Aplikace matematiky

Anton Huťa; Karl Strehmel Construction of explicit and generalized Runge-Kutta formulas of arbitrary order with rational parameters

Aplikace matematiky, Vol. 27 (1982), No. 4, 259-276

Persistent URL: http://dml.cz/dmlcz/103971

Terms of use:

© Institute of Mathematics AS CR, 1982

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.



This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-GZ: The Czech Digital Mathematics Library* http://dml.cz

CONSTRUCTION OF EXPLICIT AND GENERALIZED RUNGE-KUTTA FORMULAS OF ARBITRARY ORDER WITH RATIONAL PARAMETERS

Anton Huřa and Karl Strehmel (Received October 23, 1980)

1. INTRODUCTION

For the numerical solution of initial value problems for ordinary differential equations of the first order

(1.1)
$$y'(x) = f(x, y(x)), \quad y(x_0) = y_0 \in \mathbf{R}^v, \quad x \in [x_0, x_N]$$

with $f: [x_0, x_N] \times \mathbf{R}^v \to \mathbf{R}$, methods of discretization are the only ones applied at present. We assume that f is Lipschitz continuous in the strip $S:=\{(x,y)|x_0 \le x \le x_N, y \in \mathbf{R}^v\}$, which is known to guarantee a unique solution of (1.1). Essential in choosing a numerical method is its order of consistency and its numerical stability which has been treated by numerous authors since the fundamental work of Dahlquist [3] appeared.

In the present article we shall first of all give a general principle for the construction of explicit Runge-Kutta formulas of n-th order (RK-methods), where the solution of the generally nonlinear conditional equations for the parameters is exact and rational. Up to now, the problem of the construction of RK-methods has been solved only for $n \le 10$. However, for greater n (see Curtis [2]) some formulas only hold approximately, i.e. the residues of the conditional equations are different from zero. The reason for this lies in the nonlinearity of the system of conditional equations, which increases strongly with the increasing order.

Since the explicit RK-methods are known to have a closed domain of stability they are not suitable for "stiff" systems. These systems can be characterized by the presence of transient components which, although negligible relative to the other components of the solution, constrain the step size of an explicit RK-method to be of the order of the smallest time constant of the problem. Therefore for this class of problems we shall derive, by using an explicit RK-method, a generalized RK-method with an adaptive stability function (ARK-method). For this purpose we need

the Jacobian matrix of the system (1.1) or its approximation and in addition the exponential matrix or its approximation.

In order to avoid a too complicated notation we restrict our presentation to scalar differential equations of the form (1.1), i.e. to the case v = 1. The generalization to v > 1 is mostly apparent, otherwise we shall give special references. As additional contributions to generalized RK-methods we mention the articles of Lawson [8], Ehle and Lawson [4], van der Houwen [6], Friedli [5] and Verwer [13], [14].

2. CONSTRUCTION OF n-th ORDER RK-FORMULAS

An s-stage RK-method is defined by the well-known expression

(2.1)
$$u_{m+1} = u_m + \sum_{l=0}^{s-1} p_l \, k_l(x_m, u_m),^{-1}$$
$$u_0 = v_0, \quad m = 0, 1, \dots,$$

where

$$k_0(x, y) = h \cdot f(x, y),$$

 $k_q(x, y) = h \cdot f(x + a_q h, y + \sum_{i=1}^q b_{qi} k_{j-1}), \quad a_q = \sum_{i=1}^q b_{qi}, \quad q = 1, 2, ..., s - 1.$

The local discretization error $\tau_h(x_m)$ has the form

(2.2)
$$\tau_h(x_m) = \frac{1}{h} \left[y(x_m + h) - y(x_m) - \sum_{l=0}^{s-1} p_l \ k_l(x_m, y(x_m)) \right].$$

For $y(x_m + h) - y(x_m)$ the asymptotical expansion

(2.3)
$$y(x_m + h) - y(x_m) = h \cdot f + \frac{h^2}{2!} Df + \frac{h^3}{3!} (D^2 f + f_1 Df) + \frac{h^4}{4!} (D^3 f + f_1 D^2 f + f_1^2 Df + 3Df Df_1) + \frac{h^5}{5!} [D^4 f + f_1 D^3 f + f_1^2 D^2 f + f_1^3 Df + 4D^2 f Df_1 + 6Df D^2 f_1 + 7f_1 Df_1 Df + 3f_2 (Df)^2] + \dots, (h \to 0)$$

holds with

$$D^{\mu}f := \sum_{j=0}^{\mu} \binom{\mu}{j}_{\mu-j} f_{j}f^{j}; \quad D^{\mu}f_{\nu} := \sum_{j=0}^{\mu} \binom{\mu}{j}_{\mu-j} f_{\nu+j}f^{j}; \quad {}_{p}f_{q} := \frac{\partial^{p+q}f}{\partial x^{p} \partial y^{q}}, \quad f_{q} := {}_{0}f_{q},$$

where one has to take every derivative at the point x_m , $y(x_m)$.

Remark 1. The general term of the right-hand side of (2.3) has the form

$$\frac{h^j}{j!} \prod_{t=0}^{v} (D^m t f_t)^{l_t}; \quad D^{\mu} f_0 := D^{\mu} f.$$

¹) u_m denotes the approximation to $y(x_m)$ for $x_m = x_0 + m$. h.

If the product is assigned the natural number

$$\sum_{t=0}^{v} l_t (t + m_t)$$

- the so-called "dimension" - then the expression with the factor $h^j/j!$ in (2.3) has the dimension j-1. This is advantageous in checking the Taylor expansion.

Now we introduce the following quantities:

$$c(i,1) = a_i = \sum_{v=1}^{i} b_{iv},$$

$$c(i,2) := c(i,2/m_0) = \sum_{v=2}^{i} c^{m_0}(v-1,1) b_{iv},$$

$$c(i,3) := c(i,3/m_0/m_1,r) = \sum_{v=3}^{i} c^{m_0}(v-1,1) c^{m_1}(v-1,2/r) b_{iv},$$

$$c(i,4) := c(i,4/m_0/m_1,n_1/m_2,n_2,n_3,r) =$$

$$= \sum_{v=4}^{i} c^{m_0}(v-1,1) c^{m_1}(v-1,2/n_1) c^{m_2}(v-1,3/n_2/n_3,r)$$

etc. by means of which we can write some conditional equations of an s-stage RK-method of n – th order in the form

(2.4)
$$[f]: \sum_{i=0}^{s-1} p_i = 1 ,$$

(2.5)
$$[D^q f] : \sum_{i=1}^{s-1} p_i c^q(i, 1) = \frac{1}{q+1} \quad \text{for} \quad q = 1, 2, ..., n-1 ,$$

(2.6)
$$\left[D^q f D^r f_1 \right] : \sum_{i=2}^{s-1} p_i c^r(i,1) c(i,2/q) = \frac{1}{(q+1)(q+r+2)}$$

for
$$q = 1, 2, ..., n - 2$$
,

$$r = 0, 1, ..., n - 3$$
 with $q + r \le n - 2$,

(2.7)
$$[(Df)^2 D^r f_2] : \sum_{i=2}^{s-1} p_i c^r(i,1) c^2(i,2/1) = \frac{1}{4(r+5)}$$
 for $r=0,1,...,n-5$,

(2.8)
$$\left[f_1^2 D^r f\right] : \sum_{i=3}^{s-1} p_i c(i, 3/0/1, r) = \frac{1}{(r+3)^{[3]}}^{1}$$
 for $r = 1, 2, ..., n-3$,

(2.9)
$$[f_1^3 D^r f]$$
: $\sum_{i=4}^{s-1} p_i c(i, 4/0/0/1, 0, 1, r) = \frac{1}{(r+4)^{[4]}}$ for $r=1, 2, ..., n-4$,

(2.10)
$$[f_1^{n-2}Df]: \sum_{i=n-1}^{s-1} p_i c(i, n) = \frac{1}{n!}.$$

 $⁽r+l)^{[l]} = (r+l)(r+l-1)...(r+1).$

The expressions in the brackets indicate the terms which yield by comparing the coefficients the corresponding condition of equations. The quantities c(i; v), v = 1, 2, ..., are called the quantities of v-th order. The equations (2.4) through (2.10) show that the upper summation bound – the so-called "height" – is always s - 1; whereas the lower summation bound – the so-called "depth", which we shall denote by g – assumes the values g = 0, 1, ..., n - 1.

Now we assign:

- 1. each weight coefficient p_i the weight $\varrho = 0$,
- 2. each parameter b_{ij} and c(i, 1) the weight $\varrho = 1$,
- 3. each parameter $c(i, 2/m_0)$ the weight $\varrho = m_0 + 1$.

Then we can determine recursively the weights of the parameters c(i,1), $l=3,\ldots,$. Hence it follows e.g. that the weight of $c(i,3/m_0/m_1,r_0)$ is $\varrho=m_0+m_1(r_0+1)+1$ and for $c(i,4/m_0/m_1,n_1/m_2,n_2,n_3,r)$ one gets $\varrho=m_0+m_1(n_1+1)+m_2$. $[n_2+n_3(r+1)+1]$. If we denote the number of the conditional equations of the depth g and of the weight ϱ by $\chi(n,g,\varrho)$, furthermore the number of the equations of the weight g by $\Psi(n,g)$ and the number of the conditional equations of a certain s-stage RK-method of the order of consistency n by $\varphi(n)$ then the following relations hold:

$$\sum_{\varrho} \chi(n, g, \varrho) = \Psi(n, g)$$
 and $\sum_{g} \Psi(n, g) = \varphi(n)$.

The number $\varkappa(s)$ of the parameters of an s-stage RK-method is

$$\varkappa(s) = \binom{s+1}{2}.$$

In practice for scalar differential equation one chooses s so that the inequality

$$\varkappa(s) \geq \varphi(n)$$

holds otherwise the system of conditional equations need not have a solution. Tables 1 and 2 show the number of conditional equations classified in order, depth and dimension and the dependence of s on n as well.

The degree of freedom $\sigma(n)$ of a certain s-stage RK-method is given by the relation

$$\sigma(n) = \varkappa(s) - \varphi(n).$$

In order to obtain a rational solution of the system of conditional equations it is required to linearize the system by means of suitable transformations. The first transformation has the general form

(2.11)
$$t(i, 1/j_1/j_2, m_1/j_3, m_2, m_3, m_4) = \sum_{\mu=i+1}^{s-1} p_{\mu} c^{j_1}(\mu, 1) c^{j_2}(\mu, 2/m_1) c^{j_3}(\mu, 3/m_2/m_3, m_4) b_{\mu, i+1}$$

Table 1: The number of the conditional equations and the dependence of s on n for v = 1.

Q										
n	g	0	1	2	3	4	5	$\Psi(n,g)$	$\varphi(n)$	s
1	0	1						1	1	1
2	0	1	,					1	2	2
	1		1					1		
	0	1						1		
3	1		1	1				2	4	3
	2			1				1		
	0	1						1		
4	1		1	1	1			3	8	4
	2			1	2			3		
	3				1			1		
	0	1						1		
	1		1	1	1	1		4		
5	2			1	2	4		7	16	6
	3				1	2		3		
	4					1		1		
	0	1						1		
	1		1	1	1	1	1	5		
6	2			1	2	4	6	13	31	8
	3				1	2	5	8		
	4					1	2	3		
	5						1	1		

and the second transformation is given by

(2.12)
$$t(i, 2|j_1|j_2, m_1|j_3, p, 1, m_2||j, 1, l) =$$

$$\sum_{\mu=i+1}^{s-2} c^{j_1}(\mu, 1) c^{j_2}(\mu, 2|m_1) c^{j_3}(\mu, 3|p/1, m_2) t(i, 1|j/1, l) b_{\mu, i+1}.$$

For the third and fourth transformation we give only the two special cases

(2.13)
$$t(i, 3/m/|1/j) = \sum_{\mu=i+1}^{s-3} c^m(\mu, 1) t(i, 2/1/|j) b_{\mu, i+1}$$

and

(2.14)
$$t(i, 4/p|/m/l, j) = \sum_{\mu=i+1}^{s-4} c^{\mu}(\mu, 1) t(i, 3/m/l, j) b_{\mu, i+1}.$$

Table 2: The number of the conditional equations and the dependence of s on n for v > 1.

						ρ									
n	g		0	1	2	3	4		5	Ψ	(n, g)		$\varphi(n)$		S
1	0		1								1		1		1
2	0		1						1				2		2
	1			1							1				
	0		1								1				
3	1			1	1						2		4		3
	2				1						1				
	0		1								1				
4	1			1	1	1					3		8		4
	2				1	2					3				
	3					1					1				
	0		1								1				
	1			1	1	1	1				4				
5	2				1	2	4				7		17		6
	3					1	3				4				
	4						1				1				
and the second	0		1								1				
	1			1	1	1	1		1		5				
6	2				1	2	4		6		13		37		9
	3					1	3		8		12				
	4						1		4		5				
	5								1		1				
	Table	3: σ((n) for	v = 1	1.					Table	e 4: σ(ι	n) for	· v > 1	•	
n	1	2	3	4	5	6	_		n	1	2	3	4	5	
$\sigma(n)$	0	1	2	2	5	5		σ	(n)	0	1	2	2	4	

In the equations (2.11) to (2.14) the transformed variables occur in the form t(i, q/.../...) so that we can write the equations of the transformation schematically in the form

$$t(i, q/.../...) = \sum_{\mu=i+1}^{s-q} \gamma(\mu) t(i, q-1/.../...) b_{\mu,i+1}$$

where i indicates the index of the transformed variable t, q the order of the transfor-

mation and the sign//separates the untransformed part from the transformed one. Now we apply the transformation (2.11) to the system of nonlinear conditional equations of an s-stage consistent RK-method of order n. To the system transformed in this way we apply the transformation (2.12) etc.. In all the transformations the heights and the depths are diminished by one so that their difference remain constant. This implies that only equations with the depth of at least two are transformable. For the equations (2.6) to (2.9) we obtain by means of the transformation (2.11) the relations

(2.15)
$$\sum_{i=1}^{s-2} c^{q}(i, 1) t(i, 1/r) = \frac{1}{(q+1)(q+r+2)} \quad \text{for} \quad q = 1, 2, ..., n-2;$$
$$r = 0, 1, ..., n-3 \quad \text{with} \quad q+r \le n-2,$$

(2.16)
$$\sum_{i=1}^{s-2} c(i,1) \ t(i.1/r/1,1) = \frac{1}{4(r+5)} \text{ for } r=0,1,...,n-5.$$

(2.17)
$$\sum_{i=2}^{s-2} c(i, 2/r) t(i, 1/0) = \frac{1}{(r+3)^{[3]}} \quad \text{for} \quad r = 1, 2, ..., n-3,$$

(2.18)
$$\sum_{i=3}^{s-2} c(i, 3/0/1, r) t(i, 1/0) = \frac{1}{(r+4)^{[4]}} \quad \text{for} \quad r=1, 2, ..., n-4,$$

By using the second transformation we hence obtain the equations

(2.19)
$$\sum_{i=1}^{s-3} c^r(i,1) \ t(i,2/0/0) = \frac{1}{(r+3)^{[3]}} \quad \text{for} \quad r=1,2,...,n-3 \, .$$

(2.20)
$$\sum_{i=2}^{s-3} c(i, 2/r) t(i, 2/0/0) = \frac{1}{(r+4)^{[4]}} \text{ for } r=1, 2, ..., n-4$$

and then the third transformation yields

(2.21)
$$\sum_{i=1}^{s-4} c^r(i,1) \ t(i,3/0)/0,0) = \frac{1}{(r+4)^{[t+1]}} \quad \text{for} \quad r=1,2,...,n-4.$$

It is seen that the order of the quantities c(i; v) decreases by every transformation, for instance in the equations (2.9) (2.18), (2.20) and (2.21) we obtain successively the quantities c(i, 4), c(i, 3), c(i, 2) and c(t, 1).

The algorithm for the determination of the rational parameters of an explicit RK-method consists of the following steps:

- 1. To a prescribed consistency order n for the numerical solution of a scalar differential equation one determines the stage number s of the RK-method as the least natural number for which the condition $\varkappa(s) \ge \varphi(n)$ holds.
- 2. After choosing some weight coefficients $p_i = 0$ for $i = 1, 2, ..., \xi$, (the chosen number ξ depends on the consistency order n) and by a suitable definition of the step

parameters a_i for $i = \xi + 1, \xi + 2, ..., s - 1$, we obtain, according to (2.5), a linear system of equations with a Vandermonde matrix for the weight coefficients p_i , $i = \xi + 1, ..., s - 1$. The parameters p_i and the step parameters a_i must be adequate to the number of the equations. Then the weight coefficient p_0 follows from (2.4). For the transformations to be executed it is advantageous to choose the value of the parameter $a_{s-1} = 1$.

3. We introduce the substitutions

(2.22)
$$c(i, 2/m) = \frac{1}{m+1} a_i^{m+1}$$
 for $i = \xi + 1$, $\xi + 2, ..., s - 1$,

$$c(i, 3/0/1, m) = \frac{1}{(m+1)(m+2)} a_i^{m+2}$$
 for $i = \xi + 1$, $\xi + 2, ..., s - 1$

and in general

$$c(i, v/0/0/.../1, ..., m) = \frac{1}{(m+v-1)^{[v-1]}} a_i^{m+v-1}.$$

Remark 2. The cases $i = \xi + 1$ and $i = \xi + 2$ in (2.22) were already used by Nyström [10].

- 4. We solve the system of linear equations for the transformed variables t(i, 1/... | ... | ... | ... |). The coefficients of this system are the quantities $c^r(i, 1)$ for r = 1, 2, ... | ..., n 3 and c(i, v/0/0/.../1, ... m)
- 5. We solve the system of linear equations for the transformed variables t(i, 2 | ... | ... | ... |); t(i, 3 | ... | ... | ... |) etc. The number of the transformations depends on the order of consistency of the RK-method.

Remark 3. a) We denote by $\eta(n)$ the number of equations derived from the above presented system of conditional equations, from the definitions of the variables c(i, l) and from the transformed equations. Tables 5 and 6 show the values of $\eta(n)$ and $\varphi(n)$ for some n.

Table 5: $\varphi(n)$ and $\eta(n)$ for v=1

п	1	2	3	4	5	6
$\varphi(n)$	1	2	4	8	16	31
$\eta(n)$	1	3	11	26	87	253

Table 6: $\varphi(n)$ and $\eta(n)$ for v > 1.

n	1	2	3	4	5	6
$\varphi(n)$	1	2	4	8	17	37
$\eta(n)$	1	3	11	26		273

Since under the number $\eta(n)$ of equations a great number of dependent equations occurs, for the determination of the parameters of a s-stage RK-method not all equations are necessary.

b) Two RK-methods with (s, n) = (8, 6) constructed by means of this algorithm have, in the notation of Butcher [1], the following form:

With regard to the construction of these methods see Huia [7].

3. CONSTRUCTION OF ARK-FORMULAS

For the derivation of an s-stage ARK-method we take as a basis an s-stage consistent RK-method (2.1), which we write in the following recursive form:

(3.1)
$$x_{m+1}^{(0)} := x_m \; ; \quad u_{m+1}^{(0)} := u_m \; ,$$

$$x_{m+1}^{(q)} := x_m + a_q h \; ; \quad u_{m+1}^{(q)} := u_m + h \sum_{j=1}^q b_{qj} f(x_{m+1}^{(j-1)}, u_{m+1}^{(j-1)}) \; ; \quad q = 1, 2, ..., s \; ,$$

$$x_{m+1}^{(q)} := x_{m+1}^{(s)} \; ; \quad u_{m+1}^{(s)} := u_{m+1}^{(s)} \quad \text{with} \quad a_s := 1 \; , \quad b_{sj}^{(s)} := p_{j-1}^{(s)} \; .$$

In principle these RK-methods are based on the formal solution of the equivalent

Voltera integral equation to (1.1)

(3.2)
$$y(x) = y(x_m) + \int_{x_m}^x f(t, y(t)) dt,$$

where in each discretization interval $[x_m, x_{m+1}^{(q)}], m = 0, 1, ..., q = 1, 2, ..., s - 1$, the function f(x, y) is approximated by a polynomial

$$f_h^{(q)}(x) = \sum_{l=0}^{\varrho_q} \alpha_{l,m}^{(q)} \cdot (x - x_m)^l$$
.

The coefficients $\alpha_{l,m}^{(q)}$ depend on the function values $f_i := f(x_{m+1}^{(i)}, u_{m+1}^{(i)}), i = 0, 1, ..., q-1$. The approximation values $u_{m+1}^{(q)}$ of the s-stage RK-method (3.1) can be written in the form

$$u_{m+1}^{(q)} = u_m + h \cdot a_q \sum_{l=0}^{q} \alpha_{l,m}^{(q)} \frac{(h \cdot a_q)^l}{l+1}.$$

First of all, for the construction of ARK-methods the given differential equation (1.1) is linearized in each discretization interval $[x_m, x_{m+1}]$ formally in the form

(3.3)
$$y' = A_m y + g(x, y),$$

where A_m is a real constant (for a system, i.e. v > 1, A_m is a constant ($v \times v$) matrix). The residue function g(x, y) in (3.3) is defined by

$$g(x, y) = f(x, y) - A_m y.$$

With the initial value $y(x_m) = y_m$ we obtain from (3.3)

(3.4)
$$y(x) = \exp(A_m(x - x_m)) \left[y_m + \int_{x_m}^x \exp(-A_m(t - x_m)) g(t, y(t)) dt \right].$$

Analogously to the RK-method we approximate the function g(x, y) in $[x_m, x_{m+1}^{(q)}]$ by a polynomial

(3.5)
$$g_h^{(q)}(x) = \sum_{l=0}^{\ell_q} \hat{\alpha}_{l,m}^{(q)} \cdot (x - x_m)^l.$$

For the determination of the coefficients $\hat{\alpha}_{l,m}^{(q)}$ the same nodes are used in $[x_m, x_{m+1}^{(q)}]$ as for the coefficients $\alpha_{l,m}^{(q)}$ of the corresponding s-stage RK-method, so that

$$\hat{\alpha}_{l,m}^{(q)} = \hat{\alpha}_{l,m}^{(l)} (g(x_{m+1}^{(0)}, u_{m+1}^{(0)}), \dots, g(x_{m+1}^{(q-1)}, u_{m+1}^{(q-1)}).$$

Using (3.5) we obtain the approximate solution $u_{m+1}^{(q)}$ for $y(x_{m+1}^{(q)})$ from (3.4) in the form

$$(3.6) u_{m+1}^{(q)} = e_0(A_m a_q h) u_m + h a_q \sum_{l=0}^{\varrho_q} e_{l+1}(A_m a_q h) \hat{\alpha}_{l,m}^{(q)}(a_q h)^l, \quad q = 1, 2, ..., s,$$

where the function $e_1(z)$, $z \in \mathbf{C}$, is defined by

$$e_0(z) := \exp(z),$$

$$e_1(z) := \begin{cases} \frac{e_0(z) - 1}{z} & \text{for } z \neq 0 \\ 1 & \text{for } z = 0 \end{cases} = \sum_{i=0}^{\infty} \frac{z^i}{(1+i)!},$$

$$e_{l+1}(z) := \begin{cases} \frac{l e_1(z) - 1}{z} & \text{for } z \neq 0 \\ \frac{1}{l+1} & \text{for } z = 0 \end{cases} = \sum_{i=0}^{\infty} \frac{z^i}{(i+l+1)!}, \quad l = 1, 2, \dots.$$

(see Nickel/Rieder [9]). For the order of consistency of an ARK-method the following theorem holds:

Theorem 1. Let $f(x, y) \in C^n(U)$; $U := \{(x, y) | x_0 \le x \le x_N, |y - y(x)| \le \varepsilon, \varepsilon > 0\}$ and

$$\int_{x}^{x_{m+h}} \left[g(t, y(t)) - g_{h}^{(s)}(t) \right] dt = 0(h^{n+1}), \quad h \to 0 ;$$

then the s-stage ARK-method has the consistency order n.

Proof. For the local discretization error $\tau_h(x_m)$ we obtain from (3.4) and (3.6)

$$\tau_h(x_m) = \frac{1}{h} \left[\int_{x_m}^{x_m+h} e_0(A_m(x_m+h-t)) g(t, y(t)) dt - h \sum_{l=0}^{e_s} e_{l+1}(A_m h) \hat{\alpha}_{l,m}^{(s)} h^l \right].$$

With regard to (3.5) it follows that

(3.7)
$$\tau_h(x_m) = \frac{1}{h} \int_{x_m}^{x_m+h} e_0(A_m(x_m+h-t)) \left[g(t,y(t)) - g_h^{(s)}(t)\right] dt.$$

By inserting the expression

$$e_0(A_m(x_m + h - t)) = 1 + A_m \cdot (x_m + h - t) + \dots$$

into the integral (3.7) the first term yields

(3.8)
$$\frac{1}{h} \int_{x_{m}}^{x_{m}+h} \left[g(t, y(t)) - g_{h}^{(s)}(t) \right] dt,$$

which, according to the assumption, is of the order n. Then the further terms of the integral (3.7) yield contributions which are of higher order in h than in (3.8). \square

Remark 4. An RK-method (2.1) has the order of consistency n if

(3.9)
$$\frac{1}{h} \int_{x_{m}}^{x_{m}+h} \left[f(t, y(t)) - f_{h}^{(s)}(t) \right] dt = 0(h^{n}), \quad h \to 0.$$

The conditions of consistency for the approximation functions $f_h^{(s)}(x)$ and $g_h^{(s)}(x)$ are proved to be of the same kind.

3.1. An example of an ARK-method

In the following section, to the 3/8 rule which is characterized by the parameter scheme

a suitable ARK-method of the same order (n = 4) is constructed. In an analogous manner to each s-stage RK-method we can derive an s-stage ARK-method.

In virtue of

$$f_h^{(1)}(x) = \alpha_{0,m}^{(1)}$$

the identity

$$u_{m+1}^{(1)} = u_m + \frac{1}{3}h\alpha_{0,m}^{(1)} = u_m + \frac{1}{3}hf_0$$

directly implies

$$\alpha_{0,m}^{(1)} = f_0.$$

Taking into account

$$f_h^{(2)}(x) = \sum_{l=0}^{2} \alpha_{l,m}^{(2)}(x - x_m)^l,$$

$$u_{m+1}^{(2)} = u_m + \int_{x_m}^{x_m + \frac{2}{3}h} f_h^{(2)}(x) dx = u_m + h(-\frac{1}{3}f_0 + f_1)$$

and the requirements on

$$f_h^{(2)}(x_m) = f_0, \quad f_h^{(2)}(x_{m+1}^{(1)}) = f_1,$$

we obtain that the coefficients $\alpha_{l,m}^{(2)}$ satisfy the linear system of equations

(3.11)
$$\alpha_{0,m}^{(2)} = f_0$$

$$\alpha_{0,m}^{(2)} + \frac{1}{3}h\alpha_{1,m}^{(2)} + \frac{1}{9}h^2\alpha_{2,m}^{(2)} = f_1,$$

$$\frac{2}{3}\alpha_{0,m}^{(2)} + \frac{2}{0}h\alpha_{1,m}^{(2)} + \frac{8}{81}h^2\alpha_{2,m}^{(2)} = -\frac{1}{3}f_0 + f_1.$$

In virtue of

$$f_h^{(3)}(x) = \sum_{l=0}^{3} \alpha_{l,m}^{(3)}(x - x_m)^l,$$

$$u_{m+1}^{(3)} = u_m + \int_{x_m}^{x_m + h} f_h^{(3)}(x) dx = u_m + h(f_0 - f_1 + f_2)$$

and because of the requirements on

$$f_h^{(3)}(x_m) = f_0 \; ; \quad f_h^{(3)}(x_{m+1}^{(1)}) = f_1 \; ,$$

$$f_h^{(3)}(x_{m+1}^{(2)}) = \sum_{l=0}^{2} \gamma_l^{(3)} f_l \quad \text{with} \quad \sum_{l=0}^{2} \gamma_l^{(3)} = 1 \; , \quad \gamma_l^{(3)} \in \mathbf{R}$$

we obtain the coefficients $\alpha_{l,m}^{(3)}$ from the system of linear equations

(3.12)
$$\alpha_{0,m}^{(3)} = f_0,$$

$$\alpha_{0,m}^{(3)} + \frac{1}{3}h\alpha_{1,m}^{(3)} + \frac{1}{9}h^2\alpha_{2,m}^{(3)} + \frac{1}{27}h^3\alpha_{3,m}^{(3)} = f_1,$$

$$\alpha_{0,m}^{(3)} + \frac{2}{3}h\alpha_{1,m}^{(3)} + \frac{4}{9}h^2\alpha_{2,m}^{(3)} + \frac{8}{27}h^3\alpha_{3,m}^{(3)} = \gamma_0^{(3)}f_0 + \gamma_1^{(3)}f_1 + \gamma_2^{(3)}f_2,$$

$$\alpha_{0,m}^{(3)} + \frac{1}{2}h\alpha_{1,m}^{(3)} + \frac{1}{2}h^2\alpha_{2,m}^{(3)} + \frac{1}{4}h^3\alpha_{3,m}^{(3)} = f_0 - f_1 + f_2.$$

Finally, taking into account

$$f_h^{(4)}(x) = \sum_{l=0}^{4} \alpha_{l,m}^{(4)}(x - x_m)^l,$$

$$u_{m+1} = u_m + \int_{x_m}^{x_m + h} f_h^{(4)}(x) \, \mathrm{d}x = u_m + \frac{h}{8} \left(f_0 + 3f_1 + 3f_2 + f_3 \right)$$

and the requirements on

$$f_h^{(4)}(x_m) = f_0; \quad f_h^{(4)}(x_{m+1}^{(1)}) = f_1; \quad f_h^{(4)}(x_{m+1}^{(2)}) = f_2;$$

$$f_h^{(4)}(x_{m+1}^{(3)}) = \sum_{l=0}^{3} \gamma_l^{(4)} f_l \quad \text{with} \quad \sum_{l=0}^{3} \gamma_l^{(4)} = 1; \quad \gamma_l^{(4)} \in \mathbf{R}$$

we obtain the coefficients $\alpha_{i,m}^{(4)}$ from the system of linear equations

$$(3.13) \ \alpha_{0,m}^{(4)} = f_0 ,$$

$$\alpha_{0,m}^{(4)} + \frac{1}{3}h\alpha_{1,m}^{(4)} + \frac{1}{9}h^2\alpha_{2,m}^{(4)} + \frac{1}{27}h^3\alpha_{3,m}^{(4)} + \frac{1}{81}h^4\alpha_{4,m}^{(4)} = f_1 ,$$

$$\alpha_{0,m}^{(4)} + \frac{2}{3}h\alpha_{1,m}^{(4)} + \frac{4}{9}h^2\alpha_{2,m}^{(4)} + \frac{8}{27}h^3\alpha_{3,m}^{(4)} + \frac{16}{81}h^4\alpha_{4,m}^{(4)} = f_2 ,$$

$$\alpha_{0,m}^{(4)} + h\alpha_{1,m}^{(4)} + h^2\alpha_{2,m}^{(4)} + h^3\alpha_{3,m}^{(4)} + h^4\alpha_{4,m}^{(4)} = \sum_{l=0}^{3} \gamma_l^{(4)}f_l ,$$

$$\alpha_{0,m}^{(4)} + \frac{1}{2}h\alpha_{1,m}^{(4)} + \frac{1}{3}h^2\alpha_{2,m}^{(4)} + \frac{1}{4}h^3\alpha_{3,m}^{(4)} + \frac{1}{5}h^4\alpha_{4,m}^{(4)} = \frac{1}{8}f_0 + \frac{3}{8}f_1 + \frac{3}{8}f_2 + \frac{1}{8}f_3 .$$

Now by replacing the function values f_i , i=0,1,...,3, by suitable function values $g_i:=g(x_{m+1}^{(i)},u_{m+1}^{(i)})$ one obtains the coefficients $\hat{\alpha}_{1,m}^{(q)}$, q=1,2,3,4, from (3.10), (3.11), (3.12) and (3.13). With these coefficients $\hat{\alpha}_{1,m}^{(q)}$ we can determine the values $u_{m+1}^{(q)}$ according to (3.6). Analogously to the 3/8 rule we can write the ARK-methods in the form

$$u_{m+1}^{(q)} = e_0(a_q h A_m) u_m + h \sum_{j=1}^q B_{qj}(a_q h A_m) g_{j-1}, \quad q = 1, 2, 3, 4,$$

which can be characterized by the parameter scheme

The parameters $B_{ai}(a_a h \cdot A_m)$ are given by

$$B_{11}(\frac{1}{3}h \cdot A_m) = \frac{1}{3}e_1(\frac{1}{3}h \cdot A_m),$$

$$B_{21}(\frac{2}{3}h \cdot A_m) = \frac{2}{3}e_1(\frac{2}{3}h \cdot A_m) + \frac{2}{3}e_2(\frac{2}{3}h \cdot A_m) - 4e_3(\frac{2}{3}h \cdot A_m),$$

$$B_{22}(\frac{2}{3}h \cdot A_m) = -\frac{2}{3}e_2(\frac{2}{3}h \cdot A_m) + 4e_3(\frac{2}{3}h \cdot A_m),$$

$$B_{31}(h.A_m) = e_1(h.A_m) + \frac{1}{2}(3 - 15\gamma_0^{(3)}) e_2(h.A_m) + \frac{1}{2}(63\gamma_0^{(3)} - 45) e_3(h.A_m) + 27(1 - \gamma_0^{(3)}) e_4(h.A_m),$$

$$B_{32}(h.A_m) = -\left(2 + \frac{15}{2}\gamma_1^{(3)}\right) e_2(h.A_m) + \left(27 + \frac{63}{2}\gamma_1^{(3)}\right) e_3(h.A_m) - 9\left(4 + 3\gamma_1^{(3)}\right) e_4(h.A_m),$$

$$B_{33}(h \cdot A_m) = \left(8 - \frac{15}{2}\gamma_2^{(3)}\right) e_2(h \cdot A_m) + \left(\frac{63}{2}\gamma_2^{(3)} - 36\right) e_3(h \cdot A_m) + 9(4 - 3\gamma_2^{(3)}) e_4(h \cdot A_m),$$

$$B_{41}(h . A_m) = e_1(h . A_m) - \left(\frac{13}{2}\gamma_0^{(4)} + \frac{11}{2}\right)e_2(h . A_m) + \left(\frac{147}{4}\gamma_0^{(4)} + 9\right)e_3(h . A_m) - \left(63\gamma_0^{(4)} + \frac{9}{2}\right)e_4(h . A_m) + \frac{135}{4}\gamma_0^{(4)}e_5(h . A_m),$$

$$\begin{split} B_{42}(h \, . \, A_m) &= \left(9 \, - \frac{13}{2} \gamma_1^{(4)}\right) e_2(h \, . \, A_m) \, + \, \frac{1}{4} \big(147 \gamma_1^{(4)} \, - \, 90\big) \, e_3(h \, . \, A_m) \, + \\ &\quad + \, \big(\frac{27}{2} \, - \, 63 \gamma_1^{(4)}\big) \, e_4(h \, . \, A_m) \, + \, \frac{135}{4} \gamma_1^{(4)} e_5(h \, . \, A_m) \, , \end{split}$$

$$B_{43}(h \cdot A_m) = -\frac{1}{2}(9 + 13\gamma_2^{(4)}) e_2(h \cdot A_m) + (18 + \frac{147}{4}\gamma_2^{(4)}) e_3(h \cdot A_m) - (\frac{27}{2} + 63\gamma_2^{(4)}) e_4(h \cdot A_m) + \frac{135}{4}\gamma_2^{(4)} e_5(h \cdot A_m),$$

$$\begin{array}{lll} B_{44}(h \, . \, A_m) & = \frac{1}{2}(15 \, - \, 13\gamma_3^{(4)}) \, e_2(h \, . \, A_m) \, - \, \frac{1}{4}(165 \, - \, 147\gamma_3^{(4)}) \, e_3(h \, . \, A_m) \, + \\ & + \, \left(\frac{135}{2} \, - \, 63\gamma_3^{(4)}\right) \, e_4(h \, . \, A_m) \, - \, \frac{135}{4}(1 \, - \, \gamma_3^{(4)}) \, e_5(h \, . \, A_m) \, . \end{array}$$

The ARK-method has the consistency order n=4, if the parameters $\gamma_l^{(q)}$, q=3,4, l=0,1,...,q-1, satisfy the conditions

$$\begin{split} & \gamma_1^{(3)} = \, -\frac{44}{9} - \, 2\gamma_0^{(3)} \, , \quad \gamma_2^{(3)} = \, \frac{53}{9} \, + \, \gamma_0^{(3)} \, , \quad \gamma_0^{(3)} \in \mathbf{R} \, , \\ & \gamma_0^{(4)} = \, -\frac{4}{11} \, , \quad \gamma_1^{(4)} = \, \frac{12}{11} \, , \quad \gamma_2^{(4)} = \, -\frac{12}{11} \, , \quad \gamma_3^{(4)} = \, \frac{15}{11} \, \end{split}$$

as can be easily shown with aid of Theorem 1.

3.2. S-stability of ARK-methods

As the A-stability of a numerical method does not guarantee a stable numerical solution when a very stiff system with a variable Jacobian matrix is given, Prothero and Robinson [11] introduced the stronger notion of the S-stability.

Definition 1. A one-step method is said to be S-stable if, for a differential equation of the form

$$(3.15) \quad y'(x) = \lambda(y(x) - r(x)) + r'(x), \quad \lambda \in \mathbf{C}, \quad \text{Re}(\lambda) < 0, r(x) \in C^1[0, x^*].$$

the method yields a set of approximations $\{u_m\}$ with the property that for any constant $\lambda_0 < 0$ there exists a constant $h_0 > 0$ such that

$$\left| \frac{u_{m+1} - r(x_{m+1})}{u_m - r(x_m)} \right| < 1 \quad \text{with} \quad x_m, x_{m+1} \in [0, x^*],$$

provided that $u_m \neq r(x_m)$, for all step-lengths $h \in (0, h_0)$ and all λ with $\text{Re}(\lambda) \leq \lambda_0$.

Definition 2. A one-step method is said to be internally S-stable, if at each q-th stage, q = 1, 2, ..., s, the corresponding scheme of stage q is S-stable (see Verwer [14]).

Let us investigate the stability behaviour of the ARK-methods with respect to the differential equation (3.15).

Applying

$$u_{m+1}^{(q)} = e_0(a_q h \cdot A_m) u_m + h \sum_{j=1}^q B_{qj}(a_q h \cdot A_m) g_{j-1}, \quad q = 1, 2, ..., s,$$

to the scalar test equation (3.15) with $A_m = \lambda$ we obtain

$$u_{m+1}^{(q)} = e_0(a_q z) u_m + h \sum_{j=1}^q B_{qj}(a_q z) \delta_{j-1},$$

where

(3.16)
$$z = h\lambda; \quad \delta_{i-1} = r'(x_{m+1}^{(j-1)}) - \lambda r(x_{m+1}^{(j-1)}).$$

For

$$\varepsilon_{m+1}^{(q)} := u_{m+1}^{(q)} - r(x_{m+1}^{(q)}), \quad q = 1, 2, ..., s,$$

at the q-th stage we obtain the difference equation

(3.17)

$$\varepsilon_{m+1}^{(q)} = e_0(a_q z) \, \varepsilon_{m+1}^{(0)} + h \left\{ \sum_{j=1}^q B_{qj}(a_q z) \, \delta_{j-1} + \frac{1}{h} \left[e_0(a_q z) \, r(x_{m+1}^{(0)}) - r(x_{m+1}^{(q)}) \right] \right\}$$

with

$$\varepsilon_{m+1}^{(0)} = u_{m+1}^{(0)} - r(x_{m+1}^{(0)}).$$

We state the following theorem:

Theorem 2. An s-stage ARK-method is internally S-stable, if and only if the step parameters a_q , q = 1, 2, ..., s - 1, satisfy the condition $0 < a_q \le 1$.

Proof. We use the following

criterion (see Verwer [14]): Let us assume that by applying an s-stage one-step method to the equation (3.15) we obtain at each stage q the error equation

$$\varepsilon_{m+1}^{(q)} = T_q(z) \, \varepsilon_{m+1}^{(0)} + h \left\{ \Psi_{m+1}^{(q)}(z, h; r) + \frac{1}{h} \left[T_q(z) \, r(x_{m+1}^{(0)}) - r(x_{m+1}^{(q)}) \right] \right\},$$

$$q = 1, 2, \dots, s.$$

Then the s-stage one-step method is internally S-stable, if and only if

- a) The stability functions $T_q(z)$, q = 1, 2, ..., s, are strongly A-acceptable, 1)
- b) a constant $\bar{h} > 0$ exists, such that the local error

$$\xi_m^{(q)} := \Psi_{m+1}^{(q)}(z, h; r) + \frac{1}{h} \left[T_q(z) \, r(x_{m+1}^{(0)}) - r(x_{m+1}^{(q)}) \right], \quad q = 1, 2, ..., s,$$

is uniformly bounded on $M := \{(h, z) \mid h \in (0, \bar{h}], \operatorname{Re}(z) < 0\}$. The mean value theorem applied to

$$\xi_m^{(q)} := \sum_{j=1}^q B_{qj}(a_q z) \, \delta_{j-1} \, + \, \frac{1}{h} \left[e_0(a_q z) \, r(x_{m+1}^{(0)}) \, - \, r(x_{m+1}^{(q)}) \right],$$

yields

(3.18)
$$\xi_{m}^{(q)} = \sum_{j=1}^{q} B_{qj}(a_{q}z) \, r'(x_{m+1}^{(j-1)}) + \frac{z}{h} \, r(x_{m+1}^{(0)}) \left[a_{q} \, e_{1}(a_{q}z) - \sum_{j=1}^{q} B_{qj}(a_{q}z) \right] - z \sum_{j=2}^{q} a_{j-1} \, B_{qj}(a_{q}z) \, r'(x_{m} + \Theta_{j-1}a_{j-1}h) - a_{q} \, r'(x_{m} + \Theta_{q}a_{q}h)$$

with $0 < \Theta_i < 1$. Because of

$$\sum_{j=1}^{q} B_{qj}(a_q z) - a_q e_1(a_q z) = 0$$

(see Section 3.1.) it follows that

(3.19)
$$\zeta_m^{(q)} = \sum_{i=1}^q B_{qj}(a_q z) \, r'(x_{m+1}^{(j-1)}) - z \sum_{j=2}^q a_{j-1} \, B_{qj}(a_q z) \, r'(x_m + \Theta_{j-1} a_{j-1} h) - a_a \, r'(x_m + \Theta_a a_a h) .$$

Because the functions $B_{qj}(a_q z)$ and z $B_{qj}(a_q z)$ for q=1,2,...,s; j=1,2,...,q are uniformly bounded on $\{z \mid Re(z) < 0\}$ for $0 < a_q \le 1$, it follows from (3.19) that $\xi_m^{(q)}$ is uniformly bounded on M. The assertion follows immediately from the criterion.

¹⁾ $T_q(z)$ is said to be *strongly A-acceptable*, if $\left|T_q(z)\right| < 1$ whenever Re(z) < 0 and $\lim \left|T_q(z)\right| < 1$ as $\text{Re}(z) \to -\infty$.

Corollary. An ARK-method is always S-stable.

Proof. This result follows immediately from Theorem 2.

Definition 3. An S-stable one-step method is said to be L-S-stable if in addition

$$\left| \frac{u_{m+1}^{(s)} - r(x_{m+1}^{(s)})}{u_m - r(x_m)} \right| \to 0 ; \quad x_m, x_{m+1} \in [0, x^*]$$

as Re $(\lambda) \to -\infty$, for all stepsizes h > 0.

Theorem 3. An ARK-method is L-S-stable, if and only if the conditions

(3.20)
$$\lim_{\text{Re}(z)\to-\infty} z \sum_{l} B_{sl}(z) = \begin{cases} 0 \text{ for } a_{l-1} = \alpha, \ 0 \le \alpha < 1 \\ -1 \text{ for } a_{l-1} = 1 \end{cases}, \quad l \in \{1, 2, ..., s\}$$

hold.

Proof. From equation (3.17) and Definition 3 we see that we have L-S-stability if and only if, for any h > 0, the local error

$$\xi_m^{(s)} = \sum_{j=1}^s B_{sj}(z) \, \delta_{j-1} + \frac{1}{h} \left[e_0(z) \, r(x_{m+1}^{(0)}) - r(x_{m+1}^{(s)}) \right]$$

tends to zero as Re $(z) \rightarrow -\infty$. Taking into account (3.16) and

$$\lim_{\text{Re}(z)\to-\infty} B_{sj}(z) = 0 \quad \text{for} \quad j = 1, 2, ..., s$$

we obtain

(3.21)
$$\lim_{\text{Re}(z) \to -\infty} \xi_m^{(s)} = -\frac{1}{h} \left[\lim_{\text{Re}(z) \to -\infty} z \sum_{j=1}^s B_{sj}(z) \, r(x_{m+1}^{(j-1)}) + r(x_m + h) \right].$$

From (3.21) we obtain (3.20) directly.

Remark 5. The ARK-method (3.14) is L-S-stable, if and only if the parameters $\gamma_1^{(4)}$, l = 0, 1, 2, 3, satisfy the conditions

$$\gamma_0^{(4)} = \gamma_1^{(4)} = \gamma_2^{(4)} = 0, \quad \gamma_3^{(4)} = 1.$$

Other L-S-stable ARK-methods are given in [12].

References

- [1] J. C. Butcher, Implicit Runge-Kutta processes. Math. Comp. 18, 50 (1964).
- [2] A. R. Curtis: An eight order Runge-Kutta process with eleven function evaluations per step. Numer. Math. 16, 268-277 (1970).
- [3] G. Dahlquist: A special stability problem for linear multistep methods. BIT 3, 27-43 (1963).
- [4] B. L. Ehle and J. D. Lawson: Generalized Runge-Kutta processes for stiff initial-value problems. J. Inst. Math. Appl. 16, No. 1, 11-21 (1975).

- [5] A. Friedli: Verallgemeinertes Runge-Kutta-Verfahren zur Lösung steifer Differentialgleichungssysteme, Lect. Notes Math. 631, 35-50 (1978).
- [6] P. J. van der Houwen: Construction of integration formulas for initial value problems. Amsterdam: North Holland Publishing Company 1976.
- [7] A. Huía: The algorithm for computation of the n-th order formula for numerical solution of initial value problem of differential equations. 5th Symposium on Algorithms, 53-61, (1979).
- [8] J. D. Lawson: Generalized Runge-Kutta processes for stable systems with large Lipschitz constants. SIAM J. Numer. Anal., Vol. 4, No. 3, 372—380 (1967).
- [9] K. Nickel und P. Rieder: Ein neues Runge-Kutta ähnliches Verfahren. In: ISNM 9, Numerische Mathematik, Differentialgleichungen, Approximationstheorie, 83—96, Basel: Birkhäuser 1968.
- [10] E. J. Nyström: Über die numerische Integration von Differentialgleichungen. Acta Soc. Sci. Fennicae, Tom 50, nr. 13, 1–55 (1925).
- [11] A. Prothero and A. Robinson: On the stability and accuracy of one-step methods for solving stiff systems of ordinary differential equations. Math. Comp. 28, 145-162 (1974).
- [12] K. Strehmel: Konstruktion von adaptiven Runge-Kutta-Methoden. ZAMM, to appear 1980.
- [13] J. G. Verwer: S-stability properties for generalized Runge-Kutta methods. Numer. Math. 27, 359-370 (1977).
- [14] J. G. Verwer: Internal S-stability for generalized Runge-Kutta methods. Report NW 21, Mathematisch Centrum, Amsterdam (1975).

Súhrn

KONŠTRUKCIA EXPLICITNÝCH ZOVŠEOBECNENÝCH RUNGOVÝCH-KUTTOVÝCH VZORCOV ĽUBOVOĽNÉHO RÁDU S RACIONÁLNYMI KOEFICIENTAMI

ANTON HUŤA, KARL STREHMEL

V článku obsahujúcom algoritmus explicitných zovšeobecnených Rungových-Kuttových vzorcov ľubovoľného rádu s racionálnymi koeficientami sú vyšetrované dva problémy vyskytujúce sa pri riešení začiatočnej úlohy obyčajných diferenc álnych rovníc, totiž určenie racionálnych koeficientov a odvodenie adaptívnej Rungovej-Kuttovej metódy. Zavedením vhodných substitucií do nelineárnej sústavy podmienkových rovníc sa obdrží sústava lineárnych rovníc, ktorá má racionálne korene. Zovšeobecnenie Rungových-Kuttových vzorcov je umožnené zavedením vhodnej symboliky. Východiskom pri zostrojení adaptívnej Rungovej-Kuttovej metódy bol konzistentný s-stupňový Rungov-Kuttov vzorec. Záverom je vyšetrená S-stabilita adaptívnej Rungovej-Kuttovej metódy.

Author's address: Prof. RNDr. Anton Huía, CSc., Ulica februárového vítazstva 131, 831 02 Bratislava, ČSSR; Doz. Dr.Sc. Karl Strehmel, Sektion Mathematik, Martin-Luther-Universität, Halle-Wittenberg, Weinbergweg 17, DDR — 4020 Halle, DDR.