SIAM J. NUMER. ANAL.
Vol. 33, No. 3, pp. 1125-1134, June 1996

014

A NOTE ON UNCONDITIONAL MAXIMUM NORM CONTRACTIVITY OF DIAGONALLY SPLIT RUNGE-KUTTA METHODS*

K. J. IN 'T HOUT†

Abstract. In this paper we consider diagonally split Runge-Kutta methods for the numerical solution of initial value problems for ordinary differential equations. This class of numerical methods was recently introduced by Bellen, Jackiewicz, and Zennaro [*SIAM J. Numer. Anal.*, 31 (1994), pp. 499-523], and comprises the well-known class of Runge-Kutta methods. Their results strongly indicate that diagonally split Runge-Kutta methods break the order barrier $p \leq 1$ for unconditional contractivity in the maximum norm. In this paper we investigate the effect of the requirement of unconditional contractivity in the maximum norm on the accuracy of a diagonally split Runge-Kutta method. Besides the classical order p, we deal with an order of accuracy r which is relevant to the case where the method is applied to dissipative initial value problems that are arbitrarily stiff. We show that if a diagonally split Runge-Kutta method is unconditionally contractive in the maximum norm, then it has orders p, r which satisfy $p \leq 4, r \leq 1$.

Key words. ordinary differential equations, dissipative initial value problems, maximum norm, diagonally split Runge-Kutta methods, Runge-Kutta methods, unconditional contractivity, accuracy

AMS subject classification. 65L20

1. Introduction.

1.1. Dissipative initial value problems. This paper deals with initial value problems for systems of ordinary differential equations

$$U'(t) = f(t, U(t)) \quad (t \geq t_0), \tag{1.1.a}$$

$$U(t_0) = u_0. \tag{1.1.b}$$

Here t_0, u_0, f are given with $t_0 \in \mathbb{R}$, $u_0 \in \mathbb{R}^s$, and $f : \mathbb{R} \times \mathbb{R}^s \to \mathbb{R}^s$, whereas $U(t)$ is unknown (for $t > t_0$). We consider problems (1.1) that are dissipative in the maximum norm.

Let $\|x\|$ denote the *maximum norm* of any vector $x \in \mathbb{R}^s$; i.e., $\|x\| = \max_{j=1,2,\ldots,s} |\xi_j|$ (whenever $x = (\xi_1, \xi_2, \ldots, \xi_s)^{\mathrm{T}} \in \mathbb{R}^s$). Consider the conditions (cf. Kraaijevanger [14])

(1.2.a) $f : \mathbb{R} \times \mathbb{R}^s \to \mathbb{R}^s$ is continuous,

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

(1.2.c) for each $t_0 \in \mathbb{R}$ any two solutions $U, \widetilde{U}$ to (1.1.a) satisfy $\|\widetilde{U}(t) - U(t)\| \leq \|\widetilde{U}(t_0) - U(t_0)\|$ (whenever $t \geq t_0$).

The class of functions f, which fulfil the conditions (1.2) for some integer $s \geq 1$, is denoted by $\mathcal{F}$. If $f \in \mathcal{F}$ then f is called *dissipative* in the maximum norm and the initial value problem (1.1) is called dissipative as well. Problems (1.1) that are dissipative in the maximum norm arise, for example, from the semidiscretization of initial-boundary value problems for partial differential equations; see, e.g., [6], [14]. A characterization for the class $\mathcal{F}$ of dissipative functions f can be found in e.g., [14, Theorem 5.1].

*Received by the editors October 6, 1993; accepted for publication (in revised form) August 18, 1994. This research was supported by the Netherlands Organization for Scientific Research (NWO).

†Department of Mathematics and Statistics, University of Auckland, Private Bag 92019, Auckland, New Zealand. Present address: Department of Mathematics and Computer Science, University of Leiden, P.O. Box 9512, 2300 RA Leiden, The Netherlands (hout@wi.leidenuniv.nl).

1.2. Diagonally split Runge–Kutta methods. For the numerical solution of (1.1) we consider in this paper *diagonally split Runge–Kutta methods*. This class of numerical methods has recently been introduced by Bellen, Jackiewicz, and Zennaro [2].

Consider the *splitting function* $F : \mathbb{R} \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$ defined by

$$F_j(t, y, z) = f_j(t, (\zeta_1, \ldots, \zeta_{j-1}, \eta_j, \zeta_{j+1}, \ldots, \zeta_s)^{\mathrm{T}}) \quad (\text{for } j = 1, 2, \ldots, s) \tag{1.3}$$
$$(\text{whenever } t \in \mathbb{R},\ y = (\eta_1, \eta_2, \ldots, \eta_s)^{\mathrm{T}} \in \mathbb{R}^s,\ z = (\zeta_1, \zeta_2, \ldots, \zeta_s)^{\mathrm{T}} \in \mathbb{R}^s),$$

where f_j, F_j stand for the jth components of f, F, respectively. Let $h > 0$ denote a given *stepsize* and let the *gridpoints* t_n be given by $t_n = t_{n-1} + h$ (for $n = 1, 2, 3, \ldots$). Then a diagonally split Runge–Kutta method generates approximations u_n to $U(t_n)$ (for $n = 1, 2, 3, \ldots$) in the following one-step fashion (cf. [2]):

$$u_n = u_{n-1} + h \sum_{j=1}^{\nu} b_j F(t_{n-1} + c_j h, y_j, z_j), \tag{1.4.a}$$

where y_i, z_i $(i = 1, 2, \ldots, \nu)$ satisfy the equations

$$y_i = u_{n-1} + h \sum_{j=1}^{\nu} a_{ij} F(t_{n-1} + c_j h, y_j, z_j), \tag{1.4.b}$$

$$z_i = u_{n-1} + h \sum_{j=1}^{\nu} w_{ij} F(t_{n-1} + c_j h, y_j, z_j). \tag{1.4.c}$$

Here, a_{ij}, b_j, c_j, w_{ij} $(i, j = 1, 2, \ldots, \nu)$ denote given real coefficients that define the diagonally split Runge–Kutta method.

We note that a method of type (1.4) can be viewed as the *limit* of certain numerical processes that arise from the solution of (1.1) by the so-called waveform relaxation approach. It is, however, not within the scope of this paper to discuss this result. For details, and for results on the accuracy of method (1.4), we refer the reader to Bellen, Jackiewicz, and Zennaro [2].

In addition to method (1.4), we consider in this paper the *underlying Runge–Kutta method* of (1.4), which is the *Runge–Kutta method* defined by the coefficients a_{ij}, b_j, c_j $(i, j = 1, 2, \ldots, \nu)$. We let $A = (a_{ij})_{i,j=1}^{\nu}$, $b = (b_1, b_2, \ldots, b_\nu)^{\mathrm{T}}$, $c = (c_1, c_2, \ldots, c_\nu)^{\mathrm{T}}$ and denote the underlying Runge–Kutta method by (A, b, c). We also use the notation by the tableau

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

The method (A, b, c) generates approximations u_n to $U(t_n)$ (for $n = 1, 2, 3, \ldots$) by (see, e.g., Dekker and Verwer [6], Hairer and Wanner [10])

$$u_n = u_{n-1} + h \sum_{j=1}^{\nu} b_j f(t_{n-1} + c_j h, y_j), \tag{1.5.a}$$

where y_i $(i = 1, 2, \ldots, \nu)$ satisfy the equations

$$y_i = u_{n-1} + h \sum_{j=1}^{\nu} a_{ij} f(t_{n-1} + c_j h, y_j). \tag{1.5.b}$$

Method (1.4) reduces to the underlying Runge–Kutta method (A, b, c) whenever the initial value problem (1.1) is scalar $(s = 1)$.

Note that if the diagonally split Runge-Kutta method (1.4) satisfies $w_{ij} = a_{ij}$ (whenever $1 \le i \le \nu$, $1 \le j \le \nu$), then it reduces to the Runge-Kutta method (A, b, c) for general problems (1.1). Consequently, it holds that the class of diagonally split Runge-Kutta methods includes the class of Runge-Kutta methods.

1.3. Unconditional contractivity. Suppose that u_{n-1} in (1.4) is replaced by a (perturbed) value $\widetilde{u}_{n-1}$, and denote the result after one step of the method by $\widetilde{u}_n$. If the initial value problem (1.1) is dissipative, then it is natural to require that $\|\widetilde{u}_n - u_n\| \le \|\widetilde{u}_{n-1} - u_{n-1}\|$.

DEFINITION 1.1. *Method* (1.4) *is called* unconditionally contractive *in the maximum norm if* $\|\widetilde{u}_n - u_n\| \le \|\widetilde{u}_{n-1} - u_{n-1}\|$ *whenever* h, f *are given with* $h > 0$, $f \in \mathcal{F}$.

Unconditional contractivity is a very favourable stability property. It implies that any (rounding) errors that arise in the numerical process will not increase with n whenever $h > 0$ and the initial value problem is dissipative. The property is useful in deriving estimates for the *global errors* $U(t_n) - u_n$ $(n = 1, 2, 3, \ldots)$.

If u_{n-1} in (1.4) is replaced by $U(t_{n-1})$, and $\widehat{u}_n$ denotes the result after one (fictitious) step of the method, then the *local error* $\delta_n(h)$ is defined by $\delta_n(h) = U(t_n) - \widehat{u}_n$. Method (1.4) is said to have *classical order* p if p is the largest integer such that for each problem (1.1) with sufficiently smooth function f,

$$\delta_n(h) = \mathcal{O}(h^{p+1}) \quad (\text{for } h \downarrow 0).$$

Spijker [15] investigated contractivity properties of numerical step-by-step processes that are obtained when numerical step-by-step methods for initial value problems (1.1) are applied in the case where the differential equation (1.1.a) is linear and autonomous. The results of Spijker [15] reveal that, unfortunately, many classes of numerical methods for problems (1.1) are subject to the severe *order barrier* $p \le 1$ for unconditional contractivity in the maximum norm. This order barrier holds, for example, for the class of Runge-Kutta methods (cf. also Dekker and Verwer [6], Hairer and Wanner [10], Kraaijevanger [14]). To date, no numerical step-by-step method for (1.1) is known that is unconditionally contractive in the maximum norm and has a classical order $p > 1$. Bellen, Jackiewicz, and Zennaro [2] recently showed, however, that diagonally split Runge-Kutta methods of classical order $p > 1$ exist which are unconditionally contractive in the maximum norm on a large subclass $\mathcal{F}'$ of $\mathcal{F}$. An example of a function $f : \mathbb{R} \times \mathbb{R}^s \to \mathbb{R}^s$ which does not belong to $\mathcal{F}'$ (cf. [2, cond. (5.1)]), but does belong to $\mathcal{F}$ (cf., e.g., Hairer, Nørsett, and Wanner [9, Thms. I.11.1, I.10.5] or Kraaijevanger [14, Thm. 5.1]), is given by

$$\begin{cases} f_1(t, y) = -2\eta_1 + \eta_2, \\ f_j(t, y) = \eta_{j-1} - 2\eta_j + \eta_{j+1} \quad (\text{for } j = 2, 3, \ldots, s-1), \\ f_s(t, y) = \eta_{s-1} - 2\eta_s \end{cases}$$

(whenever $t \in \mathbb{R}$, $y = (\eta_1, \eta_2, \ldots, \eta_s)^{\mathrm{T}} \in \mathbb{R}^s$). The results in [2] strongly suggest, however, that diagonally split Runge-Kutta methods break the order barrier $p \le 1$ for unconditional contractivity in the maximum norm.

1.4. Scope of this paper. In this paper we investigate the effect of the requirement of unconditional contractivity in the maximum norm on the accuracy of the diagonally split Runge-Kutta method (1.4). Besides the classical order p, we deal with an order concept for method (1.4) that is relevant to the case where the method is applied to dissipative problems (1.1) that are arbitrarily stiff.

Consider initial value problems (1.1) of the type

$$U'(t) = \lambda(t)(U(t) - g(t)) + g'(t) \quad (t \ge t_0), \tag{1.6.a}$$

$$U(t_0) = g(t_0), \tag{1.6.b}$$

where $g : \mathbb{R} \to \mathbb{R}$ is a given smooth function and $\lambda : \mathbb{R} \to \mathbb{R}$ is given and continuous and satisfies $\lambda(t) \leq 0$ (whenever $t \in \mathbb{R}$). Each problem of the type (1.6) is dissipative and has solution $U : [t_0, \infty) \to \mathbb{R}$ given by $U(t) = g(t)$ (for $t \geq t_0$). Method (1.4) is called *B-consistent of order r* on the problem class (1.6) (cf., e.g., Dekker and Verwer [6], Frank, Schneid, and Ueberhuber [8], Hairer and Wanner [10]) if r is the largest integer such that for each problem (1.6) with sufficiently smooth function g, the local error $\delta_n(h)$ of the method satisfies

$$\delta_n(h) = \mathcal{O}(h^{r+1}) \quad (\text{for } h > 0),$$

where the $\mathcal{O}$-constant is not affected by the stiffness of the problem; i.e., it holds uniformly in the function λ. In this definition we have tacitly assumed that method (1.4) is always feasible in the case of scalar dissipative problems (1.1); cf. §2.1. Clearly, it holds that the order of B-consistency of method (1.4) (on the problem class (1.6)) is completely determined by the coefficients of the underlying Runge–Kutta method (A, b, c).

We arrive at conclusions about the possible orders p, r of the diagonally split Runge–Kutta method (1.4) under the requirement of unconditional contractivity in the maximum norm by studying the impact of this requirement on the accuracy of underlying Runge–Kutta method (A, b, c). We consider two orders of accuracy for (A, b, c). The first is the *classical order for scalar problems* (1.1), which we denote in this paper by q. It is defined in the same way as the classical order (cf. §1.3), but with the (obvious) assumption that problem (1.1) is scalar. The second order of accuracy that we consider for (A, b, c) is the *stage order*, which is defined (see, e.g., [6], [10]) as the largest integer $\tilde{p}$ such that the conditions $B(\tilde{p})$, $C(\tilde{p})$ hold, where

$$B(\tilde{p}): \qquad \sum_{j=1}^{\nu} b_j c_j^{k-1} = \frac{1}{k} \quad (\text{for } k = 1, 2, \ldots, \tilde{p}),$$

$$C(\tilde{p}): \qquad \sum_{j=1}^{\nu} a_{ij} c_j^{k-1} = \frac{1}{k} c_i^k \quad (\text{for } i = 1, 2, \ldots, \nu \text{ and } k = 1, 2, \ldots, \tilde{p}).$$

In §2 we first derive a (stability) condition on the underlying Runge–Kutta method (A, b, c) which is necessary for unconditional contractivity in the maximum norm of method (1.4). Next, we show that if the Runge–Kutta method (A, b, c) fulfils this condition, then it has orders q, $\tilde{p}$ satisfying $q \leq 4$, $\tilde{p} \leq 1$. Finally, as a consequence of these results, we obtain that if the diagonally split Runge–Kutta method (1.4) is unconditionally contractive in the maximum norm, then it has classical order $p \leq 4$ and order of B-consistency $r \leq 1$.

2. Unconditional contractivity of diagonally split Runge–Kutta methods.

2.1. Preliminaries. In this paper e_i denotes the ith unit vector in $\mathbb{R}^\nu$ (for $i = 1, 2, \ldots, \nu$) and e denotes the vector in $\mathbb{R}^\nu$ all of whose components equal 1. Further, I denotes the $\nu \times \nu$ identity matrix, and for any given numbers x_j $(j = 1, 2, \ldots, \nu)$ we denote by $\operatorname{diag}(x_1, x_2, \ldots, x_\nu)$ the $\nu \times \nu$ diagonal matrix with diagonal entries $x_1, x_2, \ldots, x_\nu$. We always assume that $x_j \in \mathbb{R}$ (for $j = 1, 2, \ldots, \nu$). We write $\operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0$ if $x_j \leq 0$ (whenever $1 \leq j \leq \nu$).

Throughout this paper we make the following (minor) assumptions on method (1.4):

(2.1.a) $(I - AX)$ is invertible whenever $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0$,

(2.1.b) $c_i = \sum_{j=1}^{\nu} a_{ij}$ (for $i = 1, 2, \ldots, \nu$),

(2.1.c) $c_i \neq c_j$ (whenever $i \neq j$).

We put $c_1 < c_2 < \cdots \prec c_\nu$. We note that condition (2.1.a) can be regarded as an assumption on the feasibility of method (1.4) in the case of scalar dissipative problems (1.1). More precisely, it can be seen that if (2.1.c) holds, then condition (2.1.a) is equivalent to requiring that the system of equations (1.5.b) always has a unique solution $y_1, y_2, \ldots, y_\nu$ whenever $h > 0$ and f is a scalar dissipative function (cf. Hundsdorfer [12, Rem. 4.3.7]).

2.2. A necessary condition on the underlying Runge–Kutta method (A, b, c). The following theorem gives a necessary condition for unconditional contractivity in the maximum norm of the diagonally split Runge-Kutta method (1.4).

THEOREM 2.1. *If method* (1.4) *is unconditionally contractive in the maximum norm, then the underlying Runge–Kutta method* (A, b, c) *satisfies*

$$\text{(2.2)} \qquad \begin{cases} |1 + b^{\mathrm{T}}X(I - AX)^{-1}e| + |b^{\mathrm{T}}X(I - AX)^{-1}d| \le 1 \\ \text{whenever } X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \le 0,\ d \in \mathbb{R}^\nu,\ \|d\| \le 1. \end{cases}$$

Proof. Let $\lambda : \mathbb{R} \to \mathbb{R}$ and $\mu : \mathbb{R} \to \mathbb{R}$ be arbitrary, but fixed, continuous functions with $\lambda(t) + |\mu(t)| \le 0$ (whenever $t \in \mathbb{R}$). Define $f : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$ by

$$f_1(t, y) = \lambda(t)\eta_1 + \mu(t)\eta_2, \quad f_2(t, y) = 0$$

(whenever $t \in \mathbb{R}$, $y = (\eta_1, \eta_2)^{\mathrm{T}} \in \mathbb{R}^2$). Then it can be verified (apply, e.g., Hairer, Nørsett, and Wanner [9, Thms. I.11.1, I.10.5] or Kraaijevanger [14, Thm. 5.1]) that $f \in \mathcal{F}$.

We consider application of method (1.4) in the case of problem (1.1) with f given above. We have (see (1.3))

$$F_1(t, y, z) = \lambda(t)\eta_1 + \mu(t)\zeta_2, \quad F_2(t, y, z) = 0$$

(whenever $t \in \mathbb{R}$, $y = (\eta_1, \eta_2)^{\mathrm{T}} \in \mathbb{R}^2$, $z = (\zeta_1, \zeta_2)^{\mathrm{T}} \in \mathbb{R}^2$). Applying (1.4.a) yields

$$\begin{aligned} u_{n,1} &= u_{n-1,1} + h\sum_{j=1}^{\nu} b_j\{\lambda(t_{n-1} + c_j h)y_{j,1} + \mu(t_{n-1} + c_j h)z_{j,2}\}, \\ u_{n,2} &= u_{n-1,2}\ , \end{aligned}$$

and (1.4.b), (1.4.c) yield

$$\begin{aligned} y_{i,1} &= u_{n-1,1} + h\sum_{j=1}^{\nu} a_{ij}\{\lambda(t_{n-1} + c_j h)y_{j,1} + \mu(t_{n-1} + c_j h)z_{j,2}\}, \\ z_{i,2} &= u_{n-1,2} \end{aligned}$$

(for $i = 1, 2, \ldots, \nu$). Here we have written $u_n = (u_{n,1}, u_{n,2})^{\mathrm{T}} \in \mathbb{R}^2$, $y_i = (y_{i,1}, y_{i,2})^{\mathrm{T}} \in \mathbb{R}^2$, etc. Let $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu)$ and $d = (d_1, d_2, \ldots, d_\nu)^{\mathrm{T}} \in \mathbb{R}^\nu$ be such that

$$\begin{aligned} x_j &= h\lambda(t_{n-1} + c_j h) \quad \text{(whenever } 1 \le j \le \nu\text{)}, \\ d_j &= \frac{\mu(t_{n-1} + c_j h)}{\lambda(t_{n-1} + c_j h)} \quad \text{(whenever } 1 \le j \le \nu,\ \lambda(t_{n-1} + c_j h) \ne 0\text{)}. \end{aligned}$$

Then, by using that $\lambda(t) = 0 \Rightarrow \mu(t) = 0$ (for any $t \in \mathbb{R}$), we obtain

$$\begin{aligned} u_{n,1} &= (1 + b^{\mathrm{T}}X(I - AX)^{-1}e)u_{n-1,1} + b^{\mathrm{T}}X(I - AX)^{-1}du_{n-1,2}\ , \\ u_{n,2} &= u_{n-1,2}\ . \end{aligned}$$

From Definition 1.1 and (2.1.c) it follows that condition (2.2) must hold. □

Condition (2.2) appears as an assumption in the unconditional contractivity result for method (1.4) by Bellen, Jackiewicz, and Zennaro [2, Thm. 4.4]. The condition (2.2) was investigated in Bellen and Zennaro [1] and Zennaro [16]. Following the terminology in [1], [2], [16], the Runge–Kutta method (A, b, c) is called $AN_f(0)$-*stable* whenever (2.2) is fulfilled. Observe that $AN_f(0)$-stability of (A, b, c) is equivalent to unconditional contractivity in the maximum norm of (A, b, c) on the class of dissipative functions considered in the proof of Theorem 2.1.

In the following we derive an algebraic criterion for condition (2.2). We remark that such a criterion has already been obtained by Bellen and Zennaro [1, Rem. 3.22] (see also [16, Rem. 4.11]). In our derivation we shall use ideas similar to those in [1], [16], but our formulation is different from [1], [16].

First, we have (cf. [1, Prop. 3.18], [16, Prop. 4.6]) Lemma 2.2.

LEMMA 2.2. *Condition* (2.2) *is equivalent to*

$$\text{(2.3)} \qquad \begin{cases} 1 + b^{\mathrm{T}}X(I - AX)^{-1}e \geq 0 \text{ and } b^{\mathrm{T}}X(I - AX)^{-1}e_i \leq 0 \\ \text{whenever } X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0,\ i = 1, 2, \ldots, \nu. \end{cases}$$

Proof. Let $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0$. Define $v^{\mathrm{T}} = (v_1, v_2, \ldots, v_\nu) = b^{\mathrm{T}}X(I - AX)^{-1}$. We have $|1 + v^{\mathrm{T}}e| + |v^{\mathrm{T}}d| \leq 1$ (whenever $d \in \mathbb{R}^\nu$, $\|d\| \leq 1$) if and only if $|1 + v^{\mathrm{T}}e| + \sum_{i=1}^{\nu} |v_i| \leq 1$. The latter condition is equivalent to $1 + v^{\mathrm{T}}e \geq 0$, $v_i \leq 0$ (for $i = 1, 2, \ldots, \nu$). □

For any matrix $M \in \mathbb{R}^{\nu\times\nu}$, let $\det[M]$ denote the determinant of M. For any vector $v \in \mathbb{R}^\nu$ and any $i \in \{1, 2, \ldots, \nu\}$, let $M \leftarrow_i v^{\mathrm{T}}$ denote the matrix obtained from M by replacing the ith row by v^{T}. Finally, for any nonempty set $S \subset \{1, 2, \ldots, \nu\}$, let $M(S)$ denote the submatrix of M that lies in the rows and columns indexed by S. Recall that a *principal minor* of M is the determinant of a matrix $M(S)$ (whenever $S \subset \{1, 2, \ldots, \nu\}$, S nonempty). We have (cf. Berman and Plemmons [3] for related results) the following lemma.

LEMMA 2.3. *Let* $M \in \mathbb{R}^{\nu\times\nu}$, $v \in \mathbb{R}^\nu$, *and* $i \in \{1, 2, \ldots, \nu\}$. *Then*

(a) *The following statements are equivalent:*

(i) $\det[I - MX] \geq 0$ *whenever* $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0$,

(ii) *all principal minors of* M *are nonnegative.*

(b) *The following statements are equivalent:*

(i) $\det[(I - MX) \leftarrow_i (v^{\mathrm{T}}X)] \leq 0$ *whenever* $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0$,

(ii) $\det[(M \leftarrow_i v^{\mathrm{T}})(S)] \geq 0$ *whenever* $S \subset \{1, 2, \ldots, \nu\}$, $i \in S$.

Proof. The proofs of parts a and b are analogous. We confine ourselves to proving part b below.

Assume (i). Choose $X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu)$ with $x_j = 0$ (for $j \notin S$) and $x_j = x$ (for $j \in S$). Then $\det[(I - MX) \leftarrow_i (v^{\mathrm{T}}X)] = \det[((I - xM) \leftarrow_i xv^{\mathrm{T}})(S)] \leq 0$. By considering the case $x \to -\infty$, it follows that $\det[(M \leftarrow_i v^{\mathrm{T}})(S)] \geq 0$.

Assume (ii). We have $\det[(I - MX) \leftarrow_i (v^{\mathrm{T}}X)] = \sum_S \det[((-MX) \leftarrow_i (v^{\mathrm{T}}X))(S)]$, where the summation is over all sets $S \subset \{1, 2, \ldots, \nu\}$ with $i \in S$. Each term in the summation satisfies $\det[((-MX) \leftarrow_i (v^{\mathrm{T}}X))(S)] = -\det[(M \leftarrow_i v^{\mathrm{T}})(S)] \cdot \det[-X(S)] \leq 0$, and thus (i) follows. □

Combining Lemmas 2.2 and 2.3, we obtain the following criterion.

THEOREM 2.4. *Condition* (2.2) *holds if and only if all principal minors of each of the matrices* $(A - eb^{\mathrm{T}})$, $(A \leftarrow_i b^{\mathrm{T}})$ $(i = 1, 2, \ldots, \nu)$ *are nonnegative.*

Proof. Using Cramer's rule (see, e.g., Horn and Johnson [11]), it can be verified that

$$1 + b^{\mathrm{T}}X(I - AX)^{-1}e = \frac{\det[I - (A - eb^{\mathrm{T}})X]}{\det[I - AX]},$$

$$b^{\mathrm{T}}X(I-AX)^{-1}e_i = \frac{\det[(I-AX) \leftarrow_i (b^{\mathrm{T}}X)]}{\det[I-AX]}$$

(for $i = 1, 2, \ldots, \nu$). From Lemma 2.2 and assumption (2.1.a) it follows that condition (2.2) holds if and only if

$$\begin{cases} \det[I-(A-eb^{\mathrm{T}})X] \geq 0 \text{ and } \det[(I-AX) \leftarrow_i (b^{\mathrm{T}}X)] \leq 0 \\ (\text{whenever } X = \operatorname{diag}(x_1, x_2, \ldots, x_\nu) \leq 0,\ i = 1, 2, \ldots, \nu). \end{cases}$$

By Lemma 2.3 this is equivalent to

$$\begin{cases} \text{all principal minors of } (A-eb^{\mathrm{T}}) \text{ are nonnegative, and for each } i \in \{1, 2, \ldots, \nu\} \\ \text{it holds that } \det[(A \leftarrow_i b^{\mathrm{T}})(S)] \geq 0 \text{ whenever } S \subset \{1, 2, \ldots, \nu\},\ i \in S. \end{cases}$$

If $i \in \{1, 2, \ldots, \nu\}$ and $S \subset \{1, 2, \ldots, \nu\}$ are such that $i \notin S$, then $\det[(A \leftarrow_i b^{\mathrm{T}})(S)] = \det[A(S)]$. We can write $\det[A(S)] = \det[(A-eb^{\mathrm{T}})(S)] + \sum_{j\in S} \det[(A \leftarrow_j b^{\mathrm{T}})(S)]$, and the equivalence of the theorem follows. □

Remark 2.5. The criterion of Theorem 2.4 guarantees that assumption (2.1.a) is fulfilled. This can be seen from $\det[I-AX] = 1 + \sum_S \det[A(S)] \cdot \det[-X(S)]$, where the summation is over all nonempty $S \subset \{1, 2, \ldots, \nu\}$, and by using the above expression for $\det[A(S)]$.

Remark 2.6. It is easily verified that the criterion of Bellen and Zennaro [1, Rem. 3.22] is equivalent to our criterion of Theorem 2.4.

Remark 2.7. Condition (2.2) also arises in the stability analysis of Runge-Kutta methods when adapted to initial value problems for delay differential equations; see Bellen and Zennaro [1], in 't Hout [13].

2.3. Order barriers for the underlying Runge–Kutta method (A, b, c). In this section we derive conclusions about the orders $\widetilde{p}$, q of the Runge-Kutta method (A, b, c) under the assumption of condition (2.2). The following lemma (see Berman and Plemmons [3, p. 149], Fiedler and Ptak [7]) forms a key result in our analysis.

LEMMA 2.8. *Let $M \in \mathbb{R}^{\nu\times\nu}$. Then the following statements are equivalent:*

(i) *all principal minors of M are nonnegative,*

(ii) *if $v = (v_1, v_2, \ldots, v_\nu)^{\mathrm{T}} \in \mathbb{R}^\nu \setminus \{0\}$ and $w = Mv = (w_1, w_2, \ldots, w_\nu)^{\mathrm{T}}$, then there exists an index $i \in \{1, 2, \ldots, \nu\}$ such that $v_i \neq 0$ and $v_i w_i \geq 0$.*

By a combination of Theorem 2.4 and the above lemma, we arrive at Theorem 2.9.

THEOREM 2.9. *Assume* (2.2) *holds. Then the stage order of the Runge–Kutta method (A, b, c) satisfies $\widetilde{p} \leq 1$. Furthermore, if $\widetilde{p} = 1$, then $c_\nu \geq 1$.*

Proof. Suppose $\widetilde{p} \geq 2$. Let M be the matrix in $\mathbb{R}^{\nu\times\nu}$ with ith row equal to the ith row of $(A-eb^{\mathrm{T}})$ (whenever $c_i \leq 1$) and equal to b^{T} (whenever $c_i > 1$). Then from Theorem 2.4 it is easily seen that all principal minors of M are nonnegative. Next, let $v = (v_1, v_2, \ldots, v_\nu)^{\mathrm{T}} = c-e$ and $w = Mv = (w_1, w_2, \ldots, w_\nu)^{\mathrm{T}}$. From the conditions $B(2)$, $C(2)$ (see §1.4) it follows that $v \neq 0$, and $v_i w_i = \frac{1}{2}(c_i - 1)^3$ (whenever $c_i \leq 1$), $v_i w_i = -\frac{1}{2}(c_i - 1)$ (whenever $c_i > 1$). If $i \in \{1, 2, \ldots, \nu\}$ is such that $v_i \neq 0$, then $v_i w_i < 0$. By Lemma 2.8, this is a contradiction. Consequently, $\widetilde{p} \leq 1$.

If $\widetilde{p} = 1$, then a similar argument as above, but with $M = A - eb^{\mathrm{T}}$ and $v = e$, yields that $c_\nu \geq 1$. □

There exist Runge-Kutta methods that satisfy (2.1), (2.2) and have stage order $\widetilde{p} = 1$. Examples of such methods are given by

$$\begin{array}{c|c} \theta & \theta \\ \hline & 1 \end{array}, \tag{2.4}$$

where $\theta \geq 1$. This is easily verified by using the criterion of Theorem 2.4. Moreover, it follows that the condition $\theta \geq 1$ is necessary for (2.1), (2.2).

Since $q \geq 1$ implies $\tilde{p} \geq 1$, we immediately obtain Corollary 2.10.

COROLLARY 2.10. *Assume* (2.2) *holds and the classical order of the Runge–Kutta method* (A, b, c) *for scalar problems* (1.1) *satisfies* $q \geq 1$. *Then,* $c_\nu \geq 1$.

We note that, for the case where the matrix A is invertible, the result of Corollary 2.10 also follows from Zennaro [16, Thm. 4.12].

THEOREM 2.11. *Assume* (2.2) *holds and* $b_j \neq 0$ *(for* $j = 1, 2, \ldots, \nu$*). Then the classical order of the Runge–Kutta method* (A, b, c) *for scalar problems* (1.1) *satisfies* $q \leq 4$.

Proof. Suppose $q \geq 5$. From the order conditions for Runge-Kutta methods in the case of scalar problems (1.1) (see Butcher [5], Hairer, Nørsett, and Wanner [9, p. 154]), and by using assumption (2.1.b), it is easily seen that the following identity due to Butcher (see, e.g., [9, p. 186]) is still valid:

$$\sum_{j=1}^{\nu} b_j \left(\sum_{k=1}^{\nu} a_{jk} c_k - \frac{c_j^2}{2} \right)^2 = 0.$$

From Theorem 2.4 we obtain that $b_j > 0$ (for $j = 1, 2, \ldots, \nu$), and, consequently, condition $C(2)$ must hold. Further, the order conditions [5] imply that $B(2)$ must hold. Hence, the stage order $\tilde{p} \geq 2$, but this contradicts the previous theorem. Therefore, $q \leq 4$. □

We recall that the classical order of the Runge-Kutta method (A, b, c) equals the classical order q for scalar problems (1.1) whenever $q \leq 4$ (see Butcher [5]). The Runge-Kutta methods (2.4), which satisfy the conditions (2.1), (2.2), all have a classical order equal to 1. Zennaro [16] constructed the following Runge-Kutta method, which satisfies the conditions (2.1), (2.2), and has classical order 3:

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

At this moment it is not known if the upper bound $q = 4$, given by Theorem 2.11, is attained by some Runge-Kutta method.

Remark 2.12. The question arises whether the assumption in Theorem 2.11 that $b_j \neq 0$ (for $j = 1, 2, \ldots, \nu$) can be replaced by the assumption that the Runge-Kutta method (A, b, c) is DJ-irreducible. We recall that the Runge-Kutta method (A, b, c) is called *DJ-reducible* (see, e.g., [6], [10]) if there exist disjoint sets S, T with S nonempty and $S \cup T = \{1, 2, \ldots, \nu\}$ such that $b_j = 0$ (whenever $j \in S$) and $a_{ij} = 0$ (whenever $i \in T$, $j \in S$). Otherwise, the method is called *DJ-irreducible*. If a Runge-Kutta method is DJ-reducible it is equivalent to a Runge-Kutta method with a number of stages $\nu^* < \nu$. Here, ν^* is equal to the number of elements in the set T. Our question concerning Theorem 2.11 is still open. We remark that, contrary to other cases of (stability) conditions for Runge-Kutta methods (see, e.g., [6], [10], [14]), the assumption of DJ-irreducibility in the case of condition (2.2) does *not* automatically imply that $b_j \neq 0$ (for $j = 1, 2, \ldots, \nu$). This can be seen from the family of Runge-Kutta methods given by

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

where $\theta > 0$. Note that all of these methods are of a classical order at most equal to 2.

2.4. Order barriers for the diagonally split Runge–Kutta method (1.4). In the following we turn our attention to the diagonally split Runge-Kutta method (1.4). Before formulating the main result of this paper, we show Lemma 2.13.

LEMMA 2.13. *Assume method* (1.4) *is unconditionally contractive in the maximum norm and* $b_j \neq 0$ *(for* $j = 1, 2, \ldots, \nu$*). Then method* (1.4) *is B-consistent of order* r *(on the problem class* (1.6)*) with* $r = \tilde{p}$*, where* $\tilde{p}$ *is the stage order of the underlying Runge–Kutta method* (A, b, c).

Proof. The local error $\delta_n(h)$ of method (1.4), in the case of problem (1.6), can be written in the form (cf., e.g., Burrage and Hundsdorfer [4], Hairer and Wanner [10])

$$\delta_n(h) = \Delta_0(h) + b^{\mathrm{T}} X (I - AX)^{-1} \Delta(h),$$

where $X = \operatorname{diag}(h\lambda(t_{n-1}+c_1h), h\lambda(t_{n-1}+c_2h), \ldots, h\lambda(t_{n-1}+c_\nu h)) \leq 0$ and $\Delta_0(h)$, $\Delta(h) = (\Delta_1(h), \Delta_2(h), \ldots, \Delta_\nu(h))^{\mathrm{T}}$ are given by

$$g(t_n) = g(t_{n-1}) + h \sum_{j=1}^{\nu} b_j g'(t_{n-1} + c_j h) + \Delta_0(h),$$

$$g(t_{n-1} + c_i h) = g(t_{n-1}) + h \sum_{j=1}^{\nu} a_{ij} g'(t_{n-1} + c_j h) + \Delta_i(h)$$

(for $i = 1, 2, \ldots, \nu$).

Theorem 2.1 yields that $\sum_{i=1}^{\nu} |b^{\mathrm{T}} X (I - AX)^{-1} e_i| \leq 1$, and, hence,

$$|\delta_n(h)| \leq \|\Delta(h)\| + |\Delta_0(h)|.$$

From the Taylor series expansion and the conditions $B(\tilde{p})$, $C(\tilde{p})$ it follows that

$$\Delta_i(h) = \mathcal{O}(h^{\tilde{p}+1}) \quad (\text{for } h > 0,\ i = 0, 1, \ldots, \nu),$$

where the $\mathcal{O}$-constant depends only on the coefficients of the Runge-Kutta method (A, b, c) and on an upper bound for $|g^{(\tilde{p}+1)}(t)|$ $(t \in \mathbb{R})$. Consequently, $r \geq \tilde{p}$.

The fact that $r = \tilde{p}$ follows from a straightforward adaptation of ideas in Burrage and Hundsdorfer [4] by choosing $g(t) = t^{\tilde{p}+1}$ (for $t \in \mathbb{R}$). Here, the assumptions (2.1.c), $b_j \neq 0$ (for $j = 1, 2, \ldots, \nu$) are used. □

THEOREM 2.14. *Consider a given method* (1.4) *with the properties* (2.1), $b_j \neq 0$ *(for* $j = 1, 2, \ldots, \nu$*). Assume the method is unconditionally contractive in the maximum norm. Then the classical order* p *satisfies* $p \leq 4$*, and the order of B-consistency* r *(on the problem class* (1.6)*) satisfies* $r \leq 1$*. Furthermore, if* $p \geq 1$ *or* $r = 1$*, then* $c_\nu \geq 1$.

Proof. The theorem is a consequence of Theorems 2.1, 2.9, Corollary 2.10, Theorem 2.11, Lemma 2.13, and the fact that $p \leq q$, where q is the classical order of the underlying Runge-Kutta method for scalar problems (1.1). □

The bound $r \leq 1$ in Theorem 2.14 is sharp. The value $r = 1$ is attained by the diagonally split Runge-Kutta methods (1.4) that coincide with the Runge-Kutta methods given by (2.4). The result that these methods are unconditionally contractive in the maximum norm can be found in, e.g., Kraaijevanger [14].

We do not know whether the bound $p \leq 4$ is sharp. Clearly, classical order $p = 1$ can be achieved, viz. by the methods (2.4). The results by Bellen, Jackiewicz, and Zennaro [2] make plausible that also orders $p = 2$, $p = 3$ can be achieved.

Remark 2.15. Theorem 2.14 extends to a class of methods much more general than (1.4). For example, let F be any (splitting) function that satisfies the (natural) conditions

(i) $F : \mathbb{R} \times \mathbb{R}^s \times \mathbb{R}^s \to \mathbb{R}^s$,

(ii) $F(t, y, y) = f(t, y)$ (whenever $t \in \mathbb{R}$, $y \in \mathbb{R}^s$),

(iii) $f_j(t, y) \equiv 0 \Rightarrow F_j(t, y, z) \equiv 0$ (whenever $j = 1, 2, \ldots, s$),

(iv) $F_j(t, y, z)$ is independent of the jth component of z (whenever $t \in \mathbb{R}$, $y \in \mathbb{R}^s$, $z \in \mathbb{R}^s$, $j = 1, 2, \ldots, s$).

Then, it is easily seen that (1.4) still reduces to the Runge–Kutta method (A, b, c) whenever the initial value problem (1.1) is scalar. Therefore, the order relations $p \leq q$, $r = \tilde{p}$ (see Lemma 2.13) remain valid. Further, it is easily verified that the method reduces to (A, b, c) whenever (1.1) is of the type considered in the proof of Theorem 2.1. Therefore, also Theorem 2.1 remains valid, and, consequently, the result of Theorem 2.14 still holds.

A further generalization of method (1.4) is obtained by considering formulas for z_i other than (1.4.c). If the following (natural) condition is satisfied,

(v) $f_j(t, y) \equiv 0 \Rightarrow z_{i,j} = u_{n-1,j}$ (whenever $i = 1, 2, \ldots, \nu$ and $j = 1, 2, \ldots, s$),

where $u_{n-1,j}$, $z_{i,j}$ stand for the jth components of u_{n-1}, z_i, respectively, then it is easily seen that Theorem 2.14 still holds.

Acknowledgment. I would like to thank M. N. Spijker and J. F. B. M. Kraaijevanger for valuable discussions on the topic of this paper.

REFERENCES

[1] A. Bellen and M. Zennaro, *Strong contractivity properties of numerical methods for ordinary and delay differential equations*, Appl. Numer. Math., 9 (1992), pp. 321–346.

[2] A. Bellen, Z. Jackiewicz, and M. Zennaro, *Contractivity of waveform relaxation Runge–Kutta iterations and related limit methods for dissipative systems in the maximum norm*, SIAM J. Numer. Anal., 31 (1994), pp. 499–523.

[3] A. Berman and R. J. Plemmons, *Nonnegative Matrices in the Mathematical Sciences*, Academic Press, New York, 1979.

[4] K. Burrage and W. H. Hundsdorfer, *The order of B-convergence of algebraically stable Runge–Kutta methods*, BIT, 27 (1987), pp. 62–71.

[5] J. C. Butcher, *On the integration processes of A. Huta*, J. Austral. Math. Soc., 3 (1963), pp. 202–206.

[6] K. Dekker and J. G. Verwer, *Stability of Runge–Kutta Methods for Stiff Nonlinear Differential Equations*, North-Holland, Amsterdam, 1984.

[7] M. Fiedler and V. Ptak, *On matrices with non-positive off-diagonal elements and positive principal minors*, Czechoslovak. Math. J., 12 (1962), pp. 382–400.

[8] R. Frank, J. Schneid, and C. W. Ueberhuber, *Order results for implicit Runge–Kutta methods applied to stiff systems*, SIAM J. Numer. Anal., 22 (1985), pp. 515–534.

[9] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations* I, 2nd ed., Springer-Verlag, Berlin, 1993.

[10] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations* II, Springer-Verlag, Berlin, 1991.

[11] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, 1990.

[12] W. H. Hundsdorfer, *The numerical solution of nonlinear stiff initial value problems: An analysis of one step methods*, CWI Tract, 12 (1985).

[13] K. J. in 't Hout, *Runge–Kutta Methods in the Numerical Solution of Delay Differential Equations*, thesis, Stellingen, Dept. Math. and Comp. Sci., Univ. of Leiden, 1992.

[14] J. F. B. M. Kraaijevanger, *Contractivity of Runge–Kutta methods*, BIT, 31 (1991), pp. 482–528.

[15] M. N. Spijker, *Contractivity in the numerical solution of initial value problems*, Numer. Math., 42 (1983), pp. 271–290.

[16] M. Zennaro, *Contractivity of Runge–Kutta methods with respect to forcing terms*, Appl. Numer. Math., 10 (1993), pp. 321–345.