U in the full time interval $[0, T]$, the scheme is exact on (2.1), no matter how nonlinear f is, because we can first construct the linear combination of the basis functions and afterwards we replace $\dot{u}$ by $f(t, u)$. It is thus of interest to use functional fitting RK-schemes whenever there is some knowledge about the solution in advance. If the scheme is not exact, the remaining part of the solution is captured by the order of the RK-scheme; and the constants in the error expansion will be small as long as the scheme is in a small neighborhood of an exact scheme.

Given an s-stage RK scheme which is exact on $U = \text{span}\{1, v_1, \ldots, v_s\}$, it is a member of a family of nonconfluent schemes which depend on s parameters $c_1, \ldots c_s$. All these schemes are exact on the same function space U. Though all these schemes are equivalent when the scheme is used as an exact scheme, they differ in their numerical performance when the scheme is used on a problem where it is not exact. This follows from the second statement in Lemma 2.4.

2.3.2. Examples for schemes obtained with Lemma 2.3. When we apply Lemma 2.3, the resulting scheme might have constant or variable coefficients, depending on the generating functions. This is illustrated by the following examples. RK schemes with variable coefficients are nonstandard. Time-dependent coefficients are very unusual and would be inconvenient in computations. In § 2.3.3 we will show how to avoid them. Coefficients depending on the step-size τ are not quite that unusual: such coefficients are always obtained in the context of exponential fitting [18] and of evaluating the Principle of Coherence [2] and seem to be unavoidable in certain situations.

For constant-coefficient schemes for non-autonomous differential equations, it is a *convention* to satisfy eq. (2.11) when designing new schemes. Because condition (2.11) implies that $t_n + c_i\tau = U_n + c_i\tau$ for $u(t) = t$ [3, p. 56]. In the case of Lemma 2.3, condition (2.11) is satisfied if $u(t) = t$ is one of the generating functions.

Example 1 (constant coefficients): We chose $s = 2$, $v_1(t) = t$, $v_2(t) = t^2$ and use c_1, c_2 as parameters. Solving system (2.22) we find the coefficients

$$\begin{aligned} a_{i1} &= \frac{c_i^2 - 2c_ic_2}{2(c_1 - c_2)}, & a_{i2} &= c_i - a_{i1}, \quad i = 1, 2 \\ b_1 &= \frac{1 - 2c_2}{2(c_1 - c_2)}, & b_2 &= 1 - b_1. \end{aligned} \tag{2.25}$$

For varying $c_1,\ c_2$ with $c_1 \neq c_2$ this is a 2–parameter family of RK schemes. For $c_1 = 0,\ c_2 = 1$ we obtain the coefficients of the trapezoidal rule, as expected. We obtain the trapezoidal rule also for $c_1 = 1, c_2 = 0$. Though all these schemes are equivalent when exact, they differ in their order and numerical performance when not exact.

These schemes are equivalent to the collocation schemes obtained for $\{c_1, c_2\}$ via eq. (2.17). This was to be expected because of the remark on $a_{ij}^{(0)}$, $b_i^{(0)}$. Moreover, if $p(t)$ is the collocation polynomial introduced earlier, we find $p(t) \in \text{span}\{1, t, t^2\}$. We see that the coefficients are undefined in the degenerate case $c_1 = c_2$, i.e. when the collocation approach of § 2.2.2 is not applicable. $\diamond$

Example 2 (τ-dependent coefficients): With $s = 2$, $c_1 = 0$, $c_2 = 1$ and $v_1(t) = t$, $v_2(t) = \exp \lambda t$, $1/\lambda \notin [0, T]$, we obtain the coefficients

$$\begin{aligned} a_{11} &= 0 & a_{12} &= 0 \\ a_{21}(\tau) &= \frac{1 - (1 - \lambda\tau)\exp \lambda\tau}{\tau\lambda(\exp \lambda\tau - 1)}, & a_{22}(\tau) &= \frac{-1 - \lambda\tau + \exp \lambda\tau}{\tau\lambda(\exp \lambda\tau - 1)} \\ b_1(\tau) &= \frac{1 - (1 - \lambda\tau)\exp \lambda\tau}{\tau\lambda(\exp \lambda\tau - 1)}, & b_2(\tau) &= \frac{-1 - \lambda\tau + \exp \lambda\tau}{\tau\lambda(\exp \lambda\tau - 1)}\ . \end{aligned} \tag{2.26}$$

These coefficients have an apparent singularity in the limit $\tau \to 0$. The limiting values of $a_{21}(\tau), a_{22}(\tau), b_1(\tau)$ and $b_2(\tau)$ computed by L'Hopital's rule all are equal to $1/2$. This was to be expected because the scheme becomes the related collocation scheme for $\tau \to 0$. Ozawa thus recommends to use τ-expanded coefficients in practical computations. $\diamond$

Example 3 (t and τ dependent coefficients): With $s = 2$, parameters $c_1 \neq c_2$ and $v_1(t) = t$, $v_2(t) = (t + 1)^{-1}$, we obtain with $d(c) := (t + 1 + c\tau)^{-1}$, $\quad d(c)^2 = d(c) \cdot d(c)$,

$$\begin{aligned} a_{i1}(t, \tau) &= \frac{d(c_i) - d(0) + \tau c_i d(c_2)^2}{(d(c_2)^2 - d(c_1)^2)\tau}, & a_{i2} &= c_i - a_{i1}, \quad i = 1, 2 \\ b_1(t, \tau) &= \frac{d(1) - d(0) + \tau d(c_2)^2}{(d(c_2)^2 - d(c_1)^2)\tau}\ , & b_2 &= 1 - b_1\ . \end{aligned} \tag{2.27}$$

This example shows that quite simple functions $\{v_m\}$ can lead to complicated coefficients which depend on t and τ. $\diamond$

2.3.3. Conditions for obtaining constant coefficients. We now ask for general conditions such that the coefficients are constant, i.e. independent of time t and/or step-size τ.

THEOREM 2.5. *Let the assumptions of Lemma 2.3 be satisfied. Then the coefficients computed according to Lemma 2.3 are independent of* t *iff the linear space*

$$\tilde{U} := span\{1, v_1(t), \ldots, v_s(t)\}$$

ETNA
Kent State University
etna@mcs.kent.edu

is closed with respect to differentiation.

Proof. 1. necessary: Time-independent coefficients satisfy

$$\dot{a}_{ij}(t,\tau) := \frac{\partial}{\partial t} a_{ij}(t,\tau) = 0, \qquad \dot{b}_i(t,\tau) := \frac{\partial}{\partial t} b_i(t,\tau) = 0 \tag{2.28}$$

for all t. Computing the time derivative of every equation of system (2.22) and using relations (2.28) we obtain

$$\begin{array}{rcl} \dot{v}_m(t+c_i\tau) & = & \dot{v}_m(t) + \tau \sum_{j=1}^{s} a_{ij}(t,\tau)\ddot{v}_m(t+c_j\tau), \\ \dot{v}_m(t+\tau) & = & \dot{v}_m(t) + \tau \sum_{i=1}^{s} b_i(t,\tau)\ddot{v}_m(t+c_i\tau). \end{array} \tag{2.29}$$

Formally, this is the same as system (2.22), with $v_m(t)$ replaced by $\dot{v}_m(t)$. Time-independent coefficients have to satisfy both this system and system (2.22). Since system (2.22) already determines the coefficients uniquely, this is possible only if system (2.29) does not add new conditions but consists of linear combinations of equations contained in system (2.22), possibly adding the trivial equation for the constant. This means that span$\{\dot{v}_1,\ldots,\dot{v}_s\}$ is a subset of span$\{1,v_1,\ldots,v_s\}$. Thus the linear space $\tilde{U}$ is closed with respect to differentiation.
2. sufficient: If the linear space $\tilde{U}$ is closed with respect to differentiation, there are coefficients $\alpha_{m\nu}$ such that $\dot{v}_m(t) = \sum_{\nu=0}^{s} \alpha_{m\nu} v_\nu(t),\ m = 1,\ldots,s,\ v_0(t) \equiv 1$. Inserting this into the time derivative of every equation of system (2.22) and using the fact that the functions v_ν solve system (2.22), we obtain

$$\begin{aligned} \sum_{j=1}^{s} \dot{a}_{ij}(t,\tau) \sum_{\nu=0}^{s} \alpha_{m\nu} v_\nu(t+c_j\tau) &=0= \sum_{j=1}^{s} \dot{a}_{ij}(t,\tau)\dot{v}_m(t+c_j\tau), \\ \sum_{i=1}^{s} \dot{b}_i(t,\tau) \sum_{\nu=0}^{s} \alpha_{m\nu} v_\nu(t+c_i\tau) &=0= \sum_{i=1}^{s} \dot{b}_i(t,\tau)\dot{v}_m(t+c_i\tau). \end{aligned}$$

This is equivalent to the $s+1$ linear systems

$$\bar{U}(t,\tau)\begin{pmatrix} \dot{a}_{i1} \\ \vdots \\ \dot{a}_{is} \end{pmatrix} = 0, \quad i = 1,\ldots,s, \qquad \bar{U}(t,\tau)\begin{pmatrix} \dot{b}_1 \\ \vdots \\ \dot{b}_s \end{pmatrix} = 0,$$

where $\bar{U}(t,\tau)$ is the Wronskian matrix of the generating functions introduced in eq. (2.23). Nonsingularity of this matrix was essential for the proof of Lemma 2.3. From this follows that $\dot{a}_{ij}(t,\tau) \equiv 0,\ \dot{b}_i(t,\tau) \equiv 0,\ i,j = 1,\ldots,s.$ □

Examples: Each of the spaces span$\{e^{\omega t}\}$, span$\{e^{-\omega t}\}$ and span$\{\sin\omega t, \cos\omega t\}$ is closed with respect to differentiation. The linear spaces span$\{t^s, t^{s-1},\ldots,t,1\}$, $s \in I\!N$, and span$\{1,t,e^{\omega t}\}$ are also closed with respect to differentiation. This agrees with the time-independence of the coefficients given in (2.25) and (2.26).

There is no finite-dimensional linear space containing span$\{(t+c)^{-1}\}$, $c \in I\!R$, and closed with respect to differentiation. This is in agreement with the time-dependence of coefficients (2.27).

There is no finite-dimensional linear space containing span$\{(1+t)^{1/2}\}$ and closed with respect to differentiation. We thus get (quite complicated) coefficients which depend both on t and τ. As we have seen earlier, however, there *is* a 1-stage constant-coefficient RK scheme exact on $u(t) = (t+1)^{1/2}$, *when applied to differential equation (2.3) with (2.8)*: the implicit midpoint rule (put $u_0 = 1,\ a_1 = -2,\ a_3 = 0$ in eq. (2.9)). This example confirms that the properties of the nonlinearities f play an important role for the results of § 2.1. ⋄

ETNA
Kent State University
etna@mcs.kent.edu

If we procede analogously for finding general conditions on generating functions that allow τ-independent coefficients, we find that the generating functions have to satisfy both system (2.22) and

$$\begin{aligned} \dot{v}_m(t+c_i\tau)c_i &= \sum_{j=1}^s a_{ij}(t,\tau)\dot{v}_m(t+c_j\tau) + \tau\sum_{j=1}^s a_{ij}(t,\tau)\ddot{v}_m(t+c_j\tau)c_j, \\ \dot{v}_m(t+\tau) &= \sum_{i=1}^s b_i(t,\tau)\dot{v}_m(t+c_i\tau) + \tau\sum_{i=1}^s b_i(t,\tau)\ddot{v}_m(t+c_i\tau)c_i. \end{aligned} \tag{2.30}$$

From § 2.2.2 we know that the collocation schemes must satisfy these two systems. When we check *how* they do so, we find that the equations for t^m in system (2.30), $m = 1,\ldots,s$, are equivalent to the mth level of the simplifying conditions $B(s)$ and $C(s)$ introduced in eqs. (2.14) and (2.16).

2.4. The Principle of Coherence. Exact schemes for linear differential equations with constant coefficients and for systems of such equations were derived by many authors, sometimes by using the known continuous solution. A straight-forward procedure for deriving equations to be satisfied by the coefficients of exact schemes is the Principle of Coherence introduced by Hersch in 1958.

The basic idea of the Principle of Coherence was formulated by Hersch [5] as *Successive approximations should not contradict each other*. We explain this by the following example: consider

$$\ddot{z}(t) + \lambda^2 z(t) = 0, \quad \lambda > 0. \tag{2.31}$$

Using central finite differences at t with step-size τ we obtain

$$z(t-\tau) - (2-\lambda\tau^2)z(t) + z(t+\tau) = 0. \tag{2.32}$$

We write instead

$$z(t-\tau) - A(\tau)z(t) + z(t+\tau) = 0, \tag{2.33}$$

where the coefficient $A(\tau)$ is to be determined. With step-size 2τ we obtain similarly

$$z(t-2\tau) - A(2\tau)z(t) + z(t+2\tau) = 0. \tag{2.34}$$

By linear combination of three difference equations of type (2.33) centered at $t-\tau$, t and $t+\tau$ we obtain on the other hand

$$z(t-2\tau) - (A(\tau)^2 - 2)z(t) + z(t+2\tau) = 0. \tag{2.35}$$

For a coherent numerical approximation of eq. (2.31), eqs. (2.34) and (2.35) should coincide. This means $A(2\tau) = A(\tau)^2 - 2$. This is satisfied for $A(\tau) = 2\cos\kappa\tau$, $\kappa \in I\!R$. Moreover, the resulting difference scheme has to be consistent with eq. (2.31), i.e. the first-order truncation error has to vanish for $\tau \to 0$. Thus we obtain $\kappa = \lambda$. The coherent scheme therefore is

$$z(t-\tau) - 2\cos(\lambda\tau)z(t) + z(t+\tau) = 0. \tag{2.36}$$

Comparison shows that scheme (2.36) is exact on eq. (2.31), while the second-order central difference formula (2.32) uses the first two terms of the Taylor expansion of $A(\tau)$. We required coherence using three grid points, and we obtained exactness. For equations of higher order or larger systems, the derivation of exact schemes by the Principle of Coherence can become quite envolved. Denk employed the calculus of distributions and obtained schemes that are very useful in applications [1].

ETNA
Kent State University
etna@mcs.kent.edu

3. Applications. We conclude by showing that exact schemes are of practical relevance in scientific computing. We give two examples where exact schemes for simpler equations led to efficient schemes for more complicated equations.

3.1. Denk-Bulirsch schemes. Denk used the Principle of Coherence explained in § 2.4 for designing a numerical integration method which is capable of simulating highly oscillatory circuits efficiently and reliably. He treated first order systems

$$\dot{u} + Au = f(t,u), \tag{3.1}$$

where the matrix A represents the linear elements of the circuit and $f(t,u)$ assembles the nonlinear terms. Applying the Principle of Coherence with the step-sizes τ and 2τ to $\dot{z} + Az = 0$ leads to

$$\begin{aligned} Z(t_1) &= \Phi(\tau)Z(t_0+\tau) = \Phi(\tau)^2 Z(t_0), \\ &= \Phi(2\tau)Z(t_0), \end{aligned} \tag{3.2}$$

and thus to $\Phi(\tau) = \exp(-A\tau)$, as expected. For approximation of eq. (3.1), Denk combined this with a multistep approach (explicit or implicit). Explicit:

$$\begin{aligned} U(t+m\tau) - \exp(-A\tau)U(t+(m-1)\tau) = \\ \tau \sum_{i=0}^{m} \beta_i(\tau) f\left(t+(m-1)\tau, U(t+(m-1)\tau)\right), \end{aligned}$$

where $\beta_i(\tau)$ are *matrix coefficients* of the multistep scheme. In the case $m = 1$ this leads to

$$\begin{aligned} \beta_1(\tau) &= -\left(I - (I - \exp(-A\tau))(A\tau)^{-1}\right)(A\tau)^{-1}, \\ \beta_0(\tau) &= (I - \exp(-A\tau))(A\tau)^{-1} - \beta_1. \end{aligned}$$

Note that the coefficients $\beta_i(\tau)$ have an apparent singularity for $\tau \to 0$, similar to the coefficients given in (2.26). This seems to be unavoidable.

The resulting scheme is consistent of order m, $A(0)$-stable and convergent. It is A-stable without order restrictions. This is no contradiction to Dahlquist's order barriers because those barriers were shown to hold for constant-coefficient schemes. Tests with a system of five equations in a time interval corresponding to about 250 000 oscillations showed that the scheme is more efficient and more reliable than the standard code LSODE for this type of problems [2]. The code HERSCH developed using these ideas for equations more general than eq. (3.1) was tested in numerical experiments on equations modeling electric circuits. It proved to be more efficient than the codes LSODE, DOPRI5 and RADAU5 on highly oscillatory problems [1].

3.2. LeRoux schemes. Consider the parabolic problem

$$\begin{aligned} v_t - \Delta v^m &= \alpha v^m \qquad \text{for } x \in \Omega \subset I\!R^d,\ t > 0, \\ v(x,t) &= 0 \qquad \text{for } x \in \partial\Omega,\ t > 0, \\ v(x,0) &= v_0(x) > 0 \qquad \text{for } x \in \Omega, \end{aligned} \tag{3.3}$$

where Ω is a smooth bounded domain, $\alpha \geq 0$ real and $m > 1$ an integer. Let (λ_1, u_1) be the principal eigenpair of

$$-\Delta u = \lambda u, \qquad u|_{\partial\Omega} = 0,$$

ETNA
Kent State University
etna@mcs.kent.edu

satisfying $u_1(x) > 0$ in Ω and $||u_1||_{L^1(\Omega)} = 1$. Let v_1 be a real smooth function satisfying $v_1^m(x) = u_1(x)$ in Ω. Then $\theta v_1,\ \theta > 0$, is a steady-state solution of (3.3) for $\alpha = \lambda_1$. A steady-state solution for all α is $v(x,t) \equiv 0$. The time-dependent solutions of (3.3) for given initial function $v_0(x) > 0$ were investigated by Sacks and others, and the results are [10]:

(i) If $0 \le \alpha < \lambda_1$ and $v_0 \in L^q(\Omega),\ q > 1$, problem (3.3) has a solution existing for all times and decaying to zero for $t \to \infty$.

(ii) If $\alpha = \lambda_1$ and $v_0 \in L^q(\Omega),\ q > 1$, problem (3.3) has a solution existing for all times and tending to $\theta v_1 = \theta u_1^{1/m}$ for $t \to \infty$. The factor θ depends on the initial function v_0.

(iii) If $\alpha > \lambda_1$ there exists $T > 0$ such that problem (3.3) has for given v_0 a unique weak solution v on $[0,T]$ with $\lim_{t\to T_-} ||v(\cdot,t)||_{L^\infty(\Omega)} = +\infty$. Such solutions are called *blow-up solutions*. The only nonnegative solution of problem (3.3) which exists for all times is $v(x,t) \equiv 0$.

To construct a numerical scheme whose solution has similar properties as the solution of the continuous problem, Le Roux [10] used the exact scheme (1.3) for

$$\frac{dw^m}{dt} = \beta (w^m)^2, \qquad w(0) = w_0 > 0,\ \beta \in I\!R,$$

to derive the approximate semi-discrete scheme

$$\frac{1}{m-1}\left(V_n^{1-m}V_{n+1}^m - V_{n+1}\right) - \tau\Delta V_{n+1}^m = \alpha\tau V_{n+1}^m \tag{3.4}$$

for eq. (3.3). Here τ is the time step and $V_n = V(x, n\tau)$ approximates $v(x,t)$ at $t = n\tau$. Note that solving eq. (3.4) for V_{n+1} with given V_n means solving a quasilinear elliptic boundary value problem with $V_{n+1|\partial\Omega} = 0$, and this has to be done at each time step. With

$$p := \frac{1}{m}, \quad q := 1 - p, \qquad U_n := V_n^m$$

and $U_n \in I\!B := H_0^1(\Omega) \cap C^2(\bar\Omega)$ where all elements satisfy the given boundary condition, (3.4) becomes

$$\begin{aligned} \tau\Delta U_{n+1} &= \frac{p}{q}(U_{n+1}U_n^{-q} - U_{n+1}^p) - \alpha\tau U_{n+1} \\ &=: f(U_{n+1}; U_n), \end{aligned} \tag{3.5}$$

which is a standard quasilinear elliptic problem for U_{n+1}. Le Roux [10] proved existence and uniqueness of the solution of scheme (3.5) for

$$t_n = n\tau < T_1 := \frac{p}{\alpha q}||U_0||_\infty^{-q} = \frac{1}{\alpha(m-1)}||v_0^m||_\infty^{(1-m)/m},$$

and formulated conditions on Y_0 under which the iterative scheme

$$\Delta Y_{j+1} = \frac{1}{\tau} f(Y_j; U_n), \qquad Y_0 \text{ given },$$

converges for $j \to \infty$ (monotonic) to U_{n+1}. She proved stability and convergence of the time discretization for a wide class of initial conditions, gave for fixed $\tau = \Delta t$ the estimates

$$||U_n||_\infty \le \begin{cases} c t_n^{-1/q} & \text{if } \alpha < \lambda_1 \\ c(t_n^{-1/q} + ||U_0||_{p+1}) & \text{if } \alpha = \lambda_1 \end{cases}, \tag{3.6}$$

where c is a constant, $c = c(\Omega, p, \alpha, U_0)$, and showed:

(i) If $\alpha \leq \lambda_1$, then there exists a constant $\Delta t_0 > 0$ depending only on Ω, p, α, U_0 such that the numerical solution U_n exists for all $n \to \infty$ for every time step $\Delta t < \Delta t_0$. From the estimate (3.6) it then follows that $||U_n|| \to 0$ if $\alpha < \lambda_1$, as desired. We also see that the norm of the initial function U_0 specifies the numerical value of θ in the case $\alpha = \lambda_1$.

(ii) If $\alpha > \lambda_1$ then there exists T^* depending on the time step Δt and on U_0 such that the numerical solution U_n exists for $n\Delta t < T^*$ and becomes infinite at T^*. The following estimate is valid:

$$||U_n||_{p+1} \leq \left(\frac{T^*}{T^* - t_n}\right)^{1/q} ||U_0||_{p+1},$$

and this estimate has also been obtained for the exact solution.

Thus we see that, for sufficiently small $\Delta t < \Delta t_0$, scheme (3.5) produces qualitatively correct numerical solutions which satisfy the estimates known for the exact solutions.

In further work this scheme and its mathematical analysis were extended to the more general case

$$\begin{aligned} v_t - \Delta v^{1+\delta} = \alpha v^p \qquad & \text{for } x \in \Omega \subset I\!R^d,\ t > 0, \\ v(x,t) = 0 \qquad & \text{for } x \in \partial\Omega,\ t > 0, \\ v(x,0) = v_0(x) > 0 \qquad & \text{for } x \in \Omega, \end{aligned}$$

where $\Omega \subset I\!R^d$ is a smooth bounded domain, δ is a parameter describing diffusion, $\delta \in (-1, 0)$ for fast diffusion and $\delta > 0$ for slow diffusion, $\alpha \geq 0$ real and $p \geq 1 + \delta$. This mathematical work is reviewed in [9].

The usefulness of scheme (3.5) for Computational Plasma Physics is reported. In investigations of fusion plasmas, diffusion equations with slow diffusion (e.g. $\delta = 2$) are used for the density of particles, and equations with fast diffusion (e.g. $\delta = -1/2$) for their temperature. In reference [11] a coupled system for density and temperature of ions is solved for various parameter values, while in reference [12] the two equations are solved separately for various cases (decay of the solution, evolution to a constant profile, blow-up case).

REFERENCES

[1] G. DENK, *A new numerical method for the integration of highly oscillatory second-order ordinary differential equations*, Appl. Num. Math., 13 (1993), pp. 57–67.

[2] G. DENK, *The simulation of oscillatory circuits: an efficient integration scheme*, in Efficient Numerical Methods in Electronic Circuit Simulation, G. Denk, M. Günter, P. A. Selting, and O. V. Stryk, eds., Technische Universität München, TUM-M9413, 1994, pp. 1–8.

[3] K. DEKKER AND J. G. VERWER, *Stability of Runge-Kutta Methods for Stiff Nonlinear Differential Equations*, North Holland, Amsterdam, 1984.

[4] M. J. GANDER AND R. MEYER-SPASCHE, *An introduction to numerical integrators preserving physical properties*, in Applications of Nonstandard Finite Difference Schemes, R. E. Mickens, ed., Chapter 5, World Scientific, Singapore, 2000.

[5] J. HERSCH, *Contribution à la méthode des équations aux différences*, Z. Angew. Math. Phys., 2 (1958), pp. 129–180.

[6] E. HAIRER, S.P. NØRSETT, AND G. WANNER, *Solving Ordinary Differential Equations*, Vol. I, 2nd ed., Springer Verlag, Berlin, 1993.

[7] E. HAIRER AND G. WANNER, *Solving Ordinary Differential Equations*, Vol. II, Springer Verlag, Berlin, 1991.

[8] H. V. KOJOUHAROV AND B. M. CHEN, *Nonstandard Methods for AdvectionDiffusionReaction Equations*, in Applications of Nonstandard Finite Difference Schemes, R. E. Mickens, ed., Chapter 2, World Scientific, Singapore, 2000.

[9] M. N. LE ROUX, *Numerical solution of fast-diffusion or slow-diffusion equations*, J. Comput. Appl. Math., 97 (1998), pp. 121–136.

ETNA
Kent State University
etna@mcs.kent.edu

[10] M. N. LE ROUX, *Semidiscretization in time of nonlinear parabolic equations with blowup of the solution*, SIAM J. Numer. Anal., 31 (1994), pp. 170–195.

[11] M. N. LE ROUX, J. WEILAND, AND H. WILHELMSSON, *Simulation of a coupled dynamic system of temperature and density in a fusion plasma*, Phys. Scripta, 46 (1992), pp. 457–462.

[12] M. N. LE ROUX AND H. WILHELMSSON, *External boundary effects on simultaneous diffusion and reaction processes*, Phys. Scripta, 40 (1989), pp. 674–681.

[13] R. MEYER-SPASCHE, *Variable-coefficient difference schemes for quasilinear evolution problems*, in Numerical Methods and Applications, I. Dimov et al., eds., Lectrure Notes in Computer Sciences, Vol. 2542, Springer-Verlag, 2003, pp. 36–47.

[14] R. MEYER-SPASCHE, *Difference schemes of optimum degree of implicitness for a family of simple ODEs with blow-up solutions*, J. Comput. Appl. Math., 97 (1998), pp. 137–152.

[15] R. MEYER-SPASCHE AND D. DÜCHS, *A general method for obtaining unconventional and nonstandard difference schemes*, Dyn. Cont., Discr. Impulsive Systems, 3 (1997), pp. 453–467.

[16] R.E. MICKENS, ed., *Applications of Nonstandard Finite Difference Schemes*, World Scientific, Singapore, 2000.

[17] K. OZAWA, *Functional fitting Runge-Kutta method with variable coefficients*, Jap. J. Ind. Appl. Math., 18 (2001), pp. 107–130.

[18] G. VANDEN BERGHE, H. DE MEYER, M. VAN DAELE, AND T. VAN HECKE, *Exponentially-fitted explicit Runge-Kutta methods*, Comp. Phys. Comm., 123 (1999), pp. 7–15.

[19] G. VANDEN BERGHE, L. GR. IXARU, AND H. DE MEYER, *Frequency determination and step-length control for exponentially-fitted Runge-Kutta methods*, J. Comput. Appl. Math., 132 (2001), pp. 95–105.