Acta Numerica (2006), pp. 157–256
doi: 10.1017/S0962492906220014
© Cambridge University Press, 2006
Printed in the United Kingdom

General linear methods

J. C. Butcher
Department of Mathematics,
The University of Auckland, Auckland,
New Zealand
E-mail: butcher@math.auckland.ac.nz

General linear methods, as multistage multivalue methods, are the natural generalizations of linear multistep and Runge–Kutta methods. This survey contains a discussion of the traditional methods and a motivation for the general linear type of generalization. The new methods are introduced in terms of their formulation and the basic properties of consistency, stability and convergence. The order of general linear methods has to be looked at from a new point of view and it is shown how to use an algebraic structure (equivalent to B-series) to express conditions for a given order. Linear and non-linear stability for the new methods brings the theories for the classical methods into a comprehensive formulation and known results are outlined. Recently a number of subfamilies have been introduced and some of these are considered in detail. This applies in particular to methods with the property known as 'inherent Runge–Kutta stability'. These seem to have prospects of yielding useful and efficient methods, and some progress towards their practical implementation is outlined. Finally, the relationship between stability functions and order of methods is discussed in a setting wide enough to include general linear methods as well as multiderivative methods, such as Obreshkov methods. The classical barriers due to Ehle, Daniel–Moore and Dahlquist (second barrier) all fit into a common pattern and these are explored in a general setting.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

CONTENTS

1 Introduction 158
2 Motivations for general linear methods 183
3 Formulations 189
4 Order conditions 194
5 Linear and non-linear stability 202
6 Special families of methods 211
7 Methods with inherent RK-stability 220
8 Order and stability barriers 232
9 Conclusions and inconclusions 247
References 248

1. Introduction

The history of numerical methods for initial value problems up to 1965 was the history of Runge–Kutta methods and linear multistep methods. These seem to have been completely separate developments with the only meeting point being the existence of several low-order methods which simultaneously lie in each of the special classes.

Against this background, it must be asked why a more general type of method should be considered at all. Two reasons are proposed. First, the general linear method formulation is often the most natural framework for analysing the properties, even of traditional methods. Secondly, it is possible that new and potentially superior methods will arise, which could not possibly have been found as developments based on classical methods.

In this introductory section we will review the traditional methods, Euler, linear multistep and Runge–Kutta. Following this section, we will discuss some of the motivations for looking towards a more general type of method. In Section 3, we will consider the formulation of general linear methods and this is followed by a consideration of the meaning and significance of the order of a method. In the short Section 5 we will review the theories of linear and non-linear stability; a theme for this section is that the general linear method formulation is, for many questions, the most natural formulation, even in the case of classical linear multistep methods. The following two sections deal with some known new classes of methods, with a special emphasis on methods possessing the *inherent Runge–Kutta stability* (IRKS) structure. Finally, in Section 8, we study the interrelation between order and stability in the context of multivalue-multistage stability functions.

The bibliography is intended to be wider than references to the publications actually cited in this paper. There is no claim that it includes all work relevant to the development of general linear methods, but it is a start in this direction.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

1.1. Initial value problems

The standard initial value problem is written in the form

$$y'(x) = f\big(x, y(x)\big), \qquad y(x_0) = y_0,$$

where $f : \mathbb{R} \times \mathbb{R}^N \to \mathbb{R}^N$, although it will sometimes be more convenient to use an autonomous form of this problem,

$$y'(x) = f\big(y(x)\big), \qquad y(x_0) = y_0, \tag{1.1}$$

where $f : \mathbb{R}^N \to \mathbb{R}^N$. The individual components will sometimes need to be written out in full:

$$y_i'(x) = f_i\big(y_1(x), y_2(x), \ldots, y_N(x)\big), \qquad y_i(x_0) = y_{0i}, \quad i = 1, 2, \ldots, N,$$

where $y_{01}, y_{02}, \ldots, y_{0N}$ are the components of y_0

Even though many practical problems are conveniently presented in non-autonomous form, it is a simple matter to rewrite these problems as an autonomous system, possibly with N increased to $N + 1$. For example, if

$$y_i'(x) = f_i\big(x, y_1(x), y_2(x), \ldots, y_N(x)\big), \qquad y_i(x_0) = y_{0i}, \quad i = 1, 2, \ldots, N,$$

then an equivalent autonomous system would be

$$\overline{y}_i'(x) = \overline{f}_i\big(\overline{y}_0(x), \overline{y}_1(x), \ldots, \overline{y}_N(x)\big), \qquad \overline{y}_i(x_0) = \overline{y}_{0i}, \quad i = 0, 1, 2, \ldots, N,$$

where

$$\begin{aligned} \overline{f}_0\big(\overline{y}_0(x), \overline{y}_1(x), \ldots, \overline{y}_N(x)\big) &= 1, \\ \overline{f}_i\big(\overline{y}_0(x), \overline{y}_1(x), \ldots, \overline{y}_N(x)\big) &= f_i\big(\overline{y}_0(x), \overline{y}_1(x), \ldots, \overline{y}_N(x)\big), \\ \overline{y}_{0i} &= \begin{cases} x_0, & i = 0, \\ y_{0i}, & i = 1, 2, \ldots, N. \end{cases} \end{aligned}$$

The autonomous form has significant advantages in that the theory of Runge–Kutta methods is much simpler with this formulation.

It is often convenient to consider an integrated form of the basic initial value problem, that is,

$$y(x) = y_0 + \int_{x_0}^{x} f(x, y(x))\,\mathrm{d}x,$$

so that the process of numerical solution consists in approximating the integral appearing in this formulation.

1.2. The Euler method

If an approximation to the solution to an initial value problem is known at $x = x_{n-1}$, then the solution at $x_n = x_{n-1} + h$ can be written as

$$y(x_n) = y(x_{n-1}) + \int_{x_{n-1}}^{x_n} f(x, y(x))\,\mathrm{d}x, \tag{1.2}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

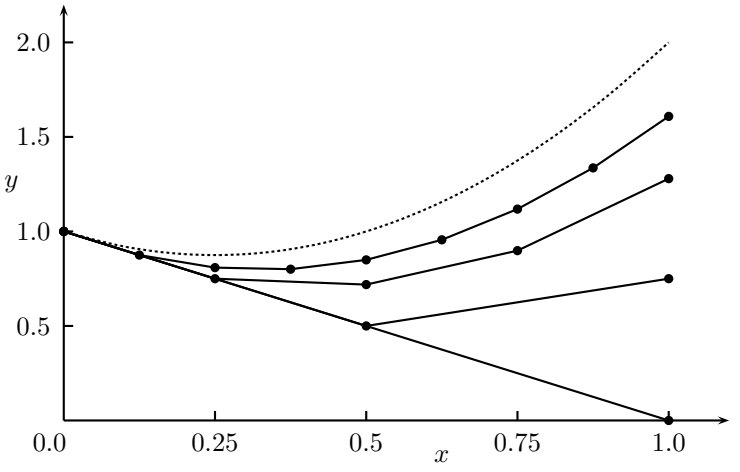


Figure 1.1. Euler method with various step-sizes and exact solution (·····).

where $y(x)$ is the trajectory defined by the initial data $y(x_{n-1}) = y_{n-1}$. Approximate the integral by the left Riemann sum

$$\int_a^b \phi(x)\,\mathrm{d}x \approx (b-a)\phi(a), \tag{1.3}$$

and we have the approximation

$$y(x_n) \approx y_{n-1} + hf(x_{n-1}, y_{n-1}).$$

Hence we obtain the basic form of the Euler method,

$$y_n = y_{n-1} + hf(x_{n-1}, y_{n-1}).$$

Because of its simplicity, the Euler method is a suitable prototype for discussing a range of questions which can also be asked about more complicated methods. Central to these considerations is the question as to when we can rely on a numerical scheme to provide arbitrarily accurate approximations, provided that sufficient computational effort is extended. This is the question of convergence. There are aspects of stability also associated with the Euler method which provide insights into corresponding questions for more general methods.

Discussion of convergence

In the computation shown in Figure 1.1, the problem

$$y'(x) = y - 2 + 5x - 2x^2, \qquad y(0) = 1,$$

is solved by the Euler method on the interval $[0, 1]$ using n steps with step-size $h = 1/n$, for $n = 1, 2, 4, 8$. It is seen that the approximations for $y(1)$ become steadily more accurate as n increases. This phenomenon is known as

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

'convergence' and is a necessary property for a numerical method to possess if it is to be used in practical computation. The precise definition and criteria for convergence is best subsumed under the corresponding theory for linear multistep methods which will be informally discussed in Section 1.4. In keeping with the intentions of this paper, we will in turn regard the linear multistep convergence theory as included in the general linear method formulation in Section 3.1.

The implicit Euler method

If instead of the left Riemann sum approximation (1.3) to (1.2), we use the *right* Riemann sum,

$$\int_a^b \phi(x)\,\mathrm{d}x \approx (b-a)\phi(b),$$

we arrive at the numerical method

$$y_n = y_{n-1} + hf(x_n, y_n).$$

This is *implicit* because y_n is not given by an explicit formula but is defined as the solution to this algebraic equation.

1.3. Stability of the Euler and implicit Euler methods

The justification for linear stability analysis is argued along the following lines. Consider an autonomous differential equation system

$$y'(x) = f(y(x)), \tag{1.4}$$

and ask how the introduction of a perturbation into the solution carries through to later times. This perturbation may be thought of as the result of computational errors caused by the inaccuracy of a numerical method, or simply as an imprecision in the initial data for the problem. Suppose the perturbation is expressed as a function $\eta(x)$ and that we can assume that η takes on sufficiently small values for the approximation

$$f(y(x) + \eta(x)) \approx f(y(x)) + f'(y(x))\eta(x)$$

to be realistic. If $y(x) + \eta(x)$ is supposed to satisfy the original differential equation system (1.4), then the development of $\eta(x)$ as time passes is approximately as the solution to the problem

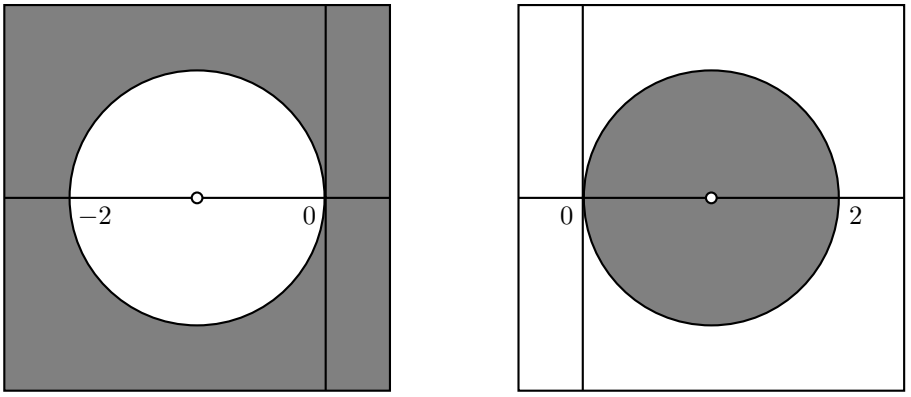
$$\eta'(x) = f'(y(x))\eta(x).$$

If the Jacobian matrix $f'(y(x))$ has an eigenvalue q, assumed to be approximately constant over a (possibly small) range of x values, then we are faced with the need to consider the linear differential equation

$$Y'(x) = qY(x), \tag{1.5}$$

as representing some aspect of the behaviour of the perturbation.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press



Euler method Implicit Euler method

Figure 1.2. Stability regions of Euler and implicit Euler methods.

A differential equation in which q can be large, with negative real part, based on a time-scale that might seem appropriate to the numerical modelling of (1.4) may cause serious difficulties, and is described as 'stiff'. The exact solution of (1.5) is a multiple of $\exp(qx)$ and dies away as x increases. However, the result computed by the Euler method, increases in magnitude by a factor $|1 + hq|$ in each time-step. Write $z = hq$ and refer to $1 + z$ as the stability function for the Euler method. The value of h that will permit stable computations to be carried out by the Euler method must be such that $|1 + z| \leq 1$. On the other hand, if it is somehow possible to compute a sequence of numerical approximations using the implicit Euler method, a similar analysis leads to the formula $(1 - z)^{-1}$ for the stability function and the set of points satisfying $|1 - z| \geq 1$ for the stability region.

The stability regions for the Euler and implicit Euler methods are shown in Figure 1.2. From this figure, we see that solving stiff problems is likely to be more successful with the implicit Euler than with the Euler method itself. Generalizations of these methods will need to take this phenomenon into account.

A method, such as the implicit Euler method, for which the stability region includes the left half-plane is said to be 'A-stable'. We will return to this concept in the context of Runge–Kutta, linear multistep and, of course, general linear methods.

Because of its good stability properties in handling linear stiff problems, it might be possible to ask how the implicit Euler methods might be expected to cope with non-linear stiff problems. We consider a test problem introduced in Dahlquist (1976) in his study of linear multistep methods. In this model problem the function $f(x, y)$ is assumed to satisfy the constraint

$$\langle u - v, f(x, u) - f(x, v)\rangle \leq 0, \qquad x \in \mathbb{R}, \quad u, v \in \mathbb{R}^N.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

The significance of this assumption is that two particular solutions, y and $\widehat{y}$, cannot drift apart, because

$$\frac{\mathrm{d}}{\mathrm{d}x}\|y(x)-\widehat{y}(x)\|^2 = 2\langle y(x)-\widehat{y}(x), f(x,y(x))-f(x,\widehat{y}(x))\rangle \leq 0.$$

A corresponding property for two numerical approximations y_n and $\widehat{y}_n$ would be that

$$\|y_n-\widehat{y}_n\|^2 \leq \|y_{n-1}-\widehat{y}_{n-1}\|^2,$$

and this is actually the case for the implicit Euler method because

$$\begin{aligned}\|y_n-\widehat{y}_n\|^2 - \|y_{n-1}-\widehat{y}_{n-1}\|^2 + \|(y_n-\widehat{y}_n)-(y_{n-1}-\widehat{y}_{n-1})\|^2 &\\ = 2\langle y_n-\widehat{y}_n, (y_n-\widehat{y}_n)-(y_{n-1}-\widehat{y}_{n-1})\rangle &\\ = 2\langle y_n-\widehat{y}_n, hf(x_n,y_n)-hf(x_n,\widehat{y}_n)\rangle &\\ \leq 0. &\end{aligned}$$

This model problem is the basis for separate studies of non-linear stability for Runge–Kutta methods, as well as for linear multistep methods. We will return to this question in Section 5, but in the more comprehensive context of general linear methods.

1.4. Linear multistep methods

Given existing approximations $y_i \approx y(x_{n-i})$, $f_i \approx y'(x_{n-i})$, $i = 1, 2, \ldots, k$, a linear k-step method is an algorithm for computing y_n and f_n so that $f_n = f(x_n, y_n)$ and

$$\sum_{i=0}^{k} \alpha_i y_{n-i} = h \sum_{i=0}^{k} \beta_i f_{n-i}. \tag{1.6}$$

In this formulation, $\alpha_0 \neq 0$, because we will want to compute the new approximation value y_n from (1.6). It is possible to rescale by multiplying (1.6) by an arbitrary nonzero factor, so that we could always assume for convenience that $\alpha_0 = 1$. However, different normalizations often lead to simplifications and we will keep the scaling of α_0 open.

Introduction of characteristic polynomials

Following the fundamental ideas of Dahlquist (1956), we introduce polynomials

$$\rho(w) = \sum_{i=0}^{k} \alpha_i w^{k-i}, \tag{1.7}$$

$$\sigma(w) = \sum_{i=0}^{k} \beta_i w^{k-i}. \tag{1.8}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

It is customary to identify the polynomial pair (ρ, σ) with the linear multistep method it represents. The polynomials ρ and σ will be assumed to have no common polynomial factor because if $\rho = \phi\widehat{\rho}$ and $\sigma = \phi\widehat{\sigma}$, where ϕ has nonzero degree $k - \widehat{k}$, then numerical results computed using $(\widehat{\rho}, \widehat{\sigma})$ would also satisfy results computed using (ρ, σ). In particular we can assume that α_k and β_k are not both zero.

Consistency, stability and convergence

A linear multistep method (ρ, σ) is said to be consistent if

$$\rho(1) = 0, \tag{1.9}$$

$$\rho'(1) = \sigma(1). \tag{1.10}$$

The significance of these assumptions is that for a consistent method, the method is able to solve any problem of the form $y'(x) = 1$ exactly over a single step, assuming that exact values of previous step values are used.

The method (ρ, σ) is said to be stable if ρ has all its zeros in the closed unit disc and repeated zeros are in the open unit disc. The significance of this assumption is that the method can not only solve problems of the form $y'(x) = 0$ exactly over many steps but it can do so even with slightly perturbed initial data. This leads to a main theorem in Dahlquist (1956) which relates convergence of a method to the method being both consistent and stable. A precise definition of convergence and further details can be found in standard textbooks. We will return to these ideas again, in the context of general linear methods, in Section 3.1.

Order of methods

A method has order p if it is capable of solving any differential equation exactly if its solution is a polynomial of degree not exceeding p. Put another way, this means that if the expression

$$\sum_{i=0}^{k} \alpha_i y_{n-i} - h \sum_{i=0}^{k} \beta_i f_{n-i}$$

is evaluated, with all y and f values replaced by the quantities they are supposed to approximate, then its formal Taylor series vanishes up to and including terms in h^p. Evaluate this series, expanding about x_{n-k}, and we obtain an expression of the form

$$\sum_{i=0}^{k} \alpha_i y(x_{n-i}) - h \sum_{i=0}^{k} \beta_i y'(x_{n-i}) = \sum_{i=0}^{\infty} C_i h^i y^{(i)}(x_{n-k}). \tag{1.11}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

From the known Taylor expansions of $y(x_{n-i})$ and $y'(x_{n-i})$ we find the following formulae for $C_0, C_1, \ldots$:

$$C_0 = \sum_{j=0}^{k} \alpha_j, \tag{1.12}$$

$$C_1 = \sum_{j=0}^{k-1} (k-j)\alpha_j - \sum_{j=0}^{k} \beta_j \tag{1.13}$$

$$C_i = \frac{1}{i!}\left(\sum_{j=0}^{k-1} (k-j)^i \alpha_j - i \sum_{j=0}^{k-1} (k-j)^{i-1} \beta_j \right), \quad i = 2, 3, \ldots. \tag{1.14}$$

Evaluate (1.11) for the special case $y(x) = \exp(z(x - x_{n-k})/h)$ (so that $y'(x) = (z/h)\exp(z(x-x_{n-k})/h)$), where z is an arbitrary complex number, and we find

$$\sum_{i=0}^{k} \alpha_i \exp((k-i)z) - z \sum_{i=0}^{k} \beta_i \exp((k-i)z) = C_0 + C_1 z + C_2 z^2 + \cdots,$$

so that

$$\rho(\exp(z)) - z\sigma(\exp(z)) = C_0 + C_1 z + C_2 z^2 + \cdots.$$

This enables us to state a convenient criterion for order.

Theorem 1.1. A linear multistep method (ρ, σ) has order p if and only if

$$\rho(\exp(z)) - z\sigma(\exp(z)) = \mathcal{O}(z^{p+1}). \tag{1.15}$$

By substituting $\log(1+z) = z - \frac{1}{2}z^2 + \frac{1}{3}z^3 - \cdots$ for z in (1.15) and rearranging we find

$$\frac{1}{\log(1+z)}\rho(1+z) - \sigma(1+z) = \mathcal{O}(z^p),$$

where, because of consistency, $\rho(1+z)$ is a multiple of z. Hence, we have the following result.

Corollary 1.2. A linear multistep method (ρ, σ) has order p if and only if

$$\frac{1}{\log(1+z)/z}\frac{\rho(1+z)}{z} - \sigma(1+z) = \mathcal{O}(z^p). \tag{1.16}$$

Adams and BDF methods

By writing $\rho(z) = z^k - z^{k-1}$ and defining σ to be the degree $k-1$ polynomial satisfying (1.16), with $p = k$, Adams–Bashforth methods are derived. By increasing the degree of σ to k and p to $k+1$, Adams–Moulton methods are found.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.1. Adams methods.

	Adams–Bashforth				Adams–Moulton				
k	β_1	β_2	β_3	β_4	β_0	β_1	β_2	β_3	β_4
1	1				$\frac{1}{2}$	$\frac{1}{2}$			
2	$\frac{3}{2}$	$-\frac{1}{2}$			$\frac{5}{12}$	$\frac{2}{3}$	$-\frac{1}{12}$		
3	$\frac{23}{12}$	$-\frac{4}{3}$	$\frac{5}{12}$		$\frac{3}{8}$	$\frac{19}{24}$	$-\frac{5}{24}$	$\frac{1}{24}$	
4	$\frac{55}{24}$	$-\frac{59}{24}$	$\frac{37}{24}$	$-\frac{3}{8}$	$\frac{251}{720}$	$\frac{323}{360}$	$-\frac{11}{30}$	$\frac{53}{360}$	$-\frac{19}{720}$

Table 1.2. BDF methods.

k	β_0	α_1	α_2	α_3	α_4
1	1	1			
2	$\frac{2}{3}$	$\frac{4}{3}$	$-\frac{1}{3}$		
3	$\frac{6}{11}$	$\frac{18}{11}$	$-\frac{9}{11}$	$\frac{2}{11}$	
4	$\frac{12}{25}$	$\frac{48}{25}$	$-\frac{36}{25}$	$\frac{16}{25}$	$-\frac{3}{25}$

The series for $z/\log(1+z)$, occurring in (1.16) is

$$\frac{z}{\log(1+z)} = 1+\tfrac{1}{2}z-\tfrac{1}{12}z^2+\tfrac{1}{24}z^3-\tfrac{19}{720}z^4+\tfrac{3}{160}z^5-\tfrac{863}{60480}z^6+\tfrac{275}{24192}z^7-\tfrac{33953}{3628800}z^8+\cdots,$$

and we readily find the first few Adams–Bashforth (order k) and Adams–Moulton (order $k+1$) methods as shown in Table 1.1. In this table the values of β_0 (for the AM method only), β_1, β_2, …, β_k, are given, assuming the scaling $\alpha_0 = -\alpha_1 = 1$.

The 'backward difference formulae' (BDF) are approximations to the derivative of $y(x)$ at x_n in terms of the values of $y(x_{n-i})$, $i = 0, 1, \ldots, k$. The corresponding linear multistep methods are referred to as BDF methods. To derive such methods of order $p = k$, write $\sigma(z) = z^k$, and find ρ of degree k from

$$\rho(1+z) = \log(1+z)(1+z)^k + \mathcal{O}(z^{p+1}).$$

For the first few BDF methods, the coefficients are shown in Table 1.2, scaled so that $\alpha_0 = 1$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Stability regions and A-stability

For a linear differential equation $y' = qy$, the difference equation (1.6) simplifies to

$$\sum_{i=0}^{k} (\alpha_i - z\beta_i) y_{n-i} = 0,$$

where here $z = hq$. The stability region is the set of points in the complex plane for which this difference equation has only bounded solution sequences. This means that, for z in the stability region,

$$\rho(w) - z\sigma(w),$$

regarded as a polynomial in w, has all its zeros in the closed unit disc and has all repeated zeros in the interior. Generalizing the discussion in Section 1.3, we will refer to a method as being A-stable if the stability region includes the left half-plane.

Dahlquist barriers

The barriers of Dahlquist (1956, 1963) state fundamental limitations on achievable order for linear multistep methods which satisfy various stability properties. The second barrier is concerned with A-stability and we will discuss this in Section 8. The first barrier is stated in the following result.

Theorem 1.3. The order p of a stable k-step method is bounded by

$$p \leq \begin{cases} k+1, & k \text{ odd}, \\ k+2, & k \text{ even}. \end{cases}$$

Proof. Substitute $\log(\frac{1+z}{1-z})$ for z in (1.15), to obtain

$$\rho\left(\frac{1+z}{1-z}\right) - \log\left(\frac{1+z}{1-z}\right)\sigma\left(\frac{1+z}{1-z}\right) = \mathcal{O}(z^{p+1}), \tag{1.17}$$

Let

$$r(z) = (1-z)^k \rho\left(\frac{1+z}{1-z}\right) = \sum_{i=0}^{k} a_i z^i, \tag{1.18}$$

$$s(z) = (1-z)^k \sigma\left(\frac{1+z}{1-z}\right) = \sum_{i=0}^{k} b_i z^i. \tag{1.19}$$

If z_0 is a zero of r, then $\rho((1+z_0)/(1-z_0)) = 0$. Hence, $(1+z_0)/(1-z_0)$ is in the closed unit disc, implying that z_0 is in the closed left half-plane. Because 1 is a zero of ρ, 0 is a zero of r and hence, $a_0 = 0$. However, because 1 is not a *repeated* zero of ρ, 0 is not a repeated zero of r. Hence, $a_1 \neq 0$. Without loss of generality, assume that $a_1 > 0$. Because all zeros of r are in the left half-plane, and because these are all real or exist in conjugate

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

pairs, r is a constant multiplied by products of the form $z+\alpha$, where $\alpha \geq 0$, or of the form $z^2+\alpha z+\beta$, where α and β are each nonnegative. Hence, no two coefficients in r can have opposite signs and therefore $a_i \geq 0$, for $i = 1, 2, \ldots, k$.

Multiply (1.17) by $(1-z)^k$ and divide by $\log(\frac{1+z}{1-z})$ and we find

$$(c_0 + c_2 z^2 + c_4 z^4 + \cdots)(a_1 + a_2 z + \cdots + a_k z^{k-1}) = b_0 + b_1 z + \cdots + b_k z^k + \mathcal{O}(z^p), \quad (1.20)$$

where

$$\big(c_0 + c_2 z^2 + c_4 z^4 + \cdots\big)\big(\tfrac{1}{z}\log\big(\tfrac{1+z}{1-z}\big)\big) = 1. \quad (1.21)$$

From the known Taylor series for $\log(\frac{1+z}{1-z})$, we see that $c_0 = \frac{1}{2}$ and $c_2 = -\frac{1}{6}$. We now prove by induction that $c_{2n} < 0$ for $n \geq 1$. From the z^{2n} and z^{2n-2} terms in the expansion of (1.21), we find

$$c_{2n} + \frac{1}{3}c_{2n-2} + \cdots + \frac{1}{2n+1}c_0 = 0, \quad (1.22)$$

$$c_{2n} + \frac{1}{3}c_{2n-4} + \cdots + \frac{1}{2n-1}c_0 = 0. \quad (1.23)$$

Multiply (1.22) by $2n+1$, subtract the result of multiplying (1.23) by $2n-1$, and rearrange to find c_{2n} as a positive linear combination of $c_2, \ldots, c_{2n-2}$, completing the proof that $c_{2n} < 0$ for all positive n.

We now need to prove that an order $p > k+1$ is impossible for k odd. If it were possible, then the coefficient of z^{k+1} in (1.20) would be zero. However, this equals

$$a_k c_2 + a_{k-2} c_4 + \cdots + a_1 c_{k+1},$$

which cannot be zero unless all the terms are zero; but this would imply $a_1 = 0$, which is impossible. If k is even and $p > k+2$, then the coefficient of z^{k+2} in (1.20) would be zero. This implies

$$a_{k-1} c_4 + a_{k-3} c_6 + \cdots + a_1 c_{k+2} = 0,$$

which again would lead to the impossible conclusion that $a_1 = 0$. □

1.5. Runge–Kutta methods

Formulation of methods

The well-known methods of Runge (1895), Heun (1900) and Kutta (1901), generalize the classical Euler method by allowing for additional functional evaluations in each time-step. Write $Y_1, Y_2, \ldots, Y_s$ for the arguments of these evaluations, and $F_1, F_2, \ldots, F_s$ for the corresponding derivative approximations. For explicit methods, as in the cited works, each Y_i is a linear combination of the hF_j values, for $j < i$ added on to the input

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

approximation, which we will write as y_{n-1} for step number n. Once the s stages and corresponding derivative approximations have been computed, the output value for the step is computed as a further linear combination of the hF_i values, for $i = 1, 2, \ldots, s$.

Putting all this together we formulate the method as

$$Y_i = y_{n-1} + h \sum_{j<i} a_{ij} F_j, \quad F_i = f(Y_i), \quad i = 1, 2, \ldots, s, \tag{1.24}$$

$$y_n = y_{n-1} + h \sum_{i=1}^{s} b_i F_i, \tag{1.25}$$

where the numbers a_{ij}, b_i are characteristic of a specific method.

Because we will also want to consider 'implicit' methods in which the sums in (1.24) extend beyond $j < i$, we will conventionally introduce a full matrix of a_{ij} coefficients where, for the explicit case we have so far considered, $a_{ij} = 0$ if $j \geq i$.

This formulation is for an autonomous problem. In the case of a non-autonomous problem, we need to take account of the point within the step to which each stage corresponds. Suppose that stage number i evaluates an approximation at $x_{n-1} + hc_i$. By considering the simple differential equation $y' = 1$ we can evaluate c_i as the sum of the elements in row number i of the a_{ij} table of coefficients. That is,

$$c_i = \sum_{j=i}^{s} a_{ij}. \tag{1.26}$$

The modification to (1.24) and (1.25) to handle the non-autonomous case is

$$Y_i = y_{n-1} + h \sum_{j<i} a_{ij} F_j, \quad F_i = f(x_{n-1} + hc_i, Y_i), \quad i = 1, 2, \ldots, s, \tag{1.27}$$

$$y_n = y_{n-1} + h \sum_{i=1}^{s} b_i F_i. \tag{1.28}$$

It is customary to write the collection of coefficients a_{ij}, b_i and c_i in a tableau thus:

$$\begin{array}{c|cccccc} 0 & & & & & & \\ c_2 & a_{21} & & & & & \\ c_3 & a_{31} & a_{32} & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ c_s & a_{s1} & a_{s2} & \cdots & a_{s,s-1} & & \\ \hline & b_1 & b_2 & \cdots & b_{s-1} & b_s \end{array},$$

where we note that $c_1 = 0$ and that we have omitted those elements of the a_{ij} array which are necessarily zero.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Sometimes we will need to introduce the full matrix of coefficients and we denote this by A. The two vectors b^T and c are also introduced. Later we will need to consider implicit methods and in this case A will be a full matrix. Using these notations, the tableau of coefficients will be written as

$$\begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

Order conditions

Order p linear multistep methods are constructed using approximations to $y(x_n)$ in terms of y evaluated at x_{n-i}, $i = 1, 2, \ldots, k$ and hy' evaluated at x_{n-i}, $i = 0, 1, \ldots, k$. To achieve the required order, the approximation must be exactly satisfied whenever y is a polynomial of degree less than p. If the values on which the approximations are based are themselves approximations found in previous steps, we can still interpret the current approximation as having order p, because the total error in $y(x_n)$ is made up from the truncation error in this approximation, together with inherited errors, all of which can be estimated in terms of $O(h^{p+1})$.

For Runge–Kutta methods, the situation is more complicated because even though the approximation (1.28) is based on the integral

$$y(x_n) = y(x_{n-1}) + h \int_0^1 y'(x_{n-1} + h\xi)\,\mathrm{d}\xi \approx y(x_{n-1}) + h \sum_{i=1}^{s} b_i y'(x_{n-1} + hc_i),$$

the values of F_i are not accurate approximations to $y'(x_{n-1} + hc_i)$, $i = 1, 2, \ldots, s$.

We deal with this complication by carrying out three steps. First we find the Taylor expansion of the exact solution; secondly we find the Taylor expansion for the approximation computed using a Runge–Kutta method. Finally, by comparing these two Taylor expansions term by term, we arrive at conditions for the difference between them to equal $O(h^{p+1})$.

To commence the first step of finding the formal Taylor expansion of y satisfying (1.1), we need formulae for the second, third, ..., derivatives for this function:

$$\begin{aligned} y'(x) &= f(y(x)), \\ y''(x) &= f'(y(x))y'(x) \\ &= f'(y(x))f(y(x)), \\ y'''(x) &= f''(y(x))(f(y(x)), y'(x)) + f'(y(x))f'(y(x))y'(x) \\ &= f''(y(x))(f(y(x)), f(y(x))) + f'(y(x))f'(y(x))f(y(x)). \end{aligned}$$

This sequence of expressions becomes increasingly complicated as we evaluate higher derivatives and we look for a systematic pattern.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.3. Tree-like structure of terms appearing in derivative formulae.

$y'(x) = \mathbf{f}$	
$y''(x) = \mathbf{f}'\mathbf{f}$	
$y'''(x) = \mathbf{f}''(\mathbf{f},\mathbf{f})$	
$+\,\mathbf{f}'\mathbf{f}'\mathbf{f}$	

Write $\mathbf{f} = f(y(x))$, $\mathbf{f}' = f'(y(x))$, $\mathbf{f}'' = f''(y(x)), \ldots$ and consider Table 1.3 where the expressions for y' and y'' and the two terms occurring in y''' are shown together with their tree-like structures.

Motivated by this structure, we introduce the set of all rooted trees and the corresponding derivative terms.

Trees and elementary differentials

Let T denote the set of rooted trees:

$$T = \left\{ \bullet, \quad \vdots, \quad \vee, \quad \vdots, \quad \curlyvee, \quad \vee\!\!\vdots, \quad \curlyvee, \quad \vdots, \quad \ldots \right\}. \tag{1.29}$$

It is convenient to introduce some notation, including a tree-building structure. We will usually omit 'rooted' and refer only to trees.

The tree $\bullet$ will be denoted by τ. Given trees $t_1, t_2, \ldots, t_m$ we consider the tree formed by joining the roots of each of these trees to a new root. This will be written as $[t_1 t_2 \ldots t_m]$. Furthermore the notation will be made more compact by denoting repeated trees within $[\cdot]$ using exponents. Repeated use of the $[\cdot]$ operation will be denoted using subscripts.

For example the first 8 trees in the sequence, listed in (1.29) are written in terms of this new notation as follows:

$$T = \left\{ \tau, [\tau], [\tau^2], [{}_2\tau]_2, [\tau^3], [\tau[\tau]], [{}_2\tau^2]_2, [{}_3\tau]_3, \ldots \right\}.$$

Trees of the form $[\tau^n]$ are sometimes referred to as 'bushy trees' and trees of the form $[{}_n\tau]_n$ as tall trees.

We can now define the elementary differentials.

Definition 1.4. The elementary differential associated with the tree t, the function f and the evaluation point y_0 is defined by

$$F(\tau)(y_0) = f(y_0),$$
$$F([t_1, t_2, \ldots, t_m])(y_0) = f^{(m)}(y_0)(F(t_1)(y_0), F(t_2)(y_0), \ldots, F(t_m)(y_0)).$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.4. Some functions on trees and an example.

Function	Name (and meaning)	Example	Construction
$r(t)$	order of t (number of vertices)	7	
$\sigma(t)$	symmetry of t (order of automorphism group)	8	
$\gamma(t)$	density of t	63	
$\alpha(t)$	(number of ways of labelling t with an ordered set)	10	$\frac{r(t)!}{\sigma(t)\gamma(t)}$
$\beta(t)$	(number of ways of labelling t with an unordered set)	630	$\frac{r(t)!}{\sigma(t)}$
$F(t)(y_0)$	elementary differential	$\mathbf{f}''\big(\mathbf{f}''(\mathbf{f},\mathbf{f}),\mathbf{f}'(\mathbf{f},\mathbf{f})\big)$	
$\Phi(t)$	elementary weight	$\sum_{i,j,k=1}^{s} b_i a_{ij} c_j^2 a_{ik} c_k^2$	

Functions on trees

The various functions we will need are summarized in Table 1.4, with a more detailed explanation available in Butcher (2003). In Table 1.4, t denotes a typical tree. Also given are examples of these functions based on the tree $t =$ , which, in terms of the notation we have introduced, can also be written as $[[\tau^2]^2]$.

The function $\Phi(t)$ will be explained below. The remaining functions are easy to compute up to order 4 trees and are shown in Table 1.5.

Taylor expansions and order conditions

The formal Taylor expansion of the solution at $x_0 + h$ is

$$y(x_0 + h) = y_0 + \sum_{t \in T} \frac{\alpha(t) h^{r(t)}}{r(t)!} F(t)(y_0).$$

Using the known formula for $\alpha(t)$, we can write this as

$$y(x_0 + h) = y_0 + \sum_{t \in T} \frac{h^{r(t)}}{\sigma(t)\gamma(t)} F(t)(y_0). \tag{1.30}$$

Our aim will now be to find a corresponding formula for the result computed by one step of a Runge–Kutta method. By comparing these formulae term by term, we will obtain conditions for a specific order of accuracy.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.5. Various functions on trees.

t	$r(t)$	$\sigma(t)$	$\gamma(t)$	$\alpha(t)$	$\beta(t)$	$F(t)$	$\Phi(t)$
	1	1	1	1	1	$\mathbf{f}$	$\sum b_i$
	2	1	2	1	2	$\mathbf{f}'\mathbf{f}$	$\sum b_i c_i$
	3	2	3	1	3	$\mathbf{f}''(\mathbf{f},\mathbf{f})$	$\sum b_i c_i^2$
	3	1	6	1	6	$\mathbf{f}'\mathbf{f}'\mathbf{f}$	$\sum b_i a_{ij} c_j$
	4	6	4	1	4	$\mathbf{f}^{(3)}(\mathbf{f},\mathbf{f},\mathbf{f})$	$\sum b_i c_i^3$
	4	1	8	3	24	$\mathbf{f}''(\mathbf{f},\mathbf{f}'\mathbf{f})$	$\sum b_i c_i a_{ij} c_j$
	4	2	12	1	12	$\mathbf{f}'\mathbf{f}''(\mathbf{f},\mathbf{f})$	$\sum b_i a_{ij} c_j^2$
	4	1	24	1	24	$\mathbf{f}'\mathbf{f}'\mathbf{f}'\mathbf{f}$	$\sum b_i a_{ij} a_{jk} c_k$

We need to evaluate various expressions, known as 'elementary weights', which depend on the tableau for a particular method. First we use the example tree we have already considered to illustrate the construction of the elementary weight $\Phi(t)$ for this tree t:

$t =$ 

The elementary weight for this tree is

$$\Phi(t) = \sum_{i,j,k,l,m,n,o=1}^{s} b_i a_{ij} a_{ik} a_{jl} a_{jm} a_{kn} a_{ko},$$

which can be simplified by summing over l, m, n, o and using (1.26):

$$\Phi(t) = \sum_{i,j,k=1}^{s} b_i a_{ij} c_j^2 a_{ik} c_k^2.$$

It is now possible to write down the formal Taylor expansion of the solution at $x_0 + h$ in the form

$$y_1 = y_0 + \sum_{t \in T} \frac{\beta(t) h^{r(t)}}{r(t)!} \Phi(t) F(t)(y_0).$$

Using the known formula for $\beta(t)$, this can be re-written as

$$y_1 = y_0 + \sum_{t \in T} \frac{h^{r(t)}}{\sigma(t)} \Phi(t) F(t)(y_0). \tag{1.31}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

If the Taylor series (1.31) is to match the Taylor series (1.30), up to h^p terms, we need to ensure that

$$\Phi(t) = \frac{1}{\gamma(t)},$$

for all trees such that

$$r(t) \leq p.$$

These are the 'order conditions'.

Low-order explicit methods

We will attempt to construct methods of order $p = s$ with s stages for $s = 1, 2, \ldots$. We will find that this is possible up to order 4 but not for $p \geq 5$. The usual approach will be to first choose $c_2, c_3, \ldots, c_s$ and then solve for $b_1, b_2, \ldots, b_s$. After this solve for those of the a_{ij} which can be found as solutions to linear equations.

Order 2. The order conditions are

$$b_1 + b_2 = 1,$$
$$b_2 c_2 = \tfrac{1}{2},$$

with solution, for arbitrary nonzero c_2,

$$\begin{array}{c|cc} 0 & & \\ c_2 & c_2 & \\ \hline & 1 - \frac{1}{2c_2} & \frac{1}{2c_2} \end{array}.$$

Choose $c_2 = \frac{1}{2}$ and $c_2 = 1$, respectively, and we obtain the two well-known special cases

$$\begin{array}{c|cc} 0 & & \\ \frac{1}{2} & \frac{1}{2} & \\ \hline & 0 & 1 \end{array}, \qquad \begin{array}{c|cc} 0 & & \\ 1 & 1 & \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

Order 3. The order conditions are

$$b_1 + b_2 + b_3 = 1,$$
$$b_2 c_2 + b_3 c_3 = \tfrac{1}{2},$$
$$b_2 c_2^2 + b_3 c_3^2 = \tfrac{1}{3},$$
$$b_3 a_{32} c_2 = \tfrac{1}{6}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Three representative special cases are

$$\begin{array}{c|ccc} 0 & & & \\ \frac{1}{2} & \frac{1}{2} & & \\ 1 & -1 & 2 & \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}, \qquad \begin{array}{c|ccc} 0 & & & \\ \frac{2}{3} & \frac{2}{3} & & \\ \frac{2}{3} & 0 & \frac{2}{3} & \\ \hline & \frac{1}{4} & \frac{3}{8} & \frac{3}{8} \end{array}, \qquad \begin{array}{c|ccc} 0 & & & \\ \frac{2}{3} & \frac{2}{3} & & \\ 0 & -1 & 1 & \\ \hline & 0 & \frac{3}{4} & \frac{1}{4} \end{array}.$$

Order 4. The order conditions are

$$b_1 + b_2 + b_3 + b_4 = 1, \tag{1.32}$$

$$b_2c_2 + b_3c_3 + b_4c_4 = \tfrac{1}{2}, \tag{1.33}$$

$$b_2c_2^2 + b_3c_3^2 + b_4c_4^2 = \tfrac{1}{3}, \tag{1.34}$$

$$b_3a_{32}c_2 + b_4a_{42}c_2 + b_4a_{43}c_3 = \tfrac{1}{6}, \tag{1.35}$$

$$b_2c_2^3 + b_3c_3^3 + b_4c_4^3 = \tfrac{1}{4}, \tag{1.36}$$

$$b_3c_3a_{32}c_2 + b_4c_4a_{42}c_2 + b_4c_4a_{43}c_3 = \tfrac{1}{8}, \tag{1.37}$$

$$b_3a_{32}c_2^2 + b_4a_{42}c_2^2 + b_4a_{43}c_3^2 = \tfrac{1}{12}, \tag{1.38}$$

$$b_4a_{43}a_{32}c_2 = \tfrac{1}{24}. \tag{1.39}$$

To solve these equations, treat c_2, c_3, c_4 as parameters, and solve for b_1, b_2, b_3, b_4 from (1.32), (1.33), (1.34), (1.36). Now solve for a_{32}, a_{42}, a_{43} from (1.35), (1.37) and (1.38). Finally, use (1.39) to obtain a consistency condition on c_2, c_3, c_4. This consistency condition is found to be $c_4 = 1$.

We will prove a stronger result in another way.

Theorem 1.5. If an explicit Runge–Kutta method with $s = 4$ has order 4, then

$$\sum_{i=j+1}^{s} b_i a_{ij} = b_j(1 - c_j), \qquad j = 1, 2, 3, 4$$

and, in particular, $c_4 = 1$.

Proof. The result $c_4 = 1$ is proved in Lemma 1.6 below. Hence, $v_4 = 0$, where $v_j = \sum_{i=j+1}^{s} b_i a_{ij} - b_j(1 - c_j)$, $j = 1, 2, 3, 4$. Also $v_3 = 0$ because $\sum_{j,k} v_j a_{jk} c_k = 0$, which we find by expanding and using the order conditions. Finally, $\sum_j v_j c_j = 0$, implying $v_2 = 0$, and $\sum_j v_j = 0$, implying $v_1 = 0$. □

Lemma 1.6. If an explicit Runge–Kutta method has order p where $s = p \geq 4$, then $c_4 = 1$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Proof. Consider the matrix formed as the result of the product

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_4 \end{bmatrix} \begin{bmatrix} b^TA^{p-3} \\ b^TA^{p-4}C \\ b^TA^{p-4} \end{bmatrix} \begin{bmatrix} Ac & Cc & c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -c_2 \end{bmatrix}.$$

This matrix has rank one because b^TA^{p-3} and $b^TA^{p-4}C-c_4b^TA^{p-4}$ are each zero except for components number $1,2,3$ and because Ac and $Cc-c_2c$ are each zero in components $1,2$. Hence, multiplying the middle two factors, we see that the product can be written as

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_4 \end{bmatrix} \begin{bmatrix} b^TA^{p-2}c & b^TA^{p-3}Cc & b^TA^{p-3}c \\ b^TA^{p-4}CAc & b^TA^{p-4}C^2c & b^TA^{p-4}Cc \\ b^TA^{p-3}c & b^TA^{p-4}Cc & b^TA^{p-4}c \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -c_2 \end{bmatrix}.$$

Evaluate the second factor by the order conditions and we obtain the result

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_4 \end{bmatrix} \begin{bmatrix} \frac{1}{p!} & \frac{2}{p!} & \frac{p}{p!} \\ \frac{3}{p!} & \frac{6}{p!} & \frac{2p}{p!} \\ \frac{p}{p!} & \frac{2p}{p!} & \frac{p(p-1)}{p!} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -c_2 \end{bmatrix}$$
$$= \frac{1}{p!} \begin{bmatrix} 1 & 2-pc_2 \\ 3-pc_4 & 6-2pc_2-2pc_4+p(p-1)c_2c_4 \end{bmatrix}.$$

Because this matrix has rank not exceeding 1, its determinant is zero. This gives the result $c_2(1-c_4)=0$. The possibility that $c_2=0$ has to be rejected because this would lead to the contradiction

$$0 = b^TA^{p-2}c = \frac{1}{p!}.$$

Hence, for any Runge–Kutta method with $s=p\geq 4$, c_4 necessarily equals 1. □

As a result of Theorem 1.5, the construction of fourth-order Runge–Kutta methods now becomes straightforward. Kutta (1901) classified all solutions to the fourth-order conditions.

In particular, we have the famous method:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}.$$

An order barrier

We will review what is achievable up to order 8. In Table 1.6, N_p is the number of order conditions to achieve this order. $M_s = s(s+1)/2$ is the

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.6. Minimum number of stages s to achieve order p.

p	N_p	s	M_s
1	1	1	1
2	2	2	3
3	4	3	6
4	8	4	10
5	17	6	21
6	37	7	28
7	115	9	45
8	200	11	66

number of free parameters to satisfy the order conditions for the required s stages.

According to Table 1.6, it is suggested that, for $p \geq 5$, it is necessary that $s > p$. We will now prove this result.

Theorem 1.7. There does not exist an explicit Runge–Kutta method with order $p = s \geq 5$.

Proof. Recall from Lemma 1.6 that $c_4 = 1$. If $s = p \geq 5$, repeat the argument but starting from the product

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -c_5 \end{bmatrix} \begin{bmatrix} b^T A^{p-4} \\ b^T A^{p-5} C \\ b^T A^{p-5} \end{bmatrix} \begin{bmatrix} A^2 c & ACc & Ac \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & -c_2 \end{bmatrix}$$

and we now find that $c_5 = 1$. If $s = p \geq 5$ and $c_4 = c_5 = 1$, we obtain the contradiction

$$0 = b^T A^{p-5}(I - C)A^2 c = \frac{1}{p!}. \qquad \square$$

1.6. Algebraic theory and B-series

We will review work first presented in Butcher (1972*a*); it has since become known under the name B-series (Hairer and Wanner 1974). A more recent account is given in Butcher (2003).

As a first step, make a slight generalization to the formulation of Runge–Kutta methods, by inserting a factor b_0 in the term y_{n-1} in (1.25). In such a generalized Runge–Kutta method, the extra coefficient can be conveniently inserted into its tableau:

$$\begin{array}{c|c} c & A \\ \hline b_0 & b^T \end{array}. \tag{1.40}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

We will conventionally add an additional 'empty tree', denoted by $\emptyset$, to the set of rooted trees to form the augmented set

$$T^{\#} = T \cup \{\emptyset\}.$$

For the generalized Runge–Kutta tableau (1.40), define the corresponding elementary weight as

$$\Phi(\emptyset) = b_0,$$

so that, for a standard Runge–Kutta method, $\Phi(\emptyset) = 1$.

We will denote by X the set of mappings $T^{\#} \to \mathbb{R}$ and by X_1 the subset for which $\emptyset \mapsto 1$. Thus, to each Runge–Kutta method, we can associate a member of X_1, such that $t \mapsto \Phi(t)$.

The order conditions can be written in terms of elementary weights, and it is natural to ask in what sense the elementary weights characterize a Runge–Kutta method. The answer is that if two Runge–Kutta methods have the same sequence of elementary weights, then they are equivalent methods in a very natural sense. For example, if two methods are equivalent then they give the same numerical result when applied to the same problem. Furthermore they are equivalent also in the sense that if unused stages are eliminated and sets of stages which give identical results are collapsed into a single stage, then the two methods have equivalent tableaux, except for the ordering of the stages. For a more detailed explanation of equivalences amongst Runge–Kutta methods, see, for example, Butcher (1996*b*).

Given that we can represent equivalence classes of Runge–Kutta methods using the sequence of elementary weights, we might ask: What is the significance of the right-hand sides of the order conditions? We will give an interpretation of these quantities in terms of a limiting type of Runge–Kutta method which can be thought of as having arbitrarily high order. For convenience we will consider a step of size h starting from $y(x_0) = y_0$.

The exact solution at the end of this step, and at points within the step, is given by the Picard integral equation

$$y(x_0 + h\xi) = y_0 + h \int_0^{\xi} f(y(x_0 + h\eta))\,\mathrm{d}\eta.$$

We can regard this as an idealized Runge–Kutta method in which the finite index set for the stages, $\{1, 2, \ldots, s\}$, is replaced by an interval $[0, 1]$. This means that for any $\xi \in [0, 1]$, we can associate a 'stage value', $Y_\xi = y(x_0 + h\xi)$, with corresponding stage derivative $f(Y_\xi)$. In this limiting interpretation, the matrix $A : \mathbb{R}^s \to \mathbb{R}^s$ is replaced by a linear operator on the set of continuous functions on $[0, 1]$. At the same time, the vector b^T, is replaced by a linear functional on the continuous functions on $[0, 1]$. More

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

specifically, the idealized A and b^T are given by

$$(A\phi)_\xi = \int_0^\xi \phi_\eta \, \mathrm{d}\eta, \quad (b^T\phi) = \int_0^1 \phi_\eta \, \mathrm{d}\eta = (A\phi)_1.$$

The elementary weights for this idealized method are found from the formula for $\Phi(t)$ and replacing various sums by integrals. For example, for the example tree used in Table 1.4, the calculation of the limiting elementary weight is as follows:

$$\int_0^1 \left(\int_0^x x^2 \, \mathrm{d}x \right)^2 \mathrm{d}x = \int_0^1 \left(\tfrac{1}{3} x^3 \right)^2 \mathrm{d}x = \tfrac{1}{63} = \tfrac{1}{\gamma(t)}.$$

The representation of this method as a member of X_1 will be denoted by E. Hence, $\mathsf{E}(t) = 1/\gamma(t)$.

If we consider equivalence classes of Runge–Kutta method as basic objects of study, then we might ask: What is the significance of the composition of two methods, one from each class? If the two methods are denoted by M_1 and M_2, with s_1 and s_2 stages, respectively, then $M_1 M_2$ will denote the combined operation of calculating the stages of the first method and the output value so that it now becomes possible to write the stages of the second method as though they were additional stages appended to the first method. Thus $M_1 M_2$ is also a Runge–Kutta method but the equivalence class to which it belongs is independent of how the representative methods M_1 and M_2 were chosen from within their classes. Furthermore we can compute the elementary weights for the product class directly from those of the classes containing M_1 and M_2.

For convenience, we will write the function on trees to elementary weights corresponding to two specific methods as α and β. For tree number i we will write $\alpha_i = \alpha(t_i)$ and $\beta_i = \beta(t_i)$. The value of $(\alpha\beta)(t_i)$ will be a function of the α and β values and formulae for these are shown in Table 1.7, up to order 4 trees. The value of β_0 which appears in this table is equal to $\beta(\emptyset)$. Restricted to $X_1 \times X_1$, the operation defined by this table generates a group. However, X also has a vector space structure and left-multiplication by a member of X_1 is a linear operator on this vector space.

We now discuss an important example of the vector space structure. The output from the Euler method, starting from initial value $y(x_0) = y_0$, gives a result $y_0 + hy'(x_0)$. Subtract y_0 from this and we obtain exactly the scaled derivative $hy'(x_0) = hf(y_0)$. We will regard the elementary weights for this scaled derivative as being exactly the same as for the Euler method, but with β_0 set equal to zero. We will denote this special generalized Runge–Kutta method by D so that $\mathsf{D}(\emptyset) = 0$, $\mathsf{D}(\tau) = 1$, $\mathsf{D}(t) = 0$ $(r(t) > 1)$.

It is quite convenient to build up elementary weights, and more complicated objects, using the D operation. If 1 is used to represent the identity element of the group. We can then write the group elements representing

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 1.7. Runge–Kutta group operation, with $\beta_0 = 1$.

i	t_i	$\alpha(t_i)$	$\beta(t_i)$	$(\alpha\beta)(t_i)$
1		α_1	β_1	$\alpha_1\beta_0 + \beta_1$
2		α_2	β_2	$\alpha_2\beta_0 + \alpha_1\beta_1 + \beta_2$
3		α_3	β_3	$\alpha_3\beta_0 + \alpha_1^2\beta_1 + 2\alpha_1\beta_2 + \beta_3$
4		α_4	β_4	$\alpha_4\beta_0 + \alpha_2\beta_1 + \alpha_1\beta_2 + \beta_4$
5		α_5	β_5	$\alpha_5\beta_0 + \alpha_1^3\beta_1 + 3\alpha_1^2\beta_2 + 3\alpha_1\beta_3 + \beta_5$
6		α_6	β_6	$\alpha_6\beta_0 + \alpha_1\alpha_2\beta_1 + (\alpha_1^2 + \alpha_2)\beta_2 + \alpha_1(\beta_3 + \beta_4) + \beta_6$
7		α_7	β_7	$\alpha_7\beta_0 + \alpha_3\beta_1 + \alpha_1^2\beta_2 + 2\alpha_1\beta_4 + \beta_7$
8		α_8	β_8	$\alpha_8\beta_0 + \alpha_4\beta_1 + \alpha_2\beta_2 + \alpha_1\beta_4 + \beta_8$

the stages by η_i, $i = 1, 2, \ldots, s$ which satisfy

$$\eta_i = 1 + \sum_{j=1}^{s} a_{ij}\eta_j D.$$

The output at the end of a Runge–Kutta method will then be

$$1 + \sum_{i=1}^{s} b_i\eta_i D.$$

Collocation methods and implicit Runge–Kutta methods

A possible approach to the solution of an initial value problem, on an interval $[x_0, x_0 + h]$, is to assume an approximation of the form

$$y(x_0 + \xi h) = P(\xi),$$

and to define the polynomial P, assumed to be of degree s, by the conditions

$$\begin{aligned} P(0) &= y_0, \\ P'(c_i) &= f(P(c_i)), \quad i = 1, 2, \ldots, s. \end{aligned}$$

In these conditions, the 'collocation points' $c_1, c_2, \ldots, c_s$, are distinct and nonzero. To obtain a step-by-step sequence of approximations, define $y_1 = P(1)$, and compute y_2 in a similar way from y_1, as the next step in the process. An attraction of such methods is the fact that an interpolated approximation is automatically available between step values.

It was pointed out in Wright (1970) that collocation methods are equivalent to implicit Runge–Kutta methods, with the abscissae identical to the collocation points. In the implicit Runge–Kutta representation, the

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

elements of A and b^T are defined by

$$\sum_{j=1} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k, \qquad k = 1, 2, \ldots, s, \quad i = 1, 2, \ldots s, \tag{1.41}$$

$$\sum_{j=1} b_j c_j^{k-1} = \frac{1}{k}, \qquad k = 1, 2, \ldots, s. \tag{1.42}$$

For example, if $c = [0, \frac{1}{2}, 1]^T$, we obtain the method

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{5}{24} & \frac{1}{3} & -\frac{1}{24} \\ 1 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \\ \hline & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array} . \tag{1.43}$$

The condition (1.41) expresses the fact that not only does the output from a step have a specific order but the stage values themselves are computed to a high precision, as measured in terms of an asymptotic error $\mathcal{O}(h^{q+1})$. Even for a method which is not based on collocation, the stage order, which we denote by q, can be close to s for implicit methods. This is regarded as an advantage in terms of the ability of the method to solve stiff problems reliably (Prothero and Robinson 1974). The next example method has stage order 2 and order 3:

$$\begin{array}{c|cc} \frac{1}{3} & \frac{5}{12} & -\frac{1}{12} \\ 1 & \frac{3}{4} & \frac{1}{4} \\ \hline & \frac{3}{4} & \frac{1}{4} \end{array} . \tag{1.44}$$

A-stable Runge–Kutta methods

As for the Euler method and linear multistep methods, the first step in assessing the stability properties of a Runge–Kutta method is to investigate its behaviour with the linear problem $y' = qy$. For this problem, the stages and final output in a single step are given by

$$Y = zAY + y_{n-1}\mathbf{1},$$

$$y_n = zb^T Y + y_{n-1},$$

where $z = hq$ and $\mathbf{1} \in \mathbb{R}^s$ has every component equal to 1. Eliminate Y and the result is

$$y_n = R(z) y_{n-1},$$

where the 'stability function' is

$$R(z) = 1 + zb^T (I - zA)^{-1} \mathbf{1}.$$

The set of z values for which $|R(z)| \leq 1$ is the 'stability region'. If this includes the left half-plane then the method is A-stable. Two examples of

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

A-stable methods can be found in (1.43) and (1.44) for which the stability functions are

$$R(z) = \frac{1 + \frac{1}{2}z + \frac{1}{12}z^2}{1 - \frac{1}{2}z + \frac{1}{12}z^2},$$

$$R(z) = \frac{1 + \frac{1}{3}z}{1 - \frac{2}{3}z + \frac{1}{12}z^2},$$

respectively. These stability functions are examples of Padé approximations and proof of the A-stability, in these particular cases, is included in Theorem 8.8.

Runge–Kutta methods for stiff problems

For stiff problems, it is not satisfactory to use explicit methods, and we need to consider methods in which A has a more complicated structure. We will consider five levels of implicitness, in terms of restrictions on the coefficients in the $s \times s$ matrix A and the vector b^T:

(i) $a_{ij} = 0$ if $j > i$,
(ii) $a_{ij} = 0$ if $j > i$; $a_{ii} = \lambda$, $i = 1, 2, \ldots, s$,
(iii) $a_{11} = 0$; $a_{ij} = 0$, $j > i$; $a_{ii} = \lambda$, $i = 2, 3, \ldots, s$; $a_{sj} = b_j$, $j = 1, 2, \ldots, s$,
(iv) $\sigma(A) = \{\lambda\}$,
(v) A an arbitrary full matrix.

The use of fully implicit methods (v) was proposed by Ceschino and Kuntzmann (1963) and Butcher (1964), with the abscissae based on Gauss–Legendre integration points. The Gauss methods and related methods based on other high-order quadrature formulae, have an important role in the solution of stiff problems. In Butcher (1964) so-called semi-implicit methods (i) were introduced but without a specific application in mind.

For efficiency reasons, there is also an interest in diagonally implicit (or DIRK) methods included within the (ii) and (iii) families. Finally, singly implicit (SIRK) methods (Burrage 1978*a*) were introduced to yield methods which not only have efficient implementation properties but have high stage order.

The following example of (i) has order 5; this would have required 6 stages if the method had been explicit. For example, the following method has order 5:

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{4} & \frac{1}{8} & \frac{1}{8} & & \\ \frac{7}{10} & -\frac{1}{100} & \frac{14}{25} & \frac{3}{20} & \\ 1 & \frac{2}{7} & 0 & \frac{5}{7} & \\ \hline & \frac{1}{14} & \frac{32}{81} & \frac{250}{567} & \frac{5}{54} \end{array}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

An example of (ii) is the following order 3 method:

$$\begin{array}{c|ccc} \lambda & \lambda & & \\ \frac{1}{2}(1+\lambda) & \frac{1}{2}(1-\lambda) & \lambda & \\ 1 & \frac{1}{4}(-6\lambda^2+16\lambda-1) & \frac{1}{4}(6\lambda^2-20\lambda+5) & \lambda \\ \hline & \frac{1}{4}(-6\lambda^2+16\lambda-1) & \frac{1}{4}(6\lambda^2-20\lambda+5) & \lambda \end{array},$$

where $\lambda \approx 0.4358665215$ satisfies $\frac{1}{6}-\frac{3}{2}\lambda+3\lambda^2-\lambda^3=0$. This method is A-stable but its stage order is only 1, making it of limited value in the solution of stiff problems.

The next method is an example of (iv) and has order and stage order 2:

$$\begin{array}{c|cc} 3-2\sqrt{2} & \frac{5}{4}-\frac{3}{4}\sqrt{2} & \frac{7}{4}-\frac{5}{4}\sqrt{2} \\ 1 & \frac{1}{4}+\frac{1}{4}\sqrt{2} & \frac{3}{4}-\frac{1}{4}\sqrt{2} \\ \hline & \frac{1}{4}+\frac{1}{4}\sqrt{2} & \frac{3}{4}-\frac{1}{4}\sqrt{2} \end{array}.$$

The method is A-stable, as are similar methods up to order 8 (with the exception of 7). Their major disadvantage is the fact that, for $s>2$, not all the abscissae lie in $[0,1]$.

It is implemented using a transformation which makes it effectively like DIRK methods, in terms of cost, at least for large problems where the overheads due to the transformations are relatively insignificant.

Finally, a simple example of (v). This is one of the Gauss–Legendre family of methods and it has order 4 and stage order 2:

$$\begin{array}{c|cc} \frac{1}{2}-\frac{1}{6}\sqrt{3} & \frac{1}{4} & \frac{1}{4}-\frac{1}{6}\sqrt{3} \\ \frac{1}{2}+\frac{1}{6}\sqrt{3} & \frac{1}{4}+\frac{1}{6}\sqrt{3} & \frac{1}{4} \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}.$$

Similar methods exist for all positive values of s with order $2s$ and stage order s. Although they are A-stable, they are difficult to implement efficiently.

2. Motivations for general linear methods

We will describe some of the circumstances and events which have led to an interest in more general methods.

The traditional methods can be regarded as generalizations, in one way or another, of the Euler method. In the terminology of general linear methods, this is a one-value ($r=1$), one-stage ($s=1$) method. Increasing the value of the integer r leads to linear multistep methods and increasing s leads to Runge–Kutta methods. It seems natural to consider methods in which both $r>1$ and $s>1$ are possible.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Thus the first motivation for studying general linear methods is that this generalization is natural and there seems to be no reason for not adopting it. Indeed even some existing methods are more naturally formulated in a general linear method ansatz.

We will look at some of the limitations of existing methods to see that the general linear method generalization is not only natural but also potentially useful. We will pursue this point of view by looking at some simple modifications of existing methods and ultimately by attempting to find some new and potentially efficient methods which do not arise naturally in any other way.

2.1. Limitations of linear multistep methods

Even though there are $2k+1$ free parameters in the specification of a linear k-step method, so that order $2k$ would seem to be possible, in fact practical methods are limited in order to $k+2$ (if k is even) and $k+1$ (if k is odd), because of the stability condition. This result, Dahlquist's first barrier (Dahlquist 1956), which we reviewed in Theorem 1.3, is coupled with the second Dahlquist barrier (Dahlquist 1963), which limits the order of A-stable methods to exactly 2. A proof, using order arrows, of the second barrier is given in Theorem 8.10. In spite of this barrier, if $A(\alpha)$-stability, with a reasonably large angle α, is regarded as acceptable, it is possible to go to at least order 4.

This applies in particular to BDF methods where we note that BDF4 is $A(0.4\pi)$-stable. For any p, it is possible to replace the factor 0.4 by a number arbitrarily close to $\frac{1}{2}$ (Widlund 1967, Grigorieff and Schroll 1978) but at the cost of impractically high error constants (Jeltsch and Nevanlinna 1982).

In addition to stability constraints, another type of limitation is the complication associated with change of step-size and change of order. Each of these requires considerable overheads.

2.2. Limitations of Runge–Kutta methods

While explicit order p Runge–Kutta methods exist with p stages, for $p = 1, 2, 3, 4$, no such methods exist for $p > 4$. Furthermore, if the minimal number of stages to achieve order p is $s(p)$, then $s(p) - p$ increases steadily as p increases, as we recall from Table 1.6.

Variable step-size and order are made difficult by the need to estimate local truncation errors in a reliable way. This is an increasingly expensive requirement as the order increases.

In the case of implicit methods, the achievable order is exactly $2s$, and methods which achieve this maximum are A-stable. This seems to be a satisfactory situation but the actual methods have two serious handicaps.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

The first of these is that they suffer from error reduction and the second is the very high implementation cost.

Even though the global truncation error is asymptotically $O(h^p)$, for step-sizes which often arise in practice, the error behaves more like $O(h^q)$, where q is the stage order.

Solving the non-linear equations defining the stage values involves a process based on the Newton method. This is much more expensive than for implicit linear multistep methods, unless the coefficient matrix A has a special structure, such as the DIRK structure. Unfortunately DIRK methods necessarily have low stage order. In the opinion of this author the only way to overcome this difficulty is to use SIRK methods, in the modified form discussed in Butcher and Chen (1998). But there seem to be better algorithms within the larger family of general linear methods.

2.3. Modifications of linear multistep methods

Many examples are known of modifications to standard methods, which somehow acquire enhanced properties. For example, by adding one or more offstep points, it is possible to give a linear multistep method a little closer to that of Runge–Kutta methods. This can break the Dahlquist barrier by permitting methods to have order greater than $2k$ and still remain stable. A class of methods in this hybrid family takes the idea of predictor–corrector pairs based on Adams–Bashforth and Adams–Moulton methods further, by including a single offstep predictor as well as the usual predictor and corrector at the end of the step. Thus for $k = 2$ the k-step PECE[1] method,

$$\begin{aligned} y_n^* &= y_{n-1} + \tfrac{3}{2}hf_{n-1} - \tfrac{1}{2}hf_{n-2},\\ y_n &= y_{n-1} + \tfrac{1}{2}hf_n^* + \tfrac{1}{2}hf_{n-1}, \end{aligned}$$

generalizes to

$$\begin{aligned} y^*_{n-\frac{1}{2}} &= y_{n-2} + \tfrac{9}{8}hf_{n-1} + \tfrac{3}{8}hf_{n-2},\\ y_n^* &= \tfrac{28}{5}y_{n-1} - \tfrac{23}{5}y_{n-2} + \tfrac{32}{15}hf^*_{n-\frac{1}{2}} - 4hf_{n-1} - \tfrac{26}{15}hf_{n-2},\\ y_n &= \tfrac{32}{31}y_{n-1} - \tfrac{1}{31}y_{n-2} + \tfrac{5}{31}hf_n^* + \tfrac{64}{93}hf^*_{n-\frac{1}{2}} + \tfrac{4}{31}hf_{n-1} - \tfrac{1}{93}hf_{n-2}. \end{aligned}$$

Note that in this discussion f_n^* denotes $f(x_n, y_n^*)$ and $f^*_{n-\frac{1}{2}}$ denotes

$$f\left(x_n - \frac{1}{2}h, y^*_{n-\frac{1}{2}}\right).$$

Even though the two predictors generate approximations only of order 3, the overall result has order 5.

[1] PECE denotes 'Predict–Evaluate–Correct–Evaluate'.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

'Hybrid' methods, as Gear named them, were introduced in Butcher (1965), Gear (1965) and Gragg and Stetter (1964).

A completely different generalization of linear multistep methods is that of cyclic composite methods, first proposed by Donelson and Hansen (1971). If we are given m linear multistep methods

$$y_n = \sum_{i=1}^{k} \alpha_i^{[j]} y_{n-i} + \sum_{i=0}^{k} \beta_i^{[j]} h f_{n-i}, \quad j = 1, \ldots, m,$$

the idea is to apply them cyclically. That is, in a sequence of m steps, use method number 1 followed by method number 2 and so on until method number m has been applied. For steps after this, the cycle is repeated.

We present just two examples. In the first we consider two methods, each based on open Newton–Cotes quadrature formulae:

$$y_n = y_{n-2} + 2h f_{n-1},$$
$$y_n = y_{n-3} + \tfrac{3}{2} h f_{n-1} + \tfrac{3}{2} h f_{n-2}.$$

Taken alone, each of these methods is 'weakly stable'. That is, regarded for convenience as 3-step methods, their zero stability matrices are

$$M_1 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{and} \quad M_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

respectively. For M_1 the eigenvalues are $\{1, -1, 0\}$ and for M_2 the eigenvalues are $\{1, \exp(2\pi i/3), \exp(-2\pi i/3)\}$. Weak stability is a consequence of the existence of eigenvalues on the unit disc, in addition to the principal eigenvalue at 1 in each case. However, when the two methods are used in alternation then the stability matrix over the pair of steps becomes

$$M_2 M_1 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix},$$

with eigenvalues $\{1, 0, 0\}$.

It is possible to go even further and to construct cycles of explicit methods which overcome the first Dahlquist barrier. For example, consider the two methods

$$y_n = -\tfrac{8}{11} y_{n-1} + \tfrac{19}{11} y_{n-2} + \tfrac{10}{11} h f_n + \tfrac{19}{11} h f_{n-1} + \tfrac{8}{11} h f_{n-2} - \tfrac{1}{33} h f_{n-3},$$
$$y_n = \tfrac{449}{240} y_{n-1} + \tfrac{19}{30} y_{n-2} - \tfrac{361}{240} y_{n-3} + \tfrac{251}{720} h f_n + \tfrac{19}{30} h f_{n-1} - \tfrac{449}{240} h f_{n-2} - \tfrac{35}{72} h f_{n-3}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Each of these methods has order 5 and each is unstable, but we will see that the corresponding cyclic method has perfect stability. To verify this remark, analyse stability using $y' = 0$:

$$y_n = -\tfrac{8}{11}y_{n-1} + \tfrac{19}{11}y_{n-2}, \tag{2.1}$$

$$y_n = \tfrac{449}{240}y_{n-1} + \tfrac{19}{30}y_{n-2} - \tfrac{361}{240}y_{n-3}. \tag{2.2}$$

The difference equation for $y_n - y_{n-1}$ is

$$\begin{bmatrix} y_n - y_{n-1} \\ y_{n-1} - y_{n-2} \end{bmatrix} = X \begin{bmatrix} y_{n-1} - y_{n-2} \\ y_{n-2} - y_{n-3} \end{bmatrix},$$

where X is

$$\begin{bmatrix} -\frac{19}{11} & 0 \\ 1 & 0 \end{bmatrix}$$

for (2.1), and

$$\begin{bmatrix} \frac{209}{240} & \frac{361}{240} \\ 1 & 0 \end{bmatrix}$$

for (2.2). Neither matrix is power-bounded but their product, corresponding to the cyclic use of the two methods, is nilpotent.

By applying the cyclic composite idea to implicit methods it is also possible to overcome the second Dahlquist barrier (Bickart and Picel 1973).

2.4. Modifications of Runge–Kutta methods

The following family of fourth-order methods is one of several such families found by Kutta:

$$\begin{array}{c|cccc} 0 & & & & \\ c_2 & c_2 & & & \\ \frac{1}{2} & \frac{1}{2} - \frac{1}{8c_2} & \frac{1}{8c_2} & & \\ 1 & \frac{1}{2c_2} - 1 & -\frac{1}{2c_2} & 2 & \\ \hline & \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \end{array}.$$

If we substitute $c_2 = -1$, it is found that

$$\begin{array}{c|cccc} 0 & & & & \\ -1 & -1 & & & \\ \frac{1}{2} & \frac{5}{8} & -\frac{1}{8} & & \\ 1 & -\frac{3}{2} & \frac{1}{2} & 2 & \\ \hline & \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \end{array}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

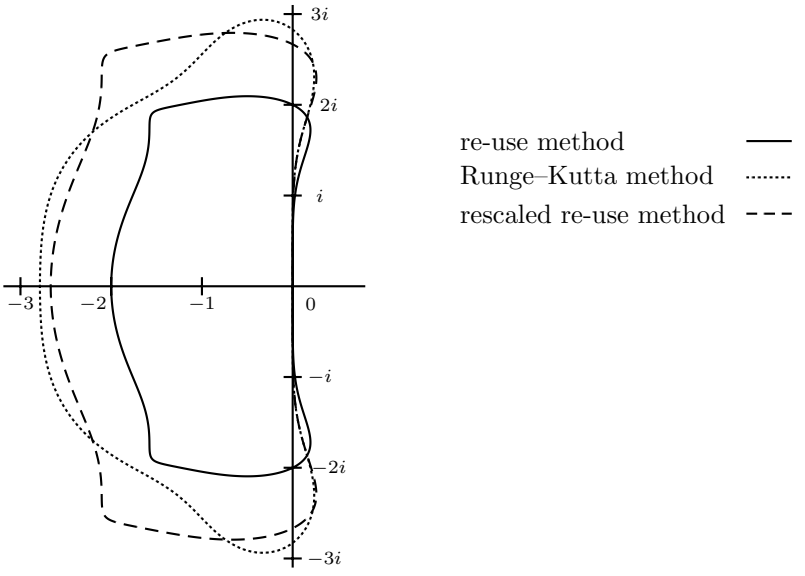


Figure 2.1. Stability region for the re-use method compared with the classical Runge–Kutta method.

We can interpret the abscissa at -1 as re-use of the derivative found at the beginning of the previous step. We then have the method

$$\begin{aligned} Y_1 &= y_{n-1} + \tfrac{5}{8}hf(y_{n-1}) - \tfrac{1}{8}hf(y_{n-2}), & F_1 &= f(Y_1),\\ Y_2 &= y_{n-1} - \tfrac{3}{2}hf(y_{n-1}) + \tfrac{1}{2}hf(y_{n-2}) + 2hF_1, & F_2 &= f(Y_2),\\ y_n &= y_{n-1} + \tfrac{1}{6}hf(y_{n-1}) + \tfrac{2}{3}hF_1 + \tfrac{1}{6}hF_2. \end{aligned}$$

Like the Runge–Kutta method on which it is based, this method retains order 4, even though it evaluates f only 3 times per time-step compared with 4 for the original method.

We can understand something about the behaviour of the new method by plotting its stability region. This is shown in Figure 2.1, with the classical fourth-order method included for comparison. Because $s = 3$ for the re-use method, rather than $s = 4$ for the Runge–Kutta method, a more appropriate comparison is achieved by the rescaling $z \mapsto \frac{4}{3}z$ in the case of the re-use method; this is also shown in the figure. Based on this comparison, there seems to be little advantage in either the Runge–Kutta method or the re-use method.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

As a general linear method, using a notation we will introduce in Section 3, the re-use method has the following matrix representation:

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ \frac{5}{8} & 0 & 0 & 1 & -\frac{1}{8} \\ -\frac{3}{2} & 2 & 0 & 1 & \frac{1}{2} \\ \hline \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{array}\right]. \tag{2.3}$$

3. Formulations

In the formulation of general linear methods given in Butcher (1966), a collection of approximations $y_i^{[n]}$, $i = 1, 2, \ldots, r$, together with derivative approximations $F_i^{[n]} = f(y_i^{[n]})$ is computed at the end of step number n in terms of the corresponding quantities available as input to the step. Making use of three matrices of coefficients, A, B, C, which characterize a specific method, the step n approximations are given by

$$y_i^{[n]} = \sum_{j=1}^{r} a_{ij} y_j^{[n-1]} + \sum_{j=1}^{r} b_{ij} h F_j^{[n]} + \sum_{j=1}^{r} c_{ij} h F_j^{[n-1]}.$$

This method was referred to as (A, B, C). It is easy to see, by raising the value of r if necessary, that C can be removed from the formulation.

An equivalent, but in many ways more convenient, formulation was introduced in Burrage and Butcher (1980) and this is now the standard way of representing general linear methods.

Denote the output approximations from step number n by $y_i^{[n]}$, $i = 1, 2, \ldots, r$, the stage values by Y_i, $i = 1, 2, \ldots, s$ and the stage derivatives by F_i, $i = 1, 2, \ldots, s$.

For convenience, write

$$y^{[n-1]} = \begin{bmatrix} y_1^{[n-1]} \\ y_2^{[n-1]} \\ \vdots \\ y_r^{[n-1]} \end{bmatrix}, \quad y^{[n]} = \begin{bmatrix} y_1^{[n]} \\ y_2^{[n]} \\ \vdots \\ y_r^{[n]} \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_s \end{bmatrix}, \quad F = \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_s \end{bmatrix}.$$

It is assumed that Y and F are related by a differential equation.

The computation of the stages and the output from step number n is

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

carried out according to the formulae

$$Y_i = \sum_{j=1}^{s} a_{ij} h F_j + \sum_{j=1}^{r} u_{ij} y_j^{[n-1]}, \qquad i = 1, 2, \ldots, s,$$

$$y_i^{[n]} = \sum_{j=1}^{s} b_{ij} h F_j + \sum_{j=1}^{r} v_{ij} y_j^{[n-1]}, \qquad i = 1, 2, \ldots, r,$$

where the matrices $A = [a_{ij}]$, $U = [u_{ij}]$, $B = [b_{ij}]$, $V = [v_{ij}]$ are characteristic of a specific method.

We can write these relations more compactly in the form

$$\begin{bmatrix} Y \\ y^{[n]} \end{bmatrix} = \begin{bmatrix} A \otimes I & U \otimes I \\ B \otimes I & V \otimes I \end{bmatrix} \begin{bmatrix} hF \\ y^{[n-1]} \end{bmatrix},$$

which we can simplify by making a harmless abuse of notation in the form

$$\begin{bmatrix} Y \\ y^{[n]} \end{bmatrix} = \begin{bmatrix} A & U \\ B & V \end{bmatrix} \begin{bmatrix} hF \\ y^{[n-1]} \end{bmatrix}. \tag{3.1}$$

An alternative formulation, of the closely related A-methods, is given in Albrecht (1985).

3.1. Consistency, stability and convergence

An idea that will be developed in Section 4 is that there is always a starting method associated with each method. For our present purpose it is enough to ask what quantity the numerical solution is supposed to approximate, at least to a first-order approximation. As a very basic requirement we ask if it is possible to approximate the solution to the problem $y'(x) = 0$, exactly. This condition has two parts. First, we want to ensure that quantities input to step number n are capable of remaining unchanged at the end of the step. Secondly we want to be able to guarantee long-term adherence to this solution, even if a slight perturbation is introduced. The first requirement will be written in terms of the existence of a 'pre-consistency vector' u which is unchanged when acted upon by V and the second that V is power-bounded. Note that for the differential equation $y' = 0$, input component number i is assumed to have the form $u_i y(x_{n-1})$ so that $Vu = u$ is the first of our conditions. In addition to this property of u we will require that $Uu = \mathbf{1}$ so that each stage gives an approximation close to $y(x_{n-1})$.

For a pre-consistent stable method, we also want to guarantee that the solution of the problem $y'(x) = 1$ has correctly advanced one step forward. We summarize these remarks with a series of definitions.

Definition 3.1. A general linear method (A, U, B, V) is 'stable' if there exists a constant C such that $\|V^n\| \leq C$ for any positive integer n.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Definition 3.2. A general linear method (A, U, B, V) is 'pre-consistent' if there exists a 'pre-consistency vector' u such that

$$Vu = u,$$
$$Uu = \mathbf{1}.$$

Definition 3.3. A general linear method (A, U, B, V) is consistent if it is pre-consistent with pre-consistency vector u and furthermore, there exists a vector v such that

$$B\mathbf{1} + Vv = u + v.$$

Given the properties embodied in these definitions it is possible to guarantee that the approximation computed by a general linear method can be found arbitrarily close to the exact solution. We express this in terms of a definition, followed by a theorem, which will not be proved in this survey.

Definition 3.4. Consider an initial value problem

$$y'(x) = f(x, y(x)), \qquad y(x_0) = y_0,$$

where $f : [x_0, \overline{x}] \times \mathbb{R}^N \to \mathbb{R}^N$ is continuous in its first variable and satisfies a Lipschitz condition in its second variable. Let (A, U, B, V) be a consistent, stable general linear method with pre-consistency and consistency vectors u and v. Let $S(h)$ denote a starting approximation which depends on h in such a way that $\lim_{h\to 0} S_i(h) = u_i y_0$, $i = 1, 2, \ldots, r$. Let $\eta(n)$ denote the value of $y^{[n]}$, computed using the given method for the given problem, with starting value defined by $y^{[0]} = S((\overline{x} - x_0)/n)$. The method is 'convergent' if, for any choice of initial value problem and starting approximation S, $\lim_{n\to\infty} \eta_i(n) = u_i y(\overline{x})$, $i = 1, 2, \ldots, r$.

Theorem 3.5. Any stable and consistent general linear method is convergent.

This result includes the theories for convergence for special methods, such as Runge–Kutta and linear multistep methods. In practice, methods will be designed to have a stage order equal to at least 0 and an order equal to at least 1. Such order conditions imply consistency so the crucial question to ask in the search for acceptable methods, is whether the method is or is not stable.

The proof of Theorem 3.5 is technical and is given in Butcher (2003), for example.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

3.2. Representation of standard methods

For a linear multistep method with input and output

$$y^{[n-1]} = \begin{bmatrix} y_{n-1} \\ y_{n-2} \\ \vdots \\ y_{n-k} \\ \hline hf(x_{n-1}, y_{n-1}) \\ hf(x_{n-2}, y_{n-2}) \\ \vdots \\ hf(x_{n-k}, y_{n-k} \end{bmatrix}, \qquad y^{[n]} = \begin{bmatrix} y_n \\ y_{n-1} \\ \vdots \\ y_{n-k+1} \\ \hline hf(x_n, y_n) \\ hf(x_{n-1}, y_{n-1}) \\ \vdots \\ hf(x_{n-k+1}, y_{n-k+1} \end{bmatrix},$$

the single stage Y_1 will be identical with the first output component, and we have the method

$$\left[\begin{array}{c|cccc|cccc|c} \beta_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{k-1} & \alpha_k & \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \\ \hline \beta_0 & \alpha_1 & \alpha_2 & \cdots & \alpha_{k-1} & \alpha_k & \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \hline 1 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 1 & 0 \end{array}\right]. \tag{3.2}$$

3.3. Transformation of methods

Because the data imported at the start of a step and exported at the end of the step is capable of being repackaged in a different way, we consider two methods (A, U, B, V) and $(A, \widehat{U}, \widehat{B}, \widehat{V})$, so related that

$$\begin{bmatrix} A & \widehat{U} \\ \widehat{B} & \widehat{V} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & T^{-1} \end{bmatrix} \begin{bmatrix} A & U \\ B & V \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & T \end{bmatrix}, \tag{3.3}$$

where T is an arbitrary $r \times r$ non-singular matrix.

If $\widehat{y}^{[n]}$ is the output from the transformed method and $y^{[n]}$ is the output from the original method then

$$\widehat{y}^{[n-1]} = (T \otimes I) y^{[n-1]}, \qquad \widehat{y}^{[n]} = (T^{-1} \otimes I) y^{[n]}.$$

The relationship between the basic properties of the two methods is expressed in the following result which is proved in a routine way.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Theorem 3.6. Let (A, U, B, V) and $(A, \widehat{U}, \widehat{B}, \widehat{V})$ be two methods related by (3.3), then if either is consistent, then so is the other, with preconsistency and consistency vectors related by

$$\widehat{u} = T^{-1}u,$$
$$\widehat{v} = T^{-1}v.$$

Furthermore, if either method is stable, then so is the other.

The significance of transformations predates by many years the introduction of general linear methods. In traditional formulations of Adams methods, for example, the input data for step number n may consist of approximations to $y(x_{n-1})$, $hy'(x_{n-1})$, $hy'(x_{n-2}), \ldots, hy'(x_{n-k})$. On the other hand, an alternative representation, popular in the days of hand computation, is to use $y(x_{n-1})$, together with backward differences, from order 0 to $k-1$, of the derivative information. The use of approximations to scaled derivatives was proposed by Nordsieck (1962) and promoted in Gear (1967, 1971).

Given a method $(A, \widehat{U}, \widehat{B}, \widehat{V})$ with $\widehat{r}$ input and output values, it may happen that a method (A, U, B, V) with $r < \widehat{r}$ input and output values might be related to it by the existence of an $r \times \widehat{r}$ matrix T, with rank r, such that

$$\widehat{U} = UT, \quad T\widehat{B} = B, \quad T\widehat{V} = VT.$$

In this case, it might be asked to what extent the method (A, U, B, V) carries out essentially the same task as the original method $(A, \widehat{U}, \widehat{B}, \widehat{V})$. To understand this question, introduce an arbitrary $(\widehat{r} - r) \times \widehat{r}$ matrix $\dot{T}$ whose rows, together with the rows of T, constitute a basis for $\mathbb{R}^{\widehat{r}}$. Now write

$$\begin{bmatrix} y^{[n]} \\ \dot{y}^{[n]} \end{bmatrix} = \begin{bmatrix} T \otimes I \\ \dot{T} \otimes I \end{bmatrix} \widehat{y}^{[n]},$$

where it is assumed that the $\widehat{y}$ sequence satisfies the original method. Transform the original $\widehat{\ }$ method to give the method

$$\left[\begin{array}{c|cc} A & U & 0 \\ \hline B & V & 0 \\ \dot{B} & \dot{V} & \ddot{V} \end{array}\right] = \begin{bmatrix} T \\ \dot{T} \end{bmatrix} \begin{bmatrix} A & \widehat{U} \\ \widehat{B} & \widehat{V} \end{bmatrix} \begin{bmatrix} T \\ \dot{T} \end{bmatrix}^{-1}.$$

It is apparent that the $y^{[n]}$ is generated by the method (A, U, B, V), without any reference to the the $\dot{y}^{[n]}$ sequence. Hence, the transformation of methods can also have the effect of reducing a method to a simpler 'reduced method'. Assuming that A has no unused stages, we refer to a method that is not capable of further reduction as irreducible.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

The most readily available example of a reducible method is for a linear multistep method written in the form (3.2). Define

$$T = \left[\begin{array}{ccccc|ccccc} \alpha_1 & \alpha_2 & \cdots & \alpha_{k-1} & \alpha_k & \beta_1 & \beta_2 & \cdots & \beta_{k-1} & \beta_k \\ \alpha_2 & \alpha_3 & \cdots & \alpha_k & 0 & \beta_2 & \beta_3 & \cdots & \beta_k & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \alpha_k & 0 & \cdots & 0 & 0 & \beta_k & 0 & \cdots & 0 & 0 \end{array}\right],$$

which has rank k because $|\alpha_k| + |\beta_k| \neq 0$.

Hence, we arrive at the r input method

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{c|cccccc} \beta_0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ \hline \beta_0\alpha_1 + \beta_1 & \alpha_1 & 1 & 0 & \cdots & 0 & 0 \\ \beta_0\alpha_2 + \beta_2 & \alpha_2 & 0 & 1 & \cdots & 0 & 0 \\ \beta_0\alpha_3 + \beta_3 & \alpha_3 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \beta_0\alpha_{k-1} + \beta_{k-1} & \alpha_{k-1} & 0 & 0 & \cdots & 0 & 1 \\ \beta_0\alpha_k + \beta_k & \alpha_k & 0 & 0 & \cdots & 0 & 0 \end{array}\right]. \tag{3.4}$$

4. Order conditions

4.1. General definition of order

In the formulation of a general linear method, there is not always a natural meaning that can be given to the quantities $y^{[n]}$ output at the end of step number n. Hence we will introduce a 'starting method', to represent the quantity we are trying to approximate. Write this as a modified type of general linear method with only a single input but with r outputs. It can also be thought of as a Runge–Kutta method with multiple outputs.

If the starting method is applied to a given initial value the output can be used as input to the first step of the main method. If $\mathcal{S}$ denotes the starting method and $\mathcal{M}$ the main method then $\mathcal{SM}$ will denote the combined operation. Similarly, $\mathcal{E}$ denotes the exact solution evaluated after a time-step h, the same as the step-size for $\mathcal{M}$ and $\mathcal{S}$, and $\mathcal{ES}$ denotes the result of applying $\mathcal{S}$ to the exact solution evaluated after a time h.

If we compare the result computed by $\mathcal{SM}$ and compare it with the result of the computation $\mathcal{ES}$, that is, two members of $\mathbb{R}^{rN}$, we have a measure of how closely $\mathcal{M}$ is able to preserve approximations to $\mathcal{S}$ applied to the exact trajectory. If $\mathcal{S}$ can be chosen so that the norm of the difference between these two results can be estimated in terms of h^{p+1}, then we say that 'the method $\mathcal{M}$ has order p relative to $\mathcal{S}$'. This enables us to state:

Definition 4.1. The method $\mathcal{M}$ has order p if there exists a starting method $\mathcal{S}$ such that $\mathcal{M}$ has order p relative to $\mathcal{S}$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

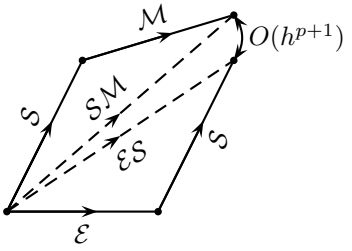


Figure 4.1. Order of general linear method.

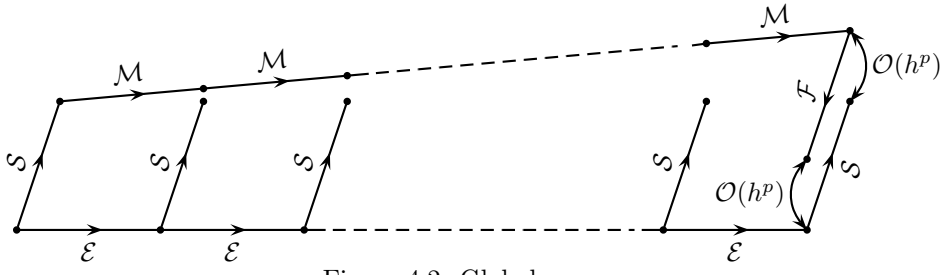


Figure 4.2. Global error.

The relationship between $\mathcal{M}$, $\mathcal{S}$ and $\mathcal{E}$ is illustrated in Figure 4.1.

According to this view of order, the method under consideration cannot be looked at in isolation but will always be related to the starting method. Looked at another way, the order is related to the *interpretation* of the quantities the method is intended to approximate. This is actually a classical point of view and reflects the fact that multivalue methods typically input approximations to specific quantities known in advance. The most well-known example is linear k-step methods in which the data input to step number n consists of approximations to $y(x_{n-i})$ and $hy'(x_{n-i})$ for $i = 1, 2, \ldots, k$. From the classical point of view, a Runge–Kutta method treats its single item of input as an approximation to $y(x_{n-1})$. However, there is no fundamental reason why we should restrict ourselves to this interpretation and it might be possible to regard $\mathcal{S}$ as an arbitrary Runge–Kutta method. Does this lead to any sort of enhancement of the concept of order, for Runge–Kutta methods? The answer is yes, as we will see in Section 4.2.

In addition to the mapping $\mathcal{S}$, we postulate the existence of a 'finishing method' $\mathcal{F}$, defined as a right-sided inverse of $\mathcal{S}$. That is, if $y^{[0]}$ is found by applying $\mathcal{S}$ to y_0, then y_0 is found by applying $\mathcal{F}$ to $y^{[0]}$.

Consider a long term integration consisting of n steps with step-size $h = (\overline{x} - x_0)/n$. Once n steps have been carried out to give the result $y^{[n]}$, $\mathcal{F}$ is applied to this to obtain an approximation to $y(\overline{x})$. Assuming that the method is stable, the errors in each of the steps combine to give an overall error, that is a global error, $n\mathcal{O}(h^{p+1}) = \mathcal{O}(h^p)$. The way this works is shown in Figure 4.2.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 4.1. Analysis for effective order 5.

i	t_i	$(\Psi\Phi)(t_i)-(E\Psi)(t_i)$
1		ϕ_1-1
2		$\phi_2-\frac{1}{2}$
3		$\phi_3-\frac{1}{3}-2\psi_2$
4		$\phi_4+\psi_2\phi_1-\frac{1}{6}-\psi_2$
5		$\phi_5-\frac{1}{4}-3\psi_2-3\psi_3$
6		$\phi_6+\phi_2\psi_2-\frac{1}{8}-\frac{3}{2}\psi_2-\psi_3-\psi_4$
7		$\phi_7+\phi_1\psi_3-\frac{1}{12}-\psi_2-2\psi_4$
8		$\phi_8+\phi_1\psi_4+\phi_2\psi_2-\frac{1}{24}-\frac{1}{2}\psi_2-\psi_4$
9		$\phi_9-\frac{1}{5}-4\psi_2-6\psi_3-4\psi_5$
10		$\phi_{10}+\phi_3\psi_2-\frac{1}{10}-2\psi_2-\frac{5}{2}\psi_3-\psi_4-\psi_5-2\psi_6$
11		$\phi_{11}+\phi_2\psi_3-\frac{1}{15}-\frac{4}{3}\psi_2-\psi_3-2\psi_4-2\psi_6-\psi_7$
12		$\phi_{12}+\phi_2\psi_4+\phi_3\psi_2-\frac{1}{30}-\frac{2}{3}\psi_2-\frac{1}{2}\psi_3-\psi_4-\psi_6-\psi_8$
13		$\phi_{13}+\phi_1\psi_2^2+2\phi_4\psi_2-\frac{1}{20}-\psi_2-\psi_3-\psi_4-2\psi_6$
14		$\phi_{14}+\phi_1\psi_5-\frac{1}{20}-\psi_2-3\psi_4-3\psi_7$
15		$\phi_{15}+\phi_1\psi_6+\phi_4\psi_2-\frac{1}{40}-\frac{1}{2}\psi_2-\frac{3}{2}\psi_4-\psi_7-\psi_8$
16		$\phi_{16}+\phi_1\psi_7+\phi_2\psi_3-\frac{1}{60}-\frac{1}{3}\psi_2-\psi_4-2\psi_8$
17		$\phi_{17}+\phi_1\psi_8+\phi_2\psi_4+\phi_4\psi_2-\frac{1}{120}-\frac{1}{6}\psi_2-\frac{1}{2}\psi_4-\psi_8$

4.2. Effective order of Runge–Kutta methods

In the case of a Runge–Kutta method, the starting method must itself be a Runge–Kutta method because it accepts a single input and yields a single output. For a given tree t, let $\Phi(t)$ denote the elementary differential associated with the main method and $\Psi(t)$ the elementary differential associated with the starting method. For convenience, we will write $\Phi(t_i)=\phi_i$ and $\Psi(t_i)=\psi_i$. Using this notation, the numbered trees are shown, together with expressions for $(\Psi\Phi)(t_i)-(E\Psi)(t_i)$ in Table 4.1. For simplification, ψ_1 has been assigned the value 0. This turns out not to limit the generality of the conditions on Ψ.

Because ψ_2, ψ_3, ψ_4, ψ_5, ψ_6 and ψ_7 can have arbitrary values, there is much more freedom on Φ for effective order 5 than for classical order. The following is a possible solution:

$$[\phi_1,\phi_2,\ldots,\phi_{17}]=[1,\tfrac{1}{2},\tfrac{1}{3},\tfrac{1}{6},\tfrac{1}{4},\tfrac{1}{8},\tfrac{1}{12},\tfrac{1}{24},\tfrac{31}{150},\tfrac{31}{300},\tfrac{1}{15},\tfrac{1}{30},\tfrac{31}{600},\tfrac{13}{300},\tfrac{13}{600},\tfrac{1}{60},\tfrac{1}{120}],$$

$$[\psi_1,\psi_2,\ldots,\psi_8]=[0,0,0,0,\tfrac{1}{600},\tfrac{1}{1200},-\tfrac{1}{600},-\tfrac{1}{1200}],$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

and the tableau for a method yielding the given Φ values is

$$\begin{array}{c|ccccc}
0 & & & & & \\
\frac{2}{5} & \frac{2}{5} & & & & \\
\frac{2}{5} & \frac{1}{5} & \frac{1}{5} & & & \\
\frac{3}{5} & \frac{3}{20} & -\frac{3}{10} & \frac{3}{4} & & \\
1 & \frac{9}{44} & \frac{5}{22} & -\frac{15}{44} & \frac{10}{11} & \\
\hline
 & \frac{11}{72} & 0 & \frac{25}{72} & \frac{25}{72} & \frac{11}{72}
\end{array}.$$

Effective order of singly implicit Runge–Kutta methods

Singly implicit Runge–Kutta methods represent an attempt to achieve the combined goals of L-stability, stage order and order equal to s and efficient implementation. It is possible to satisfy all these requirements up to order 8, with 7 the only exception, or slightly weakened requirements (A(α)-stability for α close to $\frac{1}{2}\pi$ and zero stability function at infinity) for s much higher. Unfortunately, these methods have a disadvantage that abscissae lie outside the interval $[0, 1]$, if $s > 2$. This can be overcome by applying the principle of effective order. Even the difficulty associated with variable step-size can be overcome for this type of method because the high stage order makes it possible to correct the implied starting perturbation as the solution develops. Furthermore, no finishing method is required for individual steps because one of the stages can be used for output.

4.3. Algebraic criterion for order

Let $\xi \in X^r$ represent the starting method, where X is the algebraic structure introduced in Section 1.6. Then the vector of stages, as represented by a member of X_1^s, is found from the relation

$$\eta = A(\eta D) + U\xi,$$

and the result computed at the end of the step is represented by

$$B(\eta D) + V\xi.$$

If this is to agree with $E\xi$ up to trees with order p then we have a convenient criterion for this order. We formalize this as follows.

Theorem 4.2. The general linear method (A, U, B, V) has order p if there exists $\xi \in X^r$ and $\eta \in X_1^s$, such that, for every tree t satisfying $r(t) \leq p$,

$$\eta(t) = A(\eta D)(t) + U\xi(t),$$
$$(E\xi)(t) = B(\eta D)(t) + V\xi(t).$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 4.2. Calculation of order for method (4.1).

t	ξ_1	ξ_2	η_1	$\eta_1 D$	η_2	$\eta_2 D$	η_3	$\eta_3 D$	$\widehat{\xi}_1$	$\widehat{\xi}_2$	$E\xi_1$	$E\xi_2$
1	0	-1	0	1	$\frac{1}{2}$	1	1	1	1	0	1	0
2	0	$\frac{1}{2}$	0	0	$\frac{1}{8}$	$\frac{1}{2}$	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0
3	0	$-\frac{1}{3}$	0	0	$-\frac{1}{12}$	$\frac{1}{4}$	$\frac{5}{6}$	1	$\frac{1}{3}$	0	$\frac{1}{3}$	0
4	0	$-\frac{1}{6}$	0	0	$-\frac{1}{24}$	$\frac{1}{8}$	$\frac{5}{12}$	$\frac{1}{2}$	$\frac{1}{6}$	0	$\frac{1}{6}$	0
5	0	$\frac{1}{4}$	0	0	$\frac{1}{16}$	$\frac{1}{8}$	0	1	$\frac{1}{4}$	0	$\frac{1}{4}$	0
6	0	$\frac{1}{8}$	0	0	$\frac{1}{32}$	$\frac{1}{16}$	0	$\frac{1}{2}$	$\frac{1}{8}$	0	$\frac{1}{8}$	0
7	0	$\frac{1}{12}$	0	0	$\frac{1}{48}$	$-\frac{1}{12}$	$-\frac{1}{4}$	$\frac{5}{6}$	$\frac{1}{12}$	0	$\frac{1}{12}$	0
8	0	$\frac{1}{24}$	0	0	$\frac{1}{96}$	$-\frac{1}{24}$	$-\frac{1}{8}$	$\frac{5}{12}$	$\frac{1}{24}$	0	$\frac{1}{24}$	0
9	0	$-\frac{1}{5}$	0	0	$-\frac{1}{20}$	$\frac{1}{16}$	$\frac{13}{40}$	1	$\frac{5}{24}$	0	$\frac{1}{5}$	0
10	0	$-\frac{1}{10}$	0	0	$-\frac{1}{40}$	$\frac{1}{32}$	$\frac{13}{80}$	$\frac{1}{2}$	$\frac{5}{48}$	0	$\frac{1}{10}$	0
11	0	$-\frac{1}{15}$	0	0	$-\frac{1}{60}$	$-\frac{1}{24}$	$-\frac{1}{60}$	$\frac{5}{6}$	$\frac{1}{9}$	0	$\frac{1}{15}$	0
12	0	$-\frac{1}{30}$	0	0	$-\frac{1}{120}$	$-\frac{1}{48}$	$-\frac{1}{120}$	$\frac{5}{12}$	$\frac{1}{18}$	0	$\frac{1}{30}$	0
13	0	$-\frac{1}{20}$	0	0	$-\frac{1}{80}$	$\frac{1}{64}$	$\frac{13}{160}$	$\frac{1}{4}$	$\frac{5}{96}$	0	$\frac{1}{20}$	0
14	0	$-\frac{1}{20}$	0	0	$-\frac{1}{80}$	$\frac{1}{16}$	$\frac{7}{40}$	0	$\frac{1}{24}$	0	$\frac{1}{20}$	0
15	0	$-\frac{1}{40}$	0	0	$-\frac{1}{160}$	$\frac{1}{32}$	$\frac{7}{80}$	0	$\frac{1}{48}$	0	$\frac{1}{40}$	0
16	0	$-\frac{1}{60}$	0	0	$-\frac{1}{240}$	$\frac{1}{48}$	$\frac{7}{120}$	$-\frac{1}{4}$	$-\frac{1}{36}$	0	$\frac{1}{60}$	0
17	0	$-\frac{1}{120}$	0	0	$-\frac{1}{480}$	$\frac{1}{96}$	$\frac{7}{240}$	$-\frac{1}{8}$	$-\frac{1}{72}$	0	$\frac{1}{120}$	0

An example of order calculation

Consider the general linear method

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] = \left[\begin{array}{ccc|cc} 0 & 0 & 0 & 1 & 0 \\ \frac{3}{4} & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ -2 & 2 & 0 & 2 & -1 \\ \hline \frac{1}{6} & \frac{2}{3} & \frac{1}{6} & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{array}\right]. \tag{4.1}$$

Let η_1, η_2, η_3 denote the group elements representing the stages and assume the starting method is given by 1 for the first component and E^{-1} for the second component. The calculation of the various quantities needed to establish order are given in Table 4.2. The columns headed $\widehat{\xi}_1$ and $\widehat{\xi}_2$ represent the output computed by the method. For order 5, these would equal the $E\xi_1$ and $E\xi_2$ columns, respectively. Note that the trees are numbered in the same order as for Table 4.1.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Because of differences between $\widehat{\xi}_1$ and $E\xi_1$, the method has order only 4.

4.4. Methods with high stage order

In the formal definition of order based on Figure 4.1, the existence of $\mathcal{S}$ so that this diagram commutes to within $\mathcal{O}(h^{p+1})$ was required for the method to have order p. If this starting methods exists so that, in addition, the stages approximate y at specific points related to the current step to within $\mathcal{O}(h^{q+1})$, where $q \leq p$, then the method is said to have stage order q. If $q \geq p-1$, the combined criteria for order and stage order become much simpler.

Denote the tall tree $[_k\tau]_k$ by t_k. In the special case $k=0$, t_k will be the empty tree. Suppose that, for a starting method which gives order p and stage order q, $\xi(t_k) = w_k$. The order conditions now give

$$\eta(t_k) = A\eta_{k-1} + Uw_k, \qquad k = 1, 2, \ldots, q,$$
$$\sum_{l=0}^{k} \frac{1}{l!} w_{k-l} = B\eta_{k-1} + Vw_k, \qquad k = 1, 2, \ldots, p.$$

By the stage order conditions, $\eta(t_k)$ is the vector $c^k/k!$, where c^k denotes the component-by-component power. Furthermore, $(\eta D)(t_k) = c^{k-1}/(k-1)!$.

Write

$$C = \begin{bmatrix} 1 & c_1 & \frac{1}{2!}c_1^2 & \cdots & \frac{1}{p!}c_1^p \\ 1 & c_2 & \frac{1}{2!}c_2^2 & \cdots & \frac{1}{p!}c_2^p \\ 1 & c_3 & \frac{1}{2!}c_3^2 & \cdots & \frac{1}{p!}c_3^p \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & c_s & \frac{1}{2!}c_s^2 & \cdots & \frac{1}{p!}c_s^p \end{bmatrix}, \qquad K = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix},$$

and the *necessary* order and stage order p conditions become

$$C = ACK + UW, \tag{4.2}$$
$$WE = BCK + VW, \tag{4.3}$$

where W and E are the matrices

$$W = \begin{bmatrix} w_0 & w_1 & w_2 & \ldots & w_p \end{bmatrix}, \qquad E = \begin{bmatrix} 1 & 1 & \frac{1}{2!} & \frac{1}{3!} & \cdots & \frac{1}{p!} \\ 0 & 1 & 1 & \frac{1}{2!} & \cdots & \frac{1}{(p-1)!} \\ 0 & 0 & 1 & 1 & \cdots & \frac{1}{(p-2)!} \\ 0 & 0 & 0 & 1 & \cdots & \frac{1}{(p-3)!} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{bmatrix}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

If the stage order is only required to be $q = p - 1$, then the last column is ignored in (4.2). We will show that these are actually *sufficient* order conditions. First, however, we remark that (4.2) and (4.3) can be expressed in a different form.

Theorem 4.3. Necessary conditions for order p and stage order $q \geq p-1$ conditions are the existence of a polynomial-valued vector $\phi(z)$ such that

$$\exp(cz) = zA\exp(cz) + U\phi(z) + \mathcal{O}(z^{q+1}), \tag{4.4}$$

$$\exp(z)\phi(z) = zB\exp(cz) + V\phi(z) + \mathcal{O}(z^{p+1}). \tag{4.5}$$

Proof. Let $\phi(z) = \sum_{i=0}^{p} z^i w_i$. Add the columns of (4.2) and (4.3), in each case multiplying by the sequence of factors, 1, z, $z^2, \ldots, z^p$ and the result follows. □

Theorem 4.4. The conditions given in Theorem 4.3 are sufficient for order p and stage order q.

Note that, because of the equivalence of (4.2) (with the last column ignored if $q = p - 1$) and (4.3) with (4.4) and (4.5), we will actually prove the sufficiency of the former conditions.

Proof of Theorem 4.4. Define the starting method so that

$$y^{[0]} = \sum_{i=0}^{p} w_i h^i y^{(i)}(x_0).$$

We will first show the stage order property. That is, the stage values are

$$Y_i = \sum_{k=0}^{q} h^k c_i^k y^{(k)}(x_0)/k! + \mathcal{O}(h^{q+1}), \qquad i = 1, 2, \ldots, s.$$

This is verified by noting that these stage values imply that

$$hF_j = \sum_{k=1}^{q} h^k c_j^{k-1} y^{(k-1)}(x_0)/(k-1)! + \mathcal{O}(h^{q+2}), \qquad j = 1, 2, \ldots, s,$$

and that, because of (4.2),

$$\begin{aligned} Y_i - \sum_{j=1}^{s} a_{ij} hF_j &= \sum_{k=0}^{q} \big(C - ACK\big)_{ik} h^k y^{(k)}(x_0) + \mathcal{O}(h^{q+1}) \\ &= \sum_{k=0}^{q} (UW)_{ik} h^k y^{(k)}(x_0) + \mathcal{O}(h^{q+1}) \\ &= \sum_{j=1}^{r} u_{ij} y_j^{[0]} + \mathcal{O}(h^{q+1}). \end{aligned}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

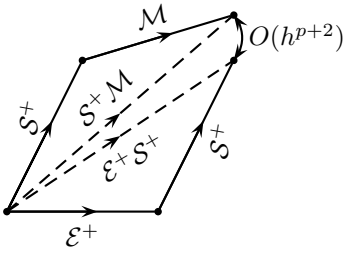


Figure 4.3. Iterative structure of underlying one-step method.

Multiplying $y^{[0]}$ by E is equivalent to replacing x_0 by $x_0 + h$ in the formula for the starting method components, to within $\mathcal{O}(h^{p+1})$. The order result for the method now follows in a similar way to the stage order result. □

4.5. The underlying one-step method

In Figure 4.1 denote the error term by ϕ. This is the Taylor expansion of $\mathcal{S}$ applied to $y(x_0 + h)$ minus the composition $\mathcal{SM}$ applied to $y(x_0)$. Resolve ϕ into two terms,

$$\phi = \epsilon u + (I - V)\delta, \tag{4.6}$$

where u is the preconsistency vector which satisfies $(I - V)u = 0$. Now construct a new diagram in which $\mathcal{E}$ is replaced by $\mathcal{E}^+ = \mathcal{E} - \epsilon$ and $\mathcal{S}$ is replaced by $\mathcal{S}^+ = \mathcal{S} - \delta$.

The meanings of $\mathcal{E}^+$ and $\mathcal{S}^+$ require some explanation. In the case of $\mathcal{E}^+$, this represents a perturbation of the flow of the differential equation in which the value of ϵ, evaluated at $y(x_0)$, is subtracted from $y(x_0 + h)$. Similarly, $\mathcal{S}^+$ represents the unperturbed starting method $\mathcal{S}$ applied to $y(x_0)$, with the vector-valued error term δ, evaluated at $y(x_0)$, subtracted from the result.

The diagram in Figure 4.1 is now replaced by Figure 4.3 so that the order, in the sense of this diagram, has been increased to $p+1$. This process can be repeated to obtain a sequence of one-step methods which we can denote by $\mathcal{E}_p = \mathcal{E}, \mathcal{E}_{p+1} = \mathcal{E}^+, \mathcal{E}_{p+2}, \ldots$, together with corresponding starting methods $\mathcal{S}_p = \mathcal{S}, \mathcal{S}_{p+1} = \mathcal{S}^+, \mathcal{S}_{p+2}, \ldots$. According to the construction of these various operations, the two compositions

$$\mathcal{S}_i\mathcal{M} \quad \text{and} \quad \mathcal{E}_i S_i$$

commute to within order i for $i = p, p + 1, p + 2, \ldots$. The underlying one-step method is the notional limit of the $\mathcal{E}_i$ sequence. This construction was proposed, in the case of linear multistep methods by Kirchgraber (1986) and extended to the case of general linear methods by Stoffer (1993).

Denote the underlying one-step method by $\mathcal{E}^\star$ and the limit of the sequence of iterated starting methods by $\mathcal{S}^\star$ and we have a new diagram corresponding to Figure 4.1 given by Figure 4.4. Now the diagram exactly

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

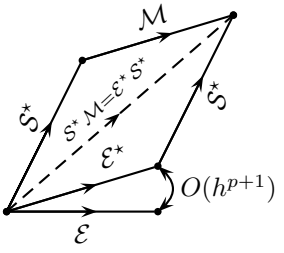


Figure 4.4. Underlying one-step method.

commutes but $\mathcal{E}$ is replaced by $\mathcal{E}^\star$. Thus, we can interpret this to mean that the general linear method behaves exactly like a one-step method. Consequently, error analysis is reduced to understanding how well $\mathcal{E}^\star$ approximates $\mathcal{E}$.

As an example of the computations involved in the analysis of the underlying one-step method, return to the method given in (4.1) and the information given in Table 4.2. From this table we see that the coefficients of $\mathcal{E}_i S_i - \mathcal{S}_i \mathcal{M}$ of the elementary differentials associated with the fifth-order trees $t_9, t_{10}, \ldots, t_{17}$ are

$$[\tfrac{1}{5}, \tfrac{1}{10}, \tfrac{1}{15}, \tfrac{1}{30}, \tfrac{1}{20}, \tfrac{1}{20}, \tfrac{1}{40}, \tfrac{1}{60}, \tfrac{1}{120}] - [\tfrac{5}{24}, \tfrac{5}{48}, \tfrac{1}{9}, \tfrac{1}{18}, \tfrac{5}{96}, \tfrac{1}{24}, \tfrac{1}{48}, -\tfrac{1}{36}, -\tfrac{1}{72}]$$
$$= [-\tfrac{1}{120}, -\tfrac{1}{240}, -\tfrac{2}{45}, -\tfrac{1}{45}, -\tfrac{1}{480}, \tfrac{1}{120}, \tfrac{1}{240}, \tfrac{2}{45}, \tfrac{1}{45}].$$

Let $\phi_1 = \sum_{i=9}^{17} \sigma(t_i)^{-1} C_i h^5 F(t_i)(y_0)$ denote the corresponding error terms, where C_i is found from this array. Because there are no error terms of this order in the second output component, we find ϵ and δ by solving the equation

$$\begin{bmatrix} \phi_1 \\ 0 \end{bmatrix} = \epsilon \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix},$$

where we have inserted the actual values of u and $I - V$. A possible solution is $\epsilon = \delta_1 = \phi_1$, $\delta_2 = 0$.

5. Linear and non-linear stability

5.1. Linear stability

The linear stability analysis of numerical methods, exemplified in Figure 1.2 is based on the test problem

$$y'(x) = qy(x). \tag{5.1}$$

The aim of this type of analysis is to investigate the existence of bounds on the step-size to achieve stable numerical behaviour for a stable problem. We will want to identify methods for which there is never any such restriction because stable behaviour will occur whenever $\operatorname{Re} hq \leq 0$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

For a general linear method (A, U, B, V), the linear problem (5.1) will produce output which satisfies the equations

$$Y = hqAY + Uy^{[n-1]}, \tag{5.2}$$

$$y^{[n]} = hqBY + Vy^{[n-1]}. \tag{5.3}$$

For convenience write $z = hq$ and solve for Y from (5.2), to give $Y = (I - zA)^{-1}Uy^{[n-1]}$. Substitute into (5.3) to give $y^{[n]} = M(z)y^{[n-1]}$, where

$$M(z) = V + zB(I - zA)^{-1}U. \tag{5.4}$$

The matrix-valued function $M(z)$ is the 'stability function' and its characteristic polynomial

$$\Phi(w, z) = \det(wI - M(z)),$$

determines its linear stability properties.

Definition 5.1. A general linear method (A, U, B, V) is A-stable if $M(z)$ is power bounded whenever $\operatorname{Re} z < 0$.

The test problem on which this definition is based can be made more realistic, but of course more difficult to analyse if q is allowed to be time-dependant. That is, we might consider the problem

$$y'(x) = q(x)y(x). \tag{5.5}$$

In this case, we need to evaluate $z = hq(x)$ at each stage value and we write the collection of all values of this quantity that occur in the step in the form of a diagonal matrix:

$$Z = \operatorname{diag}(hq(x_{n-1} + hc_1),\ hq(x_{n-1} + hc_2), \ldots, hq(x_{n-1} + hc_s)).$$

For simplicity we will assume that the c components are distinct so that there is no reason to suppose that two diagonal elements of Z are necessarily equal. The stage derivative vector is now ZY and the stage values and output values are now

$$Y = (I - AZ)^{-1}Uy^{[n-1]}, \qquad y^{[n]} = \big(V + BZ(I - AZ)^{-1}U\big)y^{[n-1]}. \tag{5.6}$$

Thus for a problem of the form (5.5), it is natural in stability considerations to replace use of $M(z)$ given by (5.4) by the matrix-valued function of r complex variables, given by

$$\widetilde{M}(Z) = V + BZ(I - AZ)^{-1}U.$$

This leads to the following definition.

Definition 5.2. A general linear method (A, U, B, V) is AN-stable if $\widetilde{M}(Z)$ given by (5.6) is power-bounded for $Z = \operatorname{diag}(z_1, z_2, \ldots, z_s)$ if

$$\operatorname{Re} z_i < 0, \qquad i = 1, 2, \ldots, s.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

5.2. Non-linear stability

To analyse stable behaviour for non-linear problems it is necessary to find a problem class for which stability of the exact solution is assured. We reintroduce (from Section 1.3) a problem made famous in Dahlquist (1976):

$$y'(x) = f(x, y(x)), \quad \langle u - v, f(x, u) - f(x, v)\rangle \leq 0. \tag{5.7}$$

Consideration of this problem led to the construction of 'one-leg methods' and to the definition of G-stability.

For a linear multistep method (ρ, σ), normalized so that $\sigma(1) = 1$, in which $\widehat{y}_n$ is computed as the solution to the equation

$$\sum_{i=0}^{k} \alpha_i \widehat{y}_{n+i} = h \sum_{i=0}^{k} \beta_i f(\widehat{x}_{n+i}, \widehat{y}_{n+i}), \tag{5.8}$$

the one-leg counterpart defines y_n as the solution to the equation

$$\sum_{i=0}^{k} \alpha_i y_{n-i} = hf\left(x_n, \sum_{i=0}^{k} \beta_i y_{n-i}\right). \tag{5.9}$$

From a linear stability point of view, these two methods have the same stability function:

$$\rho(w) - z\sigma(w) = 0.$$

Furthermore, stable behaviour, even for solutions to the non-linear problem (5.7), is closely related in the sense that if the y sequence satisfies (5.9) and

$$\widehat{x}_n = \sum_{i=0}^{k} \beta_i x_{n+i}, \qquad \widehat{y}_n = \sum_{i=0}^{k} \beta_i y_{n+i},$$

then $\widehat{y}$ satisfies (5.8). The crucial result in Dahlquist (1976) is a condition on (ρ, σ) such that, if two sequences y and $\overline{y}$ satisfy (5.9), then, for a norm $\|\cdot\|_G$, defined in the paper,

$$\|y^{[n]} - \overline{y}^{[n]}\|_G \leq \|y^{[n-1]} - \overline{y}^{[n-1]}\|_G, \tag{5.10}$$

if f satisfies the condition (5.7).

Why, it might be asked, was it necessary to introduce one-leg methods, rather than carry out an analysis of non-linear stability directly in terms of linear multistep methods?

From the general linear methods point of view, the answer is simple. A linear multistep method, in its standard formulation, is reducible, as we discussed in Section 3.3. With the irreducible formulation given by (3.4), non-linear stability for linear multistep methods can be analysed in their own right (Butcher and Hill 2006).

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Non-linear stability of Runge–Kutta methods was introduced in Butcher (1975) where it was shown that certain implicit methods, such as Gauss methods and Radau IIA methods have the property known as B-stability. A method is B-stable if, for a method satisfying (5.7),

$$\|y_n - \overline{y}_n\| \leq \|y_{n-1} - \overline{y}_{n-1}\|,$$

for two approximation sequences y and $\overline{y}$. This definition was originally introduced for an autonomous version of (5.7), but BN-stability, introduced in Burrage and Butcher (1979) made use of the general non-autonomous version of this model problem. At the same time AN-stability, where a non-autonomous linear problem is used, was introduced.

The necessary and sufficient conditions for BN-stability (and incidentally for AN-stability), at least for non-confluent methods, was shown to hinge on a matrix M given by

$$\mathsf{M} = \operatorname{diag}(b)A + A^T \operatorname{diag}(b) - bb^T.$$

It was shown in Burrage and Butcher (1979) and Crouzeix (1979) that a method is B-stable if M and $\operatorname{diag}(b)$ are each positive semi-definite. In an unpublished report, Dahlquist and Jeltsch showed that the elements of b must actually be positive for B-stability, or the method can be reduced to a simpler method, with fewer stages.

Linear and non-linear stability were analysed and inter-related in a series of papers (Burrage and Butcher 1980, Burrage 1980, Butcher 1981*b*, 1987*b*). The criterion, referred to as algebraic stability, which generalizes Dahlquist's G-stability criterion and the Runge–Kutta criterion based on M, makes use of the matrix

$$\widehat{\mathsf{M}} = \begin{bmatrix} DA + A^T D - B^T GB & DU - B^T GV \\ U^T D - V^T GB & G - V^T GV \end{bmatrix}. \tag{5.11}$$

In (5.11), G is a positive-definite matrix and D is a diagonal matrix of positive numbers. Under various conditions, it was shown in Butcher (1987*c*) that $\widehat{\mathsf{M}}$ positive semi-definite is equivalent to AN-stability and to stable behaviour in the sense of (5.10), for problems satisfying (5.7).

An interesting example of an algebraically stable method is the following, discovered by Dekker (1981), and shown by him to have this property:

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|cc} \frac{2}{3} & 0 & 1 & -\frac{7}{6} \\ \frac{2}{3} & \frac{2}{3} & 1 & \frac{1}{6} \\ \hline \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ -\frac{1}{11} & \frac{7}{11} & 0 & \frac{5}{11} \end{array}\right]. \tag{5.12}$$

To verify algebraic stability, substitute $G = \operatorname{diag}(1, \frac{11}{12})$, $D = \operatorname{diag}(\frac{1}{2}, \frac{1}{2})$ in (5.11). Like Runge–Kutta methods, but unlike linear multistep methods,

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

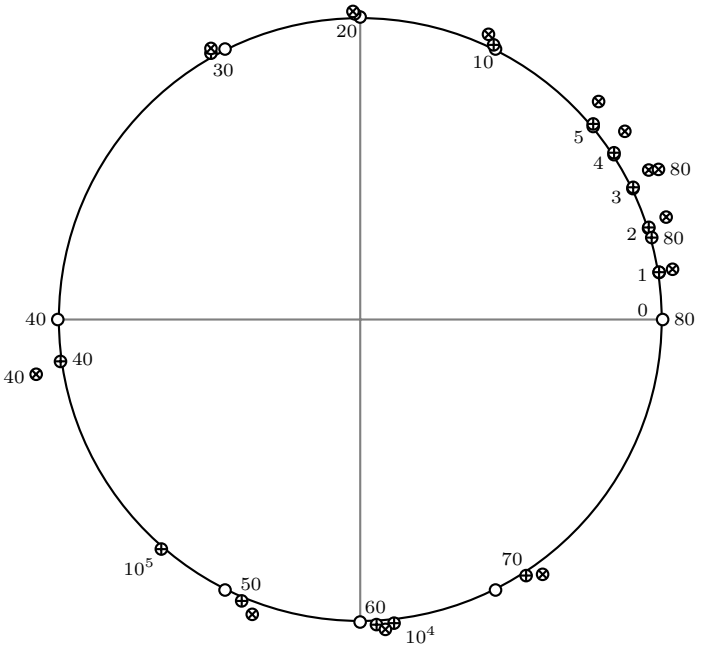


Figure 5.1. Solution of (5.15) using (5.12) (⊕) and (5.13) (⊗), compared with the exact solution (○).

algebraic stability is *not* equivalent to A-stability. For example, compare with (5.12) the method

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|cc} \frac{2}{3} & 0 & 1 & -\frac{7}{6} \\ \frac{2}{3} & \frac{2}{3} & 1 & \frac{1}{6} \\ \hline \frac{179}{88} & -\frac{19}{88} & 1 & -\frac{9}{11} \\ \frac{23}{44} & \frac{1}{44} & 0 & \frac{5}{11} \end{array}\right]. \tag{5.13}$$

The two methods have the same stability functions, and are thus each A-stable. Furthermore, the stage abscissa vector is $[-\frac{1}{2}, \frac{3}{2}]$ in each case. However, (5.13) is not algebraically stable.

As an example, it should be expected that a problem of the form

$$y'(x) = L(x, y(x))y(x), \tag{5.14}$$

where L takes values on the set of $N \times N$ matrices such that the symmetric part of $-L$ is positive semi-definite, will exhibit stable behaviour when solved using (5.12) but not when solved using (5.13).

In particular consider the initial value problem

$$\begin{bmatrix} y_1'(x) \\ y_2'(x) \end{bmatrix} = \left(y_1(x)^2 + \tfrac{1}{4}y_2(x)^2\right) \begin{bmatrix} -y_2(x) \\ y_1(x) \end{bmatrix}, \quad \begin{bmatrix} y_1(0) \\ y_2(0) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{5.15}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

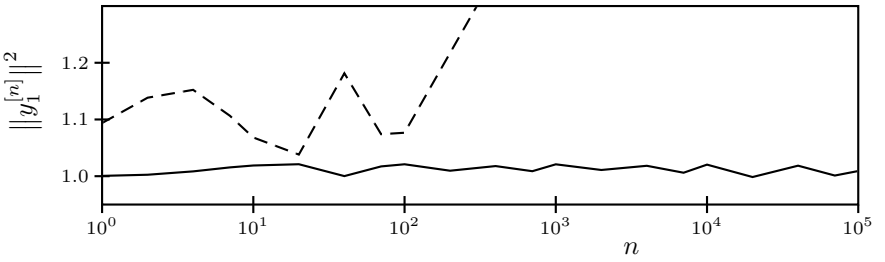


Figure 5.2. $\|y_1^{[n]}\|^2$ after n steps,
using (5.12) (——) and (5.13) (- - -).

For this problem, L takes on skew-symmetric values, so that $\|y(x)\|^2$ is not simply bounded, but is invariant.

To solve (5.15), using either (5.12) or (5.13), an appropriate starting value is $y_1^{[0]} = [1,0]^T$, $y_2^{[0]} = [0,h]^T$. Results found from these methods using constant step-size $h = \pi/20$ are presented in Figure 5.1 for n steps up to $n = 80$, which corresponds to a single period in the exact solution. In addition, $n = 10^4$ and 10^5 are also given in the case of (5.12). Even though there is a considerable phase shift, in accordance with the low accuracy of the numerical approximation, it is seen that the solutions using (5.12) adhere closely to the $\|y(x)\|^2 = 1$ manifold. This is explored further in Figure 5.2, where $\|y_1^{[n]}\|^2$ is evaluated for (5.12) up to $n = 10^5$ and for (5.13) until the value of this quantity drifts too far away. Also computed, but not explicitly shown, is the value of $\|y_1^{[n]}\|^2 + \frac{11}{12}\|y_2^{[n]}\|^2$, which is virtually constant.

We will now give a partial explanation of this phenomenon. First we note that, for (5.12), with the values of G and D that have been proposed, the partitioned matrix in (5.11) has the value

$$\widehat{\mathsf{M}} = \tfrac{8}{11}\begin{bmatrix} \frac{3}{4} \\ \frac{1}{4} \\ 0 \\ -1 \end{bmatrix}\begin{bmatrix} \frac{3}{4} & \frac{1}{4} & 0 & -1 \end{bmatrix},$$

so that we can evaluate the following inner product:

$$\begin{bmatrix} hF^T & (y^{[n-1]})^T \end{bmatrix}\widehat{\mathsf{M}}\begin{bmatrix} hF \\ y^{[n-1]} \end{bmatrix} = \tfrac{8}{11}\|\tfrac{3}{4}hF_1 + \tfrac{1}{4}hF_2 - y_2^{[n-1]}\|^2. \qquad (5.16)$$

Doing the calculation another way we find that (5.16) can be written as

$$\begin{aligned} hF^T D(hAF + Uy^{[n-1]}) + (hAF + Uy^{[n-1]})^T DhF \\ - (hBF + Vy^{[n-1]})^T G(hBF + Vy^{[n-1]}) \\ + (y^{[n-1]})^T G(y^{[n-1]}). \end{aligned} \qquad (5.17)$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Write v^TGv as $\|v\|_G^2$ with the corresponding inner product equal to $\langle u, v\rangle = u^TGv$ and (5.17) simplifies to

$$2\sum_{i=1}^{s} d_i\langle hF_i, Y_i\rangle + \|y^{[n]}\|_G^2 - \|y^{[n-1]}\|_G^2.$$

With the assumed form of the differential equation, we deduce

$$\|y^{[n]}\|_G = \|y^{[n-1]}\|_G - \tfrac{8}{11}\|\tfrac{3}{4}hF_1 + \tfrac{1}{4}hF_2 - y_2^{[n-1]}\|^2,$$

so that we cannot expect $\|y^{[n]}\|_G$ to be invariant. However, for the problem we are considering, $\frac{3}{4}hF_1 + \frac{1}{4}hF_2 - y_2^{[n-1]}$ has a small norm which decreases rapidly if $\|y^{[n]}\|_G$ becomes small. If we replace the method (5.12) by one in which, for an appropriate D and G, $\widehat{\mathsf{M}}$ is the zero matrix, then we can expect precise invariance of $\|y^{[n]}\|_G$. For example,

$$\left[\begin{array}{cc} A & U \\ B & V \end{array}\right] = \left[\begin{array}{cc|cc} \frac{1}{2} & 0 & 1 & -\frac{1}{2} \\ 1 & \frac{1}{2} & 1 & -\frac{1}{2} \\ \hline \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ 1 & 1 & 0 & -1 \end{array}\right], \tag{5.18}$$

for which we need to use $G = \mathrm{diag}(1, \frac{1}{4})$, $D = \mathrm{diag}(\frac{1}{2}, \frac{1}{2})$. The invariant behaviour of $\|y^{[n]}\|_G^2 = (y_1^{[n]})^2 + \frac{1}{4}(y_2^{[n]})^2$ is verified by numerical experiment.

This method has order only 2 and does not seem to have any real advantages over the implicit mid-point rule method given by

$$\left[\begin{array}{cc} A & U \\ B & V \end{array}\right] = \left[\begin{array}{c|c} \frac{1}{2} & 1 \\ \hline 1 & 1 \end{array}\right],$$

However, it is possible to construct more accurate methods such as

$$\left[\begin{array}{cc} A & U \\ B & V \end{array}\right] = \left[\begin{array}{cc|cc} \frac{3+\sqrt{3}}{6} & 0 & 1 & -\frac{3+2\sqrt{3}}{3} \\ -\frac{\sqrt{3}}{3} & \frac{3+\sqrt{3}}{6} & 1 & \frac{3+2\sqrt{3}}{3} \\ \hline \frac{1}{2} & \frac{1}{2} & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 0 & -1 \end{array}\right], \tag{5.19}$$

which has order 4, as we see below. As for (5.18), $\widehat{\mathsf{M}} = 0$; but in this case we use $G = \mathrm{diag}(1, 1 + \frac{2\sqrt{3}}{3})$, $D = \mathrm{diag}(\frac{1}{2}, \frac{1}{2})$.

Because (5.19) is symplectic, in a slightly more general sense than applies to Runge–Kutta methods, it has a potential role in structure-preserving algorithms. Before we discuss this question, we verify its order.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Theorem 5.3. The order of (5.19) is 4.

Proof. Given an input approximation

$$y^{[0]} = \begin{bmatrix} y(x_0) \\ \frac{\sqrt{3}}{12}h^2y''(x_0) - \frac{\sqrt{3}}{108}h^4y^{(4)}(x_0) + \frac{9+5\sqrt{3}}{216}h^4\frac{\partial f}{\partial y}y^{(4)}(x_0) \end{bmatrix}, \tag{5.20}$$

we need to verify that the output is

$$y^{[1]} = \begin{bmatrix} y(x_0)+hy'(x_0)+\frac{1}{2}h^2y''(x_0)+\frac{1}{6}h^3y^{(3)}+\frac{1}{24}h^4y^{(4)}+\mathcal{O}(h^5) \\ \frac{\sqrt{3}}{12}h^2y''(x_0) + \frac{\sqrt{3}}{12}h^3y^{(3)}(x_0)+ \\ \qquad \frac{7\sqrt{3}}{216}h^4y^{(4)}(x_0) + \frac{9+5\sqrt{3}}{216}h^4\frac{\partial f}{\partial y}y^{(4)}(x_0) + \mathcal{O}(h^5) \end{bmatrix}, \tag{5.21}$$

found by replacing x_0 by $x_1 = x_0 + h$ and expanding about x_0. By Taylor expansions we find

$$\begin{aligned} Y_1 &= y\big(x_0 + h\tfrac{3+\sqrt{3}}{6}\big) + \tfrac{9+5\sqrt{3}}{108}h^3y^{(3)}(x_0) + \mathcal{O}(h^4), \\ hF_1 &= hy'\big(x_0 + h\tfrac{3+\sqrt{3}}{6}\big) + \tfrac{9+5\sqrt{3}}{108}h^4\tfrac{\partial f}{\partial y}y^{(3)}(x_0) + \mathcal{O}(h^5), \end{aligned} \tag{5.22}$$

$$\begin{aligned} Y_2 &= y\big(x_0 + h\tfrac{3-\sqrt{3}}{6}\big) - \tfrac{9+5\sqrt{3}}{108}h^3y^{(3)}(x_0) + \mathcal{O}(h^4), \\ hF_2 &= hy'\big(x_0 + h\tfrac{3-\sqrt{3}}{6}\big) - \tfrac{9+5\sqrt{3}}{108}h^4\tfrac{\partial f}{\partial y}y^{(3)}(x_0) + \mathcal{O}(h^5). \end{aligned} \tag{5.23}$$

Using (5.22) and (5.23), evaluate $y^{[1]} = hAF + Vy^{[0]}$ by Taylor expansions, to obtain agreement with (5.21). □

5.3. Experiments with a Hamiltonian problem

Consider the simple-pendulum problem

$$\begin{aligned} \dot{p} &= -\sin(q), & p(0) &= 1, \\ \dot{q} &= p, & q(0) &= 0. \end{aligned} \tag{5.24}$$

This is based on the Hamiltonian $H(p,q) = \frac{1}{2}p^2 - \cos(q)$, where we note that

$$\dot{p} = -\frac{\partial H}{\partial q}, \qquad \dot{q} = \frac{\partial H}{\partial p}.$$

Because of the initial values assigned to (5.24), the dependent variables lie in the intervals

$$p \in [-1, 1], \quad q \in [-\tfrac{1}{3}\pi, \tfrac{1}{3}\pi]$$

and the period is calculated to be $T = 6.743001419251$.

Attempts to solve this problem using the Euler and implicit Euler methods, are shown in Figure 5.3, with the exact solution also given for

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

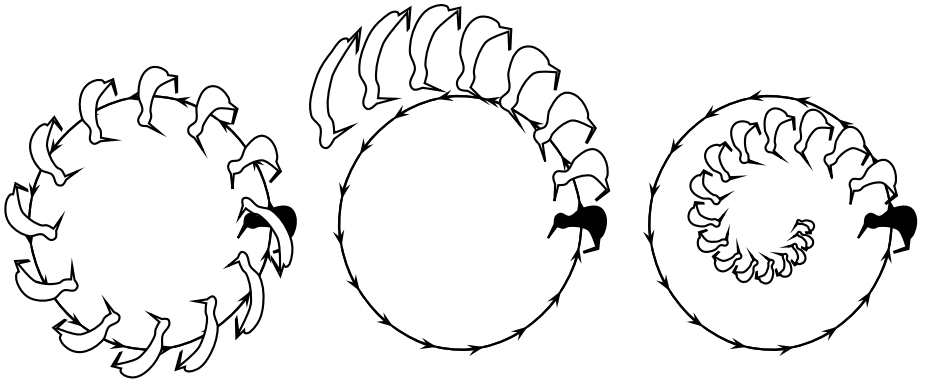


Figure 5.3. Left to right: exact solution, Euler method, implicit Euler method.

comparison. A step-size $\frac{1}{16}T$ is used and the computations are confined to the interval $[0, T]$, except for the Euler case which leaves the field of view after only 7 time-steps.

To illustrate symplectic behaviour for the exact solution, and to indicate that it does not occur for the two numerical approximations, a set of initial points, shown in black, is used at time zero. For the exact symplectic result, even though the set of points has its shape distorted, the area remains unchanged. For the Euler and implicit Euler methods, however, not only do the computed results drift rapidly away from the correct trajectory, but the areas change in size.

We will consider the use of three alternative methods. These are:

(i) the order 4 Gauss Runge–Kutta method with defining matrices

$$\left[\begin{array}{cc} A & U \\ B & V \end{array}\right] = \left[\begin{array}{cc|c} \frac{1}{4} & \frac{1}{4}-\frac{1}{6}\sqrt{3} & 1 \\ \frac{1}{4}+\frac{1}{6}\sqrt{3} & \frac{1}{4} & 1 \\ \hline \frac{1}{2} & \frac{1}{2} & 1 \end{array}\right];$$

(ii) the order 2 Gauss method, usually referred to as 'the mid-point rule method', defined by

$$\left[\begin{array}{cc} A & U \\ B & V \end{array}\right] = \left[\begin{array}{c|c} \frac{1}{2} & 1 \\ \hline 1 & 1 \end{array}\right];$$

(iii) the general linear method given by (5.19).

The general linear method requires a starting procedure to produce input

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

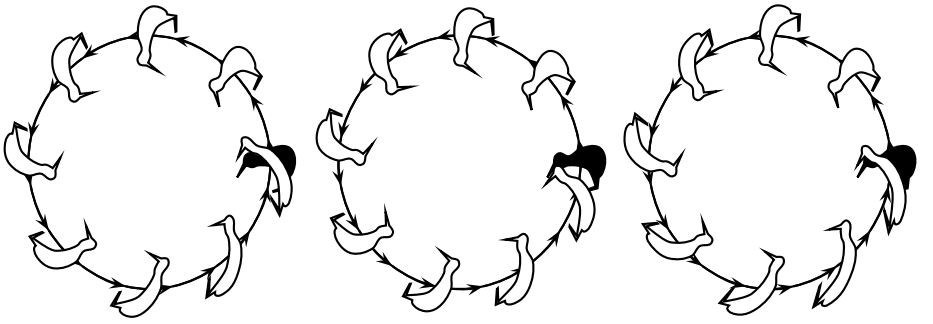


Figure 5.4. Left to right: exact solution, mid-point rule, symplectic general linear method.

for the first step of the method and it is proposed to use

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|c} \frac{3+\sqrt{3}}{6} & 0 & 1 \\ -\frac{3+\sqrt{3}}{3} & \frac{3+\sqrt{3}}{6} & 1 \\ \hline 0 & 0 & 1 \\ \frac{\sqrt{3}-1}{8} & \frac{1-\sqrt{3}}{8} & 0 \end{array}\right].$$

This simply passes through the initial value $y(x_0)$ as the first component $y_1^{[0]}$ and produces an approximation to $\frac{1}{12}\sqrt{3}h^2y''(x_0)$ as the value of $y_2^{[0]}$.

To evaluate the behaviour of these methods, a greater step-size than was used in Figure 5.3 is needed, otherwise the errors will be imperceptible; we will use a step-size $\frac{1}{8}T$. Even with steps this large, it is impossible to distinguish method (i) from the exact solution and we therefore omit this case from the results we present. The exact result and methods (ii) and (iii) are shown in Figure 5.4. We see that the mid-point rule exhibits inexact behaviour but, because it is symplectic, the areas of the solution sets do not change. Similarly, the general linear method performs very well and also appears to preserve areas.

6. Special families of methods

6.1. Re-use methods and two-step Runge–Kutta methods

The idea of using derivative approximations, computed in a previous step, as contributing to the computation of the current step, was proposed in Butcher (1966). It has recently been developed under the name 'two-step Runge–Kutta methods'. Although we will not attempt to survey this large body of work in detail, we will derive order conditions for a family of these methods.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

As a general linear method, the inputs to step n are the computed approximations to $y(x_{n-2})$ and $y(x_{n-1})$ together with the s scaled stage derivatives computed within step number $n-1$. Hence we write

$$y^{[n-1]} = \begin{bmatrix} y_{n-1} \\ y_{n-2} \\ hF_1^{[n-1]} \\ hF_2^{[n-1]} \\ \vdots \\ hF_s^{[n-1]} \end{bmatrix}, \quad y^{[n]} = \begin{bmatrix} y_n \\ y_{n-1} \\ hF_1^{[n]} \\ hF_2^{[n]} \\ \vdots \\ hF_s^{[n]} \end{bmatrix}, \quad Y^{[n]} = \begin{bmatrix} Y_1^{[n]} \\ Y_2^{[n]} \\ \vdots \\ Y_s^{[n]} \end{bmatrix}, \quad F^{[n]} = \begin{bmatrix} F_1^{[n]} \\ F_2^{[n]} \\ \vdots \\ F_s^{[n]} \end{bmatrix}$$

and we write the coefficient matrix in the partitioned form

$$\left[\begin{array}{c|c} A & U \\ \hline B & V \end{array}\right] = \left[\begin{array}{c|cc|c} A & u & e-u & \overline{A} \\ \hline b^T & \theta & 1-\theta & \overline{b}^T \\ 0 & 1 & 0 & 0 \\ \hline I & 0 & 0 & 0 \end{array}\right].$$

This method is also conveniently written using an extension of the standard tableau for Runge–Kutta methods,

$$\begin{array}{c|c|c|c} c & u & \overline{A} & A \\ \hline & \theta & \overline{b}^T & b^T \end{array},$$

indicating that $Y_i^{[n]}$, $i = 1, 2, \ldots, s$ and y_n are computed using the formulae

$$Y_i^{[n]} = u_i y_{n-2} + (1-u_i) y_{n-1} + h \sum_{j=1}^{s} (\overline{a}_{ij} F_j^{[n-1]} + a_{ij} F_j^{[n]}), \tag{6.1}$$

$$F_i^{[n]} = f(Y_i^{[n]}), \tag{6.2}$$

$$y_n = \theta y_{n-2} + (1-\theta) y_{n-1} + h \sum_{i=1}^{s} (\overline{b}_i F_i^{[n-1]} + b_i F_i^{[n]}). \tag{6.3}$$

To find the order conditions, write $\eta \in X_1^s$, to represent the stage values. We then have

$$\eta = uE^{-1} + (\mathbf{1} - u) + \overline{A}E^{-1}\eta D + A\eta D. \tag{6.4}$$

The values of $\eta_i(t)$, $i = 1, 2, \ldots, s$ are found recursively and these are then substituted into the order equation

$$E(t) = \theta E^{-1}(t) + \overline{b}^T (E^{-1}\eta D)(t) + b^T (\eta D)(t), \qquad r(t) \le p.$$

If we are going to require that $\eta(\tau) = c$, with c prescribed in advance, then

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 6.1. Analysis of re-use method.

$\eta(\bullet)$	c
$(\eta D)(\bullet)$	$\mathbf{1}$
$(E^{-1}\eta)(\bullet)$	$c-\mathbf{1}$
$(E^{-1}\eta D)(\bullet)$	$\mathbf{1}$
$\eta(\text{I})$	$\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac$
$(\eta D)(\text{I})$	c
$(E^{-1}\eta)(\text{I})$	$\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac-c+\frac{1}{2}\mathbf{1}$
$(E^{-1}\eta D)(\text{I})$	$c-\mathbf{1}$
$\eta(\text{V})$	$-\frac{1}{3}u+\overline{A}(c-\mathbf{1})^2+Ac^2$
$(\eta D)(\text{V})$	c^2
$(E^{-1}\eta)(\text{V})$	$2c-\frac{1}{3}\mathbf{1}-\frac{4}{3}u+\overline{A}(c^2-4c+3\mathbf{1})+A(c^2-2c)$
$(E^{-1}\eta D)(\text{V})$	$(c-\mathbf{1})^2$
$\eta(\text{Y})$	$-\frac{1}{6}u+\overline{A}(\frac{1}{2}u+\overline{A}(c-1)+Ac-c+\frac{1}{2}\mathbf{1})$ $+A(\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac)$
$(\eta D)(\text{Y})$	$\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac$
$(E^{-1}\eta)(\text{Y})$	$\frac{1}{2}c-\frac{1}{6}\mathbf{1}-\frac{2}{3}u+\overline{A}(\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac-2c+\frac{3}{2}\mathbf{1})$ $+A(\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac-c)$
$(E^{-1}\eta D)(\text{Y})$	$\frac{1}{2}u+\overline{A}(c-\mathbf{1})+Ac-c+\frac{1}{2}\mathbf{1}$

the conditions become simpler. The additional condition is equivalent to

$$c=-u+(\overline{A}+A)\mathbf{1} \tag{6.5}$$

and this can always be satisfied by the choice of u.

We will present in Table 6.1 formulae for $\eta(t)$ and associated quantities up to order 3. It will be assumed throughout that c is defined by (6.5). It is now a routine task to compute the order conditions up to order 4. However, we will avoid the full generality of this task and settle for the case defined by $s=2$, $u=0$, $\theta=0$ and $c=[\frac{1}{2},1]^T$. The order conditions associated with the trees $\bullet$, I, V, W enable b^T and $\overline{b}^T$ to be found. These are

$$\overline{b}^T=[0,\tfrac{1}{6}],\qquad b^T=[\tfrac{2}{3},\tfrac{1}{6}].$$

The order conditions associated with the trees Y, U and T become

$$\begin{aligned}-\tfrac{1}{3}\overline{a}_{11}-\tfrac{1}{6}\overline{a}_{21}+\tfrac{1}{6}a_{21}&=\tfrac{1}{4},\\ -\tfrac{1}{6}\overline{a}_{11}-\tfrac{1}{12}\overline{a}_{21}+\tfrac{1}{12}a_{21}&=\tfrac{1}{8},\\ -\tfrac{1}{6}\overline{a}_{11}-\tfrac{1}{4}\overline{a}_{21}+\tfrac{1}{12}a_{21}&=\tfrac{7}{36}.\end{aligned}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

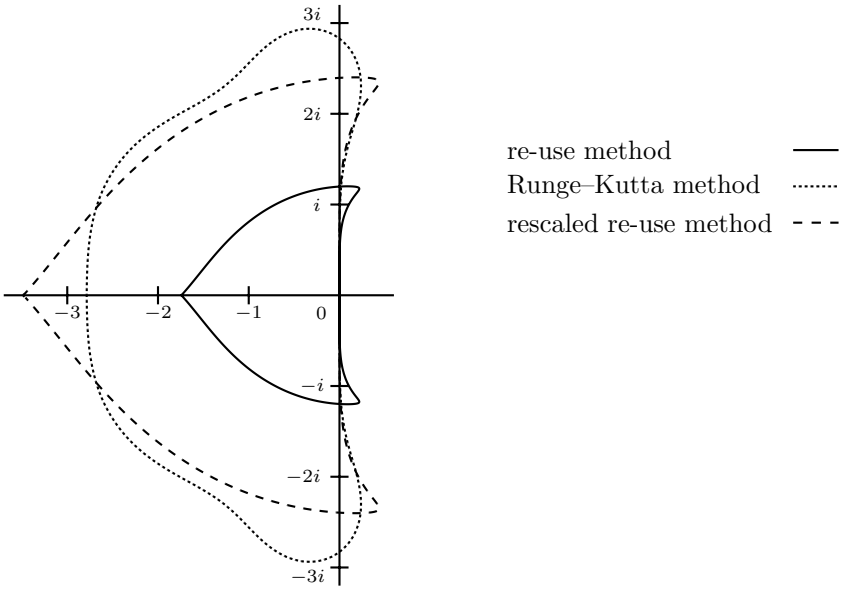


Figure 6.1. Stability region for re-use method.

These have solution

$$\overline{a}_{11} = \tfrac{1}{2}a_{21} - \tfrac{13}{24}, \qquad \overline{a}_{21} = -\tfrac{5}{12},$$

and the order condition associated with the remaining tree $\left\{\right.$ gives $a_{21} = \frac{13}{12}$ or $a_{21} = \frac{11}{12}$. Choose the second of these and we find the tableau for the method to be

$$\begin{array}{c|cc|cc|cc} \frac{1}{2} & 0 & & -\frac{1}{12} & \frac{7}{12} & 0 & 0 \\ 1 & 0 & & -\frac{5}{12} & \frac{1}{2} & \frac{11}{12} & 0 \\ \hline & 0 & & 0 & \frac{1}{6} & \frac{2}{3} & \frac{1}{6} \end{array}.$$

The stability region for this method is shown in Figure 6.1. As in Figure 2.1, we also give a rescaled stability region to allow for the fact that this method has only two stages, compared with four for the Runge–Kutta method.

Recently methods based on re-use of quantities computed in the previous step have been investigated under the name 'two-step Runge–Kutta methods'. Basic references on these methods are Jackiewicz and Tracogna (1995, 1996) and Bartoszewski and Jackiewicz (1998).

6.2. ARK methods

The aim of 'almost Runge–Kutta' or ARK methods is impose on re-use methods the additional requirement expressed in the following.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Definition 6.1. A general linear method is RK-stable (possesses Runge–Kutta stability) if its stability matrix has only a single nonzero eigenvalue.

This means that, for a method with r values passed from step to step and stability matrix $M(z)$,

$$\det(wI - M(z)) = w^{r-1}(w - R(z)).$$

The function $R(z)$, because it equals the trace of $M(z)$, is a rational function, and in the case of explicit methods a polynomial, and will be referred to as the stability function for the method.

Before formulating ARK methods, we will make a brief remark about traditional Runge–Kutta methods. The classical Kutta method, as a general linear method, is

$$\left[\begin{array}{cccc|c} 0 & 0 & 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 & 0 & 1 \\ 0 & \frac{1}{2} & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 \\ \hline \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & 1 \end{array}\right]. \tag{6.6}$$

It is also possible to formulate this as a multivalue method with $r = 2$:

$$\left[\begin{array}{cccc|cc} 0 & 0 & 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{2} & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} \\ \hline \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} \\ 0 & 0 & 0 & 1 & 1 & 0 \end{array}\right], \tag{6.7}$$

where the derivative from the first stage in (6.6) is now computed as the *last* stage of (6.7). The two output quantities are approximations to the first terms in the Taylor series at the start of the following step:

$$y_1^{[n]} \approx y(x_n), \qquad y_2^{[n]} \approx hy'(x_n).$$

Make a similar change to the re-use method (2.3) and renumber the existing $y_2^{[n]}$ as $y_3^{[n]}$:

$$\left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & \frac{5}{8} & -\frac{1}{8} \\ 2 & 0 & 0 & 1 & -\frac{3}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ \hline \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{array}\right]. \tag{6.8}$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Instead of passing on $y_3^{[n]} \approx hy'(x_{n-1})$, in addition to $y(x_n)$, and $hy'(x_n)$, it is possible to replace $y_3^{[n]}$ by an approximation to $h^2y''(x_n)$ which can be found as the difference between approximations to $hy'(x_n)$ and $hy'(x_{n-1})$. This gives the equivalent but more convenient formulation

$$\left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & \frac{1}{2} & \frac{1}{8} \\ 2 & 0 & 0 & 1 & -1 & -\frac{1}{2} \\ \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ \hline \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{array}\right]. \tag{6.9}$$

Even though we now pass on approximations to $y(x), hy'(x), h^2y''(x)$ from step to step, the third of these is accurate only to within $\mathcal{O}(h^3)$. This is not a serious handicap because the consequence of this inaccuracy cancels out to within $\mathcal{O}(h^5)$ because $y_3^{[n]}$ appears only within the arguments of the first and second scaled stage derivatives and because the first row of B is orthogonal to the last column of U. Properties like this are referred to as 'annihilation conditions' and are crucial to the design of ARK methods.

This method cannot possibly possess RK stability but if we restore the value of s to 4, this does become possible. The derivation of methods up to order 4 is given in Butcher (1997*b*). The following example is based on the abscissae $c = [1, \frac{1}{2}, 1, 1]$:

$$\left[\begin{array}{cccc|ccc} 0 & 0 & 0 & 0 & 1 & 1 & \frac{1}{2} \\ \frac{1}{16} & 0 & 0 & 0 & 1 & \frac{7}{16} & \frac{1}{16} \\ -\frac{1}{4} & 2 & 0 & 0 & 1 & -\frac{3}{4} & -\frac{1}{4} \\ 0 & \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ \hline 0 & \frac{2}{3} & \frac{1}{6} & 0 & 1 & \frac{1}{6} & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ -\frac{1}{3} & 0 & -\frac{2}{3} & 2 & 0 & -1 & 0 \end{array}\right].$$

In a variable step-size implementation, when h changes to rh between steps $n-1$ and n, a simple rescaling of $y_2^{[n-1]} \approx hy'(x_{n-1})$ by a factor r and $y_3^{[n-1]} \approx h^2y''(x_{n-1})$ by a factor r^2 is adequate to preserve fourth-order behaviour.

It is possible to find a five-stage fourth-order ARK method with the special property that its error constants exactly vanish. Unfortunately, the annihilation conditions satisfied by this method are not sufficient for this method to act like a fifth-order method when implemented in a manner in which variable step-size is dealt with by simple rescaling. However, it is

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

possible to adjust things with negligible additional work so that variable h fifth-order behaviour is achieved. Methods with this effectively fifth-order behaviour are presented in Butcher and Moir (2003) and Rattenbury (2005). Along with methods of effective order five, they present a means of breaking the order barrier on explicit Runge–Kutta methods.

Although ARK methods were originally designed for non-stiff problems, it has recently been found how to adapt them for the solution of stiff problems (Butcher and Rattenbury 2005, Rattenbury 2005). Extensive numerical testing shows this type of method to be very competitive for many stiff problems.

6.3. DIMSIM methods

In the search for practical general linear methods a systematic family was sought such that $p = q = r = s$ and such that, if possible, they possessed RK stability. Because the structure of the matrix A plays a crucial role in the implementation cost in both sequential and parallel environments, it seemed to be a good design choice to consider only lower triangular matrices in this role. Furthermore there are often advantages in forcing the diagonal elements to be equal and we will assume this to be the case. Methods designed with these considerations in mind are referred to as a *diagonally implicit multi-stage integration method* or DIMSIM (Butcher 1995). Because applications are needed for both stiff and non-stiff problems and because we will want to consider parallel as well as sequential architectures, four types of methods, determined by the structure of A are considered and these are summarized in Table 6.2 (overleaf).

Type 1 and 2 methods

The following is an example of a type 1 DIMSIM with $p = q = r = s = 2$:

$$\left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 2 & 0 & 0 & 1 \\ \hline \frac{5}{4} & \frac{1}{4} & \frac{1}{2} & \frac{1}{2} \\ \frac{3}{4} & -\frac{1}{4} & \frac{1}{2} & \frac{1}{2} \end{array}\right].$$

Even though this method has the same stability region as the classical Runge–Kutta methods of Runge, it has advantages associated with stage order $q = 2$. In particular, at no additional cost it yields interpolated results, suitable for dense output or application to certain delay differential equations. Furthermore, asymptotically correct local error estimates are available. Variable step-size, of course, presents complications which are not present for the corresponding Runge–Kutta methods. However, there are satisfactory ways round these complications.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 6.2. The four DIMSIM types.

	Structure of A	Stiffness type	Architecture
type 1	$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ a_{21} & 0 & 0 & \cdots & 0 \\ a_{31} & a_{32} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{s1} & a_{s2} & a_{s3} & \cdots & 0 \end{bmatrix}$	nonstiff	sequential
type 2	$\begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ a_{21} & \lambda & 0 & \cdots & 0 \\ a_{31} & a_{32} & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{s1} & a_{s2} & a_{s3} & \cdots & \lambda \end{bmatrix}$	stiff	sequential
type 3	$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$	nonstiff	parallel
type 4	$\begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ 0 & 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix}$	stiff	parallel

A similar method, but of order and stage order 3, has the coefficient matrix (Butcher and Jackiewicz 1996)

$$\left[\begin{array}{ccc|ccc} 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ \frac{1}{4} & 1 & 0 & 0 & 0 & 1 \\ \hline \frac{5}{4} & \frac{1}{3} & \frac{1}{6} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3} \\ \frac{35}{24} & -\frac{1}{3} & \frac{1}{8} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3} \\ -\frac{17}{12} & 0 & \frac{1}{12} & -\frac{2}{3} & \frac{4}{3} & \frac{1}{3} \end{array}\right].$$

The construction of higher-order type 1 DIMSIMs becomes increasingly complicated and numerical searches have to be made (Butcher and Jackiewicz 2004, Butcher, Jackiewicz and Mittelmann 1997). However, order 4 methods have been found by Wright (2001).

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Type 2 methods with $p = q = r = s = 2$ are easy to find, for example, an L-stable method:

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|cc} \frac{2-\sqrt{2}}{2} & 0 & 1 & 0 \\ \frac{6+2\sqrt{2}}{7} & \frac{2-\sqrt{2}}{2} & 0 & 1 \\ \hline \frac{73-34\sqrt{2}}{28} & \frac{4\sqrt{2}-5}{4} & \frac{3-\sqrt{2}}{2} & \frac{\sqrt{2}-1}{2} \\ \frac{87-48\sqrt{2}}{28} & \frac{34\sqrt{2}-45}{28} & \frac{3-\sqrt{2}}{2} & \frac{\sqrt{2}-1}{2} \end{array}\right].$$

For higher orders, L-stable type 2 methods are also increasingly difficult to construct. However, the following method with $p = q = r = s = 3$ is A-stable:

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{ccc|ccc} \frac{1}{2} & 0 & 0 & 1 & 0 & 0 \\ \frac{5}{4} & \frac{1}{2} & 0 & 0 & 1 & 0 \\ \frac{7}{5} & \frac{4}{5} & \frac{1}{2} & 0 & 0 & 1 \\ \hline \frac{14}{15} & \frac{1}{5} & -\frac{1}{12} & \frac{5}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{17}{20} & \frac{7}{60} & -\frac{1}{6} & \frac{5}{6} & \frac{1}{3} & -\frac{1}{6} \\ \frac{23}{30} & \frac{2}{15} & -\frac{1}{20} & \frac{5}{6} & \frac{1}{3} & -\frac{1}{6} \end{array}\right].$$

Type 3 and 4 methods

It is impossible to obtain RK stability for high-order methods in these families. However, reasonable stability regions are possible for type 3 methods, as in the example with $p = q = r = s = 2$:

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|cc} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline -\frac{3}{8} & -\frac{3}{8} & -\frac{3}{4} & \frac{7}{4} \\ -\frac{7}{8} & \frac{9}{8} & -\frac{3}{4} & \frac{7}{4} \end{array}\right].$$

The error constant for this method has magnitude $\frac{19}{24}$, which is abnormally large, even allowing for any possible gain due to parallelism.

An example of a type 4 method also with $p = q = r = s = 2$ is

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cc|cc} \frac{3-\sqrt{3}}{2} & 0 & 1 & 0 \\ 0 & \frac{3-\sqrt{3}}{2} & 0 & 1 \\ \hline \frac{18-11\sqrt{3}}{4} & -\frac{12+7\sqrt{3}}{4} & \frac{3}{2}-\sqrt{3} & \sqrt{3}-\frac{1}{2} \\ \frac{22-13\sqrt{3}}{4} & -\frac{12+9\sqrt{3}}{4} & \frac{3}{2}-\sqrt{3} & \sqrt{3}-\frac{1}{2} \end{array}\right].$$

In this case the stability polynomial is

$$\left(1 - z\frac{3-\sqrt{3}}{2}\right)^2 w^2 - \left(1 - z\frac{3-\sqrt{3}}{2}\right) w + \frac{1-\sqrt{3}}{2} z,$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

and it is possible to verify that the method is A-stable and that it has zero spectral radius at infinity.

An experimental implementation of methods of type 4 is reported in Singh (1999).

7. Methods with inherent RK-stability

Even though DIMSIM methods of types 1 and 2 cannot be constructed in a systematic manner, it is possible, by increasing both r and s, to $p+1 = q+1$ to derive methods which possess Runge–Kutta stability purely as a result of their structure. These methods, which are said to possess inherent RK stability, can be constructed in various ways but it seems most convenient to use 'doubly companion matrices', and this is the approach we will use.

7.1. Doubly companion matrices

Consider a matrix of the form

$$X(\alpha,\beta) = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & \cdots & -\alpha_{n-1} & -\alpha_n-\beta_n \\ 1 & 0 & 0 & \cdots & 0 & -\beta_{n-1} \\ 0 & 1 & 0 & \cdots & 0 & -\beta_{n-2} \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & -\beta_2 \\ 0 & 0 & 0 & \cdots & 1 & -\beta_1 \end{bmatrix},$$

where

$$\alpha(z) = 1 + \alpha_1 z + \cdots + \alpha_n z^n,$$
$$\beta(z) = 1 + \beta_1 z + \cdots + \beta_n z^n.$$

If $\beta(z) = 1$, so that $\beta_1 = \beta_2 = \cdots = \beta_n$, or similarly if $\alpha(z) = 1$, then the characteristic polynomials can be found from

$$\det(I - zX(\alpha, 1)) = \alpha(z),$$
$$\det(I - zX(1, \beta)) = \beta(z).$$

We now consider the general case.

Theorem 7.1. The characteristic polynomial of $X(\alpha,\beta)$ is given by

$$\det(I - zX(\alpha,\beta)) = \alpha(z)\beta(z) + \mathcal{O}(z^{n+1}). \tag{7.1}$$

In (7.1), the effect of the term $\mathcal{O}(z^{n+1})$ is to simply remove from the expanded product $\alpha(z)\beta(z)$ all terms with degree greater than n. If we denote the usual characteristic polynomial $\det(zI - X(\alpha,\beta))$ by $\phi(z)$, then (7.1) can be interpreted to mean

$$\phi(z) = \big[z^{-n}\det(zI - X(\alpha,1))\det(zI - X(1,\beta))\big],$$

where, in this formula, $[\cdot]$ means that negative powers of z are omitted.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Proof. Define the vector-valued function $P(z)$ by

$$P(z) = \begin{bmatrix} \vdots \\ z^2 + \beta_1 z + \beta_2 \\ z + \beta_1 \\ 1 \end{bmatrix}. \tag{7.2}$$

A simple calculation shows that

$$X(\alpha, \beta)P(z) = zP(z) + \phi(z)e_1, \tag{7.3}$$

showing that $(\lambda, P(\lambda))$ are an eigenvalue–eigenvector pair if $\phi(\lambda) = 0$. □

In applications of this result, we will be given $\beta(z)$ and the characteristic polynomial of $X(\alpha, \beta)$, and we will need to find α. In particular we will need to consider the case $\det(I - zX(\alpha, \beta)) = (1 - \lambda z)^n$ and we find, in this case,

$$\alpha(z) = (1 - \lambda z)^n / \beta(z) + \mathcal{O}(z^{n+1}).$$

It is possible to find explicit formulae for transformation matrices Ψ^{-1} and Ψ so that $\Psi^{-1} X \Psi$ is in Jordan canonical form. We will specialize this to the case that $X(\alpha, \beta)$ has a one-point spectrum $\sigma(X(\alpha, \beta)) = \{\lambda\}$. In addition to $P(z)$ given by (7.2), we will need the vector-valued function $Q(z)$ given by

$$Q(z) = \begin{bmatrix} 1 & z + \alpha_1 & z^2 + \alpha_1 z + \alpha_2 & \cdots \end{bmatrix}.$$

For the remainder of this paper, we will write X in place of $X(\alpha, \beta)$, unless there is the possibility of ambiguity. For the trivial case in which all α_i and β_i are zero, we will write J for this value of X.

Theorem 7.2. Define

$$\Psi = \begin{bmatrix} \frac{1}{(n-1)!} P^{(n-1)}(\lambda) & \cdots & \frac{1}{2!} P''(\lambda) & P'(\lambda) & P(\lambda) \end{bmatrix},$$

then, if the characteristic polynomial of X is $(z - \lambda)^n$,

$$\Psi^{-1} X \Psi = \lambda I + J, \tag{7.4}$$

and Ψ^{-1} is given by

$$\Psi^{-1} = \begin{bmatrix} Q(\lambda) \\ Q'(\lambda) \\ \frac{1}{2!} Q''(\lambda) \\ \vdots \\ \frac{1}{(n-1)!} Q^{(n-1)}(\lambda) \end{bmatrix}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Proof. From the Taylor expansion of (7.3) about $z = \lambda$, it is found that

$$XP(\lambda) = \lambda P(\lambda),$$
$$X\frac{1}{i!}P^{(i)} = \lambda\frac{1}{i!}P^{(i)} + \frac{1}{(i-1)!}P^{(i-1)}, \quad i = 1, 2, \ldots, n-1,$$

and (7.4) follows. Similar formulae are found for $Q(\lambda)X$ and $\frac{1}{i!}Q^{(i)}(\lambda)X$ and the fact that both Ψ and the formula given for Ψ^{-1} are unit upper triangular, completes the proof. □

7.2. Formulation of IRKS methods

We will consider the construction of methods with $p = q$, and $r = s = p+1$. Without loss of generality, we can assume that the starting method corresponds to the evaluation of the scaled derivatives $h^i y^{(i)}(x_0)$, $i = 0, 1, 2, \ldots, p$. This means that the vector $\phi(z)$ in Theorem 4.3 is equal to Z given by

$$Z = \begin{bmatrix} 1 \\ z \\ z^2 \\ \vdots \\ z^p \end{bmatrix}.$$

A consequence of this assumption is that V necessarily has the form

$$V = \begin{bmatrix} 1 & v^T \\ 0 & \dot{V} \end{bmatrix},$$

and stability requires that $\rho(\dot{V}) \leq 1$. We will want to go further than this and actually assume that $\rho(\dot{V}) = 0$.

Because we want to minimize computational cost, we will consider only methods in which A has a lower triangular structure with constant diagonals:

$$A = \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ a_{21} & \lambda & 0 & \cdots & 0 \\ a_{31} & a_{32} & \lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ a_{s1} & a_{s2} & a_{s3} & \cdots & \lambda \end{bmatrix}.$$

For RK-stable methods, using A with this structure will result in a stability function of the form

$$R(z) = \frac{N(z)}{(1-\lambda z)^{p+1}},$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

where, because the order is p, $N(z)$ is given by

$$N(z) = \exp(z)(1-\lambda z)^{p+1} - \text{const} z^{p+1},$$

where const is the 'error constant'. If n_{p+1} is the coefficient of z^{p+1} in $N(z)$ then

$$n_{p+1} = \frac{1}{(p+1)!} - \frac{1}{p!}\binom{p+1}{1}\lambda + \frac{1}{(p-1)!}\binom{p+1}{2}\lambda^2 - \cdots + (-\lambda)^{p+1} - \text{const}.$$

In the construction of a specific method, the values of λ and either n_{p+1} or const are available as design choices. For example we may want to choose λ to achieve A-stability and we might require that $n_{p+1} = 0$ to obtain the additional property of L-stability. For a non-stiff option, λ would be chosen as zero, to obtain explicit methods, and const $= n_{p+1} - \frac{1}{(p+1)!}$ would be chosen to balance the requirements of accuracy and stability.

In the formulation of these new methods we will want to find the remaining (strictly lower triangular) elements of A and the elements of B as starting points, and evaluate U and V from

$$U = C - ACK, \tag{7.5}$$

$$V = E - BCK, \tag{7.6}$$

where C, K and E have the meanings introduced in Section 4.4. It will be a constraint on the choice of the elements in B to make sure that V has the correct form.

Definition 7.3. A general linear method (A, U, B, V) is said to possess inherent Runge–Kutta stability (IRKS) if it satisfies the assumptions introduced in Section 7.2 and there exists a doubly companion matrix X such that $\alpha(z)\beta(z) = (1-\lambda z)^{p+1} + \mathcal{O}(z^{s+1})$, and a vector ξ^T, such that

$$BA = XB, \tag{7.7}$$

$$BU = XV - VX + e_1\xi^T. \tag{7.8}$$

The value of the vector ξ^T will be explored below.

The significance of Definition 7.3 is summed up in the following result.

Theorem 7.4. The characteristic polynomial of a general linear method possessing the IRKS property has only a single nonzero eigenvalue.

Proof. The stability matrix is

$$M(z) = V + zB(I - zA)^{-1}U,$$

and the characteristic polynomial of $M(z)$ is the same as for the matrix

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

formed by similarity, using the transformation matrix $(I - zX)$. Evaluate this as follows:

$$\begin{aligned}(I - zX)M(z)(I - zX)^{-1} &= (I - zX)(V + zB(I - zA)^{-1}U)(I - zX)^{-1}\\ &= (I - zX)(V + z(I - zX)^{-1}BU)(I - zX)^{-1}\\ &= (V - zXV + z(XV - VX + e_1\xi^T))(I - zX)^{-1}\\ &= V + e_1\xi^T(I - zX)^{-1}. \qquad (7.9)\end{aligned}$$

This matrix has the same form as V except for the first row. Hence, p of the zeros of the characteristic polynomial are equal to zero. □

This result makes it possible to determine ξ^T. Write $\xi(z) = \xi_1 z + \xi_2 z^2 + \cdots + \xi_{p+1} z^{p+1}$ and use (7.9) to give

$$\begin{aligned}R(z) &= 1 + z\xi^T(I - zX)^{-1}e_1 \qquad (7.10)\\ &= \frac{\det(I + z(e_1\xi^T - X))}{\det(I - zX)} \qquad (7.11)\\ &= \frac{(\alpha(z) + \xi(z))\beta(z)}{\alpha(z)\beta(z)} + \mathcal{O}(z^{p+2}). \qquad (7.12)\end{aligned}$$

Because $R(z) = N(z)(1 - \lambda z)^{-p-1}$, it follows that

$$\xi(z) = (N(z) - (1 - \lambda z)^{p+1})\beta(z)^{-1} + \mathcal{O}(z^{p+2}),$$

and the coefficients in $\xi(z)$ are found as the components of ξ^T.

7.3. Construction of specific methods

We first explore consequences of (7.7) and (7.8). Substitute U and V from (7.5) and (7.6) into (7.8) and use (7.7) to find:

$$BC(I - KX) = XE - EX + e_1\xi^T. \qquad (7.13)$$

It is found that both $I - KX$ and $XE - EX + e_1\xi^T$ are zero except for their final columns. Deleting the irrelevant columns of (7.13) we find, after some manipulation,

$$BC\begin{bmatrix}\beta_p\\ \beta_{p-1}\\ \vdots\\ \beta_1\\ 1\end{bmatrix} = \begin{bmatrix}\frac{1}{(p+1)!} + \sum_{i=1}^{p}\frac{1}{(p+1-i)!}\beta_i - \text{const}\\ \frac{1}{p!} + \sum_{i=1}^{p-1}\frac{1}{(p-i)!}\beta_i\\ \vdots\\ \frac{1}{2!} + \frac{1}{1!}\beta_1\\ \frac{1}{1!}\end{bmatrix}. \qquad (7.14)$$

Define $\widetilde{B} = \Psi^{-1}B$ and rewrite (7.7) in the form

$$\widetilde{B}(A - \lambda I) = J\widetilde{B}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

It follows from this equation that $\widetilde{B}$ is lower triangular and we can write (7.14) in the form

$$\widetilde{B}C\begin{bmatrix}\beta_p\\ \beta_{p-1}\\ \vdots\\ \beta_1\\ 1\end{bmatrix}=\Psi^{-1}\begin{bmatrix}\frac{1}{(p+1)!}+\sum_{i=1}^{p}\frac{1}{(p+1-i)!}\beta_i-\text{const}\\ \frac{1}{p!}+\sum_{i=1}^{p-1}\frac{1}{(p-i)!}\beta_i\\ \vdots\\ \frac{1}{2!}+\frac{1}{1!}\beta_1\\ \frac{1}{1!}\end{bmatrix}. \tag{7.15}$$

The condition that $\rho(\dot{V})=0$ can be written as a linear constraint on B and therefore of $\widetilde{B}$.

The construction of methods now consists of the following steps.

(i) Select suitable values of λ, $c_1,\ldots,c_{p+1}$, $\beta_1,\ldots,\beta_p$ and const.

(ii) Find $\alpha_1,\ldots,\alpha_{p+1}$ from

$$\alpha(z)=(1-\lambda z)^{p+1}/\beta(z)+\mathcal{O}(z^{p+2}).$$

(iii) Construct X, Ψ and other related matrices.

(iv) Choose $\widetilde{B}$ so that (7.15) and so that $\rho(\dot{V})=0$.

(v) Find the coefficient matrices using the formulae

$$\begin{aligned}B&=\Psi\widetilde{B},\\ A&=B^{-1}XB,\\ U&=C-ACK,\\ V&=E-BCK.\end{aligned}$$

To illustrate this procedure we construct an A-stable method of order 3. In step (i) we make the following choices.

- $\lambda=\frac{1}{2}$. This was chosen for simplicity, taking into account the need for A-stable behaviour and for a reasonably small absolute error constant. Assuming that L-stability is sought, then the error constant is given by

$$\text{const}=\lambda^4-4\lambda^3+3\lambda^2-\frac{2}{3}\lambda+\frac{1}{24},$$

and is shown in Figure 7.1 for the interval $[0.223647801, 0.572816062]$, which is approximately the set of λ values yielding A-stable methods.

- The stage abscissae are chosen as $[\frac{1}{3},\frac{2}{3},1,1]$. This choice is obviously chosen for simplicity and convenience. By imposing an additional constraint, it is possible to force the first row of B to be identical to the last row of A and at the same time forcing the second row of B to be $[0,0,0,1]$. This enables the method to have properties similar to the

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

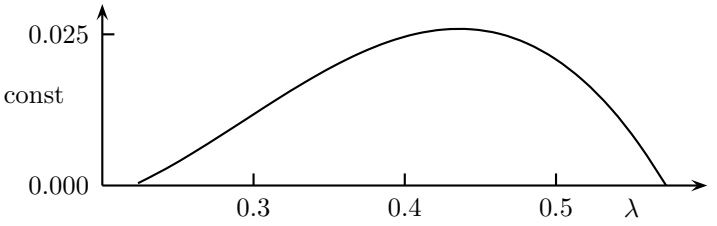


Figure 7.1. Error constant for third-order method.

so-called FSAL Runge–Kutta methods. In particular, $y_2^{[n]} = hf(y_1^{[n]})$ so that in step number $n+1$, we have available what is effectively a further stage derivative for use in interpolation and similar purposes.

- The values of $[\beta_1, \beta_2, \beta_3]$ are chosen as $[-1, \frac{1}{3}, 0]$. The zero value of β_3 is a consequence of the FSAL condition whereas β_1 and β_2 are chosen to ensure that the coefficients given as elements of $[A, U, B, V]$ have reasonably small magnitudes and are reasonably simple numbers.
- Because we want L-stability, we choose const as we have described above.

There are choices to be made in how step (iv) is to be carried out. In the present construction, we have forced $\dot{V}$ to be strictly lower triangular.

The coefficients for the method described under these choices are given by

$$\begin{bmatrix} A & U \\ B & V \end{bmatrix} = \left[\begin{array}{cccc|cccc} \frac{1}{2} & 0 & 0 & 0 & 1 & -\frac{1}{6} & -\frac{1}{9} & -\frac{7}{324} \\ \frac{36}{295} & \frac{1}{2} & 0 & 0 & 1 & \frac{79}{1770} & -\frac{403}{2655} & -\frac{1637}{23895} \\ -\frac{705}{472} & \frac{177}{160} & \frac{1}{2} & 0 & 1 & \frac{8377}{9440} & -\frac{1131}{4720} & -\frac{581}{2360} \\ \frac{39}{160} & \frac{15}{128} & \frac{5}{24} & \frac{1}{2} & 1 & -\frac{133}{1920} & -\frac{353}{960} & -\frac{109}{480} \\ \hline \frac{39}{160} & \frac{15}{128} & \frac{5}{24} & \frac{1}{2} & 1 & -\frac{133}{1920} & -\frac{353}{960} & -\frac{109}{480} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ -\frac{45}{2} & 18 & \frac{5}{2} & -6 & 0 & 8 & 0 & 0 \\ \frac{171}{4} & -\frac{531}{16} & 0 & 12 & 0 & -\frac{345}{16} & -\frac{33}{8} & 0 \end{array}\right]. \tag{7.16}$$

As a start towards approximating the underlying one-step method and finding an improved starting method, we evaluate the asymptotic error, for step number n, in the sense of Figure 4.1. This is found to be

$$\begin{bmatrix} -\frac{241}{2880}h^4y^{(4)}(x_n) + \mathcal{O}(h^5) \\ \mathcal{O}(h^5) \\ \frac{1}{3}h^4y^{(4)}(x_n) + \mathcal{O}(h^5) \\ \frac{3}{8}h^4y^{(4)}(x_n) + \mathcal{O}(h^5) \end{bmatrix}.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Denote this by ϕ and carry out the decomposition in (4.6) to obtain a solution

$$\epsilon = \frac{1}{48}h^4 y^{(4)}(x_n) + \mathcal{O}(h^5), \qquad \delta = \begin{bmatrix} 0 \\ \mathcal{O}(h^5) \\ \frac{1}{3}h^4 y^{(4)}(x_n) + \mathcal{O}(h^5) \\ -h^4 y^{(4)}(x_n) + \mathcal{O}(h^5) \end{bmatrix}.$$

Note that the coefficients of $h^4 y^{(4)}(x_n)$, in the last three components of δ, are equal to β_3, β_2, β_1, an example of a result by Wright (2002*b*). We can now construct the underlying one-step method and the corresponding modified starting method, to within $\mathcal{O}(h^5)$. For the underlying one-step method, with input $y(x_{n-1})$, simply evaluate the Taylor expansion to within this accuracy, with ϵ subtracted from it. The starting method is now a modified version of the Nordsieck vector with δ subtracted from it. Thus,

$$y^{[n-1]} = \begin{bmatrix} y(x_{n-1}) \\ hy'(x_{n-1}) + \mathcal{O}(h^5) \\ h^2 y''(x_{n-1}) - \frac{1}{3}h^4 y^{(4)}(x_{n-1}) + \mathcal{O}(h^5) \\ h^3 y'''(x_{n-1}) + h^4 y^{(4)}(x_{n-1}) + \mathcal{O}(h^5) \end{bmatrix}.$$

In Section 7.4, we will discuss the use of the modified starting method in the estimation of error information. However, the very existence of an underlying one-step method hinges on the use of constant step-size. Two approaches for dealing with variable step-size have been considered (Butcher and Jackiewicz 2002, 2003). In this paper we will emphasize the second of these, which is to 'correct' the drift away from the correct starting approximation caused by unmodified Nordsieck scaling.

7.4. Implementation issues

In the practical implementation of any method, or family of methods, it is desirable to have available an asymptotically correct error estimator together with a mechanism for adjusting the step-size. Most well-known methods have these but usually at a computational cost. Our aim in the design of general linear methods is to keep any overhead costs as low as possible. For local error estimation, the secret seems to be to insist on methods with high stage order, and for variation of step-size, the essential idea is to use the Nordsieck representation of the data passed between steps. However, a simple rescaling of the Nordsieck vector by powers of the step-size ratio is not always a satisfactory way of adjusting for a new step-size.

There are two reasons for this. The first is that a method which might exhibit stable behaviour for constant step-size might act unstably when the step-size is varied, especially if large variations are permitted. The second is that we will not only want to estimate errors for a method currently in

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

use, but we will also want to estimate errors for an alternative method of *higher* order, which is in contention as a possibly more efficient method for succeeding steps.

This leads to the idea of a 'scale and modify' scheme for step-size control. Suppose that, at the end of step n, a step-size change $h \mapsto rh$ is to be made. Also assume that in the underlying one-step method, the quantities being approximated at step number n are given by

$$\begin{bmatrix} y(x_n) + \mathcal{O}(h^{p+2}) \\ hy'(x_n) - \delta_1 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ h^2 y''(x_n) - \delta_2 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ \vdots \\ h^p y^{(p)}(x_n) - \delta_p h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \end{bmatrix}. \tag{7.17}$$

When this quantity is computed as the output to step n and an unmodified Nordsieck scaling is performed, we have as intended input for step number $n+1$, the quantities

$$\begin{bmatrix} y(x_n) + \mathcal{O}(h^{p+2}) \\ (rh)y'(x_n) - r\delta_1 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ (rh)^2 y''(x_n) - r^2\delta_2 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ \vdots \\ (rh)^p y^{(p)}(x_n) - r^p\delta_p h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \end{bmatrix},$$

which differs from what is required by

$$\begin{bmatrix} \mathcal{O}(h^{p+2}) \\ (r - r^{p+1})\delta_1 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ (r^2 - r^{p+1})\delta_2 h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \\ \vdots \\ (r^p - r^{p+1})\delta_p h^{p+1} y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2}) \end{bmatrix}.$$

The scale and modify scheme requires us to add to the scaled Nordsieck vector, an approximation to this quantity. However, we have a choice of possible approximations to $h^{p+1}y^{(p+1)}(x_n) + \mathcal{O}(h^{p+2})$ and the choice we make, which might differ from component to component, needs to take account of stability requirements.

We will illustrate how this is done using the example method (7.16). By matching Taylor expansions, we find a family of linear combinations of various quantities which give asymptotically correct approximations to $h^4 y^{(4)}$. These quantities are $hF_i = hy'(x_{n-1} + hc_i) + \mathcal{O}(h^5)$, $i = 1, 2, 3, 4$, together

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

with $y_2^{[n-1]} = hy'(x_{n-1}) + \mathcal{O}(h^5)$, $y_3^{[n-1]} = h^2y''(x_{n-1}) - \frac{1}{3}h^4y^{(4)}(x_{n-1}) + \mathcal{O}(h^5)$. Note that we do not involve $y_4^{[n-1]}$ in the error estimate because we want the modified and scaled B and V matrices to have an unchanged sparsity pattern. This will guarantee a variable step analogue of zero stability.

We have two free parameters, which we denote by C_2, C_3 (C_1 will be introduced below) and the suggested approximation is

$$\begin{aligned}(81 + 18C_3)hF_1 + (-81 - \tfrac{45}{2}C_3)hF_2 + (27 + 8C_3 - C_2)hF_3 + C_2hF_4 \\ + (-27 - \tfrac{7}{2}C_3)y_2^{[n-1]} + C_3y_3^{[n-1]} \\ = h^4y^{(4)}(x_n) + \mathcal{O}(h^5).\end{aligned}$$

We use this approximation in two places, in the modification of the scaled $y_3^{[n]}$, with C_3 replaced by zero and C_2 replaced by C_1, and in the modification to the scaled $y_4^{[n]}$. If we write $B(r)$ and $V(r)$ for the scaled and modified versions of B and V respectively, then $y^{[n]}$, as input to step number $n+1$ with step-size rh, is given in a modified version of (3.1),

$$y^{[n]} = B(r)hF + V(r)y^{[n-1]},$$

where $B(r)$ and $V(r)$ are each given as the sum of the simply scaled version plus the modifier terms:

$$B(r) = \begin{bmatrix} \frac{39}{160} & \frac{15}{128} & \frac{5}{24} & \frac{1}{2} \\ 0 & 0 & 0 & r \\ -\frac{45}{2}r^2 & 8r^2 & \frac{5}{2}r^2 & -6r^2 \\ \frac{171}{4}r^3 & -\frac{531}{16}r^3 & 0 & 12r^3 \end{bmatrix}$$

$$+ \operatorname{diag}\big(0, 0, \tfrac{r^2-r^4}{3}, -(r^3-r^4)\big) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 81 & -81 & 27 - C_1 & C_1 \\ 81{+}18C_3 & -81{-}\frac{45}{2}C_3 & 27{+}8C_3{-}C_2 & C_2 \end{bmatrix},$$

$$V(r) = \begin{bmatrix} 1 & -\frac{133}{1920} & -\frac{353}{960} & -\frac{109}{480} \\ 0 & 0 & 0 & 0 \\ 0 & 8r^2 & 0 & 0 \\ 0 & -\frac{345}{16}r^3 & -\frac{33}{8}r^3 & 0 \end{bmatrix}$$

$$+ \operatorname{diag}\big(0, 0, \tfrac{r^2-r^4}{3}, -(r^3 - r^4)\big) \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & -27 & 0 & 0 \\ 0 & -27 - \frac{7}{2}C_3 & C_3 & 0 \end{bmatrix}.$$

For arbitrary (but bounded) step-size ratios, the products of matrices $V(r)$ over many steps acts in a stable manner, simply because these are all strictly lower triangular. However, we will also seek stable behaviour

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

for infinitely stiff problems. That is, we will want to form products of matrices like

$$V(r) - B(r)A^{-1}U. \tag{7.18}$$

Although we have described C_1, C_2 and C_3 as constants, there is no reason why they should not depend on r. If we choose $C_3 = -2.6$ and define $C_1(r)$ and $C_2(r)$ so that the characteristic equation of (7.18) has only a single non-zero root, then this root is bounded in magnitude by 1 at least for $r \in [\frac{1}{2}, 2]$. This is, of course, not sufficient for acceptable variable step stability for stiff problems, but is at least an encouragement to explore this aspect of L-stable general linear methods further.

The availability of asymptotically correct error estimators, and the variable step-size adjustments we have described, provides all the equipment that is needed for reliable step-size control. However, we also want variable order. Although we will not consider adjustments to Nordsieck vector approximations when order is increased, we will discuss as the final detail on implementation, the estimation in step number n of $h^{p+2}y^{(p+2)}(x_n)$, because the asymptotic error of a method of order $p+1$ will be proportional to this quantity.

The key to estimating $h^{p+2}y^{(p+2)}(x_n)$ is the fact that

$$hF_i = hy'(x_{n-1} + hc_i) + \mathcal{O}(h^{p+2}), \quad i = 1, 2, \ldots, p+1. \tag{7.19}$$

We will consider only methods which, like the example method (7.16), have 'Property F', otherwise known as the FSAL property.

Definition 7.5. A general linear method with the IRKS property has Property F if

(i) $c_s = 1$,
(ii) $b_{1j} = a_{sj}$, $j = 1, 2, \ldots, s$,
(iii) $v_{1j} = u_{sj}$, $j = 1, 2, \ldots, r$,
(iv) $b_{2j} = \delta_{sj}$, $j = 1, 2, \ldots, s$,
(v) $v_{2j} = 0$, $j = 1, 2, \ldots, r$.

For a method with Property F, we effectively have an additional accurate derivative approximation, $hF_0 = hy'(x_0 + hc_0) + \mathcal{O}(h^{p+2})$, where $c_0 = 0$. Hence, we will regard (7.19) as holding for $i = 0, 1, \ldots, p+1$.

The FSAL property, on which Definition 7.5 is modelled, was made popular in the design of Runge–Kutta methods, by the work of Dormand and Prince (1980) because it gives an additional apparently free derivative approximation to widen options for error estimators. This is exactly how we will use Property F. For the remainder of this section we will assume this property, just as we will always assume that the scale and multiply technique is used when the step-size varies.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Suppose that the error coefficients δ_1, δ_2, are as in (7.17). Then the errors introduced into the stage values are

$$y(x_{n-1}+hc_i) - h\sum_{j=1}^{p+1} a_{ij}y'(x_{n-1}+hc_j) \\ -\sum_{j=1}^{p+1} u_{ij}\big(h^{j-1}y^{(j-1)} - \delta_{j-1}h^{p+1}y^{(p+1)}(x_{n-1})\big), \\ i = 1, 2, \ldots, p+1.$$

By Taylor's theorem this equals

$$\sigma_i h^{p+1} y^{(p+1)} y(x_{n-1}) + \mathcal{O}(h^{p+2}),$$

where

$$\sigma = \tfrac{1}{(p+1)!}c^{p+1} - \tfrac{1}{p!}Ac^p + U\delta,$$

and we find, for the corresponding error in the stage derivatives,

$$\sigma_i h^{p+2}\tfrac{\partial f}{\partial y} y^{(p+1)}(x_{n-1}) + \mathcal{O}(h^{p+3}).$$

If we contemplate using linear combinations of hF_i, $i = 0, 1, \ldots, p+1$ to estimate quantities related to errors, to within $\mathcal{O}(h^{p+3})$, we need to make use of the matrix

$$L = \begin{bmatrix} \sigma_0 & 1 & c_0 & \frac{1}{2}c_0^2 & \cdots & \frac{1}{p!}c_0^p & \frac{1}{(p+1)!}c_0^{p+1} \\ \sigma_1 & 1 & c_1 & \frac{1}{2}c_1^2 & \cdots & \frac{1}{p!}c_1^p & \frac{1}{(p+1)!}c_1^{p+1} \\ \vdots & \vdots & \vdots & \vdots & & \vdots & \vdots \\ \sigma_{p+1} & 1 & c_{p+1} & \frac{1}{2}c_{p+1}^2 & \cdots & \frac{1}{p!}c_{p+1}^p & \frac{1}{(p+1)!}c_{p+1}^{p+1} \end{bmatrix}.$$

We distinguish three cases:

(i) the c components are distinct and the first $p+2$ columns of L are linearly independent,

(ii) the c components are distinct and the first $p+2$ columns of L are linearly dependent,

(iii) $c_p = c_{p+1} = 1$ and $\sigma_p \neq \sigma_{p+1}$.

A final case, in which $c_p = 1$ but $\sigma_p = \sigma_{p+1}$, will not be explored because there does not seem to be a simple error estimator of the type we want in this case. Note that the example method (7.16) is in case (iii).

In case (i), construct a coefficient vector $\xi^T = [\xi_0, \xi_1, \ldots, \xi_{p+1}]$ such that

$$\xi^T L = e_{p+2} + \theta e_{p+3},$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

so that $\sum_{i=0}\xi_i hF_i$ equals

$$\begin{aligned}\Psi_n &= h^{p+1}y^{(p+1)}(x_{n-1}) + \theta h^{p+2}y^{(p+2)}(x_{n-1}) + \mathcal{O}(h^{p+3})\\ &= h^{p+1}y^{(p+1)}(x_{n-1}+h\theta) + \mathcal{O}(h^{p+3}).\end{aligned}$$

If the step-size used in step number $n-1$ was $r^{-1}h$, then an asymptotically correct estimate of $h^{p+2}y^{(p+2)}(x_{n-1})$ can be found from

$$\tfrac{r}{1+\theta(r-1)}\big(\Psi_n - r^{p+1}\Psi_{n-1}\big).$$

In cases (ii), it is possible to give in a single step an approximation

$$\sum_{i=0}\xi_i hF_i = h^{p+2}y^{(p+2)}(x_{n-1}) + \mathcal{O}(h^{p+3})$$

by choosing ξ^T to satisfy

$$\xi^T L = e_{p+3}.$$

Case (iii) is similar to case (i) and we will illustrate this using the example method (7.16). For this method L is found to be

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0\\ -\frac{35}{1944} & 1 & \frac{1}{3} & \frac{1}{18} & \frac{1}{162} & \frac{1}{1944}\\ \frac{10}{14337} & 1 & \frac{2}{3} & \frac{2}{9} & \frac{4}{81} & \frac{2}{243}\\ \frac{187}{2360} & 1 & 1 & \frac{1}{2} & \frac{1}{6} & \frac{1}{24}\\ \frac{1}{48} & 1 & 1 & \frac{1}{2} & \frac{1}{6} & \frac{1}{24}\end{bmatrix},$$

leading to the approximation

$$\begin{aligned}-27y_2^{[n-1]} + 81hF_1 - 81hF_2 + \tfrac{13485}{827}hF_3 + \tfrac{8844}{827}hF_4&\\ \approx h^4y^{(4)}(x_{n-1}) + \tfrac{1}{2}h^5y^{(5)}(x_{n-1}).&\end{aligned}$$

The estimation of local truncation errors is discussed in Butcher and Podhaisky (2006); this includes the estimation of $h^{p+2}y^{(p+2)}$ using the method described in this section.

8. Order and stability barriers

This discussion is relevant to multi-derivative methods, otherwise known as Obreshkov methods, as well as to general linear methods. The essential question concerns polynomial functions in two complex variables and the extent to which they can represent high-order approximations to exp and at the same time represent A-stable behaviour.

Given positive integers r, s, we consider a polynomial function of two complex variables $\Phi(w,z)$, which has degree r in w and degree s in z. Given

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

a general linear method (A, U, B, V), the stability matrix is

$$M(z) = V + zB(I - zA)^{-1}U$$

and its linear stability properties are defined in terms of the characteristic polynomial

$$\det(wI - M(z)). \tag{8.1}$$

Only when A is nilpotent, such as in an explicit method, will this expression be a polynomial. It is, however, always a polynomial in w and z divided by a polynomial in z. We can define Φ for this method by using the numerator of (8.1).

Two special cases correspond to classical methods. If $s = 1$ we will write

$$\Phi(w, z) = \rho(w) - z\sigma(w),$$

using the standard notation for linear multistep methods. On the other hand, if $r = 1$ then we will write

$$\Phi(w, z) = wD(z) - N(z),$$

corresponding to the stability function $R(z) = N(z)/D(z)$ of a Runge–Kutta method.

As much as possible, we will distance ourselves from an actual method but will consider properties of Φ in its own right.

Definition 8.1. A stability function has order p if

$$\Phi(\exp(z), z) = \mathcal{O}(z^{p+1}).$$

Note that this definition does not necessarily coincide with the order of the underlying general linear method. However, the actual order of the method cannot exceed p in Definition 8.1.

Definition 8.2. A stability function is A-stable if for every complex numbers z such that $\operatorname{Re} z \leq 0$,

(i) if w satisfies $\Phi(w, z) = 0$, then $|w| \leq 1$,

(ii) if w satisfies $\Phi(w, z) = \frac{\partial}{\partial w}\Phi(w, z) = 0$, then $|w| < 1$.

This definition is not the usual one. However, our aim will be to understand the conflict between order and stable behaviour for stiff problems and we want consistent conclusions which reasonably well make it possible to decide between suitable and unsuitable methods. Consider the approximation

$$\Phi(w, z) = \left(1 - \tfrac{5}{8}z + \tfrac{1}{8}z^2\right)w^2 - 2w + 1 + \tfrac{5}{8}z + \tfrac{1}{8}z^2.$$

According to Definition 8.1, this method has order 5 but it is not possible

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

in a numerical computation to realize the corresponding accuracy. On the other hand, according to Definition 8.2 it is not A-stable even though, for every z satisfying $\mathrm{Re}\, z < 0$, the corresponding w values are in the open unit disc. Hence, its properties are consistent with Theorem 8.11, which it should be.

For a particular approximation we might wish to refer to the polynomials in z arising as coefficients of various powers of w. Write

$$\Phi(w, z) = P_0(z)w^r + P_1(z)w^{r-1} + \cdots + P_{r-1}(z)w + P_r(z). \tag{8.2}$$

We will refer to the zeros of $P_0(z)$ as the 'poles of Φ' and the zeros of P_r as the 'zeros of Φ'.

Theorem 8.3. The approximation Φ is A-stable if and only if

(i) Φ has no poles in the left half-plane,

(ii) there do not exist complex numbers w and z such that $\mathrm{Re}\, z = 0$, $|w| > 1$ and $\Phi(w, z) = 0$,

(iii) there do not exist complex numbers w and z such that $\mathrm{Re}\, z = 0$, $|w| = 1$ and $\Phi(w, z) = \frac{\partial}{\partial w}\Phi(w, z) = 0$.

Proof. (i) is necessary because if z is near a pole there are arbitrarily high values of w satisfying $\Phi(w, z) = 0$. (ii) and (iii) are necessary because the imaginary axis is a subset of the closed left half-plane. Sufficiency follows from the maximum-modulus theorem. □

Associated with a given approximation Φ is the Riemann surface defined by $\Phi(\widehat{w}\exp(z), z) = 0$. The use of this 'relative stability function' was made famous by its use in the theory of order stars. We will use the closely related 'order arrows' in this paper to achieve many of the same goals.

In the search for A-stable methods of high order, we will consider (8.2) with the degrees

$$n_i = \deg(P_i), \qquad i = 0, 1, 2, \ldots, r, \tag{8.3}$$

specified and the coefficients chosen to maximize the order of the approximation.

Definition 8.4. Given a sequence of degrees, $n_i \geq -1$, $i = 0, 1, \ldots, r$ a generalized Padé approximation (to exp) is a sequence of polynomials $P_0, P_1, \ldots, P_r$, satisfying (8.3) with order $p = \sum_{i=0}^{r} n_i + r - 1$.

We will always assume that $n_0 \geq 0$ and we will usually assume that $n_r \geq 0$ (otherwise, r could be reduced to $r-1$). In the interpretation Definition 8.4, we will regard a polynomial of degree -1 as being the zero polynomial.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

8.1. Padé approximations

The Padé approximations, that is the generalized Padé approximations in the case $r = 1$, arise as the stability functions of certain implicit Runge–Kutta methods. If n_0 the degree of D, is written as d and n_1, the degree of N is written as n, then the (d, n) Padé approximation is given by

$$D(z)\exp(z) - N(z) = (-1)^d \frac{C}{(n+d+1)!} z^{n+d+1} + \mathcal{O}(z^{n+d+2}), \tag{8.4}$$

where the constant C is an arbitrary nonzero scale factor.

Theorem 8.5. The polynomials N and D in (8.4) are given by

$$N(z) = C \sum_{i=0}^{n} \frac{z^{n-i}}{(n-i)!} \binom{d+i}{i}, \tag{8.5}$$

$$D(z) = C \sum_{i=0}^{d} \frac{(-z)^{d-i}}{(d-i)!} \binom{n+i}{i}. \tag{8.6}$$

Proof. Operate on (8.4) by $(\frac{\mathrm{d}}{\mathrm{d}z})^{n+1}$ to obtain

$$\exp(z)(1 + \tfrac{\mathrm{d}}{\mathrm{d}z})^{n+1} D(z) = (-1)^d C \frac{z^d}{d!} + \mathcal{O}(z^{d+1}).$$

Multiply by $\exp(-z)$ and the right-hand side is unchanged. However, the left-hand side is a polynomial of degree exactly d and the $\mathcal{O}(z^{d+1})$ can be omitted. It now follows that

$$D(z) = (-1)^d C (1 + \tfrac{\mathrm{d}}{\mathrm{d}z})^{-(n+1)} \frac{z^d}{d!},$$

and (8.6) follows. Similarly, multiply (8.4) by $\exp(-z)$ and operate on the resulting equation by $(\frac{\mathrm{d}}{\mathrm{d}z})^{d+1}$, leading to (8.5). □

A convenient choice of C is $n!d!/(n+d)!$ leading to formulae in which $N(0) = D(0) = 1$:

$$N(z) = \sum_{i=0}^{n} \frac{n!(n+d-i)!}{(n-i)!(n+d)!i!} z^i, \tag{8.7}$$

$$D(z) = \sum_{i=0}^{d} \frac{d!(n+d-i)!}{(d-i)!(n+d)!i!} (-z)^i. \tag{8.8}$$

A partial table of Padé approximations to the exponential function is given in Table 8.1 (overleaf).

Recurrence relations

We will write $V_{dn}(z)$ to denote the two-dimensional vector whose first and second components are $N_{dn}(z)$ and $D_{dn}(z)$, respectively. Many relationships

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 8.1. Padé approximations to exp of degrees $[n, d]$.

d \ n	0	1	2	3
0	$\frac{1}{1}$	$\frac{1+z}{1}$	$\frac{1+z+\frac{1}{2}z^2}{1}$	$\frac{1+z+\frac{1}{2}z^2+\frac{1}{6}z^3}{1}$
1	$\frac{1}{1-z}$	$\frac{1+\frac{1}{2}z}{1-\frac{1}{2}z}$	$\frac{1+\frac{2}{3}z+\frac{1}{6}z^2}{1-\frac{1}{3}z}$	$\frac{1+\frac{3}{4}z+\frac{1}{4}z^2+\frac{1}{24}z^3}{1-\frac{1}{4}z}$
2	$\frac{1}{1-z+\frac{1}{2}z^2}$	$\frac{1+\frac{1}{3}z}{1-\frac{2}{3}z+\frac{1}{6}z^2}$	$\frac{1+\frac{1}{2}z+\frac{1}{12}z^2}{1-\frac{1}{2}z+\frac{1}{12}z^2}$	$\frac{1+\frac{3}{5}z+\frac{3}{20}z^2+\frac{1}{60}z^3}{1-\frac{2}{5}z+\frac{1}{20}z^2}$
3	$\frac{1}{1-z+\frac{1}{2}z^2-\frac{1}{6}z^3}$	$\frac{1+\frac{1}{4}z}{1-\frac{3}{4}z+\frac{1}{4}z^2-\frac{1}{24}z^3}$	$\frac{1+\frac{2}{5}z+\frac{1}{20}z^2}{1-\frac{3}{5}z+\frac{3}{20}z^2-\frac{1}{60}z^3}$	$\frac{1+\frac{1}{2}z+\frac{1}{10}z^2+\frac{1}{120}z^3}{1-\frac{1}{2}z+\frac{1}{10}z^2-\frac{1}{120}z^3}$

exist between adjacent members of the Padé table. We will here consider just one of these because of its application in this work.

Theorem 8.6. If $n \geq 2$ then

$$V_{n,n}(z) = V_{n-1,n-1}(z) + \tfrac{z^2}{4(2n-1)(2n-3)} V_{n-2,n-2}(z).$$

Proof. Because $D_{nn}(z) = N_{nn}(-z)$, it follows that $N_{nn}(z)\exp(-z/2) - \exp(z/2)D_{nn}(z)$ is an odd function; combining this with the fact that

$$\tfrac{N_{nn}(z)}{D_{nn}(z)} = \exp(z) + \mathcal{O}(z^{2n+1}),$$

we find

$$\begin{bmatrix}\exp(-z/2) & -\exp(z/2)\end{bmatrix} V_{nn}(z) = \tfrac{(-1)^{n+1}n!^2}{(2n)!(2n+1)!} z^{2n+1} + \mathcal{O}(z^{2n+3}).$$

Hence,

$$\begin{bmatrix}\exp(-z/2) & -\exp(z/2)\end{bmatrix} \left(V_{n-1,n-1}(z) + \tfrac{z^2}{4(2n-1)(2n-3)} V_{n-2,n-2}(z)\right)$$
$$= (-1)^n \theta z^{2n-1} + \mathcal{O}(z^{2n+1}),$$

where

$$\theta = -\tfrac{1}{4(2n-1)(2n-3)} \tfrac{(n-2)!^2}{(2n-4)!(2n-3)!} + \tfrac{(n-1)!^2}{(2n-2)!(2n-1)!} = 0.$$

It follows that $V_{n-1,n-1}(z) + \frac{z^2}{4(2n-1)(2n-3)} V_{n-2,n-2}(z)$ is the unique, correctly scaled vector V_{nn}. □

A-stable Padé approximations

The A-stable members of the Padé table for the exponential functions are those for which $d-n \in \{0, 1, 2\}$. The fact that approximations with $d > n+2$ cannot be A-stable will be proved in Theorem 8.9 and the corresponding result for $d < n$ is covered by the following result.

Theorem 8.7. A Padé approximation to exp with $d > n$ is not A-stable.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Proof. For $|z|$ large, we find from (8.7) and (8.8)

$$\left|\frac{N(z)}{D(z)}\right| = \frac{d!}{n!}|z|^{n-d} + \mathcal{O}(|z|^{n-d-1}),$$

and is greater than 1 for $z \in \mathbb{C}^-$ with $|z|$ sufficiently large. □

In remains to prove the result.

Theorem 8.8. If $d - n \in \{0, 1, 2\}$, the $[d, n]$ Padé approximation to exp is A-stable.

Proof. In the case $n = d$, let $\zeta_n = D_{nn}(z)/zD_{n-1,n-1}(z)$, where $\mathrm{Re}\,(z) < 0$ and deduce from the second component of (8.6) that

$$\zeta_n = \frac{1}{z} + \frac{1}{4(2n-1)(2n-3)\zeta_{n-1}}.$$

Starting from $\zeta_1 = z^{-1} - \frac{1}{2}$, we deduce that each of $\zeta_1, \zeta_2, \ldots$ is in the left half-plane, where we use the fact that the inverse of a number in the left half-plane, and the sum of two such numbers, are each also in the left half-plane. It follows that ζ_n is not zero and neither is

$$D_{nn}(z) = z^n \zeta_n \zeta_{n-1} \cdots \zeta_1.$$

Since D_{nn} has no zeros in the left half-plane, we use the maximum modulus principle to deduce that $|N_{nn}z/D_{nn}(z)|$ is bounded by 1, because the value is achieved exactly on the imaginary axis.

In the case $n = d - 1$, define the approximation $\widetilde{N}(z)/\widetilde{D}(z)$ as the components of the vector

$$\widetilde{V}(z) = (1-t)V_{nn}(z) + tV_{n,n-1}(z),$$

where the homotopy variable t moves from 0 (diagonal approximation) to 1 (sub-diagonal approximation). Because $\widetilde{N}(z) = \exp(z)\widetilde{D}(z) + \mathcal{O}(z^{2n})$ it follows that

$$|\widetilde{D}(iy)|^2 - |\widetilde{N}(iy)|^2 = C(t)y^{2n},$$

where $C(t) > 0$ for $t > 0$. Hence, $D(iy) > 0$. As t increases in $[0, 1]$, the zeros of $\widetilde{D}(z)$ move continuously and can never cross the imaginary axis, because $D(iy)$ never vanishes.

In the case $n = d - 2$ carry out a similar homotopy from $V_{n,n-1}(z)$ to $V_{n,n-2}(z)$ and obtain a similar result. □

8.2. Quadratic Padé approximations

The derivation of Padé approximations we have given can be easily generalized to the quadratic case $r = 2$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Table 8.2. Some quadratic Padé approximations to exp.

p	$[n_0,n_1,n_2]$	$P_0(z)$	$P_1(z)$	$P_2(z)$
2	$[1,0,0]$	$1-\frac{2}{3}z$	$-\frac{4}{3}$	$\frac{1}{3}$
3	$[2,0,0]$	$1-\frac{6}{7}z+\frac{2}{7}z^2$	$-\frac{8}{7}$	$\frac{1}{7}$
4	$[2,1,0]$	$1-\frac{10}{17}z+\frac{2}{17}z^2$	$-\frac{16}{17}-\frac{8}{17}z$	$-\frac{1}{17}$
4	$[2,0,1]$	$1-\frac{8}{11}z+\frac{2}{11}z^2$	$-\frac{16}{11}$	$\frac{5}{11}+\frac{2}{11}z$
4	$[3,0,0]$	$1-\frac{14}{15}z+\frac{2}{5}z^2-\frac{4}{45}z^3$	$-\frac{16}{15}$	$\frac{1}{15}$
5	$[3,1,0]$	$1-\frac{34}{49}z+\frac{10}{49}z^2-\frac{4}{147}z^3$	$-\frac{48}{49}-\frac{16}{49}z$	$-\frac{1}{49}$
5	$[3,0,1]$	$1-\frac{11}{13}z+\frac{4}{13}z^2-\frac{2}{39}z^3$	$-\frac{16}{13}$	$\frac{3}{13}+\frac{1}{13}z$
5	$[4,0,0]$	$1-\frac{30}{31}z+\frac{14}{31}z^2-\frac{4}{31}z^3+\frac{2}{93}z^4$	$-\frac{32}{31}$	$\frac{1}{31}$

Suppose $n_0\geq 0$, $n_1,n_2\geq -1$, $\min(n_1,n_2)\geq 0$ and

$$\Phi(w,z)=w^2P_0(z)+wP_1(z)+P_2(z),\qquad \deg(P_i)=n_i,\quad i=0,1,2,$$

and that $\Phi(\exp(z),z)=\mathcal{O}(z^{p+1})$, with $p=n_0+n_1+n_2+1$.

Assume for some C

$$\exp(2z)P_0(z)+\exp(z)P_1(z)+P_2(z)=C\tfrac{z^{p+1}}{(p+1)!}+\mathcal{O}(z^{p+2}).$$

To find P_0, multiply by $\exp(-2z)$ and apply $(1+\frac{\mathrm{d}}{\mathrm{d}z})^{n_1+1}(2+\frac{\mathrm{d}}{\mathrm{d}z})^{n_2+1}$ to both sides. The result is

$$\big(1+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{n_1+1}\big(2+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{n_2+1}P_0(z)=C\tfrac{z^{n_0}}{n_0!},$$

where $\mathcal{O}(z^{n_0+1})$ is omitted on the right-hand side because the left-hand side is a polynomial of degree n_0. Find similar expressions involving P_1 and P_2 and rearrange to obtain

$$\begin{aligned}
P_0(z)&=C\big(1+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_1+1)}\big(2+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_2+1)}\tfrac{z^{n_0}}{n_0!},\\
P_1(z)&=C\big(-1+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_0+1)}\big(1+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_2+1)}\tfrac{z^{n_1}}{n_1!},\\
P_2(z)&=C\big(-2+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_0+1)}\big(-1+\tfrac{\mathrm{d}}{\mathrm{d}z}\big)^{-(n_1+1)}\tfrac{z^{n_2}}{n_2!}.
\end{aligned}$$

A number of quadratic approximations are given in Table 8.2.

8.3. Order stars and order arrows

The famous theory of order stars (Wanner, Hairer and Nørsett 1978) was introduced as a means of settling some outstanding open questions. The idea is based on the observation that a rational approximation $R(z)=N(z)/D(z)$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

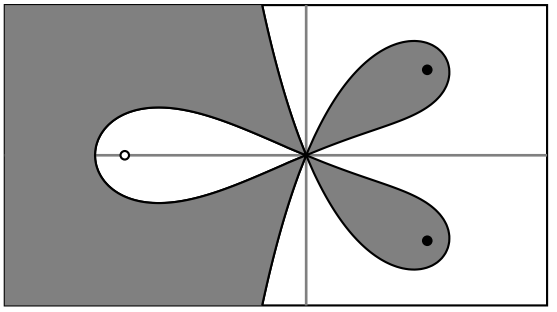


Figure 8.1. Order star for $[2,1]$ Padé approximation.

(N and D having no common factor) of order p is A-stable if and only if

(i) D has no zeros in the open left half-plane,
and
(ii) $|R(z)| \leq 1$, when $\mathrm{Re}\,(z) = 0$,

and the further observation that the criteria still hold if, in (ii), $R(z)$ is replaced by $R(z)\exp(-z)$.

The advantage of the modified form of this criterion is that the behaviour of $R(z)\exp(-z)$ is known in considerable detail when $|z|$ is small. In fact,

$$R(z)\exp(-z) = 1 - Cz^{p+1} + \mathcal{O}(z^{p+2}),$$

where the error constant C is defined by $R(z) = \exp(z) - Cz^{p+1} + \mathcal{O}(z^{p+2})$. Write $z = r\exp t\theta$ and we find

$$|R(z)\exp(-z)| = 1 - Cr^{p+1}\cos((p+1)\theta) + \mathcal{O}(r^{p+2}).$$

For arguments θ such that $C\cos((p+1)\theta) < 0$, $R(z)\exp(-z) > 1$ for sufficiently small $|z|$, and conversely, $R(z)\exp(-z) < 1$ if $C\cos((p+1)\theta) > 0$ and $|z|$ is sufficiently small.

The order star corresponding to this approximation is defined as the set of points in the complex plane such that $|R(z)\exp(-z)| > 1$ and the dual star is defined as the set of points for which $|R(z)\exp(-z)| < 1$. We have seen which points near zero lie in each of these sets. The components of the order star (respectively, dual star) close to zero are referred to as fingers (respectively, dual fingers). Further details are available in Wanner, Hairer and Nørsett (1978) and in other expositions of the theory.

The criterion for A-stability, that $|R(z)\exp(-z)| \leq 1$ on the imaginary axis, translates, in order star language, to the requirement that a finger cannot intersect the imaginary axis. Two examples are presented; first Figure 8.1 for the $[2,1]$ Padé approximation. Here two 'bounded fingers' enclose the poles and a single 'bounded dual finger' encloses the zero. The

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

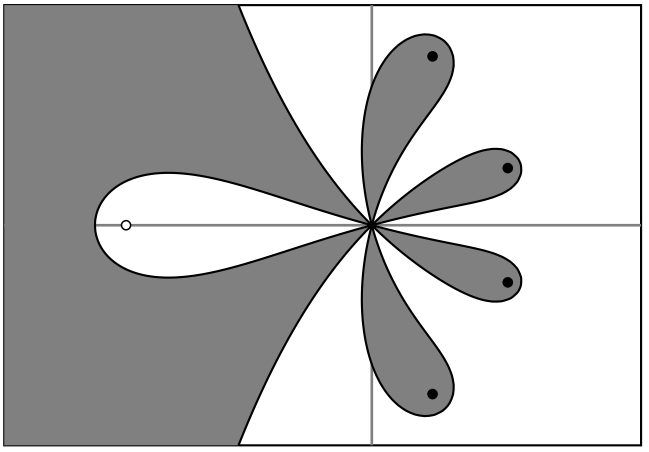


Figure 8.2. Order star for $[4,1]$ Padé approximation.

unbounded finger and the unbounded dual finger divide the parts of the complex plane distant from zero into two parts. The underlying approximation is A-stable because the two poles are in the right half-plane and there is no intersection between the order star (the shaded region) and the imaginary axis.

In contrast, we present Figure 8.2 for the $[4,1]$ Padé approximation. This is *not* A-stable, because in this case the order star intersects the imaginary axis. This is known to be the case because there are too many bounded fingers containing poles for all of them to lie entirely in the right half-plane.

As an alternative to the order star technique, 'order arrows' have been proposed. For an approximation $R(z) = N(z)/D(z)$, this also uses the modified function formed by dividing by $\exp(z)$, but considers the set of points in the complex plane for which $R(z)\exp(-z)$ is real and positive. These emanate from zero as 'up-arrows' which terminate at poles or at $-\infty$, or down-arrows which terminate at zeros or at $+\infty$. A-stability does not hold if an up-arrow leaves zero (with magnitude 1) and either crosses the imaginary axis or is tangential to it. We present the order star diagrams for the two approximations already considered. First, Figure 8.3 corresponds to the A-stable approximation $[2,1]$. In contrast, the arrow diagram for the $[4,1]$ approximation is shown in Figure 8.4. This approximation cannot be A-stable for the following reasons.

(i) Exactly four up-arrows terminate at poles, otherwise some up-arrows would cross down-arrows.

(ii) The angle subtended by the tangents at zero to two of these up-arrows is at least $4 \times \pi/4 = \pi$.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

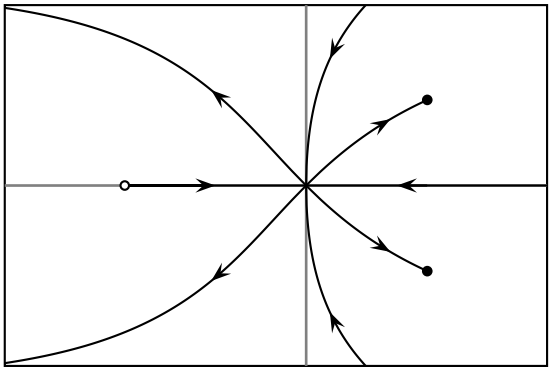


Figure 8.3. Order arrows for $[2,1]$ Padé approximation.

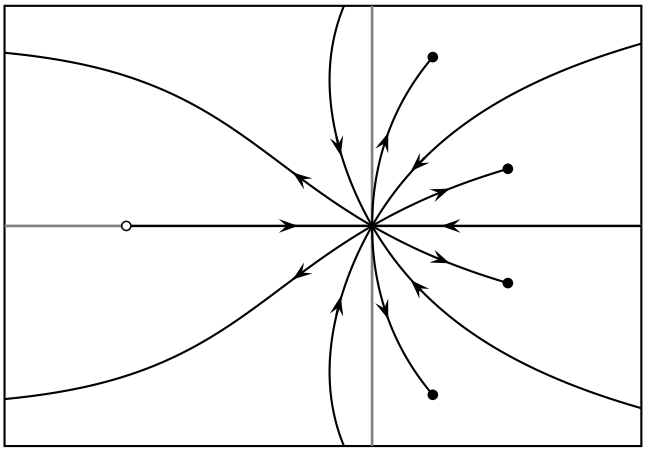


Figure 8.4. Order arrows for $[4,1]$ Padé approximation.

(iii) Hence, at least one up-arrow is either tangential to the imaginary axis or else it emanates into the left half-plane and terminates at a pole.

(iv) Hence, there is either a pole in the left half-plane or this up-arrow crosses the imaginary axis before terminating at a pole in the right half-plane.

If we define order arrows from their basic property that $\Phi(\widehat{w}\exp(z), z) = 0$ with $\widehat{w}$ real and positive, then, in addition to those arrows emanating from zero, there is an infinite family of arrows spaced approximately $2\pi i$ apart, as illustrated in Figure 8.5, for the $[1,0]$ case,

$$\Phi(w, z) = w(1 - z) - 1.$$

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

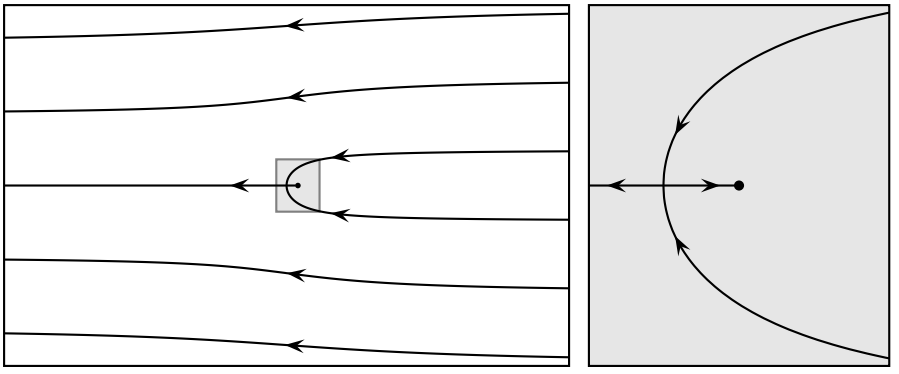


Figure 8.5. Order arrows for $[1,0]$ Padé approximation, together with magnified detail near $z = 0$.

8.4. The Ehle barrier

The result formerly known as the Ehle conjecture was one of the first successes of the order star theory (Wanner, Hairer and Nørsett 1978). Here we will present an alternative proof using order arrows.

Theorem 8.9. Let $R(z) = N(z)/D(z)$ denote the $[d,n]$ Padé approximation to exp. Then, if $d > n+2$, this approximation is not A-stable.

Proof. There are $n+d+1$ up-arrows emanating from zero, alternating with $n+d+1$ down-arrows. Suppose that $\widetilde{d}$ up-arrows terminate at poles so that $(n+d+1)-\widetilde{d}$ up-arrows terminate at $-\infty$. Suppose that $\widetilde{n}$ down-arrows terminate at zeros. These must fit into the $(n+d+1)-\widetilde{d}-1$ gaps between the up-arrows which terminate at $-\infty$. Hence

$$\widetilde{n}+\widetilde{d} \le n+d.$$

Because $\widetilde{d} \le d$ and $\widetilde{n} \le n$ it follows that $\widetilde{d} = d$ and $\widetilde{n} = n$. Since d up-arrows terminate at poles, there must be at least one up-arrow emanating from zero with tangent making an angle to the the positive real axis at least equal to $\frac{d-1}{2} \times 2\pi/(n+d+1)$. For A-stability either (i) this angle must me less than $\pi/2$ or (ii) at least one of the up-arrows terminating at a pole emanates from zero with an argument greater than $\pi/2$. Hence, in case (i),

$$\frac{d-1}{2} \cdot \frac{2\pi}{n+d+1} < \frac{\pi}{2},$$

implying $2d-2 < n+d+1$ so that $d < n+3$. In case (ii), the up-arrow referred to either terminates at a pole in the left half-plane, or crosses the imaginary axis, each of which is impossible. □

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

8.5. Order arrows on Riemann surfaces

To generalize the use of relative stability regions, by inserting the factor $\exp(-z)$ in $R(z)\exp(-z)$, we consider the modification of a generalized approximation $\Phi(w,z)$ by considering the function $\widehat{\Phi}$ defined by

$$\widehat{\Phi}(\widehat{w},z)=\Phi(\widehat{w}\exp(z),z). \tag{8.9}$$

The order star theory for this type of generalization is developed in Wanner, Hairer and Nørsett (1978) and we will discuss here only the order arrow approach.

The Riemann surface for (8.9) is the subset of $\mathbb{C}\times\mathbb{C}$ for which $\widehat{\Phi}(\widehat{w},z)=0$. It is usual to think of the Riemann surface as a multivalued function for which values of z are the arguments, and the corresponding values of $\widehat{w}$ which satisfy the equation $\widehat{\Phi}(\widehat{w},z)=0$ are the values of this function. Except at isolated points at which $(\partial\widehat{\Phi}/\partial\widehat{w})=0$, an open set in the z plane exists so that, in this open set, each value of $\widehat{w}$ acts like a function of a complex variable and satisfies the Cauchy–Riemann conditions. Except in trivial cases in which the sheets of the Riemann surface do not interact with each other, analytic extension leads to a migration onto other sheets.

We superimpose order arrows onto the Riemann surface by considering the subset for which the value $\widehat{w}$ is real and positive. Starting at a specific point on the Riemann surface for which $\widehat{w}$ has this property, up-arrows and down-arrows can be traced out. In particular, if the approximation has order p, then at $z=0$, there are $p+1$ up-arrows and $p+1$ down-arrows emanating from this point.

8.6. The Dahlquist second barrier

The famous second barrier result of Dahlquist (1963), states that an A-stable linear multistep method cannot have order greater than 2. In our context this means the following theorem.

Theorem 8.10. Let $\Phi(w,z)$ denote an $(r,1)$ A-stable approximation with order p. Then $p\leq 2$.

Although many proofs exist, we will here use an order arrow approach, if only as an example of the use of this technique. Note that the approximation is not assumed to be Padé.

Proof of Theorem 8.10. Because $p>2$, and because up-arrows cannot be tangential to the imaginary axis, there are at least 2 up-arrows leaving the origin with directions in $[-\frac{1}{2}\pi,\frac{1}{2}\pi]$. These arrows cannot terminate at $-\infty$ without crossing the imaginary axis. Hence there are at least 2 poles, contrary to the assumption that $s=1$. □

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

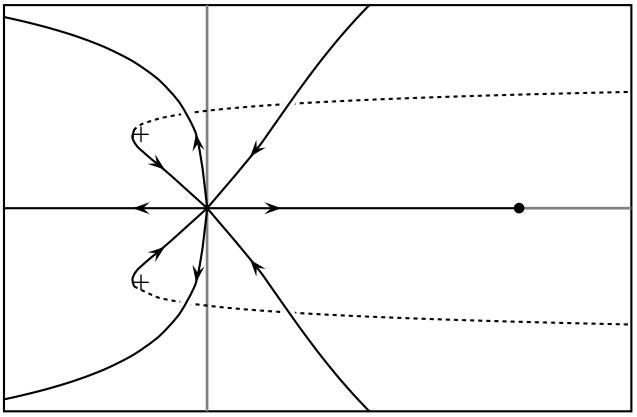


Figure 8.6. Order arrows for BDF3 (pole: • , branch points: +).

This result is illustrated in Figure 8.6 where it is seen that only a single up-arrow emanates from zero in a positive direction but two up-arrows which terminate at $-\infty$ are tangential to the imaginary axis. Note that two down-arrows which emanate in the negative direction terminate at $+\infty$ on lower sheets.

8.7. The Daniel–Moore barrier

It was conjectured in Daniel and Moore (1970) that an order $2s$, which is achieved for Gauss–Legendre Runge–Kutta methods, cannot be exceeded, except at the expense of A-stability. This was eventually proved in Wanner, Hairer and Nørsett (1978) using order stars. The proof given here uses order arrows.

Theorem 8.11. Let $\Phi(w, z)$ denote an (r, s) A-stable approximation with order p. Then $p \leq 2s$.

Proof. If the approximation is A-stable, at most s up-arrows emanate from zero in the positive direction and terminate at poles. The next up-arrow in the anticlockwise direction and the next up-arrow in the clockwise direction do not terminate at poles and must emanate in the negative direction. Because the angle between up-arrows is $2\pi/(p+1)$, it follows that

$$(s+1)\frac{2\pi}{p+1} > \pi,$$

implying that $2s \geq p$. □

We present two order arrow diagrams to illustrate this result. First the A-stable, $[2, 0, 1]$ approximation with order 4. This is given in Figure 8.7. Note that, because of the complicated behaviour on the real axis in which

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

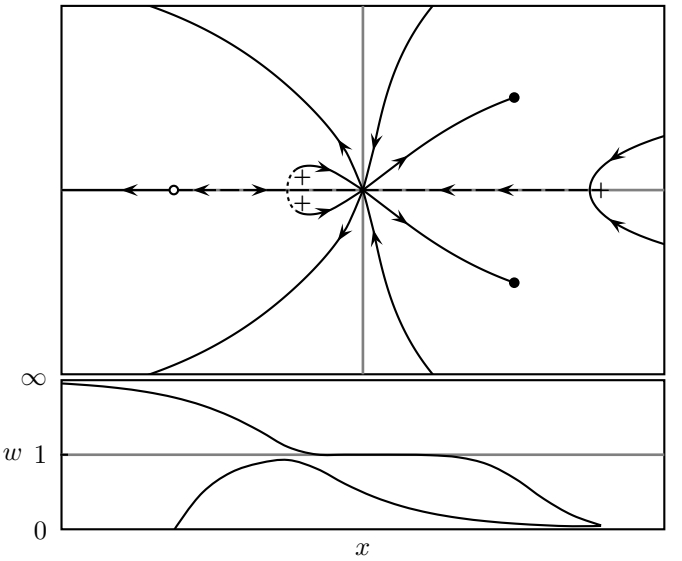


Figure 8.7. Order arrows for $[2, 0, 1]$ approximation (poles: $\bullet$, zero: $\circ$, branch points: +).

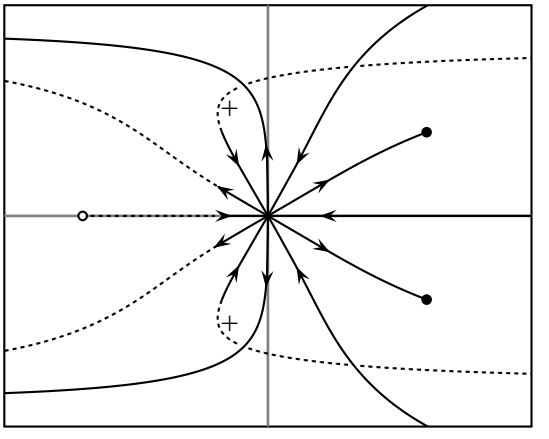


Figure 8.8. Order arrows for $[2, 1, 1]$ approximation (poles: $\bullet$, zero: $\circ$, branch points: +).

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

arrows on both sheets of the Riemann surface overlap, an additional view is given, showing w as a function of x (the real part of z). Secondly, in Figure 8.8, the $[2,1,1]$ approximation with order 5 is given. That this cannot be A-stable is seen from the up-arrows tangential to the imaginary axis at zero.

The $[2,0,1]$ and $[2,1,1]$ approximations are given respectively by

$$\Phi(w,z) = w^2\big(1 - \tfrac{8}{11}z + \tfrac{2}{11}z^2\big) - \tfrac{16}{11}w + \tfrac{5}{11} + \tfrac{2}{11}z,$$
$$\Phi(w,z) = w^2\big(1 - \tfrac{12}{23}z + \tfrac{2}{23}z^2\big) - w\big(\tfrac{16}{23} + \tfrac{16}{23}z\big) - \tfrac{7}{23} - \tfrac{2}{23}z.$$

8.8. The Butcher–Chipman conjecture

For an $[n_0, n_1, n_2]$ approximation, the value of $2n_0 - p$ seems to be related to possible A-stability. If $2n_0 - p \le 0$ then, from Theorem 8.11, A-stability is impossible. On the other hand, if $2n_0 - p > 2$, there is no known case in which A-stability occurs. For $2n_0 - p \in \{0,1,2\}$, most methods easily analysed are A-stable but not always; for example $[7,0,4]$ is not A-stable; for this approximation,

$$\Phi(w,z) = w^2\big(1 - \tfrac{1486}{1651}z + \tfrac{638}{1651}z^2 - \tfrac{512}{4953}z^3 + \tfrac{31}{1651}z^4 - \tfrac{58}{24765}z^5 + \tfrac{14}{74295}z^6 - \tfrac{4}{520065}z^7\big) - \tfrac{2048}{1651}w + \tfrac{397}{1651} + \tfrac{232}{1651}z + \tfrac{56}{1651}z^2 + \tfrac{20}{4953}z^3 + \tfrac{1}{4953}z^4. \quad (8.10)$$

Write this as $A(z)w^2 + B(z)w + C(z)$ then, because $|A(0)| > |C(0)|$, we can use the Schur criterion to find a simple necessary and sufficient condition for A-stability. This is that

$$\big(A(iy)A(-iy) - C(iy)C(-iy)\big)^2 - \big(A(iy)B(-iy) - C(-iy)B(iy)\big)\big(A(-iy)B(iy) - C(iy)B(-iy)\big) \ge 0, \quad (8.11)$$

for all real y. In the case of (8.10), (8.11) evaluates to

$$-\tfrac{3424256}{245746955354703075}y^{14} + \mathcal{O}(y^{16}).$$

After considerable numerical searching, the following statement was produced (Butcher and Chipman 1992).

Conjecture 8.12. A generalized Padé approximation with $2n_0 - p > 2$ is not A-stable.

Note that in the statement of this conjecture, the w-degree is not specified although most attempts at a proof have focused on the quadratic case. If this degree is 1, then the result is covered by Theorem 8.9. Furthermore, the method of proof for the linear case would easily generalize if it can be shown that, for each pole, there exists an up-arrow emanating from zero which terminates at this pole.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

9. Conclusions and inconclusions

If there is a single theme to the ideas presented here, it is that the general linear method formulation has reached a reasonable level of maturity. Theoretically they are quite well understood and practical methods, as well as techniques for their implementation, are starting to be developed. However, they are not simply a new class of methods, because they represent a new way of looking at earlier and more established methods.

From many points of view, the general linear formulation is more natural than the traditional way of understanding even traditional methods. A first example concerns the order conditions for general linear methods which actually give fresh insight into the significance and meaning of order for traditional methods. This is especially true in the case of Runge–Kutta methods, for which effective order is a clear-cut and useful generalization which arises naturally from a general linear point of view.

A second example of new insight coming out of general linear methods concerns non-linear stability. The irreducible formulation of linear multistep methods can stand alongside one-leg methods as a valid way of understanding G-stability and algebraic stability.

Although order arrows are not specific to general linear methods, they are featured in this paper as a tool for studying the relationship between order and stability, especially for multivalue multistage methods. They provide alternative proofs to those made available by the use of order stars and give a slightly different insight into some problems. The author would like to see the two approaches used to examine new questions, with the expectation that each of them will sometimes turn out to be the more convenient.

There is still much to be done in some of the areas identified in this paper. It would be worthwhile to know more about the consequences of algebraic stability. In particular, it would be valuable to know the extent to which general linear methods can make worthwhile contributions towards the development of structure-preserving algorithms.

However, there are already sufficiently challenging questions arising in the construction of efficient new methods. The inherent Runge–Kutta stability ansatz is promising as a source of methods but it is not yet known where to search for the best methods in this already-large family.

Finally, more detailed information on the interplay between stability and order is needed. It is a simple matter to determine in particular cases what the order of an approximation is and whether or not it is A-stable. However, there are quite likely some general patterns that can be identified and verified. The simplest outstanding question is the so-called Butcher–Chipman conjecture and, in the view of this author, is an issue capable of resolution using known techniques.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

Acknowledgements

The writing of this paper was assisted by a grant from the New Zealand Marsden Fund. I have worked, over the years, with many people on aspects of general linear methods and I wish to thank these colleagues for their collaborations. Especially I wish to acknowledge the opportunity to have worked in recent years with Zdzisław Jackiewicz, Will Wright and Helmut Podhaisky on the construction and implementation of practical methods. My interest in stability and related issues has been revived by a visit of Adrian Hill and I am grateful for the discussions he and I have had and for comments on early drafts of this paper. Steffen Voigtmann has also made constructive and helpful suggestions. Finally, I wish to thank Robert Chan, Allison Heard, and other members of the Auckland numerical analysis workshop, who have been a constant resource of support and encouragement.

REFERENCES

P. Albrecht (1978*a*) 'On the order of composite multistep methods for ordinary differential equations', *Numer. Math.* **29**, 381–396.

P. Albrecht (1978*b*) 'Explicit, optimal stability functionals and their application to cyclic discretization methods', *Computing* **19**, 233–249.

P. Albrecht (1979), *Die Numerische Behandlung Gewöhnlicher Differentialgleichungen: Eine Einführung unter Besonderer Berücksichtigung Zyklischer Verfahren*, Carl Hanser Verlag, Munich.

P. Albrecht (1985), 'Numerical treatment of ODEs: The theory of A-methods', *Numer. Math.* **47**, 59–87.

P. Albrecht (1988), 'The extension of the theory of A-methods to RK-methods', in *Numerical Treatment of Differential Equations: Halle, 1987*, Vol. 104 of *Teubner-Texte Math.*, Teubner, Leipzig, pp. 8–18.

P. Albrecht (1989), 'Elements of a general theory of composite integration methods', in *Numerical Ordinary Differential Equations: Albuquerque, NM, 1986*, *Appl. Math. Comput.* **31**, 1–17.

P. Albrecht (1996), 'The common basis of the theories of linear cyclic methods and Runge–Kutta methods', *Appl. Numer. Math.* **22**, 3–21.

R. Alexander (1977), 'Diagonally implicit Runge–Kutta methods for stiff ODEs', *SIAM J. Numer. Anal.* **14**, 1006–1021.

Z. Bartoszewski and Z. Jackiewicz (1998), 'Construction of two-step Runge–Kutta methods of high order for ordinary differential equations', *Numer. Algorithms* **18**, 51–70.

T. A. Bickart and Z. Picel (1973), 'High order stiffly stable composite multistep methods for numerical integration of stiff differential equations', *BIT* **13**, 272–286.

D. G. Brush, J. J. Kohfeld and G. T. Thompson (1967), 'Solution of ordinary differential equations using two off-step points', *J. Assoc. Comput. Mach.* **14**, 769–784.

K. Burrage (1978*a*), 'A special family of Runge–Kutta methods for solving stiff differential equations', *BIT* **18**, 22–41.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

K. Burrage (1978*b*) 'High order algebraically stable Runge–Kutta methods', *BIT* **18**, 373–383.
K. Burrage (1980), 'Non-linear stability of multivalue multiderivative methods', *BIT* **20**, 316–325.
K. Burrage (1988), 'Order properties of implicit multivalue methods for ordinary differential equations', *IMA J. Numer. Anal.* **8**, 43–69.
K. Burrage and J. C. Butcher (1979), 'Stability criteria for implicit Runge–Kutta methods', *SIAM J. Numer. Anal.* **16**, 46–57.
K. Burrage and J. C. Butcher (1980), 'Non-linear stability of a general class of differential equation methods', *BIT* **20**, 185–203.
K. Burrage and F. H. Chipman (1985), 'The stability properties of singly-implicit general linear methods', *IMA J. Numer Anal.* **5**, 287–295.
K. Burrage and F. H. Chipman (1989), 'Construction of A-stable diagonally implicit multivalue methods', *SIAM J. Numer. Anal.* **26**, 397–413.
K. Burrage and P. Moss (1980), 'Simplifying assumptions for the order of partitioned multivalue methods', *BIT* **20**, 452–465.
K. Burrage and P. W. Sharp (1994), 'A class of variable-step explicit Nordsieck multivalue methods', *SIAM J. Numer. Anal.* **31**, 1434–1451.
J. C. Butcher (1964), 'Implicit Runge–Kutta processes', *Math. Comp.* **18**, 50–64.
J. C. Butcher (1965), 'A modified multistep method for the numerical integration of ordinary differential equations', *J. Assoc. Comput. Mach.* **12**, 124–135.
J. C. Butcher (1966), 'On the convergence of numerical solutions of ordinary differential equations', *Math. Comp.* **20**, 1–10.
J. C. Butcher (1967), 'A multistep generalization of Runge–Kutta methods with four or five stages', *J. Assoc. Comput. Mach.* **14**, 84–89.
J. C. Butcher (1969), 'The effective order of Runge–Kutta methods', in *Proc. Conference on the Numerical Solution of Differential Equations: Dundee 1969* (J. L. Morris, ed.), Vol. 109 of *Lecture Notes in Mathematics*, Springer, pp. 133–139.
J. C. Butcher (1972*a*) 'An algebraic theory of integration methods', *Math. Comp.* **26**, 79–106.
J. C. Butcher (1972*b*) 'A convergence criterion for a class of integration methods', *Math. Comp.* **26**, 107–117.
J. C. Butcher (1973*a*) 'The order of numerical methods for ordinary differential equations', *Math. Comp.* **27**, 793–806.
J. C. Butcher (1973*b*) 'Order conditions for a general class of numerical methods for ordinary differential equations', in *Topics in Numerical Analysis* (J. J. H. Miller, ed.), Academic Press, London, pp. 35–40.
J. C. Butcher (1974*a*) 'The order of differential equation methods', in *Proc. Conference on the Numerical Solution of Ordinary Differential Equations: Austin 1972* (D. G. Bettis, ed.), Vol. 362 of *Lecture Notes in Mathematics*, Springer, pp. 72–75.
J. C. Butcher (1974*b*) 'Order conditions for general linear methods for ordinary differential equations', in *Numerische Methoden bei Differentialgleichungen und mit funktionalanalytischen Hilfsmitteln: Oberwolfach 1972*, Vol. 19 of *International Series of Numerical Mathematics*, Birkhäuser, pp. 77–81.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

J. C. Butcher (1975), 'A stability property of implicit Runge–Kutta methods', *BIT* **15**, 358–361.
J. C. Butcher (1981*a*) 'A generalization of singly-implicit methods', *BIT* **21**, 175–189.
J. C. Butcher (1981*b*) 'Stability properties for a general class of methods for ordinary differential equations', *SIAM J. Numer. Anal.* **18**, 37–44.
J. C. Butcher (1984), 'An application of the Runge–Kutta space', *BIT* **24**, 425–440.
J. C. Butcher (1985), 'General linear methods: A survey', *Appl. Numer. Math.* **1**, 273–284.
J. C. Butcher (1987*a*) *The Numerical Analysis of Ordinary Differential Equations: Runge–Kutta and General Linear Methods*, Wiley, Chichester.
J. C. Butcher (1987*b*) 'Linear and non-linear stability for general linear methods', *BIT* **27**, 182–189.
J. C. Butcher (1987*c*) 'The equivalence of algebraic stability and AN-stability', *BIT* **27**, 510–533.
J. C. Butcher (1988), 'On a class of matrices with real eigenvalues', *Linear Algebra Appl.* **103**, 1–12.
J. C. Butcher (1992), 'Some new hybrid methods for initial value problems', in *Computational Ordinary Differential Equations* (J. R. Cash and I. Gladwell, eds), Clarendon Press, Oxford, pp. 29–46.
J. C. Butcher (1993*a*) 'Diagonally-implicit multi-stage integration methods', *Appl. Numer. Math.* **11**, 347–363.
J. C. Butcher (1993*b*) 'General linear methods for the parallel solution of ordinary differential equations', *World Sci. Ser. Appl. Anal.* **2**, 99–111.
J. C. Butcher (1994*a*) 'The parallel solution of ordinary differential equations and some special functions', in *Approximation and Computation: West Lafayette 1993* (R. V. M. Zahar, ed.), Vol. 119 of *International Series of Numerical Mathematics*, Birkhäuser, pp. 67–76.
J. C. Butcher (1994*b*) 'A transformation for the analysis of DIMSIMs', *BIT* **34**, 25–32.
J. C. Butcher (1994*c*), 'Laguerre polynomials: Applications in numerical ordinary differential equations', in *Proc. Cornelius Lanczos International Centenary Conference* (D. Brown *et al.*, eds), SIAM, Philadelphia, PA, pp. 371–373.
J. C. Butcher (1995), 'An introduction to DIMSIMs', *Mat. Apl. Comput.* **14**, 59–72.
J. C. Butcher (1996*a*) 'General linear methods', *Comput. Math. Appl.* **31**, 105–112.
J. C. Butcher (1996*b*) 'Runge–Kutta methods as mathematical objects', in *Numerical Analysis: A. R. Mitchell 75th Birthday* (D. F. Griffiths and G. A. Watson, eds), World Scientific, Singapore, pp. 39-56.
J. C. Butcher (1997*a*) 'Order and stability of parallel methods for stiff problems', *Adv. Comput. Math.* **7**, 79–96.
J. C. Butcher (1997*b*) 'An introduction to 'Almost Runge–Kutta' methods', *Appl. Numer. Math.* **24**, 331–342.
J. C. Butcher (1998), 'ARK methods up to order five', *Numer. Algorithms* **17**, 193–221.
J. C. Butcher (2000), 'Numerical methods for ordinary differential equations in the 20th century', in *Numerical Analysis 2000*, Vol. VI: *Ordinary Differential Equations and Integral Equations*, *J. Comput. Appl. Math.* **125**, 1–29.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

J. C. Butcher (2001), 'General linear methods for stiff differential equations', *BIT,* **41**, 240–264.
J. C. Butcher (2002*a*) 'The A-stability of methods with Padé and generalized Padé stability functions', *Numer. Algorithms* **31**, 47–58.
J. C. Butcher (2002*b*) 'Software issues for ordinary differential equations', *Numer. Algorithms* **31**, 401–418.
J. C. Butcher (2003), *Numerical Methods for Ordinary Differential Equations*, Wiley, Chichester.
J. C. Butcher and J. Cash (1989), 'Some recent developments on numerical initial value problems: A survey', in *Recent Theoretical Results in Numerical Ordinary Differential Equations*, *Appl. Numer. Math.* **5**, 3–18.
J. C. Butcher and P. Chartier (1994), The construction of DIMSIMs for stiff ODEs and DAEs. Report Series, University of Auckland, New Zealand.
J. C. Butcher and P. Chartier (1995), 'Parallel general linear methods for stiff ordinary differential and differential algebraic equations', *Appl. Numer. Math.* **17**, 213–222.
J. C. Butcher and P. Chartier (1997), 'A generalization of singly-implicit Runge–Kutta methods', *Appl. Numer. Math.* **24**, 343–350.
J. C. Butcher and D. J. L. Chen (1998), 'ESIRK methods and variable stepsize', *Appl. Numer. Math.* **28**, 193–207.
J. C. Butcher and D. J. L. Chen (2001), 'On the implementation of ESIRK methods for stiff IVPs', *Numer. Algorithms* **26**, 201–218.
J. C. Butcher and J. H. Chipman (1992), 'Generalized Padé approximations to the exponential function', *BIT* **32**, 118–130.
J. C. Butcher and A. D. Heard (2002), 'Stability of numerical methods for ordinary differential equations', *Numer. Algorithms* **31**, 59–73.
J. C. Butcher and A. T. Hill (2006), 'Linear multistep methods as irreducible general linear methods', to appear in *BIT*.
J. C. Butcher and Z. Jackiewicz (1993), 'Diagonally implicit general linear methods for ordinary differential equations', *BIT* **33**, 452–472.
J. C. Butcher and Z. Jackiewicz (1996), 'Construction of diagonally implicit general linear methods of type 1 and 2 for ordinary differential equations', *Appl. Numer. Math.* **21**, 385–415.
J. C. Butcher and Z. Jackiewicz (1997*a*) 'Implementation of diagonally implicit multistage integration methods for ordinary differential equations', *SIAM J. Numer. Anal.* **34**, 2119–2141.
J. C. Butcher and Z. Jackiewicz (1997*b*) 'Construction of high order diagonally implicit multistage integration methods for ordinary differential equations', *Appl. Numer. Math.* **27**, 1–12.
J. C. Butcher and Z. Jackiewicz (2001), 'A reliable error estimation for diagonally implicit multistage integration methods', *BIT* **41**, 656–665.
J. C. Butcher and Z. Jackiewicz (2002), 'Error estimation for Nordsieck methods', *Numer. Algorithms* **31**, 75–85.
J. C. Butcher and Z. Jackiewicz (2003), 'A new approach to error estimation for general linear methods', *Numer. Math.* **95**, 487–502.
J. C. Butcher and Z. Jackiewicz (2004), 'Construction of general linear methods with Runge–Kutta stability properties', *Numer. Algorithms* **36**, 53–72.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

J. C. Butcher and N. Moir (2003), 'Experiments with a new fifth order method', *Numer. Algorithms* **33**, 137–151.
J. C. Butcher and A. E. O'Sullivan (2002), 'Nordsieck methods with an off-step point', *Numer. Algorithms* **31**, 87–101.
J. C. Butcher and H. Podhaisky (2006), 'On error estimation in general linear methods for stiff ODEs', *Appl. Numer. Math.* **56**, 345–357.
J. C. Butcher and N. Rattenbury (2005), 'ARK methods for stiff problems', *Appl. Numer. Math.* **53**, 165–181.
J. C. Butcher and A. D. Singh (2000), 'The choice of parameters in parallel general linear methods for stiff problems', *Appl. Numer. Math.* **34**, 59–84.
J. C. Butcher and S. Tracogna (1997), 'Order conditions for two-step Runge–Kutta methods', *Appl. Numer. Math.* **24**, 351–364.
J. C. Butcher and W. M. Wright (2003*a*) 'A transformation relating explicit and diagonally-implicit general linear methods', *Appl. Numer. Math.* **44**, 313–327.
J. C. Butcher and W. M. Wright (2003*b*) 'The construction of practical general linear methods', *BIT* **43**, 695–721.
J. C. Butcher, J. R. Cash and M. T. Diamantakis (1996), 'DESI methods for stiff initial value problems', *ACM Trans. Math. Software* **22**, 401–422.
J. C. Butcher, P. Chartier and Z. Jackiewicz (1997), 'Nordsieck representation of DIMSIMs', *Numer. Algorithms* **16**, 209–230.
J. C. Butcher, P. Chartier and Z. Jackiewicz (1999), 'Experiments with a variable-order type 1 DIMSIM code', *Numer. Algorithms* **22**, 237–261.
J. C. Butcher, Z. Jackiewicz and H. D. Mittelmann (1997), 'A nonlinear optimization approach to the construction of general linear methods of high order', *J. Comput. Appl. Math.* **81**, 181–196.
G. D. Byrne and R. J. Lambert (1966), 'Pseudo-Runge–Kutta methods involving two points', *J. Assoc. Comput. Mach.* **13**, 114–123.
R. Caira, C. Costabile and F. Costabile (1990), 'A class of pseudo Runge–Kutta methods', *BIT* **30**, 642–649.
J. Cash (1980), 'On the integration of stiff systems of ODEs using extended backward differentiation formulae', *Numer. Math.* **34**, 235–246.
J. Cash (1981), 'Second derivative extended backward differentiation formulas for the numerical integration of stiff systems', *SIAM J. Numer. Anal.* **18**, 21–36.
F. Ceschino and J. Kuntzmann (1963), *Problèmes Différentiels de Conditions Initiales*, Dunod, Paris.
T. M. H. Chan (1998), Algebraic structures for the analysis of numerical methods. PhD thesis, University of Auckland, New Zealand.
P. E. Chartier (1994), 'L-stable parallel one-block methods for ordinary differential equations', *SIAM J. Numer. Anal.* **31**, 552–571.
P. Chartier (1998), 'The potential of parallel multi-value methods for the simulation of large real-life problems, solving differential equations on parallel computers', *CWI Quarterly* **11**, 7–32.
G. J. Cooper (1978), 'The order of convergence of general linear methods for ordinary differential equations', *SIAM J. Numer. Anal.* **15**, 643–661.
G. J. Cooper (1981), 'Error estimates for general linear methods for ordinary differential equations', *SIAM J. Numer. Anal.* **18**, 65–82.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

M. Crouzeix (1979), 'Sur la B-stabilité des méthodes de Runge–Kutta', *Numer. Math.* **32**, 75–82.
G. Dahlquist (1956), 'Convergence and stability in the numerical integration of ordinary differential equations', *Math. Scand.* **4**, 33–53.
G. Dahlquist (1963), 'A special stability problem for linear multistep methods', *BIT* **3**, 27–43.
G. Dahlquist (1975), On stability and error analysis for stiff non-linear problems 1. Report NA 75.08, Department of Information Processing, Royal Institute of Technology, Stockholm.
G. Dahlquist (1976), 'Error analysis of a class of methods for stiff nonlinear initial value problems', in *Numerical Analysis, Dundee*, Vol. 506 of *Lecture Notes in Mathematics*, pp. 60–74.
G. Dahlquist (1978), 'G-stability is equivalent to A-stability', *BIT* **18**, 384–401.
J. W. Daniel and R. E. Moore (1970), *Computation and Theory in Ordinary Differential Equations*, Freeman.
K. Dekker (1981), Algebraic stability of general linear methods. Technical Report No. 25, Computer Science Department, University of Auckland, New Zealand.
K. Dekker (1982), Reducibility of algebraically stable general linear methods. Preprint No. NW 131/82, Mathematics Centre Amsterdam, Numerical Mathematics.
J. Donelson and E. Hansen (1971), 'Cyclic composite multistep predictor-corrector methods', *SIAM J. Numer. Anal.* **8**, 137–157.
J. R. Dormand and P. J. Prince (1980), 'A family of embedded Runge–Kutta formulae', *J. Comput. Appl. Math.* **6**, 19–26.
B. L. Ehle (1969), On Padé approximation to the exponential function and A-stable methods for the numerical solution of initial value problems. Research Report CSRR 2010, Department AACS, University of Waterloo, Canada.
B. L. Ehle (1973), 'A-stable methods and Padé approximations to the exponential', *SIAM J. Math. Anal.* **4**, 671–680.
R. Frank, J. Schneid and C. W. Ueberhuber (1981), 'The concept of B-convergence', *SIAM J. Numer. Anal.* **18**, 753–780.
C. W. Gear (1965), ' Hybrid methods for initial value problems in ordinary differential equations', *SIAM J. Numer. Anal.* **2**, 69–86.
C. W. Gear (1967), 'The numerical integration of ordinary differential equations', *Math. Comp.* **21**, 146–156.
C. W. Gear (1971), *Numerical Initial Valve Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ.
W. B. Gragg and H. J. Stetter (1964), 'Generalized multistep predictor-corrector methods', *J. Assoc. Comput. Mach.* **11**, 188–209.
R. D. Grigorieff and J. Schroll (1978), 'Über $A(\alpha)$-stabile Verfahren hoher Konsistenzordnung', *Computing* **20**, 343–350.
N. Guglielmi and N. Zennaro (2001), 'On the zero-stability of variable stepsize multistep methods: The spectral radius approach', *Numer. Math.* **88**, 445–4548.
E. Hairer and G. Wanner (1973), 'Multistep-multistage-multiderivative methods of ordinary differential equations', *Computing* **11**, 287–303.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

E. Hairer and G. Wanner (1974), 'On the Butcher group and general multi-value methods', *Computing* **13**, 1–15.
E. Hairer and G. Wanner (1996), *Solving Ordinary Differential Equations Numerically II: Stiff Problems and Differential-Algebraic Equations*, Springer, Berlin.
E. Hairer and G. Wanner (1997), 'Order conditions for general two-step Runge–Kutta methods', *SIAM J. Numer. Anal.* **34**, 2087–2089.
E. Hairer, C. Lubich and G. Wanner (2002), *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary Differential Equations*, Springer, Berlin.
E. Hairer, S. P. Nørsett and G. Wanner (1993), *Solving Ordinary Differential Equations Numerically I: Nonstiff Problems*, Springer, Berlin.
A. D. Heard (1978), The solution of the order conditions for general linear methods. Thesis, University of Auckland, New Zealand.
K. Heun (1900), 'Neue Methoden zur approximativen Integration der Differentialgleichungen einer unabhängigen Veränderlichen', *Z. Math. Phys.* **45**, 23–38.
A. T. Hill (2005), 'Nonlinear stability of general linear methods'. Submitted for publication.
A. Iserles and S. P. Nørsett (1990), 'On the theory of parallel Runge–Kutta methods', *IMA J. Numer. Anal.* **10**, 463–488.
A. Iserles and S. P. Nørsett (1991), *Order Stars*, Vol. 2 of *Applied Mathematics and Mathematical Computation*, Chapman and Hall, London.
Z. Jackiewicz and H. D. Mittelmann (1999), 'Exploiting structure in the construction of DIMSIMs', *J. Comput. Appl. Math.* **107**, 233–239.
Z. Jackiewicz and S. Tracogna (1994), 'A representation formula for two-step Runge–Kutta methods', in *Hellenic European Research on Mathematics and Informatics: Athens 1994*, Vol. 1, 2, Hellenic Mathematical Society, Athens, pp. 111–120.
Z. Jackiewicz and S. Tracogna (1995), 'A general class of two-step Runge–Kutta methods for ordinary differential equations', *SIAM J. Numer. Anal.* **32**, 1390–1427.
Z. Jackiewicz and S. Tracogna (1996), 'Variable stepsize continuous two step Runge–Kutta methods for ordinary differential equations', *Numer. Algorithms* **12**, 347–368.
Z. Jackiewicz and R. Vermiglio (1996), 'General linear methods with external stages of different orders', *BIT* **36**, 688–712.
Z. Jackiewicz and M. Zennaro (1992), 'Variable stepsize explicit two-step Runge–Kutta methods', *Math. Comp.* **59**, 421–438.
Z. Jackiewicz, R. Renaut and A. Feldstein (1991), 'Two-step Runge–Kutta methods', *SIAM J. Numer. Anal.* **28**, 1165–1182.
Z. Jackiewicz, R. Renaut and M. Zennaro (1995), 'Explicit two-step Runge–Kutta methods', *Appl. Math.* **40**, 433–456.
Z. Jackiewicz, R. Vermiglio and M. Zennaro (1995), 'Variable stepsize diagonally implicit multistage integration methods for ordinary differential equations', *Appl. Numer. Math.* **16**, 343–367.
Z. Jackiewicz, R. Vermiglio and M. Zennaro (1997), 'Regularity properties of multistage integration methods', *J. Comput. Appl. Math.* **87**, 285–302.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

R. Jeltsch (1976), 'A necessary condition for A-stability of multistep multiderivative methods', *Math. Comp.* **30**, 739–746.
R. Jeltsch and O. Nevanlinna, (1982), 'Stability and accuracy of time discretizations for initial value problems', *Numer. Math.* **40**, 245–296.
U. Kirchgraber (1986), 'Multi-step methods are essentially one-step methods', *Numer. Math.* **48**, 85–90.
J. J. Kohfeld and G. T. Thompson (1967), 'Multistep methods with modified predictors and correctors', *J. Assoc. Comput. Mach.* **14**, 155–166.
J. J. Kohfeld and G. T. Thompson (1968), 'A modification of Nordsieck's method using an off-step point', *J. Assoc. Comput. Mach.* **15**, 390–401.
W. Kutta (1901), 'Beitrag zur näherungsweisen Integration totaler Differentialgleichungen', *Z. Math. Phys.* **46**, 435–453.
M. A. Lopez-Marcos, J. M. Sanz-Serna and R. D. Skeel (1996), 'Cheap enhancement of symplectic integrators', in *Numerical Analysis*, Vol. 344 of *Pitman Res. Notes Math. Ser.*, Longman, Harlow, pp. 107–122.
M. Mihelcic (1977), 'Fast A-stable Donelson–Hansensche zyklische Verfahren zur numerischen Integration von stiff Differentialgleichungssystemen', *Ange. Inform.* **19**, 299–305.
A. Nordsieck (1962), 'On numerical integration of ordinary differential equations', *Math. Comp.* **16**, 22–49.
S. P. Nørsett (1969), 'An A-stable modification of the Adams–Bashforth methods', in *Proc. Conf. on the Numerical Solution of Differential Equations: Dundee, Scotland, June 1969*, Springer, Berlin, pp. 214–219.
N. Obreshkov (1940), 'Neue Quadraturformeln', *Abh. der Preuß. Akad. der Wiss., Math.-naturwiss. Klasse* **4**.
A. Prothero and A. Robinson (1974), 'On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations', *Math. Comp.* **28**, 145–162.
N. Rattenbury (2005), Almost Runge–Kutta methods for stiff and non-stiff problems. PhD thesis, Department of Mathematics, University of Auckland, New Zealand.
R. Renaut (1990), 'Two step Runge–Kutta methods and hyperbolic partial differential equations', *Math. Comp.* **55**, 563–579.
C. Runge (1895), 'Über die numerische Auflösung von Differentialgleichungen', *Math. Ann.* **46**, 167–178.
A. D. Singh (1999), Parallel diagonally implicit multistage integration methods for stiff ordinary differential equations. PhD thesis, University of Auckland, New Zealand.
H. M. Sloate and T. A. Bickart (1973), 'A-stable composite multistep methods', *J. Assoc. Comput. Mach.* **20**, 7–26.
D. Stoffer (1993), ' General linear methods: Connection to one step methods and invariant curves', *Numer. Math.* **64**, 395–408.
S. Tracogna (1996), 'Implementation of two-step Runge–Kutta methods for ordinary differential equations', *J. Comput. Appl. Math.* **76**, 113–136.
S. Tracogna and B. Welfert (2000), 'Two-step Runge–Kutta methods: Theory and practice', *BIT* **40**, 775–799.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press

P. J. van der Houwen and B. P. Sommeijer (1982), 'A special class of multistep Runge–Kutta methods with extended real stability interval', *IMA J. Numer. Anal.* **2**, 183–209.

G. Wanner, E. Hairer and S. P. Nørsett (1978), 'Order stars and stability theorems', *BIT* **18**, 475–489.

O. B. Widlund (1967), 'A note on unconditionally stable linear multistep methods', *BIT* **7**, 65–70.

K. Wright (1970), 'Some relationships between implicit Runge–Kutta, collocation and Lanczos τ methods and their stability properties', *BIT* **10**, 217–227.

W. M. Wright (1999), General linear methods for ordinary differential equations. MSc thesis, University of Auckland, New Zealand.

W. M. Wright (2001), 'The construction of order 4 DIMSIMs for ordinary differential equations', *Numer. Algorithms* **26**, 123–130.

W. M. Wright (2002*a*) 'Explicit general linear methods with inherent Runge–Kutta stability', *Numer. Algorithms* **31**, 381–399.

W. M. Wright (2002*b*) General linear methods with inherent Runge–Kutta stability. PhD thesis, Department of Mathematics, University of Auckland, New Zealand.

M. Zennaro (1986), 'Natural continuous extensions of Runge–Kutta methods', *Math. Comp.* **46**, 119–133.

https://doi.org/10.1017/S0962492906220014 Published online by Cambridge University Press