Journal of Computational and Applied Mathematics 20 (1987) 307–315
North-Holland

Suitability of Runge–Kutta methods

M.Z. LIU, K. DEKKER and M.N. SPIJKER
Department of Mathematics and Computer Science, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Received 8 August 1986

Abstract: In this paper we introduce the concept of suitability, which means that the nonlinear equations to be solved in a Runge–Kutta method have a unique solution. We give several results about suitability, and prove that the Butcher I, II and III, and the Lobatto IIIA and IIIB methods are suitable.

Keywords: Initial value problems, implicit Runge–Kutta methods, suitability, transposed methods.

1. Introduction

We shall deal with the initial value problem

$$dU/dt = f(t, U), \quad U(t_0) = u_0, \tag{1.1}$$

where $U_0 \in \mathbb{R}^s$ and $f(t, U)$ is a continuous function from $\mathbb{R} \times \mathbb{R}^s$ to $\mathbb{R}^s$. A numerical approximation of the solution $U(t)$ of (1.1) can be computed by the implicit Runge–Kutta method (c, b, A), which is usually denoted by the array

$$\begin{array}{c|c} c & A \\ \hline & b^{\mathrm{T}} \end{array} \qquad \begin{array}{c|ccc} c_1 & a_{11} & \cdots & a_{1m} \\ c_2 & a_{21} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ c_m & a_{m1} & \cdots & a_{mm} \\ \hline & b_1 & \cdots & b_m \end{array}$$

where c and b satisfy $c_i = \sum_{j=1}^{m} a_{ij}$, $1 \leqslant i \leqslant m$ and $\sum_{i=1}^{m} b_i = 1$. We also assume that the method is nonconfluent, i.e. the abscissae c_i, $1 \leqslant i \leqslant m$, are all different. The use of an implicit Runge–Kutta method to obtain a numerical solution of (1.1) requires the solution of a system of algebraic equations

$$y_i = u_{n-1} + h \sum_{j=1}^{m} a_{ij} f(\tau_j, y_j), \quad 1 \leqslant i \leqslant m, \tag{1.2}$$

where $\tau_i = t_{n-1} + c_i h$, $y_i \in \mathbb{R}^s$, $1 \leqslant i \leqslant m$, and $u_{n-1} \in \mathbb{R}^s$ is an approximation to the solution $U(t_{n-1})$. When system (1.2) allows a solution, and this solution has been obtained, an approximation to $U(t_n)$, $t_n = t_{n-1} + h$, can be computed from the formula

$$u_n = u_{n-1} + h \sum_{i=1}^{m} b_i f(\tau_i, y_i). \tag{1.3}$$

Recently, conditions under which system (1.2) has a unique solution have been considered by various authors, e.g. Frank, Schneid and Ueberhuber [5], Hundsdorfer and Spijker [6], Dekker and Verwer [4, Chapter 5], Di Lena and Peluso [7]. The following result, formulated by Crouzeix, Hundsdorfer and Spijker [2] will be important in our analysis.

Theorem 1.1. *If* $f(t, U)$ *satisfies*

$$\langle f(t, \xi_1) - f(t, \xi_2), \xi_1 - \xi_2 \rangle \leqslant 0, \quad \forall t, \forall \xi_1, \xi_2 \in \mathbb{R}^s, \tag{1.4}$$

and if there exists a positive definite diagonal matrix D *such that* $DA + A^{\mathrm{T}}D$ *is positive definite, then system* (1.2) *has a unique solution.*

Frank, Schneid and Ueberhuber [5] already proved that several important classes of implicit Runge–Kutta methods, viz. the Gauss–Legendre methods, the Radau IA and IIA-methods and the Lobatto IIIC-method with $m = 2$ satisfy the condition of Theorem 1.1.

In this paper we shall give results for methods which do not satisfy this condition. First, we introduce the concept of *suitability* which is equivalent to system (1.2) having a unique solution if $f(t, U)$ satisfies (1.4). Then, we prove that a lower triangular block matrix is suitable if and only if all diagonal blocks are suitable. We also prove that a matrix A is suitable if the method (c, b, A) satisfies the simplifying conditions $C(m)$ and $B(2m-2)$ (see e.g. [4, Chapter 3]) and $c_1 = 0$. Moreover, we show that these results are also applicable to the methods which are obtained by the process of transposition (cf. [8]). As a consequence system (1.2) has a unique solution for the Lobatto IIIA, IIIB and Butcher, I, II-methods (see [1]).

Finally, we prove that for the Butcher III-method [1] system (1.2) also has a unique solution.

2. Suitability

In this section we state the concept of suitability and we give the main theorem of the paper. For convenience we will denote $hf(\tau_j, y_j)$ by $f_j(y_j)$, $1 \leqslant j \leqslant m$, and we assume that $f(t, U)$ satisfies condition (1.4), so for $1 \leqslant j \leqslant m$ there holds

$$\langle f_j(\xi_1) - f_j(\xi_2), \xi_1 - \xi_2 \rangle \leqslant 0, \quad \forall \xi_1, \xi_2 \in \mathbb{R}^s. \tag{2.1}$$

Definition 2.1. A matrix A is called *suitable* if the system

$$y_i = \sum_{j=1}^{m} a_{ij} f_j(y_j), \quad 1 \leqslant i \leqslant m, \tag{2.2}$$

has a unique solution $y = (y_1, y_2, \ldots, y_m) \in (\mathbb{R}^s)^m$, whenever the functions f_j: $\mathbb{R}^s \to \mathbb{R}^s$ are continuous and satisfy (2.1), $j = 1, \ldots, m$.

Remark 2.2. Clearly suitability of A is equivalent to system (1.2) having a unique solution.

Theorem 2.3. *Suppose that after a permutation of variables a matrix* A *is of the form*

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ & \cdots & & A_k \end{pmatrix},$$

where A_i is an $r_i \times r_i$ submatrix, $1 \leqslant i \leqslant k$, and $\sum_{i=1}^{k} r_i = m$. Then A is suitable if and only if all A_i, $1 \leqslant i \leqslant k$, are suitable.

Proof. Suppose A is suitable, then obviously A_1 is suitable. Now, choose $f_j(\xi) \equiv 0$, $j = 1, \ldots, r_1$, and let $f_j(\xi)$, $r_1 + 1 \leqslant j \leqslant r_1 + r_2$, satisfy (2.1). Then the equations

$$y_i = \sum_{j=1}^{r_1} a_{ij} f_j(y_j) + \sum_{j=r_1+1}^{r_1+r_2} a_{ij} f_j(y_j) = \sum_{j=r_1+1}^{r_1+r_2} a_{ij} f_j(y_j), \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2,$$

have a unique solution, as A is suitable. Thus A_2 is suitable, and by induction it is proved that A_i is suitable, $i = 1, 2, \ldots, k$. Next suppose all A_i, $1 \leqslant i \leqslant k$, are suitable. We shall prove that system (2.2) has a unique solution for all $f_1, f_2, \ldots, f_m$ satisfying (2.1). To this end we partition (2.2) as

$$y_i = \sum_{j=1}^{r_1} a_{ij} f_j(y_j), \quad i \leqslant i \leqslant r_1, \tag{2.3}$$

$$y_i = \sum_{j=1}^{r_1} a_{ij} f_j(y_j) + \sum_{j=r_1+1}^{r_1+r_2} a_{ij} f_j(y_j), \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2, \tag{2.4}$$

$$y_i = \sum_{j=1}^{p_{k-1}} a_{ij} f_j(y_j) + \sum_{j=p_{k-1}+1}^{p_k} a_{ij} f_j(y_j), \quad p_{k-1} + 1 \leqslant i \leqslant p_k. \tag{2.5}$$

Here,

$$p_k = \sum_{i=1}^{k} r_i, \quad 1 \leqslant k \leqslant m.$$

Clearly (2.3) has a unique solution because A_1 is suitable. Now we define

$$\tilde{y}_i = y_i - e_i, \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2,$$

$$g_i(\eta) = f_i(\eta + e_i), \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2,$$

$$e_i = \sum_{j=1}^{r_1} a_{ij} f_j(y_j), \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2,$$

where the y_j, $1 \leqslant j \leqslant r_1$, are already determined by (2.3). It is easily seen that (2.4) reduces to

$$\tilde{y}_i = \sum_{j=r_1+1}^{r_1+r_2} a_{ij} g_j(\tilde{y}_j), \quad r_1 + 1 \leqslant i \leqslant r_1 + r_2,$$

which has a unique solution as A_2 is suitable and the functions $g_j(\eta)$, $r_1 + 1 \leqslant j \leqslant r_1 + r_2$, satisfy (2.1). Therefore the equations (2.4) also have a unique solution, $y_i = \tilde{y}_i + e_i$, $r_1 + 1 \leqslant i \leqslant r_1 + r_2$. Continuing this procedure, we see that A is suitable. □

Remark 2.4. Definition 2.1 and the corresponding Theorem 2.3 remain valid if the matrix $A = (a_{ij})$ is confluent (i.e. if $\sum_{k=1}^{m} a_{ik} = \sum_{k=1}^{m} a_{jk}$ for some $i \neq j$).

3. The Lobatto IIIA and Butcher I-methods

In this section we consider methods which satisfy the simplifying conditions $C(m)$ and $B(2m-2)$, i.e. (cf. e.g. [4])

$$\sum_{j=1}^{m} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k, \qquad 1 \leqslant k \leqslant m, \quad 1 \leqslant i \leqslant m, \tag{3.1}$$

$$\sum_{i=1}^{m} b_i c_i^{k-1} = \frac{1}{k}, \quad 1 \leqslant k \leqslant 2m-2. \tag{3.2}$$

We also assume that the first abscissa $c_1 = 0$. From the equations (3.1) with $i = 1$ it is then obvious that the first row of A contains zeros only. We shall prove the suitability of these matrices A in a way analogous to Dekker [3].

Theorem 3.1. *If a method (c, b, A) satisfies the simplifying conditions $C(m)$ and $B(2m-2)$, $c_1 = 0$, and the other abscissae satisfy $0 < c_i \leqslant 1$, $2 \leqslant i \leqslant m$, then A is suitable.*

Proof. The method (c, b, A) is given by

$$\begin{array}{c|c} c & A \\ \hline & b^{\mathrm{T}} \end{array} = \begin{array}{c|ccc} 0 & 0 & 0 & \cdots & 0 \\ c_2 & a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & \vdots & & \vdots \\ c_m & a_{m1} & a_{m2} & \cdots & a_{mm} \\ \hline & b_1 & b_2 & \cdots & b_m \end{array}$$

We define the $(m-1)$-stage submethod generated by

$$\begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^{\mathrm{T}} \end{array} = \begin{array}{c|ccc} c_2 & a_{22} & a_{23} & \cdots & a_{2m} \\ c_3 & a_{32} & a_{33} & \cdots & a_{3m} \\ \vdots & \vdots & \vdots & & \vdots \\ c_m & a_{m2} & a_{m3} & \cdots & a_{mm} \\ \hline & b_2 & b_3 & \cdots & b_m \end{array}$$

Let

$$\tilde{V} = \begin{pmatrix} c_2 & c_2^2 & \cdots & c_2^{m-1} \\ c_3 & c_3^2 & \cdots & c_3^{m-1} \\ \vdots & \vdots & & \vdots \\ c_m & c_m^2 & \cdots & c_m^{m-1} \end{pmatrix}, \qquad \tilde{H} = \begin{pmatrix} \frac{1}{2} & \frac{1}{3} & \cdots & \frac{1}{m} \\ \frac{1}{3} & \frac{1}{4} & \cdots & \frac{1}{m+1} \\ \vdots & \vdots & & \vdots \\ \frac{1}{m} & \frac{1}{m+1} & \cdots & \frac{1}{2m-2} \end{pmatrix},$$

$\tilde{B} = \operatorname{diag}(b_2, b_3, \ldots, b_m)$, $\tilde{C} = \operatorname{diag}(c_2, c_3, \ldots, c_m)$, $\tilde{S} = \operatorname{diag}(\frac{1}{2}, \frac{1}{3}, \ldots, 1/m)$ and let $\tilde{e}_j$ stand for the jth unit vector $(1 \leqslant j \leqslant m-1)$. The simplifying conditions $C(m)$ and $B(2m-2)$ then become

$$\tilde{A}\tilde{V} = \tilde{C}\tilde{V}\tilde{S}, \tag{3.3}$$

$$\tilde{V}^{\mathrm{T}}\tilde{B}\tilde{C}^{-1}\tilde{V} = \tilde{H}. \tag{3.4}$$

Let $\tilde{D} = \tilde{B}\tilde{C}^{-2}$, Then $\tilde{D}$ is a positive diagonal matrix. From (3.3) and (3.4) we have for $1 \leqslant i, j \leqslant m-1$

$$\tilde{e}_i^{\mathrm{T}}\tilde{V}^{\mathrm{T}}\tilde{D}\tilde{A}\tilde{V}\tilde{e}_j = \frac{1}{j+1}\frac{1}{i+j}, \qquad \tilde{e}_i^{\mathrm{T}}\tilde{V}^{\mathrm{T}}\tilde{A}^{\mathrm{T}}\tilde{D}\tilde{V}\tilde{e}_j = \frac{1}{i+1}\frac{1}{i+j},$$

$$\tilde{e}_i^{\mathrm{T}}\tilde{V}^{\mathrm{T}}\tilde{A}^{\mathrm{T}}\tilde{B}\tilde{C}^{-3}\tilde{A}\tilde{V}\tilde{e}_j = \frac{1}{i+1}\frac{1}{j+1}\frac{1}{i+j}.$$

Hence

$$\tilde{V}^{\mathrm{T}}(\tilde{A}^{\mathrm{T}}\tilde{D} + \tilde{D}\tilde{A} - 2\tilde{A}^{\mathrm{T}}\tilde{B}\tilde{C}^{-3}\tilde{A})\tilde{V} = \tilde{S}\tilde{e}\tilde{e}^{\mathrm{T}}\tilde{S} = \tilde{V}^{\mathrm{T}}\tilde{b}\tilde{b}^{\mathrm{T}}\tilde{V},$$

where $\tilde{e} = (1, 1, \ldots, 1)^{\mathrm{T}}$. Obviously, $\tilde{b}\tilde{b}^{\mathrm{T}}$ is symmetric and positive semi-definite and $\tilde{B}\tilde{C}^{-3}$ is positive definite. Therefore,

$$\tilde{D}\tilde{A} + \tilde{A}^{\mathrm{T}}\tilde{D} = \tilde{b}\tilde{b}^{\mathrm{T}} + 2\tilde{A}^{\mathrm{T}}\tilde{B}\tilde{C}^{-3}\tilde{A} \tag{3.5}$$

is positive definite. According to Theorem 1.1 $\tilde{A}$ is suitable. Let $A_1 = (0)$, $A_2 = \tilde{A}$, then it follows from Theorem 2.3 that A is suitable. □

Corollary 3.2. *For the Lobatto IIIA-methods system* (1.2) *has a unique solution.*

Proof. The Lobatto IIIA-methods satisfy the simplifying conditions $C(m)$ and $B(2m-2)$ and the abscissae are those of the Lobatto quadrature formulas, so $c_1 = 0$ and $0 < c_i \leqslant 1$, $2 \leqslant i \leqslant m$ (see e.g. [4, Chapter 3]). □

Corollary 3.3. *For the Butcher I-methods system* (1.2) *has a unique solution.*

Proof. The Butcher I-methods satisfy the conditions of Theorem 3.1 (see [1]). □

4. The Lobatto IIIB and Butcher II-methods

Scherer and Türke [8] have shown that some Runge–Kutta methods are related to each other. They established interesting interrelations between various methods by means of two mappings, viz. reflection and transposition. The latter one is a helpful tool in our analysis, so we shall give its definition. Let $B = \mathrm{diag}(b_1, b_2, \ldots, b_m)$, $C = \mathrm{diag}(c_1, c_2, \ldots, c_m)$, $e = (1, 1, \ldots, 1)^{\mathrm{T}}$.

Definition 4.1. *Let* (c, b, A) *be a Runge–Kutta method and* $c_\tau = e - c$, $A_\tau = B^{-1}A^{\mathrm{T}}B$, $b_\tau = b$, *then the Runge–Kutta method* (c_τ, b_τ, A_τ) *is called the* transposed *method.*

It is seen that the abscissae of the transposed method are ordered in reverse by this definition. We will assume that they are reordered by a suitable permutation, so that

$$c_\tau = P(e - c), \qquad A_\tau = PB^{-1}A^{\mathrm{T}}BP, \qquad b_\tau = Pb,$$

where P is the permutation matrix, and we define

$$B_\tau = PBP, \qquad C_\tau = P(I - C)P, \qquad I = \mathrm{diag}(1, 1, \ldots, 1).$$

We shall now formulate results which are relevant for the suitability of a method. In the sequel we assume that D is a positive diagonal matrix.

Theorem 4.2. *Suppose that B is a positive diagonal matrix. Then the following implications hold for $D_\tau = PD^{-1}B^2P$:*

(a) *$D_\tau A_\tau + A_\tau^{\mathrm{T}} D_\tau$ is positive (semi) definite iff $DA + A^{\mathrm{T}}D$ is positive (semi) definite.*

(b) *If A is nonsingular, then $D_\tau A_\tau^{-1} + A_\tau^{-\mathrm{T}} D_\tau$ is positive (semi) definite iff $DA^{-1} + A^{-\mathrm{T}}D$ is positive (semi) definite.*

Proof. The proof follows directly from the definitions of D_τ and B_τ, and the fact that B and D are positive:

$$\begin{aligned} D_\tau A_\tau + A_\tau^{\mathrm{T}} D_\tau &= PD^{-1}B^2PPB^{-1}A^{\mathrm{T}}BP + PBAB^{-1}PPB^2D^{-1}P \\ &= PD^{-1}BA^{\mathrm{T}}BP + PBABD^{-1}P = PD^{-1}B(A^{\mathrm{T}}D + DA)BD^{-1}P, \end{aligned}$$

and

$$D_\tau A_\tau^{-1} + A_\tau^{-1} D_\tau = PD^{-1}B(DA^{-1} + A^{-\mathrm{T}}D)BD^{-1}P. \qquad \square$$

From [8], we have the following:

Lemma 4.3. *The Butcher I and Lobatto IIIA-methods are changed into the Butcher II and Lobatto IIIB-methods by transposition respectively.*

Let us denote the Butcher I, II and Lobatto IIIA, IIIB-methods by $(c_{\mathrm{I}}, b_{\mathrm{I}}, A_{\mathrm{I}})$, $(c_{\mathrm{II}}, b_{\mathrm{II}}, A_{\mathrm{II}})$, $(c_{\mathrm{IIIA}}, b_{\mathrm{IIIA}}, A_{\mathrm{IIIA}})$, $(c_{\mathrm{IIIB}}, b_{\mathrm{IIIB}}, A_{\mathrm{IIIB}})$ respectively. Then the following relations hold

$$(c_{\mathrm{II}}, b_{\mathrm{II}}, A_{\mathrm{II}}) = \big(P(e - c_{\mathrm{I}}), Pb_{\mathrm{I}}, PB_{\mathrm{I}}^{-1}A_{\mathrm{I}}^{\mathrm{T}}B_{\mathrm{I}}P\big),$$

$$(c_{\mathrm{IIIB}}, b_{\mathrm{IIIB}}, A_{\mathrm{IIIB}}) = \big(P(e - c_{\mathrm{IIIA}}), Pb_{\mathrm{IIIA}}, PB_{\mathrm{IIIA}}^{-1}A_{\mathrm{IIIA}}^{\mathrm{T}}B_{\mathrm{IIIA}}P\big),$$

where

$$P = \begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}.$$

From these relations it is easily seen that the Butcher II (and similarly the Lobatto IIIB) methods have the form

$$\begin{array}{c|c} c_{\mathrm{II}} & A_{\mathrm{II}} \\ \hline & B_{\mathrm{II}}^{\mathrm{T}} \end{array} = \begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1m-1} & 0 \\ c_2 & a_{21} & a_{22} & \cdots & a_{2m-1} & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ c_m & a_{m1} & a_{m2} & \cdots & a_{mm-1} & 0 \\ \hline & b_1 & b_2 & \cdots & b_{m-1} & b_m \end{array} \tag{4.1}$$

with $c_m = 1$.

Theorem 4.4. *For the Butcher II-methods system* (1.2) *has a unique solution.*

Proof. Let $\tilde{D}_{\mathrm{I}} = \tilde{B}_{\mathrm{I}}\tilde{C}_{\mathrm{I}}^{-2}$ as in the proof of Theorem 3.1. Then $\tilde{D}_{\mathrm{I}}\tilde{A}_{\mathrm{I}} + \tilde{A}_{\mathrm{I}}^{\mathrm{T}}\tilde{D}_{\mathrm{I}}$ is positive definite by (3.5). Let $\tilde{A}_{\mathrm{II}}$, $\tilde{B}_{\mathrm{II}}$, $\tilde{C}_{\mathrm{II}}$, $\tilde{I}$ denote the matrices which are obtained by eliminating the last row and column from A_{II}, B_{II}, C_{II} and I, respectively. Using Lemma 4.3 we obtain

$$\tilde{A}_{\mathrm{II}} = \tilde{P}\tilde{B}_{\mathrm{I}}^{-1}\tilde{A}_{\mathrm{I}}^{\mathrm{T}}\tilde{B}_{\mathrm{I}}\tilde{P}, \qquad \tilde{B}_{\mathrm{II}} = \tilde{P}\tilde{B}_{\mathrm{I}}\tilde{P}, \qquad \tilde{C}_{\mathrm{II}} = \tilde{P}\big(\tilde{I} - \tilde{C}_{\mathrm{I}}\big)\tilde{P},$$

where $\tilde{P}$ is the matrix which originates from P by deleting the first row and last column. Now let $\tilde{D}_{\mathrm{II}} = \tilde{P}\tilde{B}_{\mathrm{I}}^{2}\tilde{D}_{\mathrm{I}}^{-1}\tilde{P} = \tilde{B}_{\mathrm{II}}(\tilde{I} - \tilde{C}_{\mathrm{II}})^{2}$. According to Theorem 4.2 $\tilde{D}_{\mathrm{II}}\tilde{A}_{\mathrm{II}} + \tilde{A}_{\mathrm{II}}^{\mathrm{T}}\tilde{D}_{\mathrm{II}}$ is positive definite, so $\tilde{A}_{\mathrm{II}}$ is suitable by Theorem 1.1. It follows from Theorem 2.3 that A_{II} is suitable. □

Theorem 4.5. *For the Lobatto IIIB-methods system* (1.2) *has a unique solution.*

Proof. The structure of the Lobatto IIIB-methods is the same as in (4.1). Let $\tilde{A}_{\mathrm{IIIB}}$, $\tilde{B}_{\mathrm{IIIB}}$, $\tilde{C}_{\mathrm{IIIB}}$ and $\tilde{I}$ be defined in a similar way as in the previous theorem, and let $\tilde{D}_{\mathrm{IIIB}} = \tilde{B}_{\mathrm{IIIB}}(\tilde{I} - \tilde{C}_{\mathrm{IIIB}})^{2}$, $\tilde{D}_{\mathrm{IIIA}} = \tilde{B}_{\mathrm{IIIA}}\tilde{C}_{\mathrm{IIIA}}^{-2}$. From the proof of Theorem 3.1 we know that $\tilde{D}_{\mathrm{IIIA}}\tilde{A}_{\mathrm{IIIA}} + \tilde{A}_{\mathrm{IIIA}}^{\mathrm{T}}\tilde{D}_{\mathrm{IIIA}}$ is positive definite. Hence, according to Lemma 4.3 and Theorem 4.2, $\tilde{D}_{\mathrm{IIIB}}\tilde{A}_{\mathrm{IIIB}} + \tilde{A}_{\mathrm{IIIB}}^{\mathrm{T}}\tilde{D}_{\mathrm{IIIB}}$ is positive definite, and $\tilde{A}_{\mathrm{IIIB}}$ is suitable by Theorem 1.1. It is seen from Theorem 2.3 that also A_{IIIB} is suitable. □

Remark 4.6. The positive definiteness of $\tilde{D}_{\mathrm{IIIB}}\tilde{A}_{\mathrm{IIIB}} + \tilde{A}_{\mathrm{IIIB}}^{\mathrm{T}}\tilde{D}_{\mathrm{IIIB}}$ could also be concluded from [3, Example 5.6]. □

5. The Butcher III-methods

The Butcher III-methods, introduced by Butcher [1], satisfy the simplifying conditions $C(m-1)$ and $B(2m-2)$, i.e.

$$\sum_{j=1}^{m} a_{ij}c_j^{k-1} = \frac{1}{k}c_i^{k}, \qquad 1 \leqslant k \leqslant m-1, \quad 1 \leqslant i \leqslant m, \tag{5.1}$$

$$\sum_{j=1}^{m} b_j c_j^{k-1} = \frac{1}{k}, \qquad 1 \leqslant k \leqslant 2m-2, \tag{5.2}$$

and they are characterized by the array

$$\begin{array}{c|c} c & A \\ \hline & b^{\mathrm{T}} \end{array} = \begin{array}{c|cccc} 0 & 0 & 0 & \cdots & 0 & 0 \\ c_2 & a_{21} & a_{22} & \cdots & a_{2m-1} & 0 \\ \vdots & \vdots & & & \vdots & \vdots \\ c_m & a_{m1} & a_{m2} & \cdots & a_{mm-1} & 0 \\ \hline & b_1 & b_2 & \cdots & b_{m-1} & b_m \end{array},$$

with $c_m = 1$.

Theorem 5.1. *For the Butcher III-methods system* (1.2) *has a unique solution.*

Proof. We define the $(m-2)$ stage submethods by

$$\begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^{\mathrm{T}} \end{array} = \begin{array}{c|ccc} c_2 & a_{22} & a_{23} & \cdots & a_{2m-1} \\ c_3 & a_{32} & a_{33} & \cdots & a_{3m-1} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{m-1} & a_{m-12} & a_{m-13} & \cdots & a_{m-1m-1} \\ \hline & b_2 & b_3 & \cdots & b_{m-1} \end{array}$$

Let $\tilde{B} = \mathrm{diag}(b_2, b_3, \ldots, b_{m-1})$, $\tilde{C} = \mathrm{diag}(c_2, c_3, \ldots, c_{m-1})$, $\tilde{e}_j$ stand for the jth unit vector, $1 \leqslant j \leqslant m-2$, and

$$\tilde{V} = \begin{pmatrix} c_2 & c_2^2 & \cdots & c_2^{m-2} \\ c_3 & c_3^2 & \cdots & c_3^{m-2} \\ \vdots & \vdots & & \vdots \\ c_{m-1} & c_{m-1}^2 & \cdots & c_{m-1}^{m-2} \end{pmatrix}.$$

Then we have from (5.1) and (5.2),

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} \tilde{B} \tilde{A} \tilde{V} \tilde{e}_j = \frac{1}{j+1}\left(\frac{1}{i+j+2} - b_m\right), \qquad 1 \leqslant i, \quad j \leqslant m-2,$$

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} \tilde{B} \tilde{C}^{-1} \tilde{A} \tilde{V} \tilde{e}_j = \frac{1}{j+1}\left(\frac{1}{i+j+1} - b_m\right), \quad 1 \leqslant i, \quad j \leqslant m-2,$$

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} \tilde{B} \tilde{C}^{-2} \tilde{A} \tilde{V} \tilde{e}_j = \frac{1}{j+1}\left(\frac{1}{i+j} - b_m\right), \qquad 1 \leqslant i, \quad j \leqslant m-2.$$

Let $\tilde{D} = \tilde{B}(\tilde{C}^{-1} - \tilde{I})^2$, then again $\tilde{D}$ is a positive diagonal matrix, because the abscissae c_i, $2 \leqslant i \leqslant m-1$, all lie in the interval (0, 1), and we have

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} (\tilde{D}\tilde{A} + \tilde{A}^{\mathrm{T}} \tilde{D}) \tilde{V} \tilde{e}_j = \frac{2}{(i+1)(j+1)(i+j)(i+j+1)}, \qquad 1 \leqslant i, \quad j \leqslant m-2.$$

A straightforward calculation also shows that

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} \tilde{A}^{\mathrm{T}} \tilde{B} \tilde{C}^{-2} \tilde{A} \tilde{V} \tilde{e}_j = \frac{1}{i+1}\frac{1}{j+1}\left(\frac{1}{i+j+1} - b_m\right), \qquad 1 \leqslant i, \quad j \leqslant m-2,$$

and

$$\tilde{e}_i^{\mathrm{T}} \tilde{V}^{\mathrm{T}} \tilde{A}^{\mathrm{T}} \tilde{B} \tilde{C}^{-3} \tilde{A} \tilde{V} \tilde{e}_j = \frac{1}{i+1}\frac{1}{j+1}\left(\frac{1}{i+j} - b_m\right), \qquad 1 \leqslant i, \quad j \leqslant m-2.$$

A combination of these equalities yields

$$\tilde{D}\tilde{A} + \tilde{A}^{\mathrm{T}} \tilde{D} = 2\tilde{A}^{\mathrm{T}} \tilde{B}(\tilde{C}^{-3} - \tilde{C}^{-2})\tilde{A}.$$

Consequently, $\tilde{D}\tilde{A} + \tilde{A}^{\mathrm{T}} \tilde{D}$ is positive definite, and there follows from Theorem 1.1 that $\tilde{A}$ is suitable. Let $A_1 = (0)$, $A_2 = \tilde{A}$ and $A_3 = (0)$, then it is obvious from Theorem 2.3 that A is suitable. □

Remark 5.2. It is also possible to choose $\tilde{D} = \tilde{B}(\tilde{C}^{-1} - \tilde{I})$ in the proof given above. It turns out that for this matrix $\tilde{D}$, $\tilde{D}\tilde{A} + \tilde{A}^{\mathrm{T}}\tilde{D}$ is also positive definite.

Remark 5.3. It is interesting to mention that for the methods considered, the Butcher I, II, III and the Lobatto IIIA, IIIB-methods, system (1.2) has a unique solution, even though they are not algebraically stable, and the Butcher I, II, III-methods are not A-stable.

References

[1] J.C. Butcher, Integration processes based on radau quadrature formulas, *Math. Comp.* **18** (1964) 223–244.
[2] M. Crouzeix, W.H. Hundsdorfer and M.N. Spijker, On the existence of solutions to the algebraic equations in implicit Runge–Kutta methods, *BIT* **23** (1983) 84–91.
[3] K. Dekker, Error bounds for the solution to the algebraic equations in Runge–Kutta methods, *BIT* **24** (1984) 347–356.
[4] K. Dekker and J.G. Verwer, *Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations* (North-Holland, Amsterdam, 1984).
[5] R. Frank, J. Schneid and C.W. Ueberhuber, Stability properties of implicit Runge–Kutta methods, *SIAM J. Numer. Anal.* **22** (1985) 497–514.
[6] W.H. Hundsdorfer, M.N. Spijker, On the algebraic equations in implicit Runge–Kutta methods, Report NM-R 8413, Centre for Mathematics and Computer Science, Amsterdam, 1984, to appear in *SIAM J. Numer. Anal.*
[7] G. di Lena and R.I. Peluso, On conditions for the existence and uniqueness of solutions to the algebraic equations in Runge–Kutta methods, *BIT* **25** (1985) 223–232.
[8] R. Scherer and H. Türke, Reflected and transposed Runge–Kutta methods, *BIT* **23** (1983) 262–266.