BIT 31 (1991), 482–528.

CONTRACTIVITY OF RUNGE-KUTTA METHODS

J. F. B. M. KRAAIJEVANGER*

Department of Mathematics and Computer Science, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands

Abstract

In this paper we present necessary and sufficient conditions for Runge-Kutta methods to be contractive. We consider not only unconditional contractivity for arbitrary dissipative initial value problems, but also conditional contractivity for initial value problems where the right hand side function satisfies a circle condition. Our results are relevant for arbitrary norms, in particular for the maximum norm.

For contractive methods, we also focus on the question whether there exists a unique solution to the algebraic equations in each step. Further we show that contractive methods have a limited order of accuracy. Various optimal methods are presented, mainly of explicit type. We provide a numerical illustration to our theoretical results by applying the method of lines to a parabolic and a hyperbolic partial differential equation.

Subject Classifications: AMS (MOS): 65L05, 65L20, 65M10.

1. Introduction.

We consider initial value problems for systems of $s \geq 1$ ordinary differential equations,

(1.1a) $$\frac{\mathrm{d}}{\mathrm{d}t} U(t) = f(t, U(t)) \qquad (t \geq t_0),$$

(1.1b) $$U(t_0) = u_0.$$

We assume here that $t_0 \in \mathbb{R}$, $u_0 \in \mathbb{R}^s$ and

(1.2a) f is a continuous function from $\mathbb{R} \times \mathbb{R}^s$ into $\mathbb{R}^s$;

(1.2b) for each $t_0 \in \mathbb{R}$ and $u_0 \in \mathbb{R}^s$ problem (1.1) has a unique solution $U: [t_0, \infty) \to \mathbb{R}^s$;

(1.2c) $\|\cdot\|$ is a norm on $\mathbb{R}^s$ such that for any $t_0 \in \mathbb{R}$ and any two solutions U, $\tilde{U}$ to (1.1a) we have $\|\tilde{U}(t) - U(t)\| \leq \|\tilde{U}(t_0) - U(t_0)\|$ (for all $t \geq t_0$).

* Research supported by the Netherlands Organization for Scientific Research (N.W.O.) and the Royal Netherlands Academy of Arts and Sciences (K.N.A.W.)

Received July 1990.

The class consisting of all pairs $(f, \|\cdot\|)$ satisfying (1.2) (for some appropriate s) is denoted by $\mathscr{F}$. If $(f, \|\cdot\|) \in \mathscr{F}$ we say that f is *dissipative* with respect to $\|\cdot\|$. In this case the initial value problem (1.1) is said to be dissipative as well.

In the present paper we study *Runge-Kutta methods* for the numerical solution of (1.1). In these methods we select a step size $h > 0$, and generate approximations $u_n \simeq U(t_n)$, $n = 1, 2, 3, \ldots$, in a step by step fashion, starting with the initial value $u_0 = U(t_0)$. Here the grid points t_n are defined by $t_n = t_{n-1} + h$. More precisely, if u_{n-1} has already been computed, u_n is defined to be

$$u_n = u_{n-1} + h \textstyle\sum_{j=1}^{m} b_j f(t_{n-1} + c_j h, y_j), \tag{1.3a}$$

where the vectors $y_1, y_2, \ldots, y_m \in \mathbb{R}^s$ are a solution to the system of equations

$$y_i = u_{n-1} + h \textstyle\sum_{j=1}^{m} a_{ij} f(t_{n-1} + c_j h, y_j) \qquad (1 \le i \le m). \tag{1.3b}$$

The coefficients a_{ij}, b_j, c_j $(i, j = 1, 2, \ldots, m)$ are real numbers specifying the Runge-Kutta method. The number m is called the *number of stages*. We always assume that

$$c_i = a_{i1} + a_{i2} + \ldots + a_{im} \qquad (1 \le i \le m).$$

In view of this assumption we can represent a Runge-Kutta method by its *coefficient scheme* (A, b), where A denotes the $m \times m$ matrix $A = (a_{ij})$ and b the m-dimensional column vector $b = (b_1, b_2, \ldots, b_m)^{\mathrm{T}}$. Usually one displays the coefficient scheme (A, b) and the values c_i in the tableau

$$\begin{array}{c|ccc} c_1 & a_{11} & \ldots & a_{1m} \\ \vdots & \vdots & & \vdots \\ c_m & a_{m1} & \ldots & a_{mm} \\ \hline & b_1 & \ldots & b_m \end{array}.$$

Suppose that, instead of u_{n-1}, we are dealing with a perturbed approximation $\tilde{u}_{n-1}$. Then the Runge-Kutta step will generate $\tilde{u}_n$ satisfying

$$\tilde{u}_n = \tilde{u}_{n-1} + h \textstyle\sum_{j=1}^{m} b_j f(t_{n-1} + c_j h, \tilde{y}_j), \tag{1.4a}$$

$$\tilde{y}_i = \tilde{u}_{n-1} + h \textstyle\sum_{j=1}^{m} a_{ij} f(t_{n-1} + c_j h, \tilde{y}_j) \qquad (1 \le i \le m). \tag{1.4b}$$

If f is dissipative with respect to a given norm $\|\cdot\|$, it is natural to require *contractivity* of the numerical method, i.e.

$$\|\tilde{u}_n - u_n\| \le \|\tilde{u}_{n-1} - u_{n-1}\|. \tag{1.5}$$

This very favourable stability property of the method has been studied by many authors. For the case where the norm $\|\cdot\|$ is generated by an inner product $\langle .,. \rangle$, i.e. $\|x\| = \langle x, x \rangle^{1/2}$ for all $x \in \mathbb{R}^s$, there exists a satisfactory theory providing necessary and sufficient conditions for contractivity. This theory comprises not only *unconditional contractivity*, i.e. contractivity for all step sizes $h > 0$, but also *conditional contractivity*, i.e. contractivity under a step size restriction $h \le H$. In the case of unconditional contractivity, arbitrary functions f are considered, which are dissipative with respect to an inner product norm, cf. [3], [6], [18] (see also [5], [9]). In the case of conditional contractivity, it is assumed that f satisfies a stronger condition than dissipativity, namely a circle condition (5.7) with a given radius $\rho > 0$, cf. [7], [8], [23] (see also [9]).

For norms not generated by an inner product, there exists no general theory for contractivity. Only for linear autonomous differential equations, necessary and sufficient conditions for unconditional and conditional contractivity are known [34] (see also Remarks 5.7 and 6.4). For more details and further results in this situation we refer to [15], [21], [22], [33], [35], [36], [37].

In this paper we shall present necessary and sufficient conditions on a Runge-Kutta method to be contractive in arbitrary norms. We will consider unconditional contractivity on the class $\mathscr{F}$ as well as conditional contractivity on subclasses of $\mathscr{F}$ (f must satisfy a circle condition (5.7)). As in [34], special attention is given to the important maximum norm. We mention that some of our results were already stated, without proof, in [24]. For related results on contractivity for linear multistep and one-leg methods we direct the reader to [26], [27], [29], [30], [31], [34], [39].

In the following we give a brief outline of the paper. Section 2 is of preliminary nature. Here we introduce the concept of absolute monotonicity for the well-known stability function φ and K-function K of a Runge-Kutta method, and also for the so-called matrix-valued K-function, denoted by $\mathbb{K}$. For the investigation of absolute monotonicity of the functions K and $\mathbb{K}$, it is convenient to consider certain algebraic conditions on the coefficient scheme (A, b). These algebraic conditions are referred to as absolute monotonicity of the scheme (A, b).

In Section 3 it is proved that absolute monotonicity of K, $\mathbb{K}$ and (A, b) are equivalent.

In Section 4, the radius of absolute monotonicity $R(A, b) \in [0, \infty]$ is defined and studied for arbitrary Runge-Kutta schemes (A, b). One of the results, cf. Theorem 4.2, is that only schemes (A, b) with $A \geq 0$ and $b > 0$ can have a non-vanishing radius $R(A, b)$.

In Section 5 we study conditional contractivity of Runge-Kutta schemes. It is assumed that the function f satisfies a circle condition (5.7), where $\|\cdot\|$ is an arbitrary norm and $\rho > 0$ is fixed. The main result, cf. Theorem 5.4, is that the maximum step size H for which we have contractivity is given by $H = R(A, b)\rho^{-1}$.

In Section 6 we study unconditional contractivity on the class $\mathscr{F}$, which is proved to be equivalent to $R(A, b) = \infty$, cf. Theorem 6.1.

In Section 7 it is proved that the system of Runge-Kutta equations (1.3b) has a unique solution whenever the conditions that guarantee contractivity are fulfilled. Moreover it is shown that there is stability with respect to perturbations of these equations.

In Section 8 we consider the impact of the conditions $R(A, b) > 0$ and $R(A, b) = \infty$ on the order of accuracy of the method. A distinction is made between the classical order of consistency p and the stage order $\tilde{p}$. We mention the negative result

$\tilde{p} \leq p \leq 1$ for unconditionally contractive methods, i.e. methods with $R(A, b) = \infty$. This order barrier was already derived in [34] by considering only linear autonomous problems. Further we have for conditionally contractive methods, i.e. methods with $R(A, b) > 0$, the order barriers $p \leq 6$ and $\tilde{p} \leq 2$. Explicit methods, for which always $\tilde{p} \leq 1$, suffer even from an order barrier $p \leq 4$ in case $R(A, b) > 0$. This last negative result was also found in [7], where only inner product norms were considered. We also present convergence estimates for methods with $R(A, b) > 0$, cf. Theorem 8.1.

In Section 9 we study, for given integers m and p, the maximum of $R(A, b)$ on the class of explicit Runge-Kutta methods with m stages and classical order at least p. Several optimal explicit methods are presented.

In Section 10 a numerical illustration is given. Heun's third order scheme (with $R(A, b) = 0$) is compared with the optimal third order 3-stage scheme derived in Section 9 (with $R(A, b) = 1$). Both methods are applied to three initial value problems which are dissipative with respect to the maximum norm (l_∞ norm) or (weighted) sum norm (l_1 norm). The latter two problems come from a space discretization of a parabolic and a hyperbolic partial differential equation with coefficients varying in space and time.

2. Preliminaries.

2.1. *Definition of the functions φ, K and $\mathbb{K}$.*

Let a Runge-Kutta method be given with $m \geq 1$ stages and coefficient scheme (A, b). If we apply this method to the linear scalar autonomous test problem

$$\text{(2.1)} \qquad \frac{\mathrm{d}}{\mathrm{d}t} U(t) = \lambda U(t) \quad (t \geq t_0), \qquad U(t_0) = u_0,$$

it is well known (cf. e.g. [5], [9], [38]) that (1.3) reduces to the simple recurrence relation $u_n = \varphi(h\lambda)u_{n-1}$ $(n \geq 1)$, where φ is the so-called *stability function* of the method. The stability function φ is a rational function in one complex variable z with numerator $\det(I - z(A - eb^{\mathrm{T}}))$ and denominator $\det(I - zA)$, where $e = (1, 1, \ldots, 1)^{\mathrm{T}} \in \mathbb{R}^m$ and I stands for the $m \times m$ identity matrix. Although we shall deal with differential equations in real vector spaces only, implying that the argument $z = h\lambda$ is real, it is convenient to define $\varphi(z)$ for complex values z as well. We note that it is possible that φ has removable singularities, namely if the numerator and denominator of φ have a common zero (in $\mathbb{C}$). Finally, it is well known that $\varphi(z)$ can be written as

$$\varphi(z) = 1 + zb^{\mathrm{T}}(I - zA)^{-1}e \qquad (\text{if } I - zA \text{ is nonsingular}).$$

If we apply the Runge-Kutta method to the more general, linear scalar *non-autonomous* test problem

$$(2.2)\qquad \frac{d}{dt}U(t) = \lambda(t)U(t) \qquad (t \geq t_0), \quad U(t_0) = u_0,$$

it is well known (cf. e.g. [3], [5], [9]) that (1.3) reduces to the recurrence relation $u_n = K(z_1, z_2, \ldots, z_m)u_{n-1}$ $(n \geq 1)$, where K is the so-called (*scalar*) *K-function* of the method and $z_i = h\lambda(t_{n-1} + c_i h)$, $i = 1, 2, \ldots, m$. The K-function is a rational function in the complex variables $z_1, z_2, \ldots, z_m$ with numerator $\det(I - (A - eb^T)Z)$ and denominator $\det(I - AZ)$, where Z is the diagonal matrix $Z = \mathrm{diag}(z_1, z_2, \ldots, z_m)$. We shall use both notations $K(z_1, z_2, \ldots, z_m)$ and $K(Z)$. It is well known that

$$K(Z) = 1 + b^T Z(I - AZ)^{-1}e \qquad (\text{if } I - AZ \text{ is nonsingular}).$$

We note that the K-function reduces to the stability function φ if we take $z_1 = z_2 = \ldots = z_m$. Further we emphasize that all variables $z_1, z_2, \ldots, z_m$ of the K-function are considered to be independent, although in the application of the method to (2.2) we always have $z_i = z_j$ whenever $c_i = c_j$.

Let s be a positive integer and consider the linear non-autonomous test problem

$$(2.3)\qquad \frac{d}{dt}U(t) = L(t)U(t) \quad (t \geq t_0), \qquad U(t_0) = u_0,$$

where $L(t)$ is a real or complex $s \times s$ matrix depending on t. In this case (1.3) reduces to the recurrence relation $u_n = \mathbb{K}(Z_1, Z_2, \ldots, Z_m)u_{n-1}$ $(n \geq 1)$, where $\mathbb{K}$ is a matrix-valued function and $Z_i = hL(t_{n-1} + c_i h)$, $i = 1, 2, \ldots, m$. We define $\mathbb{Z}$ to be the block diagonal matrix $\mathbb{Z} = \mathrm{diag}(Z_1, Z_2, \ldots, Z_m)$. Both notations $\mathbb{K}(Z_1, Z_2, \ldots, Z_m)$ and $\mathbb{K}(\mathbb{Z})$ will be used. One easily verifies that

$$\mathbb{K}(\mathbb{Z}) = \mathbb{1} + \mathbb{b}^T \mathbb{Z}(\mathbb{I} - \mathbb{A}\mathbb{Z})^{-1}\mathbb{e} \qquad (\text{if } \mathbb{I} - \mathbb{A}\mathbb{Z} \text{ is nonsingular}),$$

where $\mathbb{A} = A \otimes I_s$, $\mathbb{b} = b \otimes I_s$, $\mathbb{e} = e \otimes I_s$, $\mathbb{I} = I \otimes I_s$ and $\mathbb{1} = (1) \otimes I_s = I_s$. Here I_s stands for the $s \times s$ identity matrix and $\otimes$ for the Kronecker product (cf. e.g. [25], [9]). We shall refer to $\mathbb{K}$ as the *matrix-valued K-function* of the method. This function is defined for all block diagonal matrices $\mathbb{Z} = \mathrm{diag}(Z_1, Z_2, \ldots, Z_m)$ for which $\mathbb{I} - \mathbb{A}\mathbb{Z}$ is nonsingular and where the blocks Z_i are square matrices of the same (but arbitrary) order $s \geq 1$. If $s = 1$, the blocks become scalars, $Z_i = (z_i)$ $(i = 1, 2, \ldots, m)$, and the block diagonal matrix $\mathbb{Z}$ becomes a diagonal matrix $Z = \mathrm{diag}(z_1, z_2, \ldots, z_m)$. In this case the matrix-valued K-function $\mathbb{K}(\mathbb{Z})$ reduces to the scalar K-function $K(Z)$.

2.2. *Absolute monotonicity of* φ, K, $\mathbb{K}$ *and* (A, b).

In this subsection the functions φ, K and $\mathbb{K}$ are as in Subsection 2.1. For these three functions, and for the coefficient scheme (A, b) itself, we shall define the concept of absolute monotonicity. We begin with defining absolute monotonicity for rational functions $\psi = P/Q$, where P and Q are polynomials in the complex variable z, both with real coefficients. Note that the stability function φ is of this type.

DEFINITION 2.1. ψ is said to be *absolutely monotonic* at a given point $\xi \in \mathbb{R}$ if $Q(\xi) \neq 0$ and $(\mathrm{d}^k\psi/\mathrm{d}z^k)(\xi) \geq 0$, $k = 0, 1, 2, \ldots$. Further, ψ is said to be absolutely monotonic on a given set $\Omega \subset \mathbb{R}$ if ψ is absolutely monotonic at each $\xi \in \Omega$.

For the (scalar) K-function K, depending on the complex variables $z_1, z_2, \ldots, z_m$, the definition of absolute monotonicity is as follows.

DEFINITION 2.2. K is said to be *absolutely monotonic* at a given point $\xi \in \mathbb{R}$ if $I - \xi A$ is nonsingular and $(\partial^{i_1+i_2+\cdots+i_m}K/\partial z_1^{i_1}\partial z_2^{i_2}\ldots\partial z_m^{i_m})(\xi, \xi, \ldots, \xi) \geq 0$ for all nonnegative integers $i_1, i_2, \ldots, i_m$. Further, K is said to be absolutely monotonic on a given set $\Omega \subset \mathbb{R}$ if K is absolutely monotonic at each $\xi \in \Omega$.

Note that in the two above definitions, absolute monotonicity of the scalar functions φ and K at a real point ξ amounts to the nonnegativity of all coefficients of the Taylor expansion about $z = \xi$ and $Z = \xi I$, respectively. In this light it is natural to define absolute monotonicity of the matrix-valued function $\mathbb{K}$ at a point $\xi \in \mathbb{R}$ as the nonnegativity of all coefficients of the expansion of $\mathbb{K}(\mathbb{Z})$ about $\mathbb{Z} = \xi\mathbb{I}$. In order to find the coefficients of this expansion, we introduce the following notation, provided that $I - \xi A$ is nonsingular,

(2.4a) $$A(\xi) = (\alpha_{ij}(\xi)) = A(I - \xi A)^{-1},$$

(2.4b) $$b(\xi)^{\mathrm{T}} = (\beta_1(\xi), \beta_2(\xi), \ldots, \beta_m(\xi)) = b^{\mathrm{T}}(I - \xi A)^{-1},$$

(2.4c) $$e(\xi) = (\varepsilon_1(\xi), \varepsilon_2(\xi), \ldots, \varepsilon_m(\xi))^{\mathrm{T}} = (I - \xi A)^{-1}e.$$

Let s be a positive integer and write $\mathbb{A}(\xi) = A(\xi) \otimes I_s$, $\mathbb{b}(\xi) = b(\xi) \otimes I_s$, $\mathbb{e}(\xi) = e(\xi) \otimes I_s$. Suppose $\mathbb{W}$ is a block diagonal matrix $\mathbb{W} = \mathrm{diag}(W_1, W_2, \ldots, W_m)$, where each W_i is a real (or complex) $s \times s$ matrix. If $\mathbb{Z} = \xi\mathbb{I} + \mathbb{W}$, and $\mathbb{W}$ is sufficiently close to zero, it is easy to prove that $\mathbb{I} - \mathbb{A}\mathbb{Z}$ is nonsingular with inverse

(2.5) $$(\mathbb{I} - \mathbb{A}\mathbb{Z})^{-1} = \sum_{k=0}^{\infty} [\mathbb{A}(\xi)\mathbb{W}]^k [\mathbb{I} - \xi\mathbb{A}]^{-1}.$$

A straightforward computation shows that this leads to

(2.6) $$\mathbb{K}(\mathbb{Z}) = \varphi(\xi)I_s + \sum_{k=1}^{\infty} \mathbb{b}(\xi)^{\mathrm{T}}\mathbb{W}[\mathbb{A}(\xi)\mathbb{W}]^{k-1}\mathbb{e}(\xi) =$$

$$= \varphi(\xi)I_s + \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k} \beta_{i_1}\alpha_{i_1 i_2}\alpha_{i_2 i_3}\ldots\alpha_{i_{k-1} i_k}\varepsilon_{i_k} W_{i_1} W_{i_2}\ldots W_{i_k},$$

where we have suppressed the dependence of the entries α_{ij}, β_i and ε_i on the parameter ξ for shortness of notation. This expansion gives rise to the following definition.

DEFINITION 2.3. $\mathbb{K}$ is said to be *absolutely monotonic* at a given point $\xi \in \mathbb{R}$ if $I - \xi A$ is nonsingular, $\varphi(\xi) \geq 0$ and $\beta_{i_1}\alpha_{i_1 i_2}\alpha_{i_2 i_3} \cdots \alpha_{i_{k-1} i_k}\varepsilon_{i_k} \geq 0$ for all $k \geq 1$ and $i_1, i_2, \ldots, i_k$. Further, $\mathbb{K}$ is said to be absolutely monotonic on a given set $\Omega \subset \mathbb{R}$ if $\mathbb{K}$ is absolutely monotonic at each $\xi \in \Omega$.

For the investigation of absolute monotonicity of the functions K and $\mathbb{K}$, the following property of the coefficient scheme (A, b) is of great importance.

DEFINITION 2.4. The coefficient scheme (A, b) is said to be *absolutely monotonic* at a given point $\xi \in \mathbb{R}$ if $I - \xi A$ is nonsingular, $\varphi(\xi) \geq 0$, $A(\xi) \geq 0$, $b(\xi) \geq 0$ and $e(\xi) \geq 0$. Further, the coefficient scheme is said to be absolutely monotonic on a given set $\Omega \subset \mathbb{R}$ if it is absolutely monotonic at each $\xi \in \Omega$.

In the above definition the inequalities involving $A(\xi)$, $b(\xi)$ and $e(\xi)$ should be interpreted component-wise.

A relation between the four concepts, introduced in the definitions above, is given in the following lemma. Its easy proof is omitted.

LEMMA 2.5. *Suppose that Ω is a subset of $\mathbb{R}$. Then, for any coefficient scheme (A, b), we have (a) $\Rightarrow$ (b) $\Rightarrow$ (c) $\Rightarrow$ (d), where*
(a) *(A, b) is absolutely monotonic on Ω;*
(b) *$\mathbb{K}$ is absolutely monotonic on Ω;*
(c) *K is absolutely monotonic on Ω;*
(d) *φ is absolutely monotonic on Ω.*

We conclude this subsection with a remark.

REMARK 2.6. As is done in [24], it is possible, and probably more natural, to define absolute monotonicity of the K-function on sets $\Omega' \subset \mathbb{R}^m$ rather than – as is done in Definition 2.2 – on sets $\Omega \subset \mathbb{R}$. In this alternative definition, K is said to be absolutely monotonic on a given set $\Omega' \subset \mathbb{R}^m$ if the Taylor expansion of K about each point in Ω' exists and has nonnegative coefficients. The reason why we did not adopt this definition is that the presentation of many of our results would have become considerably more complicated. This disadvantage is even more pronounced for similar adaptations of Definition 2.3. Further, with these alternative definitions we would not arrive at stronger results than those obtained with the present definitions.

2.3. *Reducibility of* (A, b).

Let a Runge-Kutta method be given with $m \geq 1$ stages and coefficient scheme (A, b). Following [7], [8] the coefficient scheme is said to be *DJ-reducible* if there exist disjoint index sets $\mathcal{M}_1$ and $\mathcal{M}_2$ with $\mathcal{M}_1 \cup \mathcal{M}_2 = \{1, 2, \ldots, m\}$ and $\mathcal{M}_2 \neq \emptyset$ such that $b_i = 0$ (if $i \in \mathcal{M}_2$) and $a_{ij} = 0$ (if $i \in \mathcal{M}_1$ and $j \in \mathcal{M}_2$). In case of DJ-reducibility, the method makes no use of the stages with index in $\mathcal{M}_2$, and is therefore equivalent to a Runge-Kutta method with m' stages, where m' is the number of elements in $\mathcal{M}_1$. The following lemma will be useful in Section 3.

LEMMA 2.7. *Let a Runge-Kutta method be given with* $m \geq 1$ *stages and coefficient scheme* (A, b). *Suppose that there exists an index* $i \in \{1, 2, \ldots, m\}$ *such that*

(2.7) $b_{i_1} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{k-1} i_k} = 0$ (*for all* $k \geq 1$ *and all indices* $i_1, i_2, \ldots, i_k$ *with* $i_k = i$).

Then (A, b) *is* DJ-*reducible*.

PROOF. Define $\mathcal{M}_2$ as the index set containing all i with property (2.7), and let $\mathcal{M}_1 = \{1, 2, \ldots, m\} \backslash \mathcal{M}_2$. ■

A different reducibility concept was introduced in [18] (see also [8], [9]). The coefficient scheme (A, b) is said to be *HS-reducible* if for some integer r with $1 \leq r < m$ and some nonempty pairwise disjoint index sets $\mathcal{M}_1, \mathcal{M}_2, \ldots, \mathcal{M}_r$ with $\mathcal{M}_1 \cup \mathcal{M}_2 \cup \ldots \cup \mathcal{M}_r = \{1, 2, \ldots, m\}$ we have

$$\sum_{k \in \mathcal{M}_\sigma} a_{ik} = \sum_{k \in \mathcal{M}_\sigma} a_{jk}$$

for all σ with $1 \leq \sigma \leq r$ and all i, j belonging to the same index set $\mathcal{M}_\rho$ with $1 \leq \rho \leq r$. In case of HS-reducibility, all vectors y_i in (1.3b) with an index i belonging to the same index set $\mathcal{M}_\rho$ are considered to be equal, leading to a Runge-Kutta method with $r < m$ stages. In Section 5 we make use of the following lemma due to W. H. Hundsdorfer, the proof of which can be found in [23].

LEMMA 2.8. *Let a Runge-Kutta method be given with* $m \geq 1$ *stages and coefficient scheme* (A, b). *Suppose* (A, b) *is not HS-reducible. Then, for any real* γ, *there exist vectors* $p = (p_1, p_2, \ldots, p_m)^T$ and $q = (q_1, q_2, \ldots, q_m)^T$ *in* $\mathbb{R}^m$ *such that* $q = Ap$ *and*

$$p_i \neq p_j, \quad (q_i - q_j)/(p_i - p_j) < \gamma \quad (\textit{for all } i, j \textit{ with } i \neq j).$$

A unified approach combining the two reducibility concepts above is presented in [8]. Following [8] we call a coefficient scheme (A, b) *reducible* if it is DJ-reducible and/or HS-reducible, and *irreducible* otherwise.

3. Equivalence of absolute monotonicity of K, $\mathbb{K}$ and (A, b).

3.1. *Introduction.*

Throughout this section (A, b) is a given coefficient scheme of a Runge-Kutta method with $m \geq 1$ stages. The corresponding functions φ, K and $\mathbb{K}$ are defined in Subsection 2.1.

The main purpose of this section is to prove for any real interval $[-r, 0]$ with $r \geq 0$, the equivalence of the following three propositions,

(3.1) K is absolutely monotonic on $[-r, 0]$;

(3.2) $\mathbb{K}$ is absolutely monotonic on $[-r, 0]$;

(3.3) (A, b) is absolutely monotonic on $[-r, 0]$.

In addition to the notation of Section 2, the index set $\{1, 2, \ldots, m\}$ is denoted by $\mathscr{M}$. For real matrices (or vectors) $F = (f_{ij})$ and $G = (g_{ij})$ we write $F \geq G$ if $f_{ij} \geq g_{ij}$ for all i and j, and $F > G$ if $f_{ij} > g_{ij}$ for all i and j. The reverse relations $\leq$ and $<$ are defined similarly. The matrix with entries $|f_{ij}|$ is denoted by $|F|$. If F is a square matrix, its *spectral radius* is denoted by $\operatorname{spr}(F)$.

3.2. *Equivalence of absolute monotonicity of K and $\mathbb{K}$.*

In this subsection we prove that, for any $\xi \in \mathbb{R}$, absolute monotonicity at ξ of K and $\mathbb{K}$ are equivalent. We start with an auxiliary lemma.

LEMMA 3.1. *Let $\psi = P/Q$ be a rational function in the complex variable z, where P and Q are polynomials with real coefficients. Suppose ψ is absolutely monotonic at a given point $\xi \in \mathbb{R}$. Then ψ is absolutely monotonic on the interval $[\xi, \eta)$, where η satisfies $\xi < \eta \leq \infty$ and is defined by $\eta = \inf\{t \mid t \in (\xi, \infty) \text{ and } Q(t) = 0\}$. Further, the Taylor series of ψ about $z = \xi$ has a radius of convergence $\geq \eta - \xi$.*

PROOF. The Taylor series of ψ about $z = \xi$ is given by $\sum_{k=0}^{\infty} \gamma_k (z - \xi)^k$, where $\gamma_k = \psi^{(k)}(\xi)/k! \geq 0$. Suppose that the radius of convergence ρ is smaller than $\eta - \xi$, where η is defined as above. Then, for all complex z with $|z - \xi| < \rho$ and $Q(z) \neq 0$ we have

$$|\psi(z)| = \left| \sum_{k=0}^{\infty} \gamma_k (z - \xi)^k \right| \leq \sum_{k=0}^{\infty} \gamma_k |z - \xi|^k \leq \lim_{\lambda \uparrow \rho} \sum_{k=0}^{\infty} \gamma_k \lambda^k = \lim_{\lambda \uparrow \rho} \psi(\xi + \lambda) = \psi(\xi + \rho).$$

Since this uniform bound for $|\psi(z)|$ is in contradiction with the existence of a pole of ψ on the circle $\{z \mid z \in \mathbb{C}, |z - \xi| = \rho\}$, we conclude that $\rho \geq \eta - \xi$. Hence $\psi(z)$ can be represented by the above Taylor series for all $z \in [\xi, \eta)$. Term by term differentiation shows that ψ is absolutely monotonic on $[\xi, \eta)$. ■

A similar result can be proved for the (scalar) K-function.

LEMMA 3.2. *If K is absolutely monotonic at a given point $\xi \in \mathbb{R}$, then K is absolutely monotonic on the interval $[\xi, \eta)$, where η satisfies $\xi < \eta \leq \infty$ and is defined by $\eta = \inf\{t \mid t \in (\xi, \infty)$ and $I - tA$ is singular$\}$.*

PROOF. From the definition of the K-function it is not difficult to see that each partial derivative $\partial^{i_1+i_2+\ldots+i_m}K/\partial z_1^{i_1}\partial z_2^{i_2}\ldots\partial z_m^{i_m}$ is a rational function in the complex variables $z_1, z_2, \ldots, z_m$ with denominator $\det(I - AZ)^n$, where $n = 1 + i_1 + i_2 + \ldots + i_m$. If we evaluate this partial derivative at $Z = tI$ for real values t, we obtain a rational function ψ in the variable t with denominator $\det(I - tA)^n$. Since the absolute monotonicity of K at ξ implies the absolute monotonicity of ψ at ξ, it follows from Lemma 3.1 that ψ is absolutely monotonic on $[\xi, \eta)$ with η as above. In particular it follows that ψ is nonnegative on $[\xi, \eta)$. Since ψ corresponds to an arbitrarily chosen partial derivative of K, we conclude that K is absolutely monotonic on $[\xi, \eta)$. ■

The following lemma deals with the sets $\mathscr{B}$ and $\mathscr{E}$, defined by

(3.4) $\mathscr{B} = \{\xi \mid \xi \in \mathbb{R}, I - \xi A$ is nonsingular, there exists an index i with $\beta_i(\xi) = 0\}$;

(3.5) $\mathscr{E} = \{\xi \mid \xi \in \mathbb{R}, I - \xi A$ is nonsingular, there exists an index i with $\varepsilon_i(\xi) = 0\}$.

We recall that $\beta_i(\xi)$ and $\varepsilon_i(\xi)$ are defined in (2.4).

LEMMA 3.3. *Let the sets $\mathscr{B}$ and $\mathscr{E}$ be defined by* (3.4) *and* (3.5). *Then $\mathscr{E}$ is finite. For irreducible coefficient schemes the set $\mathscr{B}$ is finite if for at least one point $\xi \in \mathbb{R}$ the function K is absolutely monotonic at ξ.*

PROOF. 1. First we show that $\mathscr{E}$ is finite. For real ξ such that $I - \xi A$ is nonsingular, Cramer's rule yields $\varepsilon_i(\xi) = p_i(\xi)/\det(I - \xi A)$, where p_i is a polynomial of degree $\leq m - 1$. Since $p_i(0) = 1$, p_i can only have a finite number of zeros. Hence $\mathscr{E}$ must be finite.

2. Suppose that for at least one point $\xi \in \mathbb{R}$ the function K is absolutely monotonic at ξ, and that $\mathscr{B}$ is infinite. We shall prove that (A, b) is DJ-reducible. First note that, in view of Lemma 3.2, we may assume that K is absolutely monotonic on some nonempty open interval $\mathscr{I} \subset \mathbb{R}$. Since each β_i is a rational function in one variable, the assumption that $\mathscr{B}$ is infinite implies that there exists an index i such that $\beta_i(\xi) = 0$ for all $\xi \in \mathscr{I}$. Hence for this index and for all $\xi \in \mathscr{I}$ we arrive at $\beta_i(\xi) = \beta_i'(\xi) = \beta_i''(\xi) = \ldots = 0$. Using the fact that

$$\frac{d^k}{d\xi^k} b(\xi)^T = k!\, b(\xi)^T A(\xi)^k \qquad \text{(for all integers } k \geq 0 \text{ and all } \xi \in \mathscr{I}\text{)},$$

we obtain

$$\sum_{(i_1, i_2, \ldots, i_k) \in \mathcal{M}^k:\, i_k = i} \beta_{i_1}(\xi)\alpha_{i_1 i_2}(\xi)\alpha_{i_2 i_3}(\xi)\ldots\alpha_{i_{k-1} i_k}(\xi) = 0 \tag{3.6}$$

for all integers $k \geq 1$ and all $\xi \in \mathscr{I}$. By using induction on k we shall prove that this implies for all integers $k \geq 1$ and all $\xi \in \mathscr{I}$:

$$\beta_{i_1}(\xi)\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_{k-1} i_k}(\xi)\varepsilon_{i_k}(\xi) = 0 \quad (\text{for all } (i_1, i_2, \ldots, i_k) \in \mathcal{M}^k \text{ with } i_k = i). \tag{3.7}$$

Let $\xi \in \mathscr{I}$ be fixed. For $k = 1$, (3.7) immediately follows from (3.6). Now assume that, for some given $n \geq 1$, we have proved (3.7) for $k = 1, 2, \ldots, n$. We shall prove (3.7) for $k = n + 1$. Note that we may assume that $\varepsilon_i(\xi) \neq 0$ since otherwise (3.7) trivially holds. From the Taylor series of K about $Z = \xi I$, which can be obtained by taking $s = 1$ in (2.6), it is easy to see that absolute monotonicity of K at ξ implies

$$\sum_{(i_1, \ldots, i_n, i_{n+1}) \text{ is permutation of } (j_1, \ldots, j_n, i)} \beta_{i_1}(\xi)\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_n i_{n+1}}(\xi)\varepsilon_{i_{n+1}}(\xi) \geq 0$$

for all $(j_1, j_2, \ldots, j_n) \in \mathcal{M}^n$. In view of (3.7) (with $k = 1, 2, \ldots, n$) and $\varepsilon_i(\xi) \neq 0$ this amounts to

$$\sum_{(i_1, i_2, \ldots, i_n) \text{ is permutation of } (j_1, j_2, \ldots, j_n)} \beta_{i_1}(\xi)\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_n i}(\xi)\varepsilon_i(\xi) \geq 0 \tag{3.8}$$

for all $(j_1, j_2, \ldots, j_n) \in \mathcal{M}^n$. In combination with (3.6) (for $k = n + 1$) and the fact that

$$\sum_{(i_1, i_2, \ldots, i_n) \in \mathcal{M}^n} \ldots = \sum_{\substack{(j_1, j_2, \ldots, j_n) \in \mathcal{M}^n \\ j_1 \leq j_2 \leq \ldots \leq j_n}} \; \sum_{\substack{(i_1, i_2, \ldots, i_n) \text{ is permutation} \\ \text{of } (j_1, j_2, \ldots, j_n)}} \ldots,$$

it follows that equality holds in (3.8) for all $(j_1, j_2, \ldots, j_n) \in \mathcal{M}^n$. Now suppose that (3.7) does not hold for $k = n + 1$. Then there exists $(j_1, j_2, \ldots, j_n) \in \mathcal{M}^n$ such that

$$\beta_{j_1}(\xi)\alpha_{j_1 j_2}(\xi)\ldots\alpha_{j_n i}(\xi)\varepsilon_i(\xi) \neq 0. \tag{3.9}$$

Since we proved that equality holds in (3.8), there exists at least one permutation $(i_1, i_2, \ldots, i_n)$ of $(j_1, j_2, \ldots, j_n)$ with $(i_1, i_2, \ldots, i_n) \neq (j_1, j_2, \ldots, j_n)$ and

$$\beta_{i_1}(\xi)\alpha_{i_1, i_2}(\xi)\ldots\alpha_{i_n i}(\xi)\varepsilon_i(\xi) \neq 0. \tag{3.10}$$

Let λ denote the smallest index with $i_\lambda \neq j_\lambda$. Then $i_\lambda = j_\mu$ for some $\mu > \lambda$. It follows from (3.9) and (3.10) that

$$\beta_{i_1}(\xi)\alpha_{i_1, i_2}(\xi)\ldots\alpha_{i_{\lambda-1} i_\lambda}(\xi)\alpha_{j_\mu j_{\mu+1}}(\xi)\ldots\alpha_{j_n i}(\xi)\varepsilon_i(\xi) \neq 0.$$

Since this contradicts our induction hypothesis (3.7) (with $k = n + 1 - \mu + \lambda$), we have proved (3.7) for $k = n + 1$. This concludes the proof by induction of (3.7).

Note that the left hand side of (3.7) is a rational function in ξ with denominator $\det(I - \xi A)^{k+1}$. Since this rational function exists and vanishes for all $\xi \in \mathscr{I}$, we conclude that it must vanish at $\xi = 0$ as well, i.e. we have proved (2.7). An application of Lemma 2.7 shows that (A, b) is DJ-reducible, which was to be proved. ■

Lemmas 3.2 and 3.3 are used in the proof of the following theorem, which constitutes the main result of this subsection.

THEOREM 3.4. *Let $\xi \in \mathbb{R}$ be arbitrary. If the coefficient scheme is irreducible, then K is absolutely monotonic at ξ if and only if $\mathbb{K}$ is absolutely monotonic at ξ.*

PROOF. For the "if-part" we refer to Lemma 2.5. In order to prove the "only-if-part", we assume that K is absolutely monotonic at ξ. From Lemma 3.2 it follows that K is absolutely monotonic on an interval $\mathscr{I} = [\xi, \eta)$ with $\eta > \xi$. Hence, for any $t \in \mathscr{I}$, the Taylor expansion of K about $Z = tI$ exists and has all coefficients nonnegative. By considering (2.6) with $s = 1$ (and ξ replaced by t) we arrive at

$$\varphi(t) \geq 0, \tag{3.11}$$

$$\sum_{(j_1, j_2, \ldots, j_k) \text{ is permutation of } (i_1, i_2, \ldots, i_k)} \beta_{j_1}(t)\alpha_{j_1 j_2}(t) \ldots \alpha_{j_{k-1} j_k}(t)\varepsilon_{j_k}(t) \geq 0 \tag{3.12}$$

for all $t \in \mathscr{I}$, all integers $k \geq 1$ and all $(i_1, i_2, \ldots, i_k) \in \mathscr{M}^k$.

Define the set Γ by

$$\Gamma = \{t \mid t \in \mathscr{I} \text{ and } \beta_i(t)\varepsilon_i(t) = 0 \text{ for some } i \in \mathscr{M}\}.$$

Let $t \in \mathscr{I} \backslash \Gamma$ be given. By considering (3.12) with $k = 1$ we see that

$$\beta_i(t)\varepsilon_i(t) > 0 \quad (\text{for all } i \in \mathscr{M}). \tag{3.13}$$

Consideration of (3.12) with $k = 2$ and $i_1 = i_2$ yields $\beta_i(t)\alpha_{ii}(t)\varepsilon_i(t) \geq 0$ (for all $i \in \mathscr{M}$). In view of (3.13) this implies

$$\alpha_{ii}(t) \geq 0 \quad (\text{for all } i \in \mathscr{M}). \tag{3.14}$$

Now let $i, j \in \mathscr{M}$ be given with $i \neq j$. By considering (3.12) with $(i_1, i_2, \ldots, i_k) = (i, i, \ldots, i, j)$ we obtain for all $k \geq 2$ the inequality

$$\beta_i \alpha_{ii}^{k-2} \alpha_{ij} \varepsilon_j + \beta_j \alpha_{ji} \alpha_{ii}^{k-2} \varepsilon_i + (k-2)\beta_i \alpha_{ij} \alpha_{ji} \alpha_{ii}^{k-3} \varepsilon_i \geq 0,$$

where the dependence on t has been suppressed for shortness of notation. By taking $k = 2$ it follows that

$$\beta_i \alpha_{ij} \varepsilon_j + \beta_j \alpha_{ji} \varepsilon_i \geq 0. \tag{3.15}$$

By taking $k = 3$ if $\alpha_{ii} = 0$, or considering $k \to \infty$ if $\alpha_{ii} > 0$, it follows that $\beta_i \alpha_{ij} \alpha_{ji} \varepsilon_i \geq 0$. In view of (3.13), multiplication by $\beta_j \varepsilon_j$ yields $(\beta_i \alpha_{ij} \varepsilon_j)(\beta_j \alpha_{ji} \varepsilon_i) \geq 0$. Combination of this inequality with (3.15) shows that

$$\beta_i(t)\alpha_{ij}(t)\varepsilon_j(t) \geq 0 \quad (\text{for all } i, j \in \mathscr{M} \text{ with } i \neq j). \tag{3.16}$$

Using (3.13), (3.14) and (3.16) it is easy to deduce that

$$\beta_{i_1}(t)\alpha_{i_1 i_2}(t) \cdots \alpha_{i_{k-1} i_k}(t)\varepsilon_{i_k}(t) \geq 0 \quad (\text{for all } k \geq 1 \text{ and all } (i_1, i_2, \ldots, i_k) \in \mathscr{M}^k). \tag{3.17}$$

Thus we have proved (3.17) for all $t \in \mathscr{I} \backslash \Gamma$. Since by Lemma 3.3 the set Γ is finite, it follows from continuity arguments that (3.17) must hold for all $t \in \mathscr{I}$. Combined with (3.11) this shows that $\mathbb{K}$ is absolutely monotonic on $\mathscr{I}$, and in particular at ξ. ■

3.3. *Equivalence of* (3.1), (3.2) *and* (3.3).

In this subsection we will prove the equivalence of (3.1), (3.2) and (3.3) for arbitrary $r \geq 0$. We start with two lemmas. The easy proof of the first lemma is left to the reader.

LEMMA 3.5. *Let ρ and σ be real numbers with $\rho \geq 0$. Then the following two propositions are equivalent.*

(3.18) $\quad I - zA$ *is nonsingular for all complex z with* $|z - \sigma| \leq \rho$;

(3.19) $\quad I - \sigma A$ *is nonsingular and* $\operatorname{spr}(\rho A(I - \sigma A)^{-1}) < 1$.

LEMMA 3.6. *Let $\psi = P/Q$ be a rational function in the complex variable z, where P and Q are polynomials with real coefficients and no common zeros. Further, let λ and μ be real numbers with $\lambda < \mu$. Suppose that ψ is absolutely monotonic at all but a finite number of the points in $[\lambda, \mu)$. Then ψ is absolutely monotonic on all of $[\lambda, \mu)$.*

PROOF. If $[\lambda, \mu)$ contains no poles of ψ, the assertion easily follows from a limit argument. Therefore it is sufficient to prove that $[\xi, \eta)$ contains no poles of ψ. Suppose $\gamma \in [\xi, \eta)$ is a pole of ψ. Then either $\psi(z) < 0$ or $\psi'(z) < 0$ for all (real) z in a right neighbourhood of γ. In both cases we have a contradiction with the absolute monotonicity of ψ. ■

The two lemmas above are used in the proof of the following result.

LEMMA 3.7. *Let λ and μ be real numbers with $\lambda < \mu$. Suppose that for a given irreducible coefficient scheme (A, b) the function K is absolutely monotonic at all but a finite number of the points in the interval $[\lambda, \mu)$. Then K is absolutely monotonic on all of $[\lambda, \mu)$. Further we have* $\operatorname{spr}(|A(\xi)|) \leq (\mu - \xi)^{-1}$ *for all $\xi \in [\lambda, \mu)$.*

PROOF. According to the assumptions, K is absolutely monotonic on $[\lambda, \mu) \backslash \Gamma$, where Γ is a finite subset of $[\lambda, \mu)$. Note that it follows from Lemma 3.3 that we may assume, without loss of generality, that

(3.20) $\quad \beta_i(\xi)\varepsilon_i(\xi) \neq 0 \quad$ (for all $\xi \in [\lambda, \mu) \backslash \Gamma$ and all $i \in \mathscr{M}$).

Let ψ denote the rational function that is obtained from φ by removing all removable singularities. Since K is absolutely monotonic on $[\lambda, \mu) \backslash \Gamma$, it follows from Lemma 2.5 that φ, and therefore also ψ, is absolutely monotonic on $[\lambda, \mu)\backslash\Gamma$. Using Lemma 3.6 we see that ψ is absolutely monotonic on $[\lambda, \mu)$. From Lemma 3.1 we conclude that for any $\xi \in [\lambda, \mu)$ the Taylor series of ψ about $z = \xi$ has a radius of convergence $\geq \mu - \xi$. As the Taylor series of φ about $z = \xi$ is easily obtained by taking scalar $W_1 = W_2 = \ldots = W_m$ in (2.6), we see that

$$\psi(\xi + \delta) = \varphi(\xi) + \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k} \beta_{i_1}(\xi)\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_{k-1} i_k}(\xi)\varepsilon_{i_k}(\xi)\delta^k$$

for all $\xi \in [\lambda, \mu)\backslash\Gamma$ and all $\delta \in [0, \mu - \xi)$. Note that it follows from Theorem 3.4 that $\mathbb{K}$ is absolutely monotonic on $[\lambda, \mu)\backslash\Gamma$, so that all terms of the above series are nonnegative. But then we have

$$\sum_{i,j} \left\{ |\beta_i(\xi)\varepsilon_j(\xi)| \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k : i_1 = i, i_k = j} |\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_{k-1} i_k}(\xi)|\, \delta^{k-1} \right\} < \infty$$

for all $\xi \in [\lambda, \mu)\backslash\Gamma$ and all $\delta \in [0, \mu - \xi)$. In view of (3.20) this implies

$$\sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k : i_1 = i, i_k = j} |\alpha_{i_1 i_2}(\xi)\ldots\alpha_{i_{k-1} i_k}(\xi)|\, \delta^{k-1} < \infty$$

for all $\xi \in [\lambda, \mu)\backslash\Gamma$, $\delta \in [0, \mu - \xi)$ and all $i, j \in \mathcal{M}$. Noting that the left hand side of the above inequality is the (i,j)-th element of the series

$$I + \delta\,|A(\xi)| + \delta^2\,|A(\xi)|^2 + \ldots,$$

we conclude that this series converges for all $\xi \in [\lambda, \mu)\backslash\Gamma$ and all $\delta \in [0, \mu - \xi)$. This is known to be equivalent to

(3.21) $\qquad \operatorname{spr}(\delta\,|A(\xi)|) < 1 \quad$ (for all $\xi \in [\lambda, \mu)\backslash\Gamma$ and all $\delta \in [0, \mu - \xi)$),

implying $\operatorname{spr}(\delta A(\xi)) < 1$ (for all $\xi \in [\lambda, \mu)\backslash\Gamma$ and all $\delta \in [0, \mu - \xi)$). Lemma 3.5 shows that the latter property is equivalent to the nonsingularity of $I - zA$ for all complex z such that $|z - \xi| \leq \delta$ for some $\xi \in [\lambda, \mu)\backslash\Gamma$ and $\delta \in [0, \mu - \xi)$. Since Γ is finite, we can take ξ arbitrarily close to λ here, which leads to the nonsingularity of $I - zA$ for all complex z with $|z - \lambda| < \mu - \lambda$. In particular we have proved that $I - \xi A$ is nonsingular for all $\xi \in [\lambda, \mu)$. Hence all partial derivatives of K exist and are continuous at all points $Z = \xi I$ with $\xi \in [\lambda, \mu)$. By using this continuity it follows that K is not merely absolutely monotonic on $[\lambda, \mu)\backslash\Gamma$ but on all of $[\lambda, \mu)$. Further, it follows from (3.21), the finiteness of Γ and the nonsingularity of $I - \xi A$ for all $\xi \in [\lambda, \mu)$ that $\operatorname{spr}(\delta\,|A(\xi)|) \leq 1$ for all $\xi \in [\lambda, \mu)$ and $\delta \in [0, \mu - \xi)$, i.e. $\operatorname{spr}(|A(\xi)|) \leq (\mu - \xi)^{-1}$ for all $\xi \in [\lambda, \mu)$. ∎

This lemma will be used in Sections 4 and 5. It also plays an important role in the proof of the following result, which is the main result of this section.

THEOREM 3.8. *Let (A, b) be an irreducible coefficient scheme and r a nonnegative real number. Then* (3.1), (3.2) *and* (3.3) *are equivalent.*

PROOF. In view of Lemma 2.5 it is sufficient to prove (3.1) $\Rightarrow$ (3.3).

1. We first consider the case that $r = 0$. Hence we may assume that K is absolutely monotonic at 0 and have to prove that (A, b) is absolutely monotonic at 0, i.e. $A \geq 0$ and $b \geq 0$. First note that we may assume that $\mathbb{K}$ is absolutely monotonic at 0 by Theorem 3.4, i.e. $b_{i_1} a_{i_1 i_2} \cdots a_{i_{k-1} i_k} \geq 0$ for all $k \geq 1$ and all $i_1, i_2, \ldots, i_k$. Consideration of $k = 1$ immediately leads to $b \geq 0$. In order to prove $A \geq 0$ we choose arbitrary $i, j \in \mathcal{M}$. Since the method is irreducible, it follows from Lemma 2.7 that for some $k \geq 1$ and some indices $i_1, i_2, \ldots, i_k$ with $i_k = i$ we have $b_{i_1} a_{i_1 i_2} \cdots a_{i_{k-1} i_k} > 0$. Using the fact that $b_{i_1} a_{i_1 i_2} \cdots a_{i_{k-1} i_k} a_{ij} \geq 0$ we arrive at $a_{ij} \geq 0$.

2. Now let $r > 0$ and assume (3.1). We shall prove (3.3). From Lemma 3.2 we see that K is absolutely monotonic on $[-r, \eta)$ for some positive η. In view of Lemma 3.7 this implies $\operatorname{spr}(|A(\xi)|) \leq (\eta - \xi)^{-1}$ for all $\xi \in [-r, \eta)$, and hence

$$\operatorname{spr}(-\xi A(\xi)) \leq \operatorname{spr}(|-\xi A(\xi)|) \leq -\xi(\eta - \xi)^{-1} < 1 \quad (\text{for all } \xi \in [-r, 0]).$$

Combined with $[I + \xi A(\xi)]^{-1} = I - \xi A$ we may conclude that

$$-\xi A = -\xi A(\xi) + \xi^2 A(\xi)^2 - \xi^3 A(\xi)^3 + \ldots \quad (\text{for all } \xi \in [-r, 0]). \tag{3.22}$$

Next define the set $\Omega = \{\xi \mid \xi \in [-r, 0],\ \beta_i(\xi)\varepsilon_i(\xi) \neq 0 \text{ for all } i \in \mathcal{M}\}$. Let $\xi \in \Omega$ be fixed, and define the index sets $\mathcal{M}_1$ and $\mathcal{M}_2$ by $\mathcal{M}_1 = \{i \mid i \in \mathcal{M},\ \beta_i(\xi) > 0\}$ and $\mathcal{M}_2 = \mathcal{M} \setminus \mathcal{M}_1$. Without loss of generality we may assume that $i < j$ whenever $i \in \mathcal{M}_1$ and $j \in \mathcal{M}_2$. Since $\mathbb{K}$ is absolutely monotonic at ξ by Theorem 3.4, it follows that $b(\xi)$, $e(\xi)$ and $A(\xi)$ must have the form

$$b(\xi) = \begin{matrix} \mathcal{M}_1 \\ \mathcal{M}_2 \end{matrix} \begin{bmatrix} > 0 \\ < 0 \end{bmatrix}, \quad e(\xi) = \begin{matrix} \mathcal{M}_1 \\ \mathcal{M}_2 \end{matrix} \begin{bmatrix} > 0 \\ < 0 \end{bmatrix}, \quad A(\xi) = \begin{matrix} \\ \mathcal{M}_1 \\ \mathcal{M}_2 \end{matrix} \begin{matrix} \begin{matrix} \mathcal{M}_1 & \mathcal{M}_2 \end{matrix} \\ \begin{bmatrix} \geq 0 & \leq 0 \\ \leq 0 & \geq 0 \end{bmatrix} \end{matrix},$$

where, for example, '> 0' in $b(\xi)$ means that $\beta_i(\xi) > 0$ for all $i \in \mathcal{M}_1$. In view of (3.22) this shows that $-\xi A$ is of the same form as $A(\xi)$. But then

$$e = (I - \xi A)e(\xi) = e(\xi) - \xi A e(\xi) = \begin{bmatrix} > 0 \\ < 0 \end{bmatrix} + \begin{bmatrix} \geq 0 & \leq 0 \\ \leq 0 & \geq 0 \end{bmatrix} \begin{bmatrix} > 0 \\ < 0 \end{bmatrix} = \begin{bmatrix} > 0 \\ < 0 \end{bmatrix},$$

which is only possible if $\mathcal{M}_2 = \emptyset$. Hence we have proved that $b(\xi) > 0$, $e(\xi) > 0$ and $A(\xi) \geq 0$ for all $\xi \in \Omega$. Further, $I - \xi A$ is nonsingular and $\varphi(\xi) \geq 0$ for all $\xi \in [-r, 0]$ by the absolute monotonicity of K on $[-r, 0]$. Since Lemma 3.3 implies that $[-r, 0] \setminus \Omega$ is finite, it follows from a continuity argument that (A, b) is absolutely monotonic on $[-r, 0]$. ■

4. Investigating absolute monotonicity of (A, b).

4.1. *The radius of absolute monotonicity.*

In this section we study for coefficient schemes (A, b) the so-called *radius of absolute monotonicity* $R(A, b)$, defined by

(4.1) $R(A,b) = \sup\{r \mid r \in \mathbb{R} \text{ and } (A,b) \text{ is absolutely monotonic on } [-r,0]\}$.

In this subsection we present two general results on $R(A,b)$. In Subsection 4.2 we concentrate on the special case $R(A,b) = \infty$.

Our first general result on $R(A,b)$ is the characterization of all irreducible coefficient schemes with $R(A,b) > 0$. For the formulation and proof of this result it is convenient to give the following definition (cf. [1]).

DEFINITION 4.1. For a given matrix $F = (f_{ij})$ we define its *incidence matrix* $\mathrm{Inc}(F) = (g_{ij})$ by $g_{ij} = 1$ if $f_{ij} \neq 0$ and $g_{ij} = 0$ if $f_{ij} = 0$.

THEOREM 4.2. *For irreducible coefficient schemes* (A,b) *we have* $R(A,b) > 0$ *if and only if* $A \geq 0$, $b > 0$ *and* $\mathrm{Inc}(A^2) \leq \mathrm{Inc}(A)$.

PROOF. Note that for real ξ close to zero the matrix $I - \xi A$ is nonsingular and $\varphi(\xi) > 0$, $e(\xi) > 0$. Further, for real ξ close to zero we also have

$$A(\xi) = A + \xi A^2 + \xi^2 A^3 + \dots. \tag{4.2}$$

We see immediately from (4.2) that $A \geq 0$ and $\mathrm{Inc}(A^2) \leq \mathrm{Inc}(A)$ are necessary conditions for $A(\xi) \geq 0$ to hold in a left neighbourhood of $\xi = 0$. To see that these two conditions are also sufficient for the latter property, we note that they imply $\mathrm{Inc}(A^k) \leq \mathrm{Inc}(A)$ for all $k \geq 2$. From these inequalities, combined with $A \geq 0$ and (4.2), the desired property easily follows.

Finally, for real ξ close to zero we have

$$b(\xi)^{\mathrm{T}} = b^{\mathrm{T}} + \xi b^{\mathrm{T}} A + \xi^2 b^{\mathrm{T}} A^2 + \dots. \tag{4.3}$$

We see immediately from (4.3) that $b > 0$ implies that $b(\xi) > 0$ for all ξ in a left neighbourhood of $\xi = 0$. We conclude the proof of the theorem by showing that $R(A,b) > 0$ implies $b > 0$. Therefore, assume $R(A,b) > 0$ and note that we have already proved that $A \geq 0$ must hold then. From (4.3) we obtain $b \geq 0$ and the implication $b_j = 0 \Rightarrow \sum_i b_i a_{ij} = 0$ for all $j \in \mathcal{M}$. In view of $b \geq 0$ and $A \geq 0$ we even have $b_j = 0 \Rightarrow b_i a_{ij} = 0$ for all $i,j \in \mathcal{M}$. If we define the sets $\mathcal{M}_1$ and $\mathcal{M}_2$ by $\mathcal{M}_1 = \{i \mid i \in \mathcal{M} \text{ and } b_i > 0\}$ and $\mathcal{M}_2 = \mathcal{M} \setminus \mathcal{M}_1$ we see immediately that $a_{ij} = 0$ for all $i \in \mathcal{M}_1$ and $j \in \mathcal{M}_2$. From the irreducibility of (A,b) we conclude (cf. Subsection 2.3) that $\mathcal{M}_2 = \emptyset$, i.e. $b > 0$. ■

In the proof of the following lemma we make use of M-matrices, which are defined as follows (cf. [25]).

DEFINITION 4.3. A real square matrix F is said to be an *M-matrix* if F is nonsingular, $F^{-1} \geq 0$ and all the off-diagonal elements of F are nonpositive.

LEMMA 4.4. *Let* (A,b) *be an irreducible coefficient scheme and* r *a positive real* *number. Then $R(A,b) \geq r$ if and only if (A,b) is absolutely monotonic at $\xi = -r$ and $A \geq 0$.*

PROOF. 1. Suppose $R(A,b) \geq r$. Then (A,b) is absolutely monotonic on $(-r,0]$. By Lemma 2.5, K is absolutely monotonic on $(-r,0]$, and by Lemma 3.7 even on $[-r,0]$. Theorem 3.8 shows that (A,b) is absolutely monotonic on $[-r,0]$. In particular, (A,b) is absolutely monotonic at $\xi = 0$ (implying $A \geq 0$) and at $\xi = -r$.

2. Suppose (A,b) is absolutely monotonic at $\xi = -r$ and $A \geq 0$. From $A \geq 0$ and $A(-r) = A(I + rA)^{-1} \geq 0$ we see that the matrix $(I + rA)^{-1} = I - rA(I + rA)^{-1}$ is an M-matrix. From [25, p. 531] it follows that $\operatorname{spr}(rA(I + rA)^{-1}) < 1$. In view of Lemma 3.5 we may conclude that $I - \xi A$ is nonsingular for all $\xi \in [-r,0]$. Since by Lemma 2.5 K is absolutely monotonic at $\xi = -r$, we can apply Lemma 3.2 now to obtain absolute monotonicity of K on $[-r,0]$. By using Theorem 3.8 we arrive at $R(A,b) \geq r$. ■

The above lemma is very useful from a computational point of view. It says that for checking absolute monotonicity of (A,b) on a given interval $[-r,0]$ it is sufficient to consider the left endpoint $\xi = -r$ only.

4.2. *Absolute monotonicity on $(-\infty,0]$.*

In this subsection we characterize all coefficient schemes with $R(A,b) = \infty$. We begin with the case where A is nonsingular.

LEMMA 4.5. *Let (A,b) be an arbitrary coefficient scheme. Suppose that A is nonsingular. Then $R(A,b) = \infty$ if and only if*

(4.4a) $\quad A^{-1}$ *is an M-matrix,*

(4.4b) $\quad A^{-1}e \geq 0,$

(4.4c) $\quad b^{T}A^{-1} \geq 0,$

(4.4d) $\quad b^{T}A^{-1}e \leq 1.$

PROOF. 1. Suppose $R(A,b) = \infty$. Then $A \geq 0$ by the absolute monotonicity of (A,b) at $\xi = 0$. Now (4.4) easily follows from the absolute monotonicity of (A,b) at large negative ξ since we have the following expansions as $\xi \to -\infty$,

$$A(\xi) = -\xi^{-1}I - \xi^{-2}A^{-1} + O(\xi^{-3}),$$

$$e(\xi) = -\xi^{-1}A^{-1}e + O(\xi^{-2}),$$

$$b(\xi)^{T} = -\xi^{-1}b^{T}A^{-1} + O(\xi^{-2}),$$

$$\varphi(\xi) = 1 - b^{T}A^{-1}e + O(\xi^{-1}).$$

2. Suppose that (4.4) holds and let $\xi \in (-\infty, 0]$ be arbitrary. We will show that (A, b) is absolutely monotonic at ξ. First note that (4.4a) implies that $A^{-1} - \xi I$ is an M-matrix (cf. e.g. [25, p. 532]). Hence $I - \xi A$ is nonsingular and $A(\xi) = A(I - \xi A)^{-1} = (A^{-1} - \xi I)^{-1} \geq 0$. Using (4.4) we further obtain $e(\xi) = (I - \xi A)^{-1}e = A(\xi)A^{-1}e \geq 0$, $b(\xi)^{\mathrm{T}} = b^{\mathrm{T}}(I - \xi A)^{-1} = b^{\mathrm{T}}A^{-1}A(\xi) \geq 0$ and $\varphi(\xi) = 1 + \xi b^{\mathrm{T}}(I - \xi A)^{-1}e = 1 - b^{\mathrm{T}}A^{-1}e + b^{\mathrm{T}}A^{-1}A(\xi)A^{-1}e \geq 0$. ■

The case where A is singular is considered in the next lemma.

LEMMA 4.6. *There exists no irreducible coefficient scheme (A, b) such that A is singular and $R(A, b) = \infty$.*

PROOF. Suppose (A, b) is an irreducible coefficient scheme with $m \geq 1$ stages such that A is singular and $R(A, b) = \infty$. We will obtain a contradiction by showing that A must have at least two identical rows, i.e. (A, b) is HS-reducible (cf. Subsection 2.3).

First note that $b > 0$ by Theorem 4.2. Further, with the new variable $\lambda = -\xi^{-1}$, absolute monotonicity of (A, b) on $(-\infty, 0]$ yields

(4.5a) $\quad A + \lambda I$ is nonsingular (for all $\lambda > 0$),

(4.5b) $\quad A(A + \lambda I)^{-1} \geq 0$ (for all $\lambda > 0$),

(4.5c) $\quad b^{\mathrm{T}}(A + \lambda I)^{-1} \geq 0$ (for all $\lambda > 0$),

(4.5d) $\quad (A + \lambda I)^{-1}e \geq 0$ (for all $\lambda > 0$),

(4.5e) $\quad b^{\mathrm{T}}(A + \lambda I)^{-1}e \leq 1$ (for all $\lambda > 0$).

From the singularity of A it follows that there exist an integer $k \geq 1$ and $m \times m$ matrices V and W with $V \neq 0$ such that

(4.6) $$(A + \lambda I)^{-1} = \lambda^{-k}V + \lambda^{-k+1}W + O(\lambda^{-k+2}) \quad (\lambda \to 0).$$

Consideration of (4.5c), (4.5d) and (4.5e) for small $\lambda > 0$ shows that $b^{\mathrm{T}}V \geq 0$, $Ve \geq 0$ and $b^{\mathrm{T}}Ve \leq 0$. As $b > 0$, this leads to

(4.7a) $$b^{\mathrm{T}}V = 0,$$

(4.7b) $$Ve = 0.$$

Realizing that $A(A + \lambda I)^{-1} = (A + \lambda I)^{-1}A = I - \lambda(A + \lambda I)^{-1}$, we see from (4.6) that

(4.8) $$A(A + \lambda I)^{-1} = \lambda^{-k}VA + \lambda^{-k+1}WA + O(\lambda^{-k+2}) = I - \lambda^{-k+1}V + O(\lambda^{-k+2}) \quad (\lambda \to 0).$$

Combining (4.8) and (4.5b) we obtain

(4.9a) $$VA = 0,$$

(4.9b) $$WA \geq 0.$$

Suppose $k > 1$. Then (4.8) and (4.9a) show that $WA = -V$. By using (4.7b) and (4.9b) this leads to $V = 0$, which is a contradiction. Hence we may assume that $k = 1$, so that (4.8) and (4.9a) yield

$$WA = I - V. \tag{4.10}$$

Note that the matrix WA satisfies

$$WAx = 0 \quad (\text{for all } x \in \text{Ker}(A)), \tag{4.11a}$$

$$WAx = x \quad (\text{for all } x \in \text{Range}(A)), \tag{4.11b}$$

where we have used (4.9a) and (4.10) to obtain (4.11b). It follows from (4.11) that $\text{Ker}(A) \cap \text{Range}(A) = \{0\}$, so that

$$\mathbb{R}^m = \text{Ker}(A) \oplus \text{Range}(A), \tag{4.12}$$

where $\oplus$ denotes the direct sum (cf. e.g. [25, p. 89]). From (4.11) and (4.12) we see that WA is idempotent, i.e. $(WA)^2 = WA$. Further it follows from (4.7b), (4.9b) and (4.10) that WA is stochastic, i.e. $WA \geq 0$ and $WAe = e$. Finally, since (4.7a), (4.10) and $b > 0$ imply $b^T WA = b^T > 0$, the matrix WA has no zero columns. Hence we have proved that WA is an idempotent stochastic matrix without zero columns. Using the canonical form for idempotent stochastic matrices presented in [1, p. 66], it follows that there exists a permutation matrix P such that $PWAP^T$ is a block diagonal matrix $\text{diag}(S_1, S_2, \ldots, S_r)$, where each block S_i is an idempotent stochastic matrix all of whose rows are identical. Note that the number of blocks r satisfies $r = \text{rank}(PWAP^T) < m$, so that $PWAP^T$ and therefore also WA has at least two identical rows. Since (4.11b) implies $A = (WA)A$, we conclude that A has at least two identical rows as well. ■

Combining Lemmas 4.5 and 4.6 we arrive at the following theorem.

THEOREM 4.7. *Let (A, b) be an irreducible coefficient scheme. Then $R(A, b) = \infty$ if and only if A is nonsingular and* (4.4) *holds.*

5. Conditional contractivity.

5.1. *Dissipative initial value problems.*

In Section 1 we defined $\mathscr{F}$ to be the class of all pairs $(f, \|\cdot\|)$ such that f is dissipative with respect to $\|\cdot\|$. In this subsection we will replace the defining property (1.2) of the class $\mathscr{F}$ by equivalent conditions, which are easier to verify.

Suppose that an integer $s \geq 1$ and a norm $\|\cdot\|$ on $\mathbb{R}^s$ are given. For arbitrary $\tau \in \mathbb{R}\backslash\{0\}$ and $x, y \in \mathbb{R}^s$ we define

$$m_\tau[x, y] = \tau^{-1}[\|x + \tau y\| - \|x\|].$$

It is easy to verify (cf. e.g. [28, p. 37]) that

(5.1) $m_\tau[x, y] \le m_\sigma[x, y]$ for all $x, y \in \mathbb{R}^s$ and all $\tau, \sigma \in \mathbb{R}\backslash\{0\}$ with $\tau < \sigma$.

This guarantees the existence of the one-sided limits

$$m_+[x, y] = \lim_{\tau \downarrow 0} m_\tau[x, y], \quad m_-[x, y] = \lim_{\tau \uparrow 0} m_\tau[x, y]$$

for all $x, y \in \mathbb{R}^s$. Now we can formulate the following well-known theorem, which can be proved by using the material in [28].

THEOREM 5.1. *Suppose that $s \ge 1$ is an integer, $\|\cdot\|$ a norm on $\mathbb{R}^s$ and f a continuous function from $\mathbb{R} \times \mathbb{R}^s$ into $\mathbb{R}^s$. Then the following four propositions are equivalent.*

(5.2) $(f, \|\cdot\|) \in \mathscr{F}$;

(5.3) $m_\tau[\tilde{x} - x, f(t, \tilde{x}) - f(t, x)] \le 0$ *for all* $\tau < 0, t \in \mathbb{R}$ *and* $x, \tilde{x} \in \mathbb{R}^s$;

(5.4) $m_-[\tilde{x} - x, f(t, \tilde{x}) - f(t, x)] \le 0$ *for all* $t \in \mathbb{R}$ *and* $x, \tilde{x} \in \mathbb{R}^s$;

(5.5) $m_+[\tilde{x} - x, f(t, \tilde{x}) - f(t, x)] \le 0$ *for all* $t \in \mathbb{R}$ *and* $x, \tilde{x} \in \mathbb{R}^s$.

Note that it follows from (5.1) that $m_+[x, y] \le m_\tau[x, y]$ for all $\tau > 0$ and all $x, y \in \mathbb{R}^s$. Hence the condition

(5.6) $m_\tau[\tilde{x} - x, f(t, \tilde{x}) - f(t, x)] \le 0$ for all $t \in \mathbb{R}$ and $x, \tilde{x} \in \mathbb{R}^s$

is at least as strong as (5.5) for any fixed $\tau > 0$. Introducing $\rho = \tau^{-1}$ we can reformulate (5.6) as a so-called *circle condition* (cf. [7], [33], [39], [30], [31]), i.e.

(5.7) $\|f(t, \tilde{x}) - f(t, x) + \rho(\tilde{x} - x)\| \le \rho\|\tilde{x} - x\|$ for all $t \in \mathbb{R}$ and $x, \tilde{x} \in \mathbb{R}^s$.

DEFINITION 5.2. For given $\rho \in (0, \infty)$ we define $\mathscr{F}(\rho) \subset \mathscr{F}$ as the class of all pairs $(f, \|\cdot\|)$ satisfying (5.7), where f is a continuous function from $\mathbb{R} \times \mathbb{R}^s$ into $\mathbb{R}^s$, $s \ge 1$ and $\|\cdot\|$ is a norm on $\mathbb{R}^s$.

Contractivity properties of Runge-Kutta methods on the class $\mathscr{F}(\rho)$ will be studied in this section. In Section 6 we will consider the class $\mathscr{F}$.

5.2. *Main theorem on conditional contractivity.*

In this subsection we present necessary and sufficient conditions on (A, b) to be contractive on $\mathscr{F}(\rho)$ under a step size restriction $h \le H$.

DEFINITION 5.3. A coefficient scheme (A, b) is said to be *contractive* for the step size h and the pair $(f, \|\cdot\|)$ if (1.5) holds whenever (1.3) and (1.4) are fulfilled.

For each integer $s \ge 1$ the *maximum norm* on $\mathbb{R}^s$ is denoted by $\|\cdot\|_\infty$.

THEOREM 5.4. *Let ρ, $H \in (0, \infty)$ be given. Then, for any irreducible coefficient scheme (A, b), the following three propositions are equivalent.*

(P1) $R(A, b) \geq \rho H$;

(P2) *(A, b) is contractive for all step sizes $h \leq H$ and all pairs $(f, \|\cdot\|) \in \mathscr{F}(\rho)$;*

(P3) *(A, b) is contractive for all step sizes $h \leq H$ and all pairs $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ with $\|\cdot\| = \|\cdot\|_\infty$ and with a function f not depending on t.*

The proof of the above theorem follows from the implications (P1) $\Rightarrow$ (P2), (P2) $\Rightarrow$ (P3) and (P3) $\Rightarrow$ (P1). The first implication will be proved in Subsection 5.3, the second is trivial and the third will be proved in Subsection 5.4.

REMARK 5.5. We emphasize that Lemma 4.4 provides a simple algebraic characterization of all irreducible coefficient schemes with property (P1).

REMARK 5.6. In [7] and [8] (see also [23] for an extension to irreducible schemes with $c_i = c_j$ for some $i \neq j$) Dahlquist and Jeltsch studied property (P2), confining themselves to the case where the norms are generated by an inner product. They arrived at a criterion weaker than (P1), namely that the K-function must satisfy $|K(Z)| \leq 1$ for all $Z = \text{diag}(z_1, z_2, \ldots, z_m)$ with $z_j \in \mathbb{C}$, $|z_j + \rho H| \leq \rho H$ and $I - AZ$ nonsingular. For an algebraic characterization of this property we refer to loc. cit.

REMARK 5.7. In [34] Spijker studied properties (P2) and (P3), confining himself to the case in which f has the form $f(t, x) \equiv Lx$, where L is a square matrix. He proved that both properties are equivalent to a property weaker than (P1), viz. absolute monotonicity of the stability function φ on the interval $[-\rho H, 0]$. See also [15], [21], [22], [24], [33], [35], [36], [37].

REMARK 5.8. Note that it follows from the above theorem that an irreducible method is conditionally contractive on a given class $\mathscr{F}(\rho)$ if and only if $R(A, b) > 0$. In view of Theorem 4.2 this is equivalent to the conditions $A \geq 0$, $b > 0$ and $\text{Inc}(A^2) \leq \text{Inc}(A)$. It is interesting to note that in the framework of inner product norms (cf. Remark 5.6) Dahlquist and Jeltsch [7] arrived at the weaker criterion $b > 0$.

5.3. *Absolute monotonicity implies contractivity.*

In this subsection we prove the implication (P1) $\Rightarrow$ (P2) of Theorem 5.4. We extend the notation of the previous sections by writing $[x_i]$ for the vector $(x_1^{\mathrm{T}}, x_2^{\mathrm{T}}, \ldots, x_m^{\mathrm{T}})^{\mathrm{T}}$ whenever $x_1, x_2, \ldots, x_m$ are given vectors in $\mathbb{R}^k$ for some $k \geq 1$.

Suppose that (P1) holds and let $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ and $h \leq H$ be given. Define $r = \rho h$.

Since $R(A, b) \geq r$, it follows from Lemma 4.4 that (A, b) is absolutely monotonic at $\xi = -r$ and $A \geq 0$. Now assume that relations (1.3) and (1.4) are fulfilled. Subtracting (1.3) from (1.4) we obtain

$$d_n = d_{n-1} + \sum_{j=1}^{m} b_j w_j, \tag{5.8a}$$

$$v_i = d_{n-1} + \sum_{j=1}^{m} a_{ij} w_j \quad (1 \leq i \leq m), \tag{5.8b}$$

where $d_n = \tilde{u}_n - u_n$, $d_{n-1} = \tilde{u}_{n-1} - u_{n-1}$, $v_i = \tilde{y}_i - y_i$ and $w_i = hf(t_{n-1} + c_i h, \tilde{y}_i) - hf(t_{n-1} + c_i h, y_i)$. From $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ we have (cf. Definition 5.2)

$$\|w_i + rv_i\| \leq r\|v_i\| \quad (1 \leq i \leq m). \tag{5.9}$$

Introducing $v = [v_i] \in \mathbb{R}^{ms}$ and $w = [w_i] \in \mathbb{R}^{ms}$ we can rewrite (5.8) as (cf. Subsection 2.1 for notation)

$$d_n = d_{n-1} + \mathbb{b}^T w, \tag{5.10a}$$

$$v = e \otimes d_{n-1} + \mathbb{A} w. \tag{5.10b}$$

From (5.10b) it follows that

$$(\mathbb{I} + r\mathbb{A})v = e \otimes d_{n-1} + \mathbb{A}(w + rv).$$

In view of the nonsingularity of $I + rA$ this implies

$$v = ((I + rA)^{-1}e) \otimes d_{n-1} + \mathbb{A}(\mathbb{I} + r\mathbb{A})^{-1}(w + rv). \tag{5.11}$$

Consequently, using $(I + rA)^{-1}e \geq 0$, $A(I + rA)^{-1} \geq 0$ and (5.9),

$$[\|v_i\|] \leq \|d_{n-1}\|(I + rA)^{-1}e + A(I + rA)^{-1}[\|w_i + rv_i\|] \leq$$
$$\leq \|d_{n-1}\|(I + rA)^{-1}e + rA(I + rA)^{-1}[\|v_i\|], \quad \text{i.e.}$$
$$(I + rA)^{-1}[\|v_i\|] \leq \|d_{n-1}\|(I + rA)^{-1}e.$$

Since $A \geq 0$ implies $I + rA \geq 0$, we have $[\|v_i\|] \leq \|d_{n-1}\| e$. Hence we have proved

$$\|v_i\| \leq \|d_{n-1}\| \quad (1 \leq i \leq m). \tag{5.12}$$

Further, (5.10a) and (5.11) yield

$$d_n = d_{n-1} + \mathbb{b}^T w = d_{n-1} - r\mathbb{b}^T v + \mathbb{b}^T(w + rv) =$$
$$= d_{n-1} - r\mathbb{b}^T\{((I + rA)^{-1}e) \otimes d_{n-1} + \mathbb{A}(\mathbb{I} + r\mathbb{A})^{-1}(w + rv)\} + \mathbb{b}^T(w + rv) =$$
$$= (1 - rb^T(I + rA)^{-1}e)d_{n-1} + \mathbb{b}^T(\mathbb{I} + r\mathbb{A})^{-1}(w + rv).$$

In view of $\varphi(-r) \geq 0, b^T(I + rA)^{-1} \geq 0$, (5.9) and (5.12) this implies

$$\|d_n\| \leq (1 - rb^T(I + rA)^{-1}e)\|d_{n-1}\| + b^T(I + rA)^{-1}[\|w_i + rv_i\|] \leq$$
$$\leq (1 - rb^T(I + rA)^{-1}e)\|d_{n-1}\| + rb^T(I + rA)^{-1}[\|v_i\|] \leq$$
$$\leq (1 - rb^T(I + rA)^{-1}e)\|d_{n-1}\| + (rb^T(I + rA)^{-1}e)\|d_{n-1}\| = \|d_{n-1}\|.$$

Thus we have shown contractivity (1.5) and the proof of (P1) ⇒ (P2) is complete.

5.4. *Contractivity implies absolute monotonicity.*

In this subsection we prove the implication (P3) ⇒ (P1) of Theorem 5.4. For any real $s \times s$ matrix $F = (f_{ij})$ we define the matrix norm $\|F\|_\infty$ by $\|F\|_\infty = \max\{\|Fx\|_\infty \,|\, x \in \mathbb{R}^s, \|x\|_\infty = 1\}$. It is well known that

$$\|F\|_\infty = \max_i \sum_j |f_{ij}|. \tag{5.13}$$

Further, for any $r \in (0, \infty)$, the set $\mathscr{D}(r)$ is defined as the set containing all block diagonal matrices $\mathbb{Z} = \operatorname{diag}(Z_1, Z_2, \ldots, Z_m)$, where the blocks Z_i are real square matrices of the same (but arbitrary) order $s \geq 1$, such that $\mathbb{I} - \mathbb{A}\mathbb{Z}$ is nonsingular and $\|Z_i + rI_s\|_\infty \leq r$ $(1 \leq i \leq m)$.

In order to show (P3) ⇒ (P1) we start with a lemma which has the same flavour as [18, Lemma 3.2] and [23, Lemma 3.6]. Roughly speaking, this lemma states that any matrix $\mathbb{K}(\mathbb{Z})$ with $\mathbb{Z} \in \mathscr{D}(r)$ will occur as 'error propagation matrix' in the numerical solution of a suitably chosen nonlinear autonomous system of differential equations (1.1a).

LEMMA 5.9. *Suppose (A, b) is an irreducible m-stage coefficient scheme. Let $r \in (0, \infty)$, an integer $s \geq 1$, an $ms \times ms$ matrix $\mathbb{Z} \in \mathscr{D}(r)$ and vectors $u_0, \tilde{u}_0 \in \mathbb{R}^s$ be given with $u_0 \neq \tilde{u}_0$. Then there exist vectors $y_i, \tilde{y}_i$ $(1 \leq i \leq m)$ and a mapping g: $\mathbb{R}^s \to \mathbb{R}^s$ such that*

$$y_i = u_0 + \sum_j a_{ij} g(y_j), \quad \tilde{y}_i = \tilde{u}_0 + \sum_j a_{ij} g(\tilde{y}_j) \quad (1 \leq i \leq m), \tag{5.14}$$

$$\|g(\tilde{x}) - g(x) + r(\tilde{x} - x)\|_\infty \leq r\|\tilde{x} - x\|_\infty \quad (\textit{for all } x, \tilde{x} \in \mathbb{R}^s), \tag{5.15}$$

and such that the vectors u_1 and $\tilde{u}_1$, defined by

$$u_1 = u_0 + \sum_j b_j g(y_j), \quad \tilde{u}_1 = \tilde{u}_0 + \sum_j b_j g(\tilde{y}_j), \tag{5.16}$$

satisfy

$$\tilde{u}_1 - u_1 = \mathbb{K}(\mathbb{Z})(\tilde{u}_0 - u_0). \tag{5.17}$$

PROOF. Define $v = [v_i] \in \mathbb{R}^{ms}$ by $v = (\mathbb{I} - \mathbb{A}\mathbb{Z})^{-1}(e \otimes (\tilde{u}_0 - u_0))$, i.e.

$$v_i = \tilde{u}_0 - u_0 + \sum_j a_{ij} Z_j v_j \quad (1 \le i \le m).$$

Let $\varepsilon > 0$ be defined as $\varepsilon = \max_i \|v_i\|_\infty$. Since (A, b) is irreducible, it follows from Lemma 2.8 with $\gamma = (-2r)^{-1}$ that there exist vectors $p = (p_1, p_2, \ldots, p_m)^T$ and $q = (q_1, q_2, \ldots, q_m)^T$ in $\mathbb{R}^m$ such that $q = Ap$ and

$$q_i \ne q_j, \quad |(p_i - p_j)/(q_i - q_j) + r| < r \quad (\text{for all } i, j \text{ with } i \ne j).$$

Let d be any vector in $\mathbb{R}^s$ with $\|d\|_\infty = 1$. We define the vectors $y = [y_i]$, $\tilde{y} = [\tilde{y}_i]$ and $w = [w_i]$ in $\mathbb{R}^{ms}$ by

$$y = e \otimes u_0 + \lambda q \otimes d, \quad \tilde{y} = y + v, \quad w = \lambda p \otimes d,$$

and the balls B_i in $\mathbb{R}^s$ by

$$B_i = \{x \mid x \in \mathbb{R}^s, \quad \|x - y_i\|_\infty \le \varepsilon\} \quad (1 \le i \le m),$$

where $\lambda \in (0, \infty)$ is so large that these balls are pairwise disjoint and that

$$\lambda \cdot \min_{i \ne j} |q_i - q_j| \{r - |(p_i - p_j)/(q_i - q_j) + r|\} \ge 4\varepsilon r.$$

Further, we define a mapping g from $V = B_1 \cup B_2 \cup \ldots \cup B_m$ into $\mathbb{R}^s$ by

$$g(x) = Z_i(x - y_i) + w_i \quad (\text{for all } i \text{ and all } x \in B_i).$$

From the above definitions it immediately follows that $y_i, \tilde{y}_i \in B_i$ for all i, so that $g(y_i)$ and $g(\tilde{y}_i)$ are well defined. A straightforward calculation shows that (5.14) is fulfilled, that the vectors u_1 and $\tilde{u}_1$, defined in (5.16), satisfy (5.17), and that

$$\|g(\tilde{x}) - g(x) + r(\tilde{x} - x)\|_\infty \le r\|\tilde{x} - x\|_\infty \quad (\text{for all } x, \tilde{x} \in V). \tag{5.18}$$

The proof of the lemma is completed by noting that the domain of g can be extended from V to all of $\mathbb{R}^s$ in such a way that (5.18) becomes (5.15) (see e.g. [32, Corollary 3.9] or [40, p. 145]). ■

Our next lemma is a result on the maximum value of the matrix norm $\|\mathbb{K}(\mathbb{Z})\|_\infty$ for $\mathbb{Z} \in \mathscr{D}(r)$.

LEMMA 5.10. *Let (A, b) be an arbitrary m-stage coefficient scheme and $r \in (0, \infty)$. Suppose that $I + rA$ is nonsingular. Then*

$$\sup\{\|\mathbb{K}(\mathbb{Z})\|_\infty \mid \mathbb{Z} \in \mathscr{D}(r)\} \ge$$

$$|\varphi(-r)| + \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k} |\beta_{i_1}(-r)\alpha_{i_1 i_2}(-r) \ldots \alpha_{i_{k-1} i_k}(-r)\varepsilon_{i_k}(-r)| r^k,$$

where the functions α_{ij}, β_i and ε_i are defined in (2.4).

PROOF. Let $n \geq 1$ be an arbitrary integer. We will construct a directed tree Γ_n with $s = 1 + m + m^2 + \ldots + m^n$ vertices $P_1, P_2, \ldots, P_s$ and with each edge $P_i \to P_j$ labelled with one of the labels $\omega_1, \omega_2, \ldots, \omega_m$. The construction starts with the vertex P_1, the root of Γ_n, and proceeds in n steps to the graph Γ_n. In the first step we add the m vertices $P_2, P_3, \ldots, P_{m+1}$ and the m edges $P_1 \to P_2$, $P_1 \to P_3, \ldots, P_1 \to P_{m+1}$, labelled with $\omega_1, \omega_2, \ldots, \omega_m$, respectively. We call these m new vertices the vertices at level one. In the second step of the process we attach to each vertex at level one, m new edges, labelled with $\omega_1, \omega_2, \ldots, \omega_m$, and pointing to m new vertices. This leads to m^2 new vertices, the vertices at level two. Proceeding in this way we finally add m^n new vertices and the same number of edges in the n-th step to obtain Γ_n (see Figure 1).

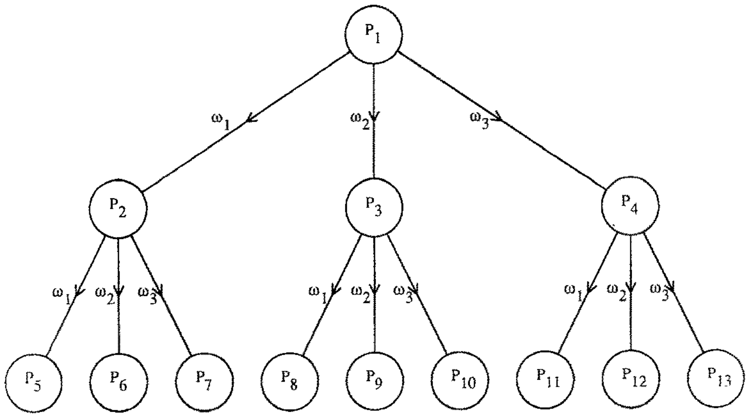

Fig. 1. The graph Γ_2 if $m = 3$.

For $i = 1, 2, \ldots, s$ the ith standard basis vector in $\mathbb{R}^s$ is denoted by e_i. We define the $s \times s$ matrices $W_1, W_2, \ldots, W_m$ by

$$(5.19) \qquad W_l = \sum_{i,j:\,\text{edge } P_i \to P_j \text{ occurs in } \Gamma_n \text{ with label } \omega_l} r e_i e_j^T$$

for all $l = 1, 2, \ldots, m$. One easily verifies that

(5.20a) $\|W_l\|_\infty = r \quad$ for all $l \in \mathcal{M}$,

(5.20b) $W_{i_1} W_{i_2} \ldots W_{i_{n+1}} = 0 \quad$ for all $i_1, i_2, \ldots, i_{n+1} \in \mathcal{M}$,

(5.20c) $e_1^T W_{i_1} W_{i_2} \ldots W_{i_k} = r^k e_j^T \quad$ for all $k = 1, 2, \ldots, n$ and all $i_1, i_2, \ldots, i_k$, where j is the index of the vertex P_j you arrive at when you start in P_1 and follow the path labelled $\omega_{i_1}, \omega_{i_2}, \ldots, \omega_{i_k}$.

If we define the block diagonal matrix $\mathbb{Z} = \mathrm{diag}(Z_1, Z_2, \ldots, Z_m)$ by $\mathbb{Z} = -r\mathbb{I} + \mathbb{W}$, where $\mathbb{W}$ is the block diagonal matrix $\mathbb{W} = \mathrm{diag}(W_1, W_2, \ldots, W_m)$, it follows from the nonsingularity of $I + rA$ and from (5.20b) that $\mathbb{I} - \mathbb{A}\mathbb{Z}$ is nonsingular with inverse given by (2.5) (Take $\xi = -r$ here and note that the series in (2.5) is in fact finite with k running from 0 to n). Hence, in combination with (5.20a), we have proved that $\mathbb{Z} \in \mathscr{D}(r)$. Further it is easy to see now that $\mathbb{K}(\mathbb{Z})$ is given by (2.6) (Take again $\xi = -r$ and note that only values $k \leq n$ are relevant), so that the first row of $\mathbb{K}(\mathbb{Z})$ is given by

$$e_1^{\mathrm{T}}\mathbb{K}(\mathbb{Z}) = \varphi(-r)e_1^{\mathrm{T}} + \sum_{k=1}^{n} \sum_{i_1, i_2, \ldots, i_k} \beta_{i_1}\alpha_{i_1 i_2} \ldots \alpha_{i_{k-1} i_k}\varepsilon_{i_k} e_1^{\mathrm{T}} W_{i_1} W_{i_2} \ldots W_{i_k},$$

where the entries α_{ij}, β_i and ε_i are evaluated at $\xi = -r$. Since different choices of $k = 1, 2, \ldots, n$ and $i_1, i_2, \ldots, i_k \in \mathscr{M}$ lead to different values of the index j in (5.20c), it follows from (5.13), (5.20c) and the above expression for $e_1^{\mathrm{T}}\mathbb{K}(\mathbb{Z})$ that

$$\|\mathbb{K}(\mathbb{Z})\|_\infty \geq |\varphi(-r)| + \sum_{k=1}^{n} \sum_{i_1, i_2, \ldots, i_k} |\beta_{i_1}\alpha_{i_1 i_2} \ldots \alpha_{i_{k-1} i_k}\varepsilon_{i_k}| r^k.$$

The fact that $n \geq 1$ was arbitrarily chosen concludes the proof. ■

With the help of Lemmas 5.9 and 5.10 we will prove (P3) ⇒ (P1).

We assume that (P3) holds. Let r be an arbitrary number in $(0, \rho H]$ such that $I + rA$ is nonsingular. Further, let an integer $s \geq 1$, an $ms \times ms$ matrix $\mathbb{Z} \in \mathscr{D}(r)$ and vectors $u_0, \tilde{u}_0 \in \mathbb{R}^s$ be given with $u_0 \neq \tilde{u}_0$. We define the mapping g and the vectors $u_1, \tilde{u}_1, y_i, \tilde{y}_i$ as in Lemma 5.9. One easily verifies that relations (1.3) and (1.4) are fulfilled if we take $n = 1$, $h = r/\rho$, $t_0 = 0$ and f defined by

$$f(t, x) = h^{-1}g(x) \quad (\text{for all } t \in \mathbb{R} \text{ and } x \in \mathbb{R}^s).$$

Since it follows from (5.15) that $(f, \|\cdot\|_\infty) \in \mathscr{F}(\rho)$ and from $r \in (0, \rho H]$ that $h \leq H$, an application of (P3) yields $\|\tilde{u}_1 - u_1\|_\infty \leq \|\tilde{u}_0 - u_0\|_\infty$. In view of (5.17) and the fact that u_0 and $\tilde{u}_0$ were arbitrarily chosen (with $u_0 \neq \tilde{u}_0$), we may conclude that $\|\mathbb{K}(\mathbb{Z})\|_\infty \leq 1$. Since also $\mathbb{Z} \in \mathscr{D}(r)$ was arbitrarily chosen, it follows from Lemma 5.10 that

$$(5.21) \quad |\varphi(-r)| + \sum_{k=1}^{\infty} \sum_{i_1, i_2, \ldots, i_k} |\beta_{i_1}(-r)\alpha_{i_1 i_2}(-r) \ldots \alpha_{i_{k-1} i_k}(-r)\varepsilon_{i_k}(-r)| r^k \leq 1.$$

Further, by taking $\xi = -r$ and scalar $W_1 = W_2 = \ldots = W_m$ in (2.6) it follows that the Taylor series of φ about $z = -r$ is given by $\sum_{k=0}^{\infty} \gamma_k (z + r)^k$, where

$$(5.22a) \qquad \gamma_0 = \varphi(-r),$$

$$(5.22b) \qquad \gamma_k = \sum_{i_1, i_2, \ldots, i_k} \beta_{i_1}(-r)\alpha_{i_1 i_2}(-r) \ldots \alpha_{i_{k-1} i_k}(-r)\varepsilon_{i_k}(-r) \quad (\text{for all } k \geq 1).$$

In view of (5.21) we have $\sum_{k=0}^{\infty} |\gamma_k| r^k \leq 1$, so that the radius of convergence of the Taylor series is larger than r. Thus we have

$$\sum_{k=0}^{\infty} \gamma_k r^k = \varphi(0) = 1.$$

In combination with (5.21) and (5.22) this implies that $\varphi(-r)$ and all terms of the sum in the right hand side of (5.22b) are nonnegative (for all $k \geq 1$), i.e. we have proved absolute monotonicity of $\mathbb{K}$ at $\xi = -r$ (cf. Definition 2.3). By Lemma 2.5, K is absolutely monotonic at $\xi = -r$ as well. Since r was an arbitrary number in $(0, \rho H]$ such that $I + rA$ is nonsingular, it follows from Lemma 3.7 that K is absolutely monotonic on $[-\rho H, 0)$. Using Lemma 3.2 we see that K is absolutely monotonic on $[-\rho H, 0]$. An application of Theorem 3.8 shows that $R(A, b) \geq \rho H$, i.e. (P1) holds. This completes the proof of (P3) $\Rightarrow$ (P1).

6. Unconditional contractivity.

In this section we will give necessary and sufficient conditions on (A, b) to be contractive on $\mathscr{F}$ for all step sizes $h > 0$.

THEOREM 6.1. *For any irreducible coefficient scheme (A, b) the following three propositions are equivalent.*

(Q1) $R(A, b) = \infty$;

(Q2) *(A, b) is contractive for all step sizes $h > 0$ and all pairs $(f, \|\cdot\|) \in \mathscr{F}$;*

(Q3) *(A, b) is contractive for all step sizes $h > 0$ and all pairs $(f, \|\cdot\|) \in \mathscr{F}$ with $\|\cdot\| = \|\cdot\|_\infty$ and with a function f not depending on t.*

PROOF. The implication (Q2) $\Rightarrow$ (Q3) is trivial. To prove the implication (Q3) $\Rightarrow$ (Q1), we note that (Q3) implies (P3) for all $\rho, H \in (0, \infty)$. The implication (P3) $\Rightarrow$ (P1) in Theorem 5.4 then establishes (Q3) $\Rightarrow$ (Q1).

We complete the proof of the theorem by showing (Q1) $\Rightarrow$ (Q2). Assume that (Q1) holds. It follows from Theorem 4.7 that A is nonsingular, A^{-1} is an M-matrix, $A^{-1}e \geq 0$, $b^{\mathrm{T}}A^{-1} \geq 0$ and $b^{\mathrm{T}}A^{-1}e \leq 1$. We choose a real number $\lambda > 0$, which is so large that

$$\lambda I - A^{-1} \geq 0. \tag{6.1}$$

To prove (Q2) we assume that $(f, \|\cdot\|) \in \mathscr{F}$ and $h > 0$ are given, and that relations (1.3) and (1.4) are fulfilled. Using the same notation as in Subsection 5.3 we see that (5.10) holds. In view of Theorem 5.1 we have (5.3), so that

$$\|w_i - \lambda v_i\| \geq \lambda \|v_i\| \quad (1 \leq i \leq m). \tag{6.2}$$

From (5.10b) it follows that

$$w - \lambda v = -(A^{-1}e) \otimes d_{n-1} - (\lambda \mathbb{I} - \mathbb{A}^{-1})v.$$

In combination with (6.2), $A^{-1}e \geq 0$ and (6.1) this implies

$$\lambda[\|v_i\|] \leq \|d_{n-1}\| A^{-1}e + (\lambda I - A^{-1})[\|v_i\|], \text{ i.e.}$$

$$A^{-1}[\|v_i\|] \leq \|d_{n-1}\| A^{-1}e.$$

Since $A \geq 0$ we have $[\|v_i\|] \leq \|d_{n-1}\|e$. Hence we have proved

$$\|v_i\| \leq \|d_{n-1}\| \quad (1 \leq i \leq m). \tag{6.3}$$

Further, (5.10) yields

$$d_n = d_{n-1} + \mathbb{b}^{\mathrm{T}}(\mathbb{A}^{-1}v - (A^{-1}e) \otimes d_{n-1}) = (1 - b^{\mathrm{T}}A^{-1}e)d_{n-1} + \mathbb{b}^{\mathrm{T}}\mathbb{A}^{-1}v.$$

Using $b^{\mathrm{T}}A^{-1}e \leq 1, b^{\mathrm{T}}A^{-1} \geq 0$ and (6.3) we obtain

$$\|d_n\| \leq (1 - b^{\mathrm{T}}A^{-1}e)\,\|d_{n-1}\| + b^{\mathrm{T}}A^{-1}[\|v_i\|] \leq$$
$$\leq (1 - b^{\mathrm{T}}A^{-1}e)\,\|d_{n-1}\| + (b^{\mathrm{T}}A^{-1}e)\,\|d_{n-1}\| = \|d_{n-1}\|.$$

Thus we have shown contractivity (1.5) and the proof of (Q1) ⇒ (Q2) is complete. ∎

REMARK 6.2. We emphasize that Theorem 4.7 provides a simple algebraic characterization of all irreducible coefficient schemes with property (Q1).

REMARK 6.3. Burrage and Butcher [3] and Crouzeix [6] (see also [18] for an extension to irreducible schemes with $c_i = c_j$ for some $i \neq j$) studied property (Q2), confining themselves to the case where the norms are generated by an inner product. They arrived at a criterion weaker than (Q1), namely that the K-function must satisfy $|K(Z)| \leq 1$ for all $Z = \mathrm{diag}(z_1, z_2, \ldots, z_m)$ with $z_j \in \mathbb{C}, \mathrm{Re}(z_j) \leq 0$ and $I - AZ$ nonsingular (cf. also Remark 5.6). For an algebraic characterization of this property we refer to loc. cit.

REMARK 6.4. Spijker [34] studied properties (Q2) and (Q3), confining himself to the case in which f has the form $f(t, x) \equiv Lx$, where L is a square matrix. He proved that both properties are equivalent to a property weaker than (Q1), viz. absolute monotonicity of the stability function φ on the interval $(-\infty, 0]$ (cf. also Remark 5.7).

7. Solvability of the systems of equations.

In Theorems 5.4 and 6.1 we presented necessary and sufficient conditions for conditional contractivity on $\mathscr{F}(\rho)$ and unconditional contractivity on $\mathscr{F}$, respectively. It should be realized that the property of contractivity, as defined in Definition 5.3, does *not* comprise the (unique) solvability of the systems of equations (1.3b) and (1.4b). Contractivity only means that (1.5) holds *whenever* (1.3) and (1.4) are fulfilled. Obviously, the mere property of contractivity, without solvability of the systems of equations, is of little value. It is therefore a lucky circumstance that the conditions which are necessary and sufficient for contractivity, turn out to be sufficient for unique solvability of the systems of equations.

THEOREM 7.1. *Let (A, b) be a given irreducible coefficient scheme.*
(a) *If $\rho, h \in (0, \infty)$ and $R(A, b) \geq \rho h$, then (1.3b) has a unique solution $y = [y_i]$ whenever $(f, \|\cdot\|) \in \mathscr{F}(\rho)$;*
(b) *If $R(A, b) = \infty$, then (1.3b) has a unique solution $y = [y_i]$ whenever $(f, \|\cdot\|) \in \mathscr{F}$ and $h > 0$.*

PROOF. (a) Let $\rho, h \in (0, \infty)$, $(f, \|\cdot\|) \in \mathscr{F}(\rho)$, $t_{n-1} \in \mathbb{R}$ and $u_{n-1} \in \mathbb{R}^s$ be given and assume $R(A, b) \geq \rho h$. We will prove that (1.3b) has a unique solution $y = [y_i] \in \mathbb{R}^{ms}$ by showing that $G(y) = 0$ has a unique solution. Here G: $\mathbb{R}^{ms} \to \mathbb{R}^{ms}$ is defined by

$$G(y) = [G_i(y)] = y - e \otimes u_{n-1} - \mathbb{A}F(y) \quad (\text{for all } y = [y_i] \in \mathbb{R}^{ms}),$$

where F: $\mathbb{R}^{ms} \to \mathbb{R}^{ms}$ is defined by

$$F(y) = [hf(t_{n-1} + c_i h, y_i)] \qquad (\text{for all } y = [y_i] \in \mathbb{R}^{ms}).$$

Suppose that $y = [y_i]$, $\hat{y} = [\hat{y}_i]$ and $d = [d_i]$ are given in $\mathbb{R}^{ms}$ such that

$$G(\hat{y}) - G(y) = d.$$

Introducing $v = [v_i] = \hat{y} - y$ and $w = [w_i] = F(\hat{y}) - F(y)$, we find

$$d = v - \mathbb{A}w = (\mathbb{I} + r\mathbb{A})v - \mathbb{A}(w + rv), \tag{7.1}$$

where r is defined as $r = \rho h$. As in the proof given in Subsection 5.3, we may assume that (5.9) holds, that (A, b) is absolutely monotonic at $\xi = -r$ and that $A \geq 0$. By (7.1) and the nonsingularity of $I + rA$,

$$v = (\mathbb{I} + r\mathbb{A})^{-1}d + \mathbb{A}(\mathbb{I} + r\mathbb{A})^{-1}(w + rv).$$

In view of $A(I + rA)^{-1} \geq 0$ and (5.9) this implies

$$[\|v_i\|] \leq |(I + rA)^{-1}| [\|d_i\|] + rA(I + rA)^{-1}[\|v_i\|], \quad \text{i.e.}$$

$$(I + rA)^{-1}[\|v_i\|] \leq |(I + rA)^{-1}| [\|d_i\|].$$

Using $I + rA \geq 0$ we obtain

$$[\|v_i\|] \leq (I + rA) |(I + rA)^{-1}| [\|d_i\|]. \tag{7.2}$$

Hence we have proved that for all $y, \hat{y} \in \mathbb{R}^{ms}$ we have

$$|||\hat{y} - y||| \leq \gamma |||G(\hat{y}) - G(y)|||, \tag{7.3}$$

where $\gamma = \|(I + rA) |(I + rA)^{-1}| \|_\infty$ (cf. (5.13)) and where the norm $|||\cdot|||$ on $\mathbb{R}^{ms}$ is defined by $|||z||| = \max_{1 \leq i \leq m} \|z_i\|$ for all $z = [z_i] \in \mathbb{R}^{ms}$. It follows from [19, Lemma 4.2] (take $\phi(z; t) \equiv \gamma t$ here) that $G(y) = 0$ has a unique solution $y \in \mathbb{R}^{ms}$. This completes the proof of part (a).

(b) Let $(f, \|\cdot\|) \in \mathscr{F}$, $h > 0$, $t_{n-1} \in \mathbb{R}$ and $u_{n-1} \in \mathbb{R}^s$ be given and assume $R(A, b) = \infty$. Suppose that y, $\hat{y}$, d, v, w and G are defined as above. As in the proof of Theorem 6.1 we may assume that A is nonsingular, that $A \geq 0$ and that (6.1) and (6.2) hold for some real number $\lambda > 0$. Since (7.1) holds with $r = -\lambda$, it follows from the nonsingularity of A that

$$w - \lambda v = -\mathbb{A}^{-1}d - (\lambda \mathbb{I} - \mathbb{A}^{-1})v.$$

Using (6.2) and (6.1) we obtain

$$\lambda[\|v_i\|] \leq |A^{-1}|[\|d_i\|] + (\lambda I - A^{-1})[\|v_i\|], \quad \text{i.e.}$$

$$A^{-1}[\|v_i\|] \leq |A^{-1}|[\|d_i\|].$$

But then $A \geq 0$ yields

$$[\|v_i\|] \leq A|A^{-1}|[\|d_i\|], \tag{7.4}$$

so that (7.3) follows with $\gamma = \|A|A^{-1}|\|_\infty$. As in the proof of part (a), this implies that $G(y) = 0$ has a unique solution $y \in \mathbb{R}^{ms}$. This completes the proof of part (b). ■

For a given approximation u_{n-1} it is usually impossible in practical computations to find $u_n, y_1, y_2, \ldots, y_m$ such that the relations (1.3a) and (1.3b) are satisfied exactly. Instead we find $\hat{u}_n, \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_m$ satisfying

$$\hat{u}_n = u_{n-1} + h \sum_{j=1}^{m} b_j f(t_{n-1} + c_j h, \hat{y}_j) + \delta, \tag{7.5a}$$

$$\hat{y}_i = u_{n-1} + h \sum_{j=1}^{m} a_{ij} f(t_{n-1} + c_j h, \hat{y}_j) + d_i \quad (1 \leq i \leq m). \tag{7.5b}$$

It is important to know whether the effect of the residuals δ and d_i is moderate. We are interested in bounds for $\|\hat{y}_i - y_i\|$ and $\|\hat{u}_n - u_n\|$.

THEOREM 7.2. *Let (A, b) be a given irreducible coefficient scheme.*

(a) *Suppose that $\rho, h \in (0, \infty)$ satisfy $\rho h \leq R(A, b)$ and that $(f, \|\cdot\|) \in \mathscr{F}(\rho)$. If (1.3) and (7.5) are fulfilled, then we have the bounds*

$$[\|\hat{y}_i - y_i\|] \leq (I + rA)|(I + rA)^{-1}|[\|d_i\|], \tag{7.6a}$$

$$\|\hat{u}_n - u_n\| \leq \|\delta\| + rb^{\mathrm{T}}\{(I + rA)^{-1} + |(I + rA)^{-1}|\}[\|d_i\|], \tag{7.6b}$$

where $r = \rho h$.

(b) *Suppose that $(f, \|\cdot\|) \in \mathscr{F}$, $h > 0$ and $R(A, b) = \infty$. If (1.3) and (7.5) are fulfilled, then we have the bounds*

$$[\|\hat{y}_i - y_i\|] \leq A|A^{-1}|[\|d_i\|], \tag{7.7a}$$

$$\|\hat{u}_n - u_n\| \leq \|\delta\| + b^{\mathrm{T}}\{A^{-1} + |A^{-1}|\}[\|d_i\|]. \tag{7.7b}$$

PROOF. (a) Inequality (7.6a) follows immediately from (7.2). Further, using the same notation as in the proof of Theorem 7.1, we have

$$\hat{u}_n - u_n = \delta + \mathbb{b}^{\mathrm{T}} w = \delta + \mathbb{b}^{\mathrm{T}}(\mathbb{I} + r\mathbb{A})^{-1}(w + rv) - r\mathbb{b}^{\mathrm{T}}(\mathbb{I} + r\mathbb{A})^{-1}d.$$

In view of $b^{\mathrm{T}}(I + rA)^{-1} \geq 0$, (5.9) and (7.2) this implies

$$\|\hat{u}_n - u_n\| \leq \|\delta\| + rb^{\mathrm{T}}(I + rA)^{-1}[\|v_i\|] + rb^{\mathrm{T}}(I + rA)^{-1}[\|d_i\|] \leq$$

$$\leq \|\delta\| + rb^{\mathrm{T}}\,|(I + rA)^{-1}|\,[\|d_i\|] + rb^{\mathrm{T}}(I + rA)^{-1}[\|d_i\|],$$

which establishes (7.6b).

(b) Inequality (7.7a) follows immediately from (7.4). Further, using the same notation as in the proof of Theorem 7.1, we have

$$\hat{u}_n - u_n = \delta + \mathbb{b}^{\mathrm{T}} w = \delta + \mathbb{b}^{\mathrm{T}}\mathbb{A}^{-1}v - \mathbb{b}^{\mathrm{T}}\mathbb{A}^{-1}d.$$

In view of $b^{\mathrm{T}}A^{-1} \geq 0$ and (7.4) this implies

$$\|\hat{u}_n - u_n\| \leq \|\delta\| + b^{\mathrm{T}}A^{-1}[\|v_i\|] + b^{\mathrm{T}}A^{-1}[\|d_i\|] \leq$$

$$\leq \|\delta\| + b^{\mathrm{T}}\,|A^{-1}|\,[\|d_i\|] + b^{\mathrm{T}}A^{-1}[\|d_i\|],$$

and the proof of the theorem is complete. ■

REMARK 7.3. We mention that the error bounds (7.6) and (7.7) are best possible in the sense that all entries of $(I + rA)\,|(I + rA)^{-1}|$, $rb^{\mathrm{T}}\{(I + rA)^{-1} + |(I + rA)^{-1}|\}$, $A\,|A^{-1}|$ and $b^{\mathrm{T}}\{A^{-1} + |A^{-1}|\}$ are minimal. This can be proved by adapting the results on $\mathbb{K}(\mathbb{Z})$, presented in Lemmas 5.9 and 5.10, to the case of $(\mathbb{I} - \mathbb{A}\mathbb{Z})^{-1}$ and $\mathbb{b}^{\mathrm{T}}\mathbb{Z}(\mathbb{I} - \mathbb{A}\mathbb{Z})^{-1}$. Note that it follows from the sharpness of the bounds (7.6a) and (7.6b), combined with the fact that $\mathscr{F}(\rho) \subset \mathscr{F}(\sigma)$ whenever $0 < \rho \leq \sigma < \infty$, that the entries of $(I + rA)\,|(I + rA)^{-1}|$ and $rb^{\mathrm{T}}\{(I + rA)^{-1} + |(I + rA)^{-1}|\}$ are nondecreasing functions of r. This can also be proved directly, by exploiting the absolute monotonicity of (A, b).

REMARK 7.4. We emphasize that (7.6) and (7.7) are error bounds that hold *uniformly* for all $(f, \|\cdot\|)$ in $\mathscr{F}(\rho)$ or $\mathscr{F}$, respectively. Following the terminology introduced by Frank, Schneid and Ueberhuber in [12] (see also [9]), the error bounds (7.6a) and (7.7a) can be regarded as *BSI*-stability bounds on the classes $\mathscr{F}(\rho)$ and $\mathscr{F}$, respectively, and the error bounds (7.6b) and (7.7b) as *BS*-stability bounds on these classes.

REMARK 7.5. Note that in the situation of part (b) of Theorem 7.2 the matrix A^{-1} is an M-matrix (cf. Theorem 4.7), so that we can write $A^{-1} = \mathrm{diag}(A^{-1}) - E$, where $\mathrm{diag}(A^{-1}) \geq 0$ is the diagonal of A^{-1} and $E \geq 0$ has zero diagonal. Hence the error bounds (7.7a) and (7.7b) can be reformulated by using $A\,|A^{-1}| = 2A\,\mathrm{diag}(A^{-1}) - I$ and $A^{-1} + |A^{-1}| = 2\,\mathrm{diag}(A^{-1})$. Similarly, the error bounds (7.6a) and (7.6b) can be reformulated by using $(I + rA)\,|(I + rA)^{-1}| = 2(I + rA)\,\mathrm{diag}((I + rA)^{-1}) - I$ and

$(I + rA)^{-1} + |(I + rA)^{-1}| = 2\,\mathrm{diag}((I + rA)^{-1})$, since $(I + rA)^{-1}$ is an M-matrix as well (cf. the proof of Lemma 4.4).

8. Orders of accuracy.

8.1. *Classical order p and stage order $\tilde{p}$.*

In this section we consider the impact of the conditions $R(A, b) > 0$ (for conditional contractivity, cf. Theorem 5.4) and $R(A, b) = \infty$ (for unconditional contractivity, cf. Theorem 6.1) on the order of accuracy. We will distinguish between two different orders of accuracy.

The first order of accuracy we will consider is the *classical order of consistency p*, which is defined as the maximum integer such that for all problems (1.1) with a sufficiently smooth function f (all partial derivatives of order at most $p + 1$ must exist and be continuous in a neighbourhood of the solution trajectory), the error after the (fictitious) step $\hat{u}_{n-1} = U(t_{n-1}) \to \hat{u}_n$ can be bounded by

$$\|U(t_n) - \hat{u}_n\| \leq Ch^{p+1} \quad (\text{for all } h \in (0, H]). \tag{8.1}$$

Here the constants C and H may depend on the problem, in fact on the magnitude of the partial derivatives of f. For more details we refer to the books [5], [17] and [38]. In the situation that $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ or $(f, \|\cdot\|) \in \mathscr{F}$, we can combine contractivity (cf. Theorems 5.4 and 6.1) and consistency of order p (cf. (8.1)) in a standard way (cf. e.g. [11, pp. 761–762]) to obtain a bound for the global error,

$$\|U(t_n) - u_n\| \leq (t_n - t_0)Ch^p \quad (\text{for all } h \in (0, H]). \tag{8.2}$$

If the function f is not smooth, the constant C in (8.1) and (8.2) may become very large (and the constant H very small), even if the exact solution U is smooth. In this case the above error bounds are useless, and we have to replace them by robust error bounds, which are not affected by a lack of smoothness of f. Robust error bounds can be obtained by considering the *stage order* $\tilde{p}$ of (A, b), which is defined as the maximum integer l such that $B(l)$ and $C(l)$ hold. Here the so-called simplifying conditions $B(l)$ and $C(l)$ are defined by

$$\sum_{i=1}^{m} b_i c_i^{k-1} = \frac{1}{k} \quad (\text{for all } k = 1, 2, \ldots, l) \quad \text{and}$$

$$\sum_{j=1}^{m} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k \quad (\text{for all } i = 1, 2, \ldots, m \text{ and } k = 1, 2, \ldots, l),$$

respectively. It is well known that $\tilde{p} \leq p$ (cf. e.g. [9]). Combining the results on contractivity (Theorems 5.4 and 6.1) with those on BS-stability (Theorem 7.2) in a standard way (cf. e.g. [9]), we arrive at the following theorem.

THEOREM 8.1. *Let (A, b) be a given irreducible coefficient scheme with stage order $\tilde{p}$.*

(a) *Suppose that $\rho, H \in (0, \infty)$ are such that $\rho H \leq R(A, b)$. Then, for all problems* (1.1) *with $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ and with a $\tilde{p} + 1$ times continuously differentiable solution U, the error bounds* (8.1) *and* (8.2) *hold with p replaced by $\tilde{p}$ and with $C = \alpha\mu$, where α is a constant depending only on (A, b) and μ is an upper bound for $\|U^{(\tilde{p}+1)}(t)\|$ for all $t \geq t_0$.*

(b) *Suppose that* $R(A, b) = \infty$ and $H \in (0, \infty]$. *Then, for all problems* (1.1) *with $(f, \|\cdot\|) \in \mathscr{F}$ and with a $\tilde{p} + 1$ times continuously differentiable solution U, the error bounds* (8.1) *and* (8.2) *hold with p replaced by $\tilde{p}$ and with C as in part* (*a*).

Clearly the concept of stage order fits very well in our framework, and we have derived an error bound (8.2), where the constant C depends only on the smoothness of the exact solution U, and where the maximum step size in case (a) is given by $H = R(A, b)\rho^{-1}$. We emphasize that this error bound does not require any smoothness of the function f, apart from $(f, \|\cdot\|) \in \mathscr{F}(\rho)$ or $(f, \|\cdot\|) \in \mathscr{F}$, and that the bound does not depend on problem dependent quantities like the dimension s or the norm $\|\cdot\|$. Further we mention that in case (a) the step size restriction $h \leq R(A, b)\rho^{-1}$ can become very severe if ρ is large, which is the case, for instance, if (1.1) is stiff.

REMARK 8.2. Following the terminology introduced by Frank, Schneid and Ueberhuber (cf. [13], [11], [9]), the error bounds of Theorem 8.1 can be regarded as (optimal) B-convergence bounds on the classes $\mathscr{F}(\rho)$ and $\mathscr{F}$ (cf. also Remark 7.4).

For unconditional contractivity we mention the following negative result.

THEOREM 8.3. *Let (A, b) be an arbitrary coefficient scheme with $R(A, b) = \infty$. Then $p \leq 1$ (and hence also $\tilde{p} \leq 1$).*

PROOF. If $R(A, b) = \infty$ then it follows from Lemma 2.5 that the stability function φ is absolutely monotonic on $(-\infty, 0]$. Further we have $\varphi(z) = \exp(z) + O(z^{p+1})$ as $z \to 0$ (cf. e.g. [5, pp. 241–242]). It follows from [2, Lemma 2] (see also [34, Theorem 2.5] or [9, Lemma 2.3.6]) that $p \leq 1$. ■

As an illustration to the above theorem we consider the class of methods with tableau

$$\begin{array}{c|c} \theta & \theta \\ \hline & 1 \end{array}$$

where $\theta \in \mathbb{R}$. Using Theorem 4.7 one easily verifies that $R(A, b) = \infty$ if and only if $\theta \geq 1$. In this case we have $p = \tilde{p} = 1$.

In view of the order barrier $\tilde{p} \leq p \leq 1$ for coefficient schemes with $R(A, b) = \infty$, the question arises whether there exist order barriers for coefficient schemes with $R(A, b) > 0$. This question will be answered in Subsections 8.2 and 8.3.

REMARK 8.4. The order barrier $\tilde{p} \le p \le 1$ for unconditional contractivity on $\mathscr{F}$ disappears if we would have restricted ourselves to inner product norms (cf. Remark 6.3). For example, the well-known Gaussian Runge-Kutta method with $m \ge 1$ stages is unconditionally contractive for all $(f, \|\cdot\|) \in \mathscr{F}$ where $\|\cdot\|$ is generated by an inner product, and is known to have stage order $\tilde{p} = m$ and classical order $p = 2m$ (cf. e.g. [9]).

8.2. *Barriers for the stage order $\tilde{p}$ when $R(A,b) > 0$.*

In this subsection we determine the maximum stage order $\tilde{p}$ for coefficient schemes with $R(A,b) > 0$. The following theorem was pointed out to us by J. C. Butcher.

THEOREM 8.5. (*J. C. Butcher; private communication* 1989). *Let (A,b) be an arbitrary coefficient scheme with $A \ge 0$. Then the stage order $\tilde{p}$ is at most 2. Further, if $\tilde{p} = 2$ then A has a zero row.*

PROOF. Suppose (A,b) is such that $A \ge 0$. We shall prove that $\tilde{p} \ge 2$ implies that A has a zero row, and that $\tilde{p} \ge 3$ is impossible.

1. Suppose that $\tilde{p} \ge 2$. Without loss of generality we may assume that $0 \le c_1 \le c_2 \le \ldots \le c_m$. From $C(2)$ we obtain

$$\int_0^{c_i} q(x)\,dx = \sum_{j=1}^{m} a_{ij} q(c_j) \tag{8.3}$$

for all polynomials q of degree at most one and all $i = 1, 2, \ldots, m$. Taking $i = 1$ and $q(x) \equiv x - c_1$ it follows that the right hand side of (8.3) is nonnegative. With this choice the left hand side is nonnegative only if $c_1 = 0$. Hence $c_1 = 0$, so that the first row of A must be zero.

2. Suppose that $\tilde{p} \ge 3$. Then $B(3)$ holds, so that not all c_i are zero. In view of part 1 we may therefore assume that there exists an index $i \ge 2$ such that $0 = c_1 = \ldots = c_{i-1} < c_i \le \ldots \le c_m$. From $C(3)$ it follows that (8.3) holds for all polynomials q of degree at most 2. Taking $q(x) \equiv x(x - c_i)$, the left hand side of (8.3) is seen to be negative, whereas the right hand side is nonnegative. This is a contradiction and the proof of the theorem is complete. ∎

We see immediately from the above theorem that coefficient schemes (A,b) with $R(A,b) > 0$ suffer from an order barrier $\tilde{p} \le 2$. That the order $\tilde{p} = 2$ can be attained, follows from consideration of the family

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2}\theta^{-1} & \frac{1}{4}\theta^{-1} & \frac{1}{4}\theta^{-1} \\ \hline & 1-\theta & \theta \end{array},$$

where $\theta \in (0, 1)$. All these methods have stage order $\tilde{p} = 2$, and it follows from Theorem 4.2 that their radius of absolute monotonicity $R = R(\theta)$ is positive as well. In fact we have $R(\theta) = 4\theta$ (if $0 < \theta \leq 2/3$) and $R(\theta) = 4\theta(1-\theta)(2\theta-1)^{-1}$ (if $2/3 \leq \theta < 1$). The maximum value $R = 8/3$ is attained for $\theta = 2/3$. If we select the value $\theta = 1/2$, we get the familiar trapezoidal rule with $R = 2$. A method with classical order $p = 3$, due to Hammer and Hollingsworth (cf. e.g. [17, pp. 200–201]), is obtained by choosing $\theta = 3/4$, leading to the value $R = 3/2$.

For explicit methods (i.e., methods with $a_{ij} = 0$ when $j \geq i$) we note that their stage order $\tilde{p}$ can never exceed one (cf. e.g. [7, Theorem 4.4]).

8.3. *Barriers for the classical order p when $R(A, b) > 0$.*

In this subsection we derive an upper bound for the classical order p for explicit and implicit methods with $R(A, b) > 0$.

LEMMA 8.6. *Let (A, b) be an arbitrary coefficient scheme with $b > 0$. If the classical order satisfies $p \geq 2k + 1$, for some integer $k \geq 0$, then the stage order $\tilde{p}$ is at least k.*

PROOF. It is well known that if p is the classical order, then $B(p)$ holds (cf. e.g. [4]), and hence also $B(k)$. Further it was proved in [16] that if $b > 0$ and $p \geq 2k + 1$, then $C(k)$ holds. ■

COROLLARY 8.7. *Let (A, b) be an arbitrary coefficient scheme with $A \geq 0$, $b > 0$ and classical order p. Then we have $p \leq 6$ for implicit schemes and $p \leq 4$ for explicit schemes. Further, if $p \geq 5$ then A has a zero row.*

PROOF. For explicit methods the assertion follows from Lemma 8.6 and the fact that the stage order cannot exceed one (see the end of Subsection 8.2). For implicit methods the assertions are established by a combination of Theorem 8.5 and Lemma 8.6. ■

A combination of Corollary 8.7 and Theorem 4.2 shows that the property $R(A, b) > 0$ induces an order barrier $p \leq 4$ for explicit methods, and an order barrier $p \leq 6$ for implicit methods. The bound for explicit methods is sharp, as can be seen from the method

$$
\begin{array}{c|ccccc}
0 & & & & & \\
2/5 & 2/5 & & & & \\
3/5 & 1/10 & 1/2 & & & \\
1/2 & 1/16 & 1/16 & 3/8 & & \\
1 & 1/10 & 1/10 & 4/15 & 8/15 & \\
\hline
 & 5/32 & 25/96 & 25/96 & 1/6 & 5/32
\end{array}.
$$

This method has classical order $p = 4$, while Theorem 4.2 shows that $R(A, b) > 0$. For implicit methods it is not known whether $p = 6$ can be achieved, but $p = 5$ is certainly possible, as is seen from the following method, which was constructed with the help of J. C. Butcher (private communication, 1989),

0	0				
1/4	1/8	1/8			
1/2	1/24	5/12	1/24		
3/4	1/32	7/16	5/32	1/8	
1	3/14	1/7	1/14	4/7	0
	7/90	16/45	2/15	16/45	7/90

It follows from Theorem 4.2 that this method with classical order $p = 5$ satisfies $R(A, b) > 0$.

REMARK 8.8. In case we would have restricted ourselves to inner product norms, the condition $R(A, b) > 0$ for conditional contractivity on $\mathscr{F}(\rho)$ would be weakened to the condition $b > 0$ (cf. Remark 5.8). In this case the order barrier $p \leq 4$ for explicit methods remains valid (cf. [7]), but the order barriers $\tilde{p} \leq 2$ and $p \leq 6$ for implicit methods disappear (cf. Remark 8.4).

9. Optimal explicit methods.

In this section we study, for given integers m and p, the maximum of $R(A, b)$ on the class of explicit Runge-Kutta methods with m stages and classical order at least p. In Subsections 9.1 and 9.2 we solve this optimization problem completely for the cases $p = 1$ and $p = 2$, respectively. In Subsection 9.3 we determine the optimal methods for $p = 3$ and $m = 3, 4$. Finally, in Subsection 9.4, we present results for the maximal order $p = 4$ (cf. Subsection 8.3) and $m = 4, 5$.

The following lemma will be the key result for solving our optimization problems in case $p = 1, 2, 3$.

LEMMA 9.1. *Let (A, b) be an arbitrary explicit coefficient scheme with m stages and classical order at least p, where $1 \leq p \leq m$. Then $R(A, b) \leq m - p + 1$. Further, if $R(A, b) = m - p + 1$ then, for all integers k with $p - 1 \leq k \leq m - 1$, all k-th order partial derivatives of the K-function at $Z = -(m - p + 1)I$ are zero.*

PROOF. Let (A, b) denote an explicit coefficient scheme with m stages and classical order at least p, where $1 \leq p \leq m$. Then the stability function φ is a polynomial of degree at most m satisfying $\varphi(z) = \exp(z) + O(z^{p+1})$ (as $z \to 0$) (cf. e.g. [5, pp. 241–242]). By Lemma 2.5 the stability function φ and the K-function K are absolutely monotonic on $(-r, 0]$, where $r = R(A, b)$. It follows from Theorem 2.1 in [21] that $r \leq m - p + 1$. Further it is easy to see from the proof of this theorem that $r = m - p + 1$ is only possible if $\varphi^{(k)}(-r) = 0$ for all integers k with $p - 1 \leq k \leq m - 1$. Using the absolute monotonicity of K at $\xi = -r$ and the fact that $\varphi(z) \equiv K(z, z, \ldots, z)$, we see that in this case the kth order partial derivatives of K at $Z = -rI$ must be zero as well for these integers k. ■

9.1. *Optimal methods with* $p = 1$.

In this subsection we maximize $R(A, b)$ over the class of explicit Runge-Kutta methods with a fixed number of stages m. We assume that $p \geq 1$ (or, equivalently, $\tilde{p} \geq 1$), i.e.

$$b_1 + b_2 + \ldots + b_m = 1. \tag{9.1}$$

THEOREM 9.2. *Let $m \geq 1$ be given. Then we have $R(A, b) \leq m$ for all explicit m-stage coefficient schemes (A, b) with classical order $p \geq 1$. Further we have $R(A, b) = m$ for exactly one of these coefficient schemes. This scheme is defined by $a_{ij} = 1/m$ $(1 \leq j < i \leq m)$ and $b_i = 1/m$ $(1 \leq i \leq m)$.*

PROOF. First note that the scheme specified above has order $p = 1$ and a K-function given by

$$K(Z) = (1 + z_1/m)(1 + z_2/m)\ldots(1 + z_m/m). \tag{9.2}$$

Clearly this function is absolutely monotonic on $[-m, 0]$. It follows from Theorem 3.8 that $R(A, b) \geq m$ for this scheme.

Now suppose that (A, b) is an arbitrary explicit m-stage coefficient scheme with $p \geq 1$ and $R(A, b) \geq m$. Then Lemma 9.1 shows that $R(A, b) = m$ and that the K-function has the form

$$K(Z) = \gamma(z_1 + m)(z_2 + m)\ldots(z_m + m),$$

where γ is a real (nonnegative) constant. Since $K(0) = 1$, we have $\gamma = m^{-m}$, so that K is given by (9.2), i.e.

$$K(Z) = 1 + m^{-1}\sum_i z_i + m^{-2}\sum_{i>j} z_i z_j + \ldots.$$

Comparing the latter expansion with the expansion that is obtained after substitution of $\xi = 0$ and $s = 1$ in (2.6),

$$K(Z) = 1 + \sum_i b_i z_i + \sum_{i>j} b_i a_{ij} z_i z_j + \ldots, \tag{9.3}$$

we conclude that (A, b) must be equal to the scheme specified in the theorem. ■

Note that the optimal method in Theorem 9.2 is nothing but a cyclic application of Euler's method with step size h/m. This method was found to be optimal within the framework of inner product norms as well, cf. [7, Theorem 5.1].

9.2. *Optimal methods with* $p = 2$.

In this subsection we maximize $R(A, b)$ over the class of explicit Runge-Kutta methods with a fixed number of stages m and with classical order at least 2, i.e. (9.1) holds as well as

$$b_1 c_1 + b_2 c_2 + \ldots + b_m c_m = \tfrac{1}{2}. \tag{9.4}$$

THEOREM 9.3. *Let $m \geq 2$ be given. Then we have $R(A, b) \leq m - 1$ for all explicit m-stage coefficient schemes (A, b) with classical order $p \geq 2$. Further we have $R(A, b) = m - 1$ for exactly one of these coefficient schemes. This scheme is defined by $a_{ij} = (m-1)^{-1}$ $(1 \leq j < i \leq m)$ and $b_i = m^{-1}$ $(1 \leq i \leq m)$.*

PROOF. First note that the scheme specified above has order $p = 2$ and is absolutely monotonic at $\xi = -(m-1)$. It follows from Lemma 4.4 that $R(A, b) \geq m - 1$ for this scheme.

Now suppose that (A, b) is an arbitrary explicit m-stage coefficient scheme with $p \geq 2$ and $R(A, b) \geq m - 1$. Then Lemma 9.1 shows that $R(A, b) = m - 1$ and that the K-function has the form

$$K(Z) = \gamma_1 + \gamma_2 (z_1 + m - 1)(z_2 + m - 1)\ldots(z_m + m - 1),$$

where γ_1 and γ_2 are real (nonnegative) constants. Since (9.1) and (9.3) imply $K(0) = 1$ and $\sum_i (\partial K/\partial z_i)(0) = 1$, we obtain $\gamma_1 = m^{-1}$ and $\gamma_2 = m^{-1}(m-1)^{1-m}$, so that

$$K(Z) = \frac{1}{m} + \frac{m-1}{m}\left(1 + \frac{z_1}{m-1}\right)\left(1 + \frac{z_2}{m-1}\right)\ldots\left(1 + \frac{z_m}{m-1}\right) =$$

$$= 1 + \frac{1}{m}\sum_i z_i + \frac{1}{m(m-1)}\sum_{i>j} z_i z_j + \ldots.$$

Comparing the latter expansion with (9.3) we conclude that (A, b) must be equal to the scheme specified in the theorem. ■

Note that for $m = 2$ stages the optimal method in Theorem 9.2 is the well-known improved Euler method, which is also known as the (second order) Heun method and the explicit trapezoidal method.

9.3. *Optimal methods with* $p = 3$.

In this subsection we maximize $R(A, b)$ over the class of explicit Runge-Kutta methods with a fixed number of stages m and with classical order at least 3, i.e. (9.1) and (9.4) hold and

$$b_1 c_1^2 + b_2 c_2^2 + \ldots + b_m c_m^2 = \tfrac{1}{3}, \tag{9.5a}$$

$$b^{\mathrm{T}} A^2 e = \tfrac{1}{6}. \tag{9.5b}$$

In the following two theorems we consider the cases $m = 3$ and $m = 4$, respectively. It is interesting to note that the optimal method in Theorem 9.4 was considered by Fehlberg in [10] (see also [5, p. 305] and [17, p. 170]).

THEOREM 9.4. *For all explicit* 3*-stage coefficient schemes* (A, b) *with classical order* $p = 3$ *we have* $R(A, b) \leq 1$. *Further we have* $R(A, b) = 1$ *for exactly one of these schemes. This scheme is defined by the tableau*

$$\begin{array}{c|ccc} 0 & & & \\ 1 & 1 & & \\ 1/2 & 1/4 & 1/4 & \\ \hline & 1/6 & 1/6 & 2/3 \end{array}.$$

PROOF. First note that the scheme specified above has order $p = 3$ and is absolutely monotonic at $\xi = -1$. It follows from Lemma 4.4 that $R(A, b) \geq 1$ for this scheme.

Now suppose that (A, b) is an arbitrary explicit 3-stage coefficient scheme with $p = 3$ and $R(A, b) \geq 1$. Then the order conditions (9.1), (9.4) and (9.5) reduce to

(9.6a) $$b_1 + b_2 + b_3 = 1,$$

(9.6b) $$b_2c_2 + b_3c_3 = 1/2,$$

(9.6c) $$b_2c_2^2 + b_3c_3^2 = 1/3,$$

(9.6d) $$b_3a_{32}a_{21} = 1/6.$$

Further, substitution of $\xi = 0$ and $s = 1$ in (2.6) shows that the K-function equals

$$K(Z) = 1 + b_1z_1 + b_2z_2 + b_3z_3 + b_3a_{32}z_3z_2 + b_3a_{31}z_3z_1 +$$
$$+ b_2a_{21}z_2z_1 + b_3a_{32}a_{21}z_3z_2z_1.$$

Using Lemma 9.1 we find that $R(A, b) = 1$ and that all second order partial derivatives of K at $Z = -I$ are zero, i.e. $b_3a_{32} = b_3a_{31} = b_2a_{21} = b_3a_{32}a_{21}$. In view of (9.6d) this yields $b_3a_{32} = b_3a_{31} = b_2a_{21} = 1/6$. Combined with (9.6) these relations lead immediately to the scheme specified in the theorem. ■

THEOREM 9.5. *For all explicit* 4*-stage coefficient schemes* (A, b) *with classical order* $p \geq 3$ *we have* $R(A, b) \leq 2$. *Further we have* $R(A, b) = 2$ *for exactly one of these schemes. This scheme is defined by the tableau*

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

PROOF. First note that the scheme specified above has order $p = 3$ and is absolutely monotonic at $\xi = -2$. It follows from Lemma 4.4 that $R(A,b) \geq 2$ for this scheme.

Now suppose that (A,b) is an arbitrary explicit 4-stage coefficient scheme with $p \geq 3$ and $R(A,b) \geq 2$. Then the order conditions (9.1), (9.4) and (9.5) reduce to

(9.7a) $$b_1 + b_2 + b_3 + b_4 = 1,$$

(9.7b) $$b_2c_2 + b_3c_3 + b_4c_4 = 1/2,$$

(9.7c) $$b_2c_2^2 + b_3c_3^2 + b_4c_4^2 = 1/3,$$

(9.7d) $$b_3a_{32}a_{21} + b_4a_{42}a_{21} + b_4a_{43}a_{32} + b_4a_{43}a_{31} = 1/6.$$

Further, substitution of $\xi = 0$ and $s = 1$ in (2.6) shows that the K-function equals

$$K(Z) = 1 + \sum_{k=1}^{4} \sum_{i_1 > i_2 > \ldots > i_k} b_{i_1}a_{i_1i_2}\ldots a_{i_{k-1}i_k}z_{i_1}z_{i_2}\ldots z_{i_k}.$$

Using Lemma 9.1 we find that $R(A,b) = 2$ and that all second and third order partial derivatives of K at $Z = -2I$ are zero. Combined with (9.7d), the condition on the third order partial derivatives leads to

(9.8) $$b_4a_{43}a_{32} = b_4a_{43}a_{31} = b_4a_{42}a_{21} = b_3a_{32}a_{21} = 2b_4a_{43}a_{32}a_{21} = \tfrac{1}{24}.$$

It follows from (9.8) and the condition on the second order partial derivatives that

(9.9) $$b_4a_{43} = b_4a_{42} = b_4a_{41} = b_3a_{32} = b_3a_{31} = b_2a_{21} = \tfrac{1}{12}.$$

It is easy to show that relations (9.7), (9.8), (9.9) lead to the scheme specified in the theorem. ■

In view of Theorems 9.2, 9.3, 9.4 and 9.5 one might conjecture that the maximum radius of absolute monotonicity is $R(A,b) = m - 2$ for explicit m-stage coefficient schemes with classical order $p \geq 3$. For $m \geq 5$ this conjecture is false. It follows from [21, Theorem 5.2] that $R(A,b) \leq m - \sqrt{m}$ for these schemes whenever $m \geq 5$.

9.4. *Optimal methods with $p = 4$.*

In this subsection we study the maximum of $R(A,b)$ on the class of explicit Runge-Kutta methods with a fixed number of stages m and with classical order (at least) 4. We will consider the cases $m = 4$ and $m = 5$ only.

THEOREM 9.6. *There exists no explicit 4-stage coefficient scheme (A,b) with classical order $p = 4$ and $R(A,b) > 0$.*

PROOF. Using the general form of explicit 4-stage schemes with $p = 4$, presented in [5, pp. 179–180], it can be verified that the only method of this type with $A \geq 0$ and $b \geq 0$ is the well-known classical Runge-Kutta method. For this method the only nonzero entries of A are $a_{21} = a_{32} = 1/2$ and $a_{43} = 1$, so that the condition $\text{Inc}(A^2) \leq \text{Inc}(A)$ is violated. Since this method is irreducible, it follows from Theorem 4.2 that $R(A, b) = 0$ here. ■

It follows from the theorem above that we need at least five stages to construct an explicit method with $p = 4$ and $R(A, b) > 0$. It can be shown that, under the mild assumption that all c_i are distinct and $b_4 b_5[c_3(c_3 - c_2)(4c_5 - 3) + 2c_2(3 - 4c_2 - 4c_5 + 6c_2c_5)a_{32}] \neq 0$, the general explicit 5-stage Runge-Kutta method with $p = 4$ can be obtained as follows:

Step 1: Select distinct $c_1 = 0, c_2, c_3, c_4, c_5$.

Step 2: Select b_1 arbitrarily and solve b_2, b_3, b_4, b_5 from $B(4)$ (cf. Subsection 8.1). The only restriction on b_1 is that the resulting b_4 and b_5 must not vanish.

Step 3: Select a_{53} arbitrarily and a_{32} such that
$c_3(c_3 - c_2)(4c_5 - 3) + 2c_2(3 - 4c_2 - 4c_5 + 6c_2c_5)a_{32} \neq 0.$

Step 4: Define a_{54} by

$$a_{54} = \frac{b_4(c_5 - c_4)[c_3(c_3 - c_2) + 2c_2(2c_2 - 1)a_{32}]}{b_5[c_3(c_3 - c_2)(4c_5 - 3) + 2c_2(3 - 4c_2 - 4c_5 + 6c_2c_5)a_{32}]}.$$

Step 5: Define a_{52}, a_{43} and a_{42} by

$$a_{52} = (\alpha - a_{54}c_4 - a_{53}c_3)/c_2,$$
$$a_{43} = (\beta - b_5 a_{53})/b_4,$$
$$a_{42} = \frac{b_5 a_{53} c_3 - b_3 a_{32} c_2 - \alpha b_5 - \beta c_3 + 1/6}{b_4 c_2}, \quad \text{where}$$
$$\alpha = \frac{3 - 4c_4 + 24 b_3 a_{32} c_2 (c_4 - c_3)}{24 b_5 (c_5 - c_4)}, \quad \beta = \frac{1 - 2c_2 - 12 b_5 a_{54} c_4 (c_4 - c_2)}{12 c_3 (c_3 - c_2)}.$$

Step 6: Select $a_{21}, a_{31}, a_{41}, a_{51}$ to satisfy $c_i = \sum_j a_{ij} \quad (2 \leq i \leq 5)$.

Numerical investigations seem to indicate that the maximum of $R(A, b)$ equals $r \simeq 1.50818$ and that the maximum value r, together with the seven free parameters $c_2, c_3, c_4, c_5, b_1, a_{53}, a_{32}$ are determined by the eight equations

$$b_1 = r b_2 c_2, \quad b_3 = r b_4 a_{43}, \quad a_{31} = r a_{32} c_2, \quad a_{41} = r a_{42} c_2,$$

$$a_{42} = r a_{43} a_{32}, \quad a_{51} = r a_{54} a_{41}, \quad a_{52} = r a_{54} a_{42}, \quad a_{53} = r a_{54} a_{43},$$

leading to

$r \simeq 1.50818$	00491	89837	92280	$a_{21} = c_2$				
$b_1 \simeq 0.14681$	18760	84786	44956	$a_{31} \simeq 0.21766$	90962	61169	21036
$b_2 \simeq 0.24848$	29094	44976	14757	$a_{32} \simeq 0.36841$	05930	50372	02075
$b_3 \simeq 0.10425$	88303	31980	29567	$a_{41} \simeq 0.08269$	20866	57810	75441
$b_4 \simeq 0.27443$	89009	01349	45681	$a_{42} \simeq 0.13995$	85021	91895	73938
$b_5 \simeq 0.22600$	74832	36907	65039	$a_{43} \simeq 0.25189$	17742	71692	63984
$c_2 \simeq 0.39175$	22265	71889	05833	$a_{51} \simeq 0.06796$	62836	37114	96324
$c_3 \simeq 0.58607$	96893	11541	23111	$a_{52} \simeq 0.11503$	46985	04631	99467

$c_4 \simeq 0.47454\ 23631\ 21399\ 13362$	$a_{53} \simeq 0.20703\ 48985\ 97384\ 71851$
$c_5 \simeq 0.93501\ 06309\ 67651\ 59845$	$a_{54} \simeq 0.54497\ 47502\ 28519\ 92204$

10. Numerical illustration.

In this final section of the paper we illustrate the theory with three examples. In these three examples we compare Heun's third order scheme (cf. e.g. [9, p. 193] or [17, p. 133])

$$\begin{array}{c|ccc} 0 & & & \\ 1/3 & 1/3 & & \\ 2/3 & 0 & 2/3 & \\ \hline & 1/4 & 0 & 3/4 \end{array}$$

with the optimal third order 3-stage scheme presented in Theorem 9.4. We will refer to these schemes as *Heun's scheme* and *Fehlberg's scheme*, respectively. Note that Heun's scheme has radius of absolute monotonicity $R_{\text{Heun}} = 0$ (cf. Theorem 4.2), whereas Fehlberg's scheme has radius $R_{\text{Fehlberg}} = 1$ (cf. Theorem 9.4).

10.1. *A constructed initial value problem in eight dimensions.*

Consider the linear non-autonomous test problem (2.3), where $t_0 = 0$ and $L(t)$ is a real 8×8 matrix depending continuously on t. We assume that (cf. (5.13))

$$\|L(t) + I_8\|_\infty \leq 1 \quad (\text{for all } t \in \mathbb{R}). \tag{10.1}$$

This means that, if we write the initial value problem in the form (1.1), the function f satisfies $(f, \|\cdot\|_\infty) \in \mathscr{F}(\rho)$ with $\rho = 1$ (cf. Definition 5.2), so that f is dissipative with respect to the maximum norm. It follows from Theorem 5.4 that we will have contractivity (1.5) in the maximum norm if we use Fehlberg's scheme with step size $h \leq 1$. However, if we use Heun's scheme with step size $h = 1$, we will show that we can have

$$\|\tilde{u}_1 - u_1\|_\infty \geq \tfrac{7}{3}\|\tilde{u}_0 - u_0\|_\infty \tag{10.2}$$

for some $\tilde{u}_0 \neq u_0$. To see that (10.2) can occur, we note that it follows from Subsection 2.1 that $\tilde{u}_1 - u_1 = \mathbb{K}(\mathbb{Z})(\tilde{u}_0 - u_0)$, where $\mathbb{K}$ is the matrix-valued K-function and $\mathbb{Z}$ is the block diagonal matrix $\mathbb{Z} = \text{diag}(Z_1, Z_2, Z_3)$ with $Z_1 = L(0)$, $Z_2 = L(1/3)$ and $Z_3 = L(2/3)$. Writing $Z_i = -I_8 + W_i$ $(i = 1, 2, 3)$ we obtain

$$\begin{aligned} \mathbb{K}(\mathbb{Z}) = {} & \tfrac{1}{3}I_8 + \tfrac{5}{12}W_1 - \tfrac{1}{3}W_2 + \tfrac{5}{12}W_3 - \tfrac{1}{6}W_2W_1 - \tfrac{1}{6}W_3W_1 \\ & + \tfrac{1}{3}W_3W_2 + \tfrac{1}{6}W_3W_2W_1. \end{aligned} \tag{10.3}$$

In view of (10.1) the matrices W_1, W_2, W_3 satisfy

$$\|W_i\|_\infty \leq 1 \quad (i = 1, 2, 3). \tag{10.4}$$

On the other hand, if we are given 8×8 matrices W_1, W_2, W_3 satisfying (10.4), we can easily find an interpolating mapping $t \to L(t)$ satisfying (10.1). We will consider the choice

$$W_1 = e_1 e_2^T + e_3 e_5^T + e_4 e_6^T + e_7 e_8^T,$$
$$W_2 = e_1 e_3^T + e_4 e_7^T,$$
$$W_3 = e_1 e_4^T,$$

where e_i, $i = 1, 2, \ldots, 8$, are the standard basis vectors in $\mathbb{R}^8$. Here we closely follow the proof of Lemma 5.10. In fact, our choice of W_1, W_2, W_3 is obtained from formula (5.19) (with $r = 1$ and $n = 3$), where – due to the explicitness of the scheme – the graph Γ_3 is simplified to the graph in Figure 2.

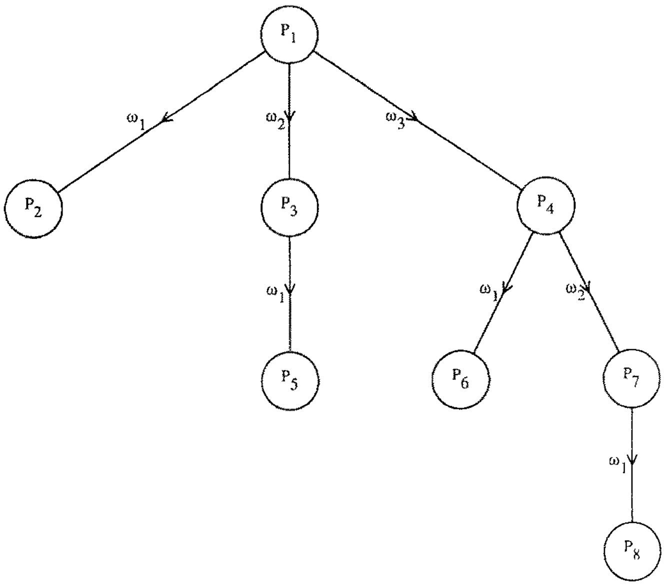

Fig. 2. The graph that defines W_1, W_2, W_3.

Using (10.3) one readily verifies that the first row of $\mathbb{K}(\mathbb{Z})$ is given by

$$e_1^T \mathbb{K}(\mathbb{Z}) = (\tfrac{1}{3}, \tfrac{5}{12}, \tfrac{-1}{3}, \tfrac{5}{12}, \tfrac{-1}{6}, \tfrac{-1}{6}, \tfrac{1}{3}, \tfrac{1}{6}).$$

Consequently, if we choose $\tilde{u}_0 - u_0 = (1, 1, -1, 1, -1, -1, 1, 1)^T$, we obtain $e_1^T(\tilde{u}_1 - u_1) = 7/3$, which proves (10.2).

10.2. *A partial differential equation of parabolic type.*

Consider the following initial-boundary value problem

$$(10.5)\qquad \begin{cases} \dfrac{\partial}{\partial t} V(\xi,t) = a(\xi,t)\dfrac{\partial^2}{\partial \xi^2} V(\xi,t), \\ V(0,t) = V(1,t) = 0, \quad V(\xi,0) = v_0(\xi) \quad (0 \le \xi \le 1,\ t \ge 0), \end{cases}$$

where $a(\xi,t) \equiv \frac{1}{50} + \frac{49}{50}\cos^2(30\xi + 800t)$. If v_0 is sufficiently smooth and satisfies the compatibility conditions $v_0(0) = v_0''(0) = v_0(1) = v_0''(1) = 0$, then this problem has a unique classical solution V (cf. [14, p. 65]). This solution has the property that

$$(10.6)\qquad \max_{0\le\xi\le 1} |V(\xi,t)| \le \max_{0\le\xi\le 1} |v_0(\xi)| \quad (\text{for all } t \ge 0).$$

By approximating the second order spatial derivative in (10.5) by a central difference quotient, we obtain an initial value problem of type (2.3) where

$$L(t) = (\Delta\xi)^{-2} \begin{pmatrix} -2a(\xi_1,t) & a(\xi_1,t) & & & 0 \\ a(\xi_2,t) & -2a(\xi_2,t) & a(\xi_2,t) & & \\ & \ddots & \ddots & \ddots & \\ & & a(\xi_{s-1},t) & -2a(\xi_{s-1},t) & a(\xi_{s-1},t) \\ 0 & & & a(\xi_s,t) & -2a(\xi_s,t) \end{pmatrix},$$

$\Delta\xi = (s+1)^{-1}$, $\xi_i = i\Delta\xi$ $(1 \le i \le s)$, $t_0 = 0$ and $u_0 = (v_0(\xi_1), v_0(\xi_2), \ldots, v_0(\xi_s))^{\mathrm{T}}$. Its solution $U(t)$ approximates $(V(\xi_1,t), V(\xi_2,t), \ldots, V(\xi_s,t))^{\mathrm{T}}$ for $t > 0$. Writing the initial value problem (2.3) in the form (1.1), we see that the function f satisfies $(f, \|\cdot\|_\infty) \in \mathscr{F}(\rho)$ with $\rho = 2(\Delta\xi)^{-2}$ (cf. Definition 5.2). Hence f is dissipative with respect to the maximum norm and it follows that the semi-discrete problem (2.3) has the property

$$(10.7)\qquad \|U(t)\|_\infty \le \|u_0\|_\infty \quad (\text{for all } t \ge 0),$$

reflecting property (10.6) of the parabolic problem (10.5). If we use Fehlberg's scheme with step size $h = \Delta t \le \frac{1}{2}(\Delta\xi)^2$ for the numerical solution of the semi-discrete problem (2.3), it follows from Theorem 5.4 that the obtained approximations $u_n \simeq U(nh)$ satisfy

$$(10.8)\qquad \|u_n\|_\infty \le \|u_0\|_\infty \quad (n \ge 1),$$

which constitutes a fully discrete analogon to properties (10.6) and (10.7). For Heun's scheme we cannot expect property (10.8) in view of Theorem 5.4 and the fact that $R_{\text{Heun}} = 0$. Indeed, if we take $s = 19$ and $h = \frac{1}{2}(\Delta\xi)^2$, the first step $u_0 \to u_1$ is given by $u_1 = Gu_0$, where the 19×19 matrix G has a norm $\|G\|_\infty \simeq 1.022$ (cf. (5.13)). Further it is interesting to note that G has some negative entries, so that we do not have the property

$$(10.9)\qquad u_0 \ge 0 \Rightarrow u_n \ge 0 \quad (n \ge 1),$$

although the corresponding semi-discrete and fully continuous versions of this property are present for problems (2.3) and (10.5), respectively. For Fehlberg's scheme, property (10.9) holds for all step sizes $h \le \frac{1}{2}(\Delta\xi)^2$. For linear autonomous problems this property was studied in [2].

10.3. *A partial differential equation of hyperbolic type.*

Consider the following initial-boundary value problem

$$(10.10) \qquad \begin{cases} \dfrac{\partial}{\partial t} V(\xi, t) = \dfrac{\partial}{\partial \xi}[a(\xi, t)V(\xi, t)], \\ V(0, t) = 0, \quad V(\xi, 0) = v_0(\xi) \quad (0 \le \xi \le 1,\ t \ge 0), \end{cases}$$

where $a(\xi, t) = -\cos^2(20\xi + 80t)$. If v_0 is continuously differentiable on $[0, 1]$ and $v_0(0) = v_0'(0) = 0$, then one can show (e.g., by making use of characteristics, cf. [20]) that this problem has a unique classical solution V. This solution has the property that

$$(10.11) \qquad \int_0^1 |V(\xi, t)|\, d\xi \le \int_0^1 |v_0(\xi)|\, d\xi \quad (\text{for all } t \ge 0).$$

By approximating the spatial derivative in (10.10) by a backward difference quotient, we obtain an initial value problem of type (2.3), where

$$L(t) = (\Delta\xi)^{-1} \begin{pmatrix} a(\xi_1, t) & & & 0 \\ -a(\xi_1, t) & a(\xi_2, t) & & \\ & -a(\xi_2, t) & a(\xi_3, t) & \\ & \ddots & \ddots & \\ 0 & & -a(\xi_{s-1}, t) & a(\xi_s, t) \end{pmatrix},$$

$\Delta\xi = s^{-1}$, $\xi_i = i\Delta\xi$ $(1 \le i \le s)$, $t_0 = 0$ and $u_0 = (v_0(\xi_1), v_0(\xi_2), \ldots, v_0(\xi_s))^{\mathrm{T}}$. Its solution $U(t)$ approximates $(V(\xi_1, t), V(\xi_2, t), \ldots, V(\xi_s, t))^{\mathrm{T}}$ for $t > 0$. Writing the initial value problem (2.3) in the form (1.1), we see that the function f satisfies $(f, \|\cdot\|_1) \in \mathscr{F}(\rho)$ with $\rho = (\Delta\xi)^{-1}$ (cf. Definition 5.2), where the norm $\|\cdot\|_1$ is the (weighted) l_1 norm, defined by

$$\|x\|_1 = \Delta\xi \sum_{i=1}^{s} |x_i| \quad (\text{for all } x = (x_1, x_2, \ldots, x_s)^{\mathrm{T}} \in \mathbb{R}^s).$$

Hence f is dissipative with respect to the (weighted) l_1 norm and it follows that the semi-discrete problem (2.3) has the property

$$(10.12) \qquad \|U(t)\|_1 \le \|u_0\|_1 \quad (\text{for all } t \ge 0),$$

reflecting property (10.11) of the hyperbolic problem (10.10). If we use Fehlberg's scheme with step size $h = \Delta t \le \Delta\xi$ for the numerical solution of the semi-discrete problem (2.3), it follows from Theorem 5.4 that the obtained approximations $u_n \simeq U(nh)$ satisfy

$$\|u_n\|_1 \leq \|u_0\|_1 \quad (n \geq 1), \tag{10.13}$$

which constitutes a fully discrete analogon to properties (10.11) and (10.12). Heun's scheme, however, does not have property (10.13) if we take $h = \Delta\xi$ and, e.g., $s = 20$. In this case we have $u_1 = Gu_0$, where the 20×20 matrix $G = (g_{ij})$ has induced matrix norm $\|G\|_1 = \max\{\|Gx\|_1 \,|\, x \in \mathbb{R}^{20}, \|x\|_1 = 1\} = \max_j \sum_i |g_{ij}| \simeq 1.511$. Further we note that, as in the previous example, Heun's scheme does not have property (10.9), whereas Fehlberg's scheme does, and that the corresponding semi-discrete and fully continuous versions of this property are present for problems (2.3) and (10.10), respectively.

Acknowledgements.

I would like to thank K. Burrage and J. C. Butcher for their help with the order conditions, J. A. van de Griend and M. N. Spijker for their constructive criticism, T. Murdoch and J. T. Ratnanather for their assistance with the numerical computations in Section 9 and R. Pluis for his assistance with the graphic illustrations. Further I would like to thank K. W. Morton for his kind hospitality during my stay at the Oxford University Computing Laboratory in the period July 1987–June 1988.

REFERENCES

1. Berman, A. and Plemmons, R. J.: *Nonnegative Matrices in the Mathematical Sciences*. New York San Francisco London: Academic Press 1979.
2. Bolley, C. and Crouzeix, M.: *Conservation de la positivité lors de la discrétisation des problèmes d'évolution paraboliques*. RAIRO Anal. Numér. 12, 237–245 (1978).
3. Burrage, K. and Butcher, J. C.: *Stability criteria for implicit Runge-Kutta methods*. SIAM J. Numer. Anal. 16, 46–57 (1979).
4. Butcher, J. C.: *Implicit Runge-Kutta processes*. Math. Comp. 18, 50–64 (1964).
5. Butcher, J. C.: *The Numerical Analysis of Ordinary Differential Equations*. Chichester New York Brisbane: John Wiley 1987.
6. Crouzeix, M.: *Sur la B-stabilité des méthodes de Runge-Kutta*. Numer. Math. 32, 75–82 (1979).
7. Dahlquist, G. and Jeltsch, R.: *Generalized disks of contractivity for explicit and implicit Runge-Kutta methods*. Report TRITA-NA-7906, Dept. of Numer. Anal. and Comp. Sci., Royal Inst. of Techn., Stockholm (1979).
8. Dahlquist, G. and Jeltsch, R.: *Shifted Runge-Kutta methods and transplanted differential equations*. In: Strehmel, K. (ed.) *Numerical Treatment of Differential Equations*, proceedings, Halle, 1987, pp. 47–56. Leipzig: Teubner 1988.
9. Dekker, K. and Verwer, J. G.: *Stability of Runge-Kutta Methods for Stiff Nonlinear Differential Equations*. Amsterdam New York Oxford: North-Holland 1984.
10. Fehlberg, E.: *Klassische Runge-Kutta-Formeln vierter und niedrigerer Ordnung mit Schrittweiten-Kontrolle und ihre Anwendung auf Wärmeleitungsprobleme*. Computing 6, 61–71 (1970).
11. Frank, R., Schneid, J. and Ueberhuber, C. W.: *The concept of B-convergence*. SIAM J. Numer. Anal. 18, 753–780 (1981).
12. Frank, R., Schneid, J. and Ueberhuber, C. W.: *Stability properties of implicit Runge-Kutta methods*. SIAM J. Numer. Anal. 22, 497–514 (1985).

13. Frank, R., Schneid, J. and Ueberhuber, C. W.: *B-convergence: a survey.* Appl. Numer. Math. 5, 51–61 (1989).
14. Friedman, A.: *Partial Differential Equations of Parabolic Type.* Englewood Cliffs, N.J.: Prentice-Hall 1964.
15. Griend, J. A. van de and Kraaijevanger, J. F. B. M.: *Absolute monotonicity of rational functions occurring in the numerical solution of initial value problems.* Numer. Math. 49, 413–424 (1986).
16. Hairer, E.: *Highest possible order of algebraically stable diagonally implicit Runge-Kutta methods.* BIT 20, 254–256 (1980).
17. Hairer, E., Nørsett, S. P. and Wanner, G.: *Solving Ordinary Differential Equations I.* Berlin Heidelberg New York: Springer 1987.
18. Hundsdorfer, W. H. and Spijker, M. N.: *A note on B-stability of Runge-Kutta methods.* Numer. Math. 36, 319–331 (1981).
19. Hundsdorfer, W. H. and Spijker, M. N.: *On the algebraic equations in implicit Runge-Kutta methods.* SIAM J. Numer. Anal. 24, 583–594 (1987).
20. John, F.: *Partial Differential Equations.* Appl. Math. Sciences, Vol. 1. New York Heidelberg Berlin: Springer 1971.
21. Kraaijevanger, J. F. B. M.: *Absolute monotonicity of polynomials occurring in the numerical solution of initial value problems.* Numer. Math. 48, 303–322 (1986).
22. Kraaijevanger, J. F. B. M., Lenferink, H. W. J. and Spijker, M. N.: *Stepsize restrictions for stability in the numerical solution of ordinary and partial differential equations.* J. Comp. Appl. Math. 20, 67–81 (1987).
23. Kraaijevanger, J. F. B. M. and Spijker, M. N.: *Algebraic stability and error propagation in Runge-Kutta methods.* Appl. Numer. Math. 5, 71–87 (1989).
24. Kraaijevanger, J. F. B. M.: *Contractivity in the maximum norm for Runge-Kutta methods.* Report TW-89-06, Dept. of Math. and Comp. Sci., Leiden Univ. 1989. To appear in the proceedings of the 1989 IMA conference on computational ordinary differential equations, London. Oxford Univ. Press (editors: Cash, J. R., Gladwell, I.).
25. Lancaster, P., Tismenetsky, M.: *The Theory of Matrices, 2nd Ed.* Orlando San Diego New York London: Academic Press 1985.
26. Lenferink, H. W. J.: *Contractivity preserving explicit linear multistep methods.* Numer. Math. 55, 213–223 (1989).
27. Lenferink, H. W. J.: *Contractivity preserving implicit linear multistep methods.* Math. Comp. 56, 177–199 (1991).
28. Martin, R. H., Jr.: *Nonlinear Operators and Differential Equations in Banach Spaces.* New York London Sydney: John Wiley 1976.
29. Nevanlinna, O. and Liniger, W.: *Contractive methods for stiff differential equations.* BIT 18, 457–474 (1978); BIT 19, 53–72 (1979).
30. Sand, J.: *Circle contractive linear multistep methods.* BIT 26, 114–122 (1986).
31. Sand, J.: *Choices in contractivity theory.* Appl. Numer. Math. 5, 105–115 (1989).
32. Schönbeck, S. O.: *On the extension of Lipschitz maps.* Arkiv för Matematik 7, 201–209 (1967).
33. Spijker, M. N.: *Contractivity of Runge-Kutta methods.* In: Dahlquist, G., Jeltsch, R. (editors) *Numerical Methods for Solving Stiff Initial Value Problems,* proceedings, Oberwolfach, 28.6–4.7.1981. Institut für Geometrie und Praktische Mathematik der RWTH Aachen, Bericht Nr. 9, 1981.
34. Spijker, M. N.: *Contractivity in the numerical solution of initial value problems.* Numer. Math. 42, 271–290 (1983).
35. Spijker, M. N.: *Numerical contractivity in the solution of initial value problems.* In: Strehmel, K. (ed.) *Numerische Behandlung von Differentialgleichungen,* proceedings, Halle, 1983, pp. 118–124. Martin-Luther-Universität, Halle-Wittenberg 1984.
36. Spijker, M. N.: *On the relation between stability and contractivity.* BIT 24, 656–666 (1984).
37. Spijker, M. N.: *Stepsize restrictions for stability of one-step methods in the numerical solution of initial value problems.* Math. Comp. 45, 377–392 (1985).
38. Stetter, H. J.: *Analysis of Discretization Methods for Ordinary Differential Equations.* Berlin Heidelberg New York: Springer 1973.
39. Vanselow, R.: *Nonlinear stability behaviour of linear multistep methods.* BIT 23, 388–396 (1983).
40. Wakker, P. P.: *Extending monotone and non-expansive mappings by optimization.* Cahiers du C.E.R.O. 27, 141–149 (1985).