MATHEMATICS OF COMPUTATION
Volume 80, Number 274, April 2011, Pages 863–886
S 0025-5718(2010)02422-6
Article electronically published on September 21, 2010

BOUNDEDNESS AND STRONG STABILITY OF RUNGE-KUTTA METHODS

W. HUNDSDORFER AND M. N. SPIJKER

Abstract. In the literature, much attention has been paid to Runge-Kutta methods (RKMs) satisfying special nonlinear stability requirements indicated by the terms total-variation-diminishing (TVD), strong stability preserving (SSP) and monotonicity. Stepsize conditions, guaranteeing these properties, were derived by Shu and Osher [J. Comput. Phys., 77 (1988) pp. 439-471] and in numerous subsequent papers. These special stability requirements imply essential boundedness properties for the numerical methods, among which the property of being total-variation-bounded. Unfortunately, for many RKMs, the above special requirements are violated, so that one cannot conclude in this way that the methods are (total-variation) bounded.

In this paper, we study stepsize-conditions for boundedness directly, rather than via the detour of the above special stability properties. We focus on stepsize-conditions which are optimal, in that they are not unnecessarily restrictive. We find that, in situations where the special stability properties mentioned above are violated, boundedness can be present only within a class of very special RKMs.

As a by-product, our analysis sheds new light on the known theory of monotonicity for RKMs. We obtain separate results for internal and external monotonicity, as well as a new proof of the fundamental relation between monotonicity and Kraaijevanger's coefficient. This proof distinguishes itself from older ones in that it is shorter and more transparent, while it requires simpler assumptions on the RKMs under consideration.

1. Introduction

1.1. Monotonicity and boundedness of Runge-Kutta methods.

1.1.1. *Runge-Kutta methods.* In this paper we deal with initial value problems, for systems of ordinary differential equations, which can be written in the form

$$\frac{d}{dt}u(t) = F(u(t)) \quad (t \geq 0), \quad u(0) = u_0. \tag{1.1}$$

We study Runge-Kutta methods (RKMs) for computing numerical approximations u_n to the true solution values $u(n\Delta t)$, where Δt denotes a positive stepsize and

Received by the editor September 28, 2009 and, in revised form, February 16, 2010.
2010 *Mathematics Subject Classification.* Primary 65L05, 65L06, 65L20, 65M20.
Key words and phrases. Initial value problem, method of lines (MOL), ordinary differential equation (ODE), Runge-Kutta method (RKM), total-variation-diminishing (TVD), strong-stability-preserving (SSP), monotonicity, total-variation-bounded (TVB), boundedness.

$n = 1, 2, 3, \ldots$. The general Runge-Kutta method can be written in the form

$$y_i^{[n]} \;=\; u_{n-1} \;+\; \Delta t \cdot \sum_{j=1}^{s} a_{ij}\, F(y_j^{[n]}) \quad (1 \le i \le s), \tag{1.2a}$$

$$u_n \;=\; u_{n-1} \;+\; \Delta t \cdot \sum_{j=1}^{s} b_j\, F(y_j^{[n]}). \tag{1.2b}$$

Here a_{ij}, b_j are parameters defining the method. Furthermore, $y_i^{[n]}$ $(1 \le i \le s)$ are *internal approximations* used for computing the *external approximation* u_n from u_{n-1} $(n = 1, 2, 3, \ldots)$; cf. e.g. Butcher [1, 2] or Hairer, Nørsett and Wanner [12]. If $a_{ij} = 0$ (for all $j \ge i$), the method is called *explicit*.

1.1.2. *Monotonicity.* In the following, $\mathbb{V}$ stands for the vector space on which the differential equation is defined, and $\|\cdot\|$ denotes a seminorm on $\mathbb{V}$ (i.e.: $\|u + v\| \le \|u\| + \|v\|$ and $\|\lambda\, v\| = |\lambda|\, \|v\|$ for all $u, v \in \mathbb{V}$ and real λ). Much attention has been paid, in the literature, to the following bounds on the internal and external approximations, respectively,

$$\|y_i^{[n]}\| \;\le\; \|u_{n-1}\| \quad (\text{for } 1 \le i \le s), \tag{1.3}$$

$$\|u_n\| \le \|u_{n-1}\|. \tag{1.4}$$

In the literature, the above bounds are often referred to by the term *monotonicity* or *strong stability*. In the following, we will distinguish between (1.3) and (1.4) by using the terms *internal monotonicity* and *external monotonicity*, respectively.

Inequalities (1.3), (1.4) are of particular importance in situations where (1.1) results from (method of lines) semidiscretizations of time-dependent partial differential equations. Choices for $\|\cdot\|$ that occur in that context include e.g. the *supremum norm* $\|x\| = \|x\|_\infty = \sup_i |\xi_i|$ and the *total variation seminorm* $\|x\| = \|x\|_{TV} = \sum_i |\xi_{i+1} - \xi_i|$ (for vectors x with components ξ_i). Numerical processes, satisfying $\|u_n\|_{TV} \le \|u_{n-1}\|_{TV}$, play a special role in the solution of hyperbolic conservation laws and are called *total-variation-diminishing* (TVD); cf. e.g. Harten [14], Shu [27], Shu and Osher [29], LeVeque [25], Hundsdorfer and Verwer [21].

In the literature, conditions on Δt can be found which guarantee (1.3), (1.4). In many papers, one starts from an assumption about F which, for given $\tau_0 > 0$, essentially amounts to

$$F :\; \mathbb{V} \to \mathbb{V}, \quad \text{with } \|v + \tau_0\, F(v)\| \;\le\; \|v\| \quad (\text{for all } v \in \mathbb{V}). \tag{1.5}$$

Assumption (1.5) means that the forward Euler method is monotonic with stepsize τ_0. It can be interpreted as a condition on the manner in which the semidiscretization is performed in the case that $\frac{d}{dt}u(t) = F(u(t))$ stands for a semidiscrete version of a partial differential equation.

For classes of RKMs, positive *stepsize-coefficients* γ were determined, such that monotonicity, in the sense of (1.3), (1.4), is present for all Δt with

$$0 < \Delta t \le \gamma \cdot \tau_0; \tag{1.6}$$

see e.g. Shu and Osher [29], Gottlieb, Shu and Tadmor [11], Shu [28], Spiteri and Ruuth [33], Ferracina and Spijker [4, 5], Higueras [15, 16], Gottlieb [8], Ruuth [26], Spijker [32, Section 3.2.1], Gottlieb, Ketcheson and Shu [8].

1.1.3. *Boundedness.* For total-variation-diminishing processes, there is trivially the *total-variation-boundedness* (TVB) property, in that a finite μ (independent of $N \geq 1$) exists such that

$$\|u_N\|_{TV} \leq \mu \cdot \|u_0\|_{TV}, \tag{1.7}$$

for the approximation u_N obtained by applying method (1.2) for $n = 1, \ldots, N$. In the solution of hyperbolic conservation laws, condition (1.7) is of crucial importance for suitable convergence properties when $\Delta t \to 0$, and it constitutes one of the underlying reasons why attention has been paid in the literature to (1.3), (1.4); cf. e.g. LeVeque [25] or Hundsdorfer and Verwer [21].

Unfortunately, there are well-known RKMs, with a record of practical success, for which there exist *no positive stepsize-coefficients* γ such that (1.5), (1.6) imply (1.4); e.g. for the Dormand-Prince formula, cf. e.g. Butcher [2, p. 194], Hairer, Nørsett and Wanner [12, p. 171]. Moreover, no second-order (implicit) RKMs exist with $\gamma = \infty$; see e.g. Spijker [30, Sections 2.2, 3.2].

These circumstances suggest that there are situations where monotonicity may be too strong a theoretical demand, and that it is worthwhile to study, along with monotonicity, also directly the weaker *boundedness requirements*

$$\|y_i^{[N]}\| \leq \mu \cdot \|u_0\| \quad (\text{for } 1 \leq i \leq s), \tag{1.8}$$

$$\|u_N\| \leq \mu \cdot \|u_0\|, \tag{1.9}$$

for vectors $y_i^{[N]}$, u_N obtained by applying method (1.2) for $n = 1, 2, \ldots, N$. Here μ stands for a finite constant (independent of $N \geq 1$) which is allowed to be greater than 1. Requirement (1.9), with $\|\cdot\| = \|\cdot\|_{TV}$, still implies the TVB-property, which highlights the importance of studying (1.9). For the just-mentioned Dormand-Prince method, e.g., it is an open question if positive γ and finite μ exist such that (1.5), (1.6) would imply (1.8) or (1.9).

1.2. **Scope of the paper.**

1.2.1. *An approach suitable for studying boundedness without monotonicity.* In this paper, we study the largest factor by which, under conditions (1.5), (1.6), the quantities $\max_i \|y_i^{[n]}\|$ and $\|u_n\|$ can exceed $\|u_{n-1}\|$. We study also the maximal factor by which, under the same conditions, the quantities $\max_i \|y_i^{[N]}\|$ and $\|u_N\|$ can exceed $\|u_0\|$ (for any $N \geq 1$).

This approach makes it possible to settle, for RKMs of a fairly general type, the interesting question of whether boundedness can be present, when there exists *no* positive γ with the property that (1.5), (1.6) imply (1.3), (1.4).

Besides being useful for studying boundedness, the study of the largest factors, mentioned above, will shed new light on the existing monotonicity theory given in the literature: we will find separate results for external and internal monotonicity, as well as a new proof of the fundamental relation between monotonicity and Kraaijevanger's coefficient (introduced in Kraaijevanger [24]). We believe this proof is more natural and shorter than the classical one; moreover, the proof requires a simpler irreducibility condition than usually imposed in the literature.

For completeness we note that the boundedness of RKMs was studied earlier in Spijker [31], Ferracina and Spijker [6]. But, the first of these papers is only relevant to seminorms $\|\cdot\|$ generated by (pseudo) inner products, excluding, e.g., the seminorm $\|\cdot\|_{TV}$, whereas in the second paper the focus is on establishing bounds under weaker conditions than (1.5). As a result, the analysis in the present paper is largely different from the one in the papers just mentioned.

1.2.2. *Organization of the paper.* After introducing in Section 2 some notation and definitions needed in the subsequent sections, we present in Section 3 two theorems which play a central role in this paper, Theorems 3.1 and 3.2.

Theorem 3.1 gives explicit expressions for the largest factors by which, under conditions (1.5), (1.6), the quantities $\max_i\|y_i^{[n]}\|$ and $\|u_n\|$, respectively, can exceed $\|u_{n-1}\|$. These expressions are denoted by $\varphi(\gamma)$ and $\psi(\gamma)$, respectively. They depend only on γ and the coefficients a_{ij}, b_j of the RKM.

Theorem 3.2 specifies, for RKMs of a general type, the maximal factors by which, under conditions (1.5), (1.6), the quantities $\max_i\|y_i^{[N]}\|$ and $\|u_N\|$ can exceed $\|u_0\|$ (for any $N \geq 1$). Explicit expressions for these factors are given in terms of $\varphi(\gamma)$ and $\psi(\gamma)$.

In Section 4, monotonicity and boundedness are studied by making use of Theorems 3.1, 3.2. Section 4.1 gives Lemma 4.1, which relates the conditions $\varphi(\gamma) = 1$ and $\psi(\gamma) = 1$ to each other. This lemma is quite useful when applying Theorems 3.1, 3.2 in the subsequent Sections 4.2, 4.3.

In Section 4.2 we give four corollaries to Theorems 3.1, 3.2. The Corollaries 4.2, 4.4 characterize stepsize-coefficients γ for monotonicity and boundedness, respectively. Corollary 4.3 states that, for irreducible RKMs, a value γ cannot be a stepsize-coefficient for external monotonicity without being at the same time a stepsize-coefficient for internal monotonicity. Corollary 4.5 is relevant to the question, mentioned above, of whether a positive stepsize-coefficient γ for boundedness can exist in situations where no positive γ exists for monotonicity. The corollary reveals the surprising fact that for all irreducible RKMs which are *not* of a very special type, any stepsize-coefficient γ for boundedness must at the same time be a stepsize-coefficient for monotonicity.

In Section 4.3, we relate the conditions $\psi(\gamma) = 1$ and $\varphi(\gamma) = 1$, respectively, to Kraaijevanger's coefficient $r(A, B)$ and to a closely related coefficient $r(A)$. Theorem 4.9 gives a new formal characterization of these coefficients, whereas Corollary 4.10 highlights the relevance of the coefficients to monotonicity and boundedness. This corollary covers and supplements the fundamental relation between external monotonicity and Kraaijevanger's coefficient, as stated earlier in the literature; cf. Ferracina and Spijker [4, 5], Higueras [15, 16], Ketcheson [22], Spijker [32, Section 3.2.1]. In our opinion, the derivation of this relation via the framework of the present paper (notably Theorem 3.1) is shorter and more transparent than the one in the literature, cf. Kraaijevanger [24], Ferracina and Spijker [4].

In Section 5, we shortly illustrate the material of Sections 3, 4, in the analysis of concrete RKMs. In Section 5.1, we settle for various RKMs without positive stepsize-coefficients γ for monotonicity, among which is the Dormand-Prince method mentioned above, the question of whether positive γ exist corresponding to boundedness. In Section 5.2, we consider a special example of an irreducible RKM, with a positive stepsize-coefficient γ for boundedness, but no positive γ for

monotonicity. Furthermore, we give counterexamples showing that various basic assumptions, made in Sections 3, 4, cannot be omitted.

In Section 6 we give proofs, needed to complete the material of the preceding sections. Parts of the proofs of Theorems 3.1, 3.2 and Lemma 4.1 are not straightforward and make use of common lemmas, viz. Lemmas 6.1, 6.2, 6.3. For this reason, and also not to interrupt the presentation in Sections 3 and 4 too much, we have collected these parts of the proofs in the separate Section 6.

In Section 7 we conclude by summarizing the main findings of the paper.

2. Preliminaries

2.1. **Notation to be used throughout the paper.** In all of the following we denote by A the $s \times s$ matrix made up of the coefficients a_{ij} of the RKM (1.2), and by B the $1 \times s$ matrix, made up of the coefficients b_j, so that the method can be characterized completely by the $(s+1) \times s$ matrix $\binom{A}{B}$.

We denote the $s \times 1$ matrix, with all entries equal to 1, by E.

For any $k \geq 1$, we denote the $k \times k$ identity matrix by I. For vectors $x \in \mathbb{R}^k$ with components ξ_i, we define $\|x\|_\infty = \max_i |\xi_i|$.

For given matrix $M = (m_{ij})$, we put $\|M\|_\infty = \max_{x \neq 0} \frac{\|Mx\|_\infty}{\|x\|_\infty}$ and we recall the well-known formula $\|M\|_\infty = \max_i \sum_j |m_{ij}|$.

We shall use the notation $M(i,:)$ and $M(:,j)$ to denote the *i-th row and j-th column*, respectively, of the matrix M. We define $|M| = (|m_{ij}|)$, and denote the *spectral radius* of square matrices M by $\mathrm{spr}(M)$.

Any inequalities between matrices with the same dimensions, say $Q = (q_{ij})$ and $R = (r_{ij})$, have to be interpreted entry-wise; i.e., $Q \leq R$ means that all $q_{ij} \leq r_{ij}$.

2.2. **Reducibility.** We shall make use of reducibility concepts for RKMs, corresponding to the following definition.

Definition 2.1 (Reducibility and irreducibility concepts)**.**

(a) The RKM (1.2) is called DJ-reducible if there exist disjoint index sets $\mathcal{M}$ and $\mathcal{N}$, where $\mathcal{N}$ is nonempty and $\mathcal{M} \cup \mathcal{N} = \{1, 2, ..., s\}$, such that $b_j = 0 \quad (\text{for } j \in \mathcal{N})$ and $a_{ij} = 0 \quad (\text{for } i \in \mathcal{M},\ j \in \mathcal{N})$. If such sets do not exist, the RKM is called DJ-irreducible.

(b) The RKM (1.2) is called row-reducible if there exist indices $i \neq j$ in $\{1, \ldots, s\}$, such that $A(i,:) = A(j,:)$. If such indices do not exist, the RKM is called row-irreducible.

(c) The RKM (1.2) is called reducible if it is DJ-reducible or row-reducible or both. If the RKM is neither DJ-reducible nor row-reducible, it is called irreducible.

In case there exist sets $\mathcal{M}$, $\mathcal{N}$ as in part (a) above, the vectors $y_j^{[n]}$ in (1.2) with $j \in \mathcal{N}$ have no influence on u_n, so that the Runge-Kutta method is equivalent to a method with less than s stages. Also in case indices i, j exist as in (b) above, the Runge-Kutta method essentially reduces to a method with less then s stages. For more details about the definition in part (a), see e.g. [2, Sect. 381], [3], [24]; for part (b), cf. e.g. [31] and [32, Sect. 3.2.1].

Clearly, from a practical point of view, it is enough to consider only Runge-Kutta methods which are irreducible (in the sense of (c)).

3. Estimating $\|y_i^{[n]}\|$ and $\|u_n\|$

3.1. **The effect of one Runge-Kutta step.** For values γ such that the inverses occurring below exist, we shall use the notation

$$C = (c_{ij}) = \gamma\, A\,(I + \gamma\, A)^{-1}, \quad D = (d_j) = \gamma\, B\,(I + \gamma\, A)^{-1}, \tag{3.1}$$
$$Z = (z_i) = (I - |C|)^{-1}|E - C\,E|.$$

Although the matrices C, D, Z depend on γ, we do not express this dependence explicitly so as to keep our notation simple; in subsequent formulas involving C, D or Z the dependence of the formulas on γ could always be made clear by substituting the expressions (3.1) into the formulas at hand.

The following functions φ, ψ will play a main role in our approach to monotonicity and boundedness:

$$\text{(3.2a)}\quad \varphi(\gamma) = \begin{cases} \|Z\|_\infty & \text{if } I + \gamma\, A \text{ is invertible, and } \mathrm{spr}(|C|) < 1, \\ \infty & \text{otherwise,} \end{cases}$$

$$\text{(3.2b)}\quad \psi(\gamma) = \begin{cases} |1 - D\,E| + |D|\,Z & \text{if } I + \gamma\, A \text{ is invertible, and } \mathrm{spr}(|C|) < 1, \\ \infty & \text{otherwise.} \end{cases}$$

Theorem 3.1, below, describes the effect on $\|y_i^{[n]}\|$ and $\|u_n\|$ of carrying out one step of method (1.2). It gives conditions on the factors α, β in order that the following two basic properties are present:

(3.3a) The estimate $\max_i \|y_i^{[n]}\| \le \alpha \cdot \|u_{n-1}\|$ holds whenever $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and $y_i^{[n]}$ is generated from u_{n-1} under conditions (1.5), (1.6).

(3.3b) The estimate $\|u_n\| \le \beta \cdot \|u_{n-1}\|$ holds whenever $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and u_n is generated from u_{n-1} under conditions (1.5), (1.6).

Theorem 3.1 (Expressing the effect of one Runge-Kutta step in terms of $\varphi(\gamma)$ and $\psi(\gamma)$)**.** *Let $\gamma > 0$ be given. Then the following two statements hold:*

(I) Property (3.3a) *is valid with $\alpha = \varphi(\gamma)$. Moreover, when the RKM is row-irreducible, then* (3.3a) *is* not *valid with any $\alpha < \varphi(\gamma)$.*

(II) Property (3.3b) *is valid with $\beta = \psi(\gamma)$. Moreover, when the RKM is irreducible, then* (3.3b) *is* not *valid with any $\beta < \psi(\gamma)$.*

In the theorem, as well as throughout the rest of the paper, we have used the convention $\infty \cdot x = \infty$ (for all real $x \ge 0$). Due to this convention, properties (3.3a) and (3.3b) make also sense (and are trivially present) in case $\alpha = \infty$ or $\beta = \infty$, respectively. Note that these cases actually occur in statement (I), (II) when $I + \gamma A$ is not invertible or $\mathrm{spr}(|C|) \ge 1$.

Clearly, for irreducible RKMs, the theorem shows that the estimates

$$\max_i \|y_i^{[n]}\| \le \varphi(\gamma) \cdot \|u_{n-1}\|, \qquad \|u_n\| \le \psi(\gamma) \cdot \|u_{n-1}\| \tag{3.4}$$

are *optimal*, in that the factors $\varphi(\gamma)$, $\psi(\gamma)$ cannot be replaced by any smaller ones in the general situation (1.5), (1.6).

In the proof of Theorem 3.1, as well as in the rest of the paper, we shall write (1.2a) and similar relations often more concisely, by using the following notation relevant to the vector space $\mathbb{V}$.

For any integer $k \geq 1$, we denote the vector in $\mathbb{V}^k$, with components $x_1, \ldots, x_k \in \mathbb{V}$, by

$$x = [x_i] = \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{V}^k.$$

Corresponding to any $p \times k$ matrix $M = (m_{ij})$, we define a linear operator $\boldsymbol{M}$ from $\mathbb{V}^k$ to $\mathbb{V}^p$ by $\boldsymbol{M}(x) = y$, with $y = [y_i] \in \mathbb{V}^p$, $y_i = \sum_{j=1}^k m_{ij}\, x_j$ $(1 \leq i \leq p)$ for $x = [x_i] \in \mathbb{V}^k$. When $F : \mathbb{V} \to \mathbb{V}$, we define $\boldsymbol{F} : \mathbb{V}^k \to \mathbb{V}^k$ by $\boldsymbol{F}(y) = [F(y_i)] \in \mathbb{V}^k$ for $y = [y_i] \in \mathbb{V}^k$.

Combining the vectors $y_i^{[n]}$, occurring in (1.2a), into the vector $y^{[n]} = [y_i^{[n]}] \in \mathbb{V}^s$, the relations (1.2) can thus be written compactly as

$$y^{[n]} = \boldsymbol{E}\, u_{n-1} + \Delta t \cdot \boldsymbol{A}\, \boldsymbol{F}(y^{[n]}), \tag{3.5a}$$

$$u_n = u_{n-1} + \Delta t \cdot \boldsymbol{B}\, \boldsymbol{F}(y^{[n]}). \tag{3.5b}$$

Below we give the proof, which is rather short, of property (3.3) with the values $\alpha = \varphi(\gamma)$, $\beta = \psi(\gamma)$. The proof of the optimality of these values under suitable irreducibility assumptions, is less straightforward and will be given in Section 6.2.

Partial proof of Theorem 3.1: *proving* (3.3) *with* $\alpha = \varphi(\gamma)$ *and* $\beta = \psi(\gamma)$. Without loss of generality, assume $I + \gamma\, A$ is invertible and $\mathrm{spr}(|C|) < 1$. We shall rewrite (3.5) in a convenient form, not much different from [24, p. 503], by introducing the vectors $z_i^{[n]} = y_i^{[n]} + \frac{\Delta t}{\gamma}\, F(y_i^{[n]})$ and $z^{[n]} = [z_i^{[n]}]$. Using (1.5), (1.6), we have

$$\begin{aligned} \|z_i^{[n]}\| &= \big\| \big(1 - \tfrac{\Delta t}{\gamma\, \tau_0}\big)\, y_i^{[n]} + \tfrac{\Delta t}{\gamma\, \tau_0} \big(y_i^{[n]} + \tau_0 F(y_i^{[n]})\big) \big\| \\ &\leq \big(1 - \tfrac{\Delta t}{\gamma\, \tau_0}\big)\, \big\|y_i^{[n]}\big\| + \tfrac{\Delta t}{\gamma\, \tau_0}\, \big\|y_i^{[n]}\big\| = \|y_i^{[n]}\|. \end{aligned}$$

Substituting $\Delta t\, \boldsymbol{F}(y^{[n]}) = \gamma \cdot (z^{[n]} - y^{[n]})$ into (3.5a), one arrives easily at the following property (3.6a). Using a similar substitution into (3.5b), combined with the expression for $y^{[n]}$ in (3.6a), it can be seen that the following relation (3.6b) is valid as well:

$$y^{[n]} = (\boldsymbol{E} - \boldsymbol{C}\boldsymbol{E})\, u_{n-1} + \boldsymbol{C}\, z^{[n]}, \quad \text{with } \|z_i^{[n]}\| \leq \|y_i^{[n]}\| \quad (\text{for } 1 \leq i \leq s), \tag{3.6a}$$

$$u_n = (1 - D\, E)\, u_{n-1} + \boldsymbol{D}\, z^{[n]}. \tag{3.6b}$$

From (3.6a) there follows $(I - |C|)\, [\|y_i^{[n]}\|] \leq |E - C\, E|\, \|u_{n-1}\|$. Multiplying this inequality by $(I - |C|)^{-1} = I + |C| + |C|^2 + \cdots \geq 0$, we obtain

$$[\|y_i^{[n]}\|] \leq Z\, \|u_{n-1}\|, \tag{3.7}$$

with Z defined in (3.1). Therefore, $\max_i \|y_i^{[n]}\|] \leq \|Z\|_\infty \|u_{n-1}\| = \alpha\, \|u_{n-1}\|$, where $\alpha = \varphi(\gamma)$.

Using (3.6b), (3.7), we obtain $\|u_n\| \leq |1 - D\, E|\, \|u_{n-1}\| + |D|\, Z\, \|u_{n-1}\| = \beta\, \|u_{n-1}\|$, where $\beta = \psi(\gamma)$. □

3.2. **The accumulated effect of consecutive Runge-Kutta steps.** Theorem 3.2, below, describes the effect of applying method (1.2) for $n = 1, 2, \ldots, N$. It

gives conditions, on the factors α_N, β_N, in order that the following two properties are present:

(3.8a) The estimate $\max_i \|y_i^{[N]}\| \le \alpha_N \cdot \|u_0\|$ holds whenever $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and $y_i^{[N]}$ is generated from u_0 under conditions (1.5), (1.6).

(3.8b) The estimate $\|u_N\| \le \beta_N \cdot \|u_0\|$ holds whenever $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and u_N is generated from u_0 under conditions (1.5), (1.6).

In the theorem, we will refer to the following condition on the RKM:

(3.9) There is *no* pair of indices i, j with $A(i,:) = 0$ and $A(j,:) = B$.

We will say that the RKM is of *general type* if (3.9) holds, and of *special type* otherwise. For some comments on (3.9), see Remark 3.3 below.

Theorem 3.2 (Expressing the accumulated effect of N consecutive Runge-Kutta steps in terms of $\varphi(\gamma)$ and $\psi(\gamma)$)**.** *Let $\gamma > 0$ and $N \ge 1$. Then the following two statements hold:*

(I) Property (3.8a) *is valid with $\alpha_N = \varphi(\gamma)\,\psi(\gamma)^{N-1}$. Moreover, when the RKM is row-irreducible and of general type (in the sense of* (3.9)*), then* (3.8a) *is* not *valid with any $\alpha_N < \varphi(\gamma)\,\psi(\gamma)^{N-1}$.*

(II) Property (3.8b) *is valid with $\beta_N = \psi(\gamma)^N$. Moreover, when the RKM is irreducible and of general type (in the sense of* (3.9)*, then* (3.8b) *is* not *valid with any $\beta_N < \psi(\gamma)^N$.*

For irreducible RKMs of general type, the theorem shows that the estimates

$$\max_i \|y_i^{[N]}\| \le \varphi(\gamma)\,\psi(\gamma)^{N-1} \cdot \|u_0\|, \qquad \|u_N\| \le \psi(\gamma)^N \cdot \|u_0\| \tag{3.10}$$

are *optimal* in the general situation (1.5), (1.6), because the factors $\varphi(\gamma)\,\psi(\gamma)^{N-1}$ and $\psi(\gamma)^N$ cannot be replaced by any smaller ones.

Below we shall prove (3.8), with $\alpha_N = \varphi(\gamma)\,\psi(\gamma)^{N-1}$, $\beta_N = \psi(\gamma)^N$. The proof of the optimality of these values (under the appropriate irreducibility conditions and condition (3.9)) is less straightforward and will be postponed to Section 6.3.

Partial proof of Theorem 3.2: *proving* (3.8) *with* $\alpha_N = \psi(\gamma)^{N-1}\varphi(\gamma)$, $\beta_N = \psi(\gamma)^N$. By Theorem 3.1, properties (3.3a), (3.3b) are present with $\alpha = \varphi(\gamma)$, $\beta = \psi(\gamma)$. Repeated application of (3.3b) yields property (3.8b) with $\beta_N = \psi(\gamma)^N$.

Using (3.3a) (with $n = N$ and $\alpha = \varphi(\gamma)$) and (3.8b) (with N replaced by $N-1$ and $\beta_{N-1} = \psi(\gamma)^{N-1}$), we obtain:

$$\max_i \|y_i^{[N]}\| \le \varphi(\gamma)\,\|u_{N-1}\| \le \varphi(\gamma)\ \psi(\gamma)^{N-1}\,\|u_0\|,$$

i.e. (3.8a) with $\alpha_N = \varphi(\gamma)\,\psi(\gamma)^{N-1}$. □

Remark 3.3 (About condition (3.9)). In proving optimality of (3.10), we shall rewrite N steps of method (1.2) as one step of a (formal) RKM with Ns stages; cf. (6.6) in Section 6.3. Our proof will need row-irreducibility of the latter RKM, which means that the original method (1.2) must satisfy (3.9).

When condition (3.9) is violated, one might say that the RKM suffers from a weak kind of reducibility, which is apparent when $N \ge 2$ consecutive steps of the RKM are carried out: only $N(s-1)+1$ different internal stages play a role,

rather than the standard value of Ns; cf. the 'FASAL property', [2]. Note that any DJ-irreducible, explicit RKM automatically satisfies (3.9).

The natural question arises of whether condition (3.9) can be omitted in the above Theorem 3.2. In Section 5.2.1, we will show by means of a counterexample that the estimates (3.10) need not be optimal if (3.9) would be omitted.

4. Monotonicity and boundedness

4.1. **The conditions $\varphi(\gamma) = 1$ and $\psi(\gamma) = 1$.** Clearly, by Theorems 3.1, 3.2, the magnitudes of $\varphi(\gamma)$ and $\psi(\gamma)$ are crucial for monotonicity and boundedness. For this reason, in the present Section 4.1, we have a closer look at the size of these two quantities.

Below, we shall frequently use that Z, defined in (3.1), satisfies

$$Z \geq E \quad \text{(when } I + \gamma A \text{ is invertible, and } \mathrm{spr}(|C|) < 1). \tag{4.1}$$

This follows from $(I - |C|)^{-1} = I + |C| + |C|^2 + \cdots \geq 0$ together with $Z \geq (I - |C|)^{-1}(E - |C|\,E) = E$.

When $I + \gamma A$ is invertible, and $\mathrm{spr}(|C|) < 1$, we see from (3.2), (4.1) that $\varphi(\gamma) \geq 1$ and $\psi(\gamma) \geq |1 - D\,E| + |D|\,E \geq 1 - |D|\,E + |D|\,E = 1$. Consequently

$$\varphi(\gamma) \geq \varphi(0) = 1, \quad \psi(\gamma) \geq \psi(0) = 1 \quad \text{(for all real } \gamma). \tag{4.2}$$

In view of (4.2) and Theorem 3.1, the following condition is crucial for property (1.3):

$$\varphi(\gamma) = 1. \tag{4.3}$$

Lemma 4.1, below, states that (4.3), combined with requirement

$$|1 - D\,E| + |D|\,E = 1, \tag{4.4}$$

implies that

$$\psi(\gamma) = 1. \tag{4.5}$$

Via (4.2) and Theorem 3.1, we see that condition (4.5) is crucial for the monotonicity property (1.4). Note that the left-hand member of (4.4) is a variant of the expression for $\psi(\gamma)$ in (3.2b).

The following lemma will be applied in Sections 4.2, 4.3.

Lemma 4.1 (Relating conditions $\varphi(\gamma) = 1$ and $\psi(\gamma) = 1$ to each other)**.** *Property* (4.3)*, combined with* (4.4)*, implies* (4.5)*. Conversely, when the RKM is DJ-irreducible, property* (4.5) *implies* (4.3) *and* (4.4)*.*

Partial proof of Lemma 4.1: *proving* (4.5) *from* (4.3)*,* (4.4)*.* A combination of (4.3) and (4.1) yields $Z = E$, so that $\psi(\gamma)$ is equal to the left-hand member of (4.4). Hence, (4.3), (4.4) entail (4.5). □

The proof of (4.3), (4.4) from (4.5), for DJ-irreducible methods, is not so straightforward and is postponed to Section 6.1.

4.2. **Stepsize-coefficients for monotonicity and boundedness.**

4.2.1. *Stepsize-coefficients for monotonicity.* Let a stepsize $\Delta t > 0$, a vector space $\mathbb{V}$ with seminorm $\|\cdot\|$ and a function $F : \mathbb{V} \to \mathbb{V}$ be given. We will say that the RKM (1.2) is *internally* or *externally monotonic*, respectively, if (1.3) or (1.4) holds whenever the vectors $y_i^{[n]},\ u_n,\ u_{n-1} \in \mathbb{V}$ satisfy (1.2).

We will say that a value $\gamma \in [0, \infty]$ is a *stepsize-coefficient for internal* or *external monotonicity*, respectively, if the RKM is internally or externally monotonic whenever $\mathbb{V}$ is vector space with seminorm $\|\cdot\|$ and (1.5), (1.6) are fulfilled.

Theorem 3.1 yields the following corollary:

Corollary 4.2 (Characterization of stepsize-coefficients for monotonicity)**.** *For $0 < \gamma < \infty$, the following statements are valid:*

(I) When the RKM is row-irreducible, then γ is a stepsize-coefficient for internal monotonicity if and only if $\varphi(\gamma) = 1$.

(II) When the RKM is irreducible, then γ is a stepsize-coefficient for external monotonicity if and only if $\psi(\gamma) = 1$.

In view of Lemma 4.1, there follows

Corollary 4.3 (External monotonicity implies internal monotonicity)**.** *Assume the RKM is irreducible. Then any stepsize-coefficient γ for external monotonicity is a stepsize-coefficient for internal monotonicity as well.*

4.2.2. *Stepsize-coefficients for boundedness.* Let a stepsize $\Delta t > 0$, a vector space $\mathbb{V}$ with seminorm $\|\cdot\|$ and a function $F : \mathbb{V} \to \mathbb{V}$ be given. We will say that the RKM (1.2) is *internally bounded with factor μ*, if for all $N \geq 1$ we have (1.8), whenever $y_i^{[N]} \in \mathbb{V}$ is generated by applying (1.2) for $n = 1, \ldots, N$. Similarly, we will say that the RKM (1.2) is *externally bounded with factor μ*, if for all $N \geq 1$ we have (1.9), whenever $u_N \in \mathbb{V}$ is generated by applying (1.2) for $n = 1, \ldots, N$.

We will call a value $\gamma \in [0, \infty]$ a *stepsize-coefficient for internal or external boundedness*, respectively, if the RKM is internally or externally bounded, with some finite factor μ, whenever $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$ and (1.5), (1.6) are fulfilled. Here, the factor μ is understood to depend only on γ and the (coefficients a_{ij}, b_j of the) RKM under consideration.

Theorem 3.2 yields the following corollary:

Corollary 4.4 (Characterization of stepsize-coefficients for boundedness)**.** *For $0 < \gamma < \infty$, the following statements are valid:*

(I) When the RKM is row-irreducible and of general type (3.9)*, then γ is a stepsize-coefficient for internal boundedness if and only if $\psi(\gamma) = 1$.*

(II) When the RKM is irreducible and of general type (3.9)*, then γ is a stepsize-coefficient for external boundedness if and only if $\psi(\gamma) = 1$.*

A combination of the last three corollaries yields

Corollary 4.5 (Boundedness implies monotonicity for RKMs of general type)**.** *Assume the RKM is irreducible and of general type* (3.9)*, and let γ be any stepsize-coefficient for internal or external boundedness. Then γ is necessarily a stepsize-coefficient for internal monotonicity as well as for external monotonicity.*

For irreducible methods of general type, the last corollary thus shows, rather surprisingly, that requiring just boundedness, instead of (the a priori stronger property of) monotonicity, *cannot* lead to a more favourable (larger) stepsize-coefficient γ.

4.3. **Maximal stepsize-coefficients γ and the coefficients $r(A)$, $r(A,B)$.**

4.3.1. *The coefficients $r(A)$ and $r(A,B)$.* Below we shall define Kraaijevanger's coefficient as well as a closely related coefficient by using the following conditions, in which γ denotes a real variable:

$$I + \gamma A \text{ is invertible,} \tag{4.6}$$

$$\gamma A\,(I+\gamma A)^{-1} \geq 0, \quad \gamma A\,(I+\gamma A)^{-1} E \leq E, \tag{4.7}$$

$$\gamma B\,(I+\gamma A)^{-1} \geq 0, \quad \gamma B\,(I+\gamma A)^{-1} E \leq 1. \tag{4.8}$$

Definition 4.6 (Coefficients $r(A)$ and $r(A,B)$)**.** We define

$$r(A) = \sup\{\gamma \,:\, \gamma \geq 0 \text{ and } (4.6), (4.7) \text{ hold}\},$$
$$r(A,\,B) = \sup\{\gamma \,:\, \gamma \geq 0 \text{ and } (4.6), (4.7), (4.8) \text{ hold}\}.$$

The value $r(A,B)$, defined above, is closely related to a quantity introduced in Kraaijevanger [24] and will be referred to as *Kraaijevanger's coefficient*. For completeness, we note that the original definition in [24] amounts essentially to

$$R = \sup\{\rho \,:\, \rho \in \mathbb{R} \text{ and } (4.6), (4.7), (4.8) \text{ hold for all } \gamma \in [0,\rho]\}.$$

The following theorem implies, among other things, that the value $r(A,\,B)$ (Definition 4.6) is equal to Kraaijevanger's value R:

Theorem 4.7 (Fulfillment of conditions (4.6), (4.7), (4.8))**.** *Any finite γ, with $0 \leq \gamma \leq r(A)$, satisfies* (4.6), (4.7)*. Similarly, any finite γ, with $0 \leq \gamma \leq r(A,\,B)$, satisfies* (4.6), (4.7), (4.8)*.*

This theorem can be viewed as a (somewhat stronger) version of earlier results about $r(A,\,B)$ in the literature; for related material, see [24, Lemma 4.4], [17, Prop. 2.11], [19, Thm. 4]. The theorem follows easily from material in [32, Thm. 2.2 (ii), Sect. 3.2.2].

Definition 4.6 implies that $r(A)$ and $r(A,\,B)$ are always nonnegative. It is relatively easy to determine whether the coefficients $r(A)$ and $r(A,\,B)$ are positive, and to compute (numerically) their actual size; cf. [24, 4, 32]. In fact, for many RKMs, explicit expressions or numerical values were obtained for $r(A,\,B)$; cf. [24, 4, 7, 23].

4.3.2. *The relevance of $r(A)$, $r(A,B)$ to monotonicity and boundedness.* The following lemma is crucial in linking monotonicity or boundedness to the coefficients defined in Section 4.3.1.

Lemma 4.8 (Relating the conditions $\varphi(\gamma)=1$ and $\psi(\gamma)=1$ to (4.6), (4.7), (4.8))**.** *For $0 < \gamma < \infty$, the following statements are valid.*

(I) Property $\varphi(\gamma)=1$ is equivalent to (4.6), (4.7)*.*

(II) Property $\psi(\gamma)=1$ follows from conditions (4.6), (4.7), (4.8)*. Conversely, when the RKM is DJ-irreducible, property $\psi(\gamma)=1$ implies* (4.6), (4.7), (4.8)*.*

Proof. About Statement (I). Assume first $\varphi(\gamma)=1$. From (3.2a) we have immediately (4.6). Because $\|Z\|_\infty = 1$, we see from (4.1) that $Z = E$. Hence, $|E - C\,E| + |C|\,E = E = (E - C\,E) + C\,E$, which implies that $(E - C\,E) = |E - C\,E| \geq 0$ and $C = |C| \geq 0$. We have thus also proved property (4.7).

Next, assume conversely (4.6), (4.7). Because $C \geq 0$, we can apply the Perron-Frobenius theory as presented e.g. in [18, p. 503], so as to conclude that $\mathrm{spr}(|C|) < 1$, provided C has no real eigenvalues $\lambda \geq 1$.

Because $C = I - (I + \gamma A)^{-1}$, the value $\lambda = 1$ is no eigenvalue of C. Moreover, we see from $|C|\,E = C\,E \le E$ that $\|C\|_\infty \le 1$, so that C has no eigenvalues $\lambda > 1$. Consequently, $\mathrm{spr}(|C|) < 1$. Because $Z = (I - C)^{-1}(E - C\,E) = E$, we thus arrive, via (3.2a), at $\varphi(\gamma) = 1$.

About Statement (II). Clearly, (4.8) implies (4.4). Conversely, (4.4) implies

$$|D|\,E = 1 - |1 - D\,E| \le D\,E \le |D|\,E,$$

so that $|D|\,E = D\,E$, $1 - D\,E = |1 - D\,E|$, which leads to (4.8). Hence,

(4.9) Property (4.4) is equivalent to (4.8) (when $I + \gamma A$ is invertible).

The proof of Statement (II) is easily completed by combining Statement (I), Lemma 4.1 and (4.9). □

The above lemma, combined with Definition 4.6, yields directly

Theorem 4.9 (Characterizing $r(A)$ and $r(A, B)$ in terms of the functions $\varphi(\gamma)$, $\psi(\gamma)$)**.** *Let the RKM be DJ-irreducible. Then*

$$r(A) = \sup\{\gamma\,:\ \gamma \ge 0 \text{ and } \varphi(\gamma) = 1\}, \quad r(A,\,B) = \sup\{\gamma\,:\ \gamma \ge 0 \text{ and } \psi(\gamma) = 1\}.$$

We do not think that the above new characterizations of $r(A)$ and $r(A, B)$ are more handy than the original ones in Definition 4.6 for actually computing these coefficients. But, unlike the characterizations in Definition 4.6, the new ones in Theorem 4.9 are clearly related to monotonicity and boundedness properties of the RKM; cf. Sections 3, 4.2. In fact, Definition 4.6 seems most suitable for actually computing $r(A)$, $r(A, B)$, whereas Theorem 4.9 gives an easy understanding of the relevance of these coefficients to monotonicity.

Our last corollary summarizes interesting conclusions, about maximal stepsize-coefficients for monotonicity and boundedness, obtainable from the above. Corollary 4.10 follows from a combination of Lemma 4.8, Theorem 4.7 and the material of Section 4.2.

Corollary 4.10 (Relating monotonicity and boundedness to the coefficients $r(A)$, $r(A,\,B)$)**.**

(Ia) When the RKM is row-irreducible, the largest stepsize-coefficient for internal monotonicity is equal to $r(A)$.

(Ib) When the RKM is irreducible, the largest stepsize-coefficient for external monotonicity is equal to $r(A, B)$.

(IIa) When the RKM is irreducible and of general type, the largest stepsize-coefficient for internal boundedness is equal to $r(A, B)$.

(IIb) When the RKM is irreducible and of general type, the largest stepsize-coefficient for external boundedness is equal to $r(A, B)$.

For any given irreducible RKM, Corollary 4.10 highlights once more that requiring boundedness instead of monotonicity may lead to a more favourable stepsize-coefficient γ only when the method violates (3.9).

Statement (Ia) is related to results obtained earlier in the literature via quite different proofs; cf. e.g. [32]. Statement (Ib) is closely related to statements in [4, 15, 22]. The irreducibility requirement demanded in these three papers is more restrictive than in the above corollary; cf. Definition 2.1. Furthermore, the proofs given (or referred to) in these papers actually make use of an ingenious but complicated construction going back to Kraaijevanger [24]; for details see [24, pp. 485-496, pp. 505-508] and [4, p. 1091]. We think the proof of Statement (Ib) in the present

paper, via the functions $\varphi(\gamma)$, $\psi(\gamma)$, is essentially shorter and more transparent than in the above references.

For completeness we note that in the existing monotonicity literature related to (Ia), (Ib), a monotonicity concept is often used that is somewhat stronger than the one defined at the start of Section 4.2.1, in that $\|\cdot\|$ is not necessarily a seminorm but is only required to satisfy the *convexity requirement*

$$\|\theta\, u + (1-\theta)v\| \le \theta\,\|u\| + (1-\theta)\|v\| \quad (\text{for all } u, v \in \mathbb{V} \text{ and } 0 \le \theta \le 1). \tag{4.10}$$

Clearly, any stepsize coefficient γ corresponding to this stronger monotonicity concept is at the same time a stepsize coefficient for monotonicity in the sense of Section 4.2.1. Hence statements (Ia), (Ib) of Corollary 4.10 have some relevance to the former monotonicity concept as well: any stepsize coefficient for internal or external monotonicity in the stronger sense cannot exceed $r(A)$ or $r(A, B)$, respectively.

Note that not all findings in the present paper would remain valid if the assumption that $\|\cdot\|$ is a seminorm would have been replaced consistently by (4.10). One easily sees, e.g., that Theorems 3.1, 3.2 cannot be true after such a replacement.

5. Examples and applications

In this section, we shall give illustrations and counterexamples corresponding to the theory of Sections 3, 4. We shall focus on values of γ, μ for which the following two properties are present:

(5.1) Restriction $0 < \Delta t \le \gamma \cdot \tau_0$ implies the bound $\max_i \|y_i^{[N]}\| \le \mu \cdot \|u_0\|$ whenever $N \ge 1$, $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and $y_i^{[N]}$ is generated from u_0 by applying (1.2), for $n = 1, \ldots, N$, under condition (1.5).

(5.2) Restriction $0 < \Delta t \le \gamma \cdot \tau_0$ implies the bound $\|u_N\| \le \mu \cdot \|u_0\|$ whenever $N \ge 1$, $\mathbb{V}$ is a vector space with seminorm $\|\cdot\|$, and u_N is generated from u_0 by applying (1.2), for $n = 1, \ldots, N$, under condition (1.5).

Clearly, γ is a stepsize-coefficient for internal or external boundedness if and only if, for some (finite) μ, we have property (5.1) or (5.2), respectively.

5.1. **Conclusions from Section 4 about boundedness for actual RKMs.**

5.1.1. *Two simple RKMs.* We consider two simple explicit RKMs with $s = 2$ stages, the nonzero coefficients a_{ij}, b_j of which are given by (5.3) and (5.4), respectively:

$$a_{2\,1} = 1, \quad b_1 = b_2 = 1/2, \tag{5.3}$$

$$a_{2\,1} = -20, \quad b_1 = 41/40, \quad b_2 = -1/40. \tag{5.4}$$

Both methods are of second order and yield identical numerical approximations when applied to linear autonomous problems.

In [10], monotonicity properties of the two methods were compared to each other, both theoretically and by a numerical experiment. Furthermore, in [20], boundedness results for the methods were obtained, via (laborious) ad-hoc calculations. Below we shall recover and extend the last-mentioned results, without any laborious calculations, by using the material of Section 4. Note that both methods are irreducible, cf. Definition 2.1, and of general type (3.9).

We find easily, from Definition 4.6, that

$$r(A) = r(A, B) = 1 \text{ (for method (5.3))}, \quad r(A) = r(A, B) = 0 \text{ (for method (5.4))}.$$

Applying Corollary 4.10, Parts (Ia), (Ib), it follows that for method (5.3) the largest stepsize-coefficient, for either *internal or external monotonicity*, is equal to $\gamma = 1$; whereas for method (5.4) it equals $\gamma = 0$.

By further applications of Corollary 4.10, we obtain conclusions about *internal and external boundedness* of the methods. For method (5.3) we find:

(5.5) For any given $\mu \geq 1$, the largest γ, with either property (5.1) or (5.2), is equal to $\gamma = 1$.

Similarly, we find for method (5.4):

(5.6) For any given $\mu \geq 1$, the largest γ, with either property (5.1) or (5.2), is equal to $\gamma = 0$.

It follows that method (5.3) is superior to (5.4) regarding both internal and external boundedness, with any factor $\mu \geq 1$. We think these conclusions neatly supplement and confirm the discussion of the methods in the two papers mentioned above.

5.1.2. *Solving the question of boundedness for some well-known RKMs.* The question of whether positive γ and finite μ exist, with either property (5.1) or (5.2), can in many cases be answered quite easily by applying Corollary 4.10, notably for the RKMs listed below:

- 4-th order "3/8−Rule" of Kutta; cf. e.g. [12, p. 137],
- 4-th order method of Gill; cf. e.g. [1, p. 183], [2, p. 167] and [12, p. 138],
- Butcher's methods of orders $5, 6, 7$, respectively, as specified e.g. in [2, pp 92, 177, 179],
- 8-th order method of Cooper and Verner; cf. e.g. [1, p. 208] and [2, p. 180],
- Radau-IA methods as specified e.g. in [1, p. 228] and [2, pp. 207, 208],
- Lobatto-III and Lobatto-IIIB methods as specified e.g. in [2, pp. 210, 211],
- 5-th order method of Higham and Hall; cf. e.g. [13, p. 27],
- 7-th order method of Fehlberg; cf. e.g. [12, p. 194] and [2, p. 192],
- 5-th order Dormand-Prince method and 8-th order Prince-Dormand method as given e.g. in [12, pp. 171, 195] and [2, p. 194].

All of the above methods have a coefficient matrix A containing some negative entry a_{ij}. It follows, via Definition 4.6 and Theorem 4.7, that $r(A) = r(A, B) = 0$. Since the methods are irreducible, we can apply Corollary 4.10, Parts (Ia), (Ib). It follows for all methods that the largest stepsize-coefficient, for either *internal or external monotonicity*, is equal to $\gamma = 0$.

Because all methods satisfy (3.9), we can also apply Parts (IIa), (IIb) of Corollary 4.10 to obtain conclusions about *internal and external boundedness*. As $r(A, B) = 0$, we arrive for all methods listed above, somewhat disappointingly, at the (negative) conclusion (5.6).

5.2. **Counterexamples.**

5.2.1. *Boundedness in the absence of monotonicity when* (3.9) *is violated.* We shall give an example showing that the restriction on γ for of RKMs can be less severe than for monotonicity, when condition (3.9) is violated. The example will prove that condition (3.9) cannot be omitted in Theorem 3.2 and in Corollaries 4.4, 4.5, 4.10.

We consider method (1.2), with $s = 2$, $a_{1,1} = a_{1,2} = 0$, $a_{2,1} = b_1 = (1-\theta)$, $a_{2,2} = b_2 = \theta$, where $\theta > 1$. This method is irreducible, but violates (3.9). Combining Definition 4.6 and Theorem 4.7, one sees that $r(A) = r(A, B) = 0$. Applying Corollary 4.10, Parts (Ia), (Ib), it follows that the largest stepsize-coefficient, for either *internal or external monotonicity*, is equal to $\gamma = 0$. This follows also from Theorem 3.1 or Corollary 4.2, because a direct computation using (3.2) yields

$$\varphi(\gamma) = \psi(\gamma) = 1 + 2\,(\theta - 1)\,\gamma \quad (\text{for } \gamma \geq 0). \tag{5.7}$$

To obtain conclusions about *internal and external boundedness* of the method, we consider $\gamma > 0$ and assume that (1.2) holds for $1 \leq n \leq N$ under the assumptions (1.5), (1.6). Defining $f_n = \Delta t F(u_n)$, we have

$$u_n = u_0 + (1-\theta) f_0 + \sum_{i=1}^{n-1} f_i + \theta\, f_n \quad (1 \leq n \leq N),$$

where for $n = 1$ we use the convention $\sum_{i=1}^{0} f_i = 0$. By introducing $v_{n-1} = u_0 + (1-\theta) f_0 + \sum_{i=1}^{n-1} f_i$ $(1 \leq n \leq N)$, there follows

$$u_n = v_{n-1} + \theta\, f_n \quad (1 \leq n \leq N), \tag{5.8a}$$

$$v_n = v_{n-1} + f_n \quad (1 \leq n \leq N-1). \tag{5.8b}$$

Using (5.8a) to eliminate f_n from (5.8b), we obtain $v_n = (1 - \theta^{-1})\, v_{n-1} + \theta^{-1}\, u_n$, and therefore

$$\|v_n\| \leq \max\{\|v_{n-1}\|, \|u_n\|\} \quad (1 \leq n \leq N-1).$$

From (5.8a) we see also that $\|v_{n-1}\| = \|\big(1 + \frac{\theta\,\Delta t}{\tau_0}\big)\, u_n - \frac{\theta\,\Delta t}{\tau_0}\,\big(u_n + \tau_0\, F(u_n)\big)\| \geq \big(1 + \frac{\theta\,\Delta t}{\tau_0}\big)\,\|u_n\| - \frac{\theta\,\Delta t}{\tau_0}\,\|u_n\| = \|u_n\|$, so that

$$\|u_n\| \leq \|v_{n-1}\| \quad (1 \leq n \leq N).$$

It follows that $\|v_n\| \leq \|v_{n-1}\|$, and therefore $\|v_n\| \leq \|v_0\|$ $(0 \leq n \leq N-1)$. We thus find $\|u_n\| \leq \|v_0\| = \|u_0 + (1-\theta)\,\Delta t\, F(u_0)\| = \|\big(1 + \frac{(\theta-1)\,\Delta\, t}{\tau_0}\big)\, u_0 - \frac{(\theta-1)\,\Delta t}{\tau_0}\,\big(u_0 + \tau_0\, F(u_0)\big)\|$. Hence,

$$\|u_n\| \leq \mu\,\|u_0\| \quad (1 \leq n \leq N), \quad \text{with} \quad \mu = 1 + 2\,(\theta - 1)\,\gamma. \tag{5.9}$$

It follows that any $\gamma > 0$ is a stepsize-coefficient for internal and external boundedness, although it is no stepsize-coefficient for internal or external monotonicity. Furthermore, it follows that condition (3.9) cannot be omitted in Theorem 3.2, because (5.7), (5.9) imply: $\mu < \varphi(\gamma)\,\psi(\gamma)^{N-1} = \psi(\gamma)^N$ (for $\gamma > 0,\ N > 1$). It is also clear from the above that (3.9) cannot be omitted in Corollaries 4.4, 4.5, 4.10.

For completeness we note that the bound (5.9) was also given in [20] (proved by a longer computation than above), and that (5.9), when combined with (5.7) and Theorem 3.1, leads to the following conclusion:

(5.10) For any given $\mu \geq 1$, the largest γ with either property (5.1) or (5.2) is equal to $\gamma = \frac{\mu - 1}{2\,(\theta - 1)}$.

5.2.2. *The necessity of row- and DJ-irreducibility.* We shall give counterexamples showing that the conditions of row- and DJ-irreducibility cannot be omitted in Sections 3, 4.

Necessity of row-irreducibility.

Consider method (1.2), with $s = 2$, $b_1 = a_{i,1} = 2$, $b_2 = a_{i,2} = -1$ (for $i = 1, 2$). One sees that the method satisfies (3.9) and is DJ-irreducible, but *not row-irreducible*. Furthermore, it is easy to see that $r(A) = r(A, B) = 0$. By (4.2) and Theorem 4.9, it follows that $\varphi(\gamma) > 1$, $\psi(\gamma) > 1$ (for $\gamma > 0$).

Clearly, the RKM is equivalent to the backward Euler method $y_1^{[n]} = u_{n-1} + \Delta t\, F(y_1^{[n]})$, $u_n = u_{n-1} + \Delta t\, F(y_1^{[n]})$. In line with Corollary 4.10, for the latter method the maximal stepsize-coefficient γ for (internal or external) monotonicity equals $\gamma = \infty$. Therefore, the same holds for the original RKM. It follows that the requirement of row-irreducibility cannot be omitted in Theorems 3.1, 3.2 and Corollaries 4.2, 4.4, 4.10.

Necessity of DJ-irreducibility.

Consider method (1.2), with $s = 2$, $a_{1,1} = -1$, $a_{1,2} = a_{2,1} = 0$, $a_{2,2} = 1$, $b_1 = 0$, $b_2 = 1$. Clearly, the method satisfies (3.9) and is row-irreducible, but *not DJ-irreducible*. We have $r(A) = r(A, B) = 0$, and from (3.2) it can be seen that $\varphi(\gamma) = (1-2\,\gamma)^{-1}$, $\psi(\gamma) = 1$ (for $0 \le \gamma < 1/2$) and $\varphi(\gamma) = \psi(\gamma) = \infty$ (for $\gamma \ge 1/2$).

The approximations u_n, generated by this RKM, can also be obtained by the backward Euler method mentioned above. Consequently, for the given RKM, any $\gamma \ge 0$ is a stepsize-coefficient for external monotonicity. Using this property, it can be concluded that none of the theorems and corollaries in Sections 3, 4, where the condition of DJ-irreducibility occurs, would remain true if that condition would be omitted.

6. Proofs related to Lemma 4.1 and Theorems 3.1, 3.2

6.1. **Completing the proof of Lemma 4.1.** The following lemma, about the matrices $C = (c_{ij})$, $D = (d_j)$ defined in (3.1), will be used in completing the proofs of Lemma 4.1 and Theorem 3.1.

Lemma 6.1 (A useful property of the matrices $C = (c_{ij})$, $D = (d_j)$)**.** *Let the RKM be DJ-irreducible, and consider $\gamma \neq 0$ with $I + \gamma\, A$ invertible. Then for each i_0 with $1 \le i_0 \le s$, the following property is present:*

(6.1) *We have $d_{i_0} \neq 0$, or there exist indices $i_n, \ldots, i_1 \in \{1, \ldots, s\}$ with $d_{i_n}\, c_{i_n i_{n-1}} \cdots c_{i_1 i_0} \neq 0$.*

Proof of Lemma 6.1. Suppose (6.1) would *not* hold for all i_0. We define $\mathcal{N}$ to be the subset of $\{1, \ldots, s\}$ consisting of all i_0 violating (6.1), and $\mathcal{M}$ the set of all remaining indices in $\{1, \ldots, s\}$. We shall first prove that the entries d_j, c_{ij} have the properties imposed on b_j, a_{ij}, respectively, in part (a) of Definition 2.1.

Let $i \in \mathcal{M}$, $j \in \mathcal{N}$. Then $d_j = 0$. Furthermore, we have $c_{ij} = 0$ if $d_i \neq 0$, because $d_i\, c_{ij} = 0$. If $d_i = 0$, we have $d_{i_n}\, c_{i_n i_{n-1}} \cdots c_{i_1 i} \neq 0$ for some indices $i_n, \ldots, i_1$. Because $j \in \mathcal{N}$, there follows: $d_{i_n}\, c_{i_n i_{n-1}} \cdots c_{i_1 i}\, c_{ij} = 0$, which implies again that $c_{ij} = 0$. We have thus proved that d_j, c_{ij} have the properties stated in Definition 2.1, part (a).

Next, we denote by v_j (with $j \in \mathcal{N}$) the column vector in $\mathbb{R}^s$, with j-th component equal to 1 and all other components equal to 0, and we denote the subspace of $\mathbb{R}^s$ spanned by all v_j (with $j \in \mathcal{N}$) by X. In view of the properties of d_j, c_{ij}, just

proved, we have $D\,v_j = 0$, $C\,v_j \in X$ $(j \in \mathcal{N})$. We define $w_k = (I + \gamma\,A)^{-1} v_k$ $(k \in \mathcal{N})$, so that

$$\gamma\,B\,w_k = 0 \quad \text{and} \quad \gamma\,A\,w_k \in X \quad (k \in \mathcal{N}).$$

Using the expression $(I + \gamma\,A)^{-1} = I - C$, we see that w_k $(k \in \mathcal{N})$ are linearly independent vectors in the subspace X. We can thus express each v_j as a linear combination of the vectors w_k, so that

$$\gamma\,B\,v_j = 0 \quad \text{and} \quad \gamma\,A\,v_j \in X \quad (j \in \mathcal{N}).$$

Because $\gamma \neq 0$, the coefficients b_j, a_{ij} satisfy the conditions imposed on them in Definition 2.1, part (a). This contradicts DJ-irreducibility, so that property (6.1) holds (for each i_0). □

Proof of (4.3), (4.4) *for DJ-irreducible RKMs satisfying* (4.5). It is enough to consider $\gamma \neq 0$, and to show that, under assumption (4.5), we have $Z = E$. We shall prove $Z = E$ from (4.5) using (6.1).

Because of $|1 - D\,E| + |D|\,Z = \psi(\gamma) = 1 = (1 - D\,E) + D\,E$ and (4.1), we have $0 \geq (I - D\,E) - |1 - D\,E| = |D|\,Z - D\,E \geq 0$, which implies that

$$|D|\,Z = D\,E, \quad \text{and} \quad d_i\,z_i = d_i \ \ (\text{for } 1 \leq i \leq s). \tag{6.2}$$

Let any index $i_0 \in \{1, \ldots, s\}$ be given. If $d_{i_0} \neq 0$, we see from (6.2) that $z_{i_0} = 1$. If, on the other hand, $d_{i_0} = 0$, there are indices $i_n, \ldots, i_1$ as in (6.1). Because $d_{i_n} \neq 0$, the equality $d_{i_n}\,z_{i_n} = d_{i_n}$ implies that $z_{i_n} = 1$.

Because $Z = (I + \cdots + |C|^{n-1})\,|E - C\,E| + |C|^n Z \geq |C|^n Z + (I + \cdots + |C|^{n-1})(I - |C|)\,E$, we have $Z \geq |C|^n\,Z + (I - |C|^n)\,E$. Hence, $|C|^n\,(Z - E) \leq Z - E$, which implies that

$$|c_{i_n i_{n-1}} \cdots c_{i_1 i_0}|\,(z_{i_0} - 1) + \sum |c_{i_n j_{n-1}} \cdots c_{j_1 j_0}|\,(z_{j_0} - 1) \quad \leq \quad z_{i_n} - 1 \ = \ 0,$$

where the summation is over all indices j_k with $(j_{n-1}, \ldots, j_0) \neq (i_{n-1}, \ldots, i_0)$. Since all $z_{j_0} \geq 1$ and $c_{i_n i_{n-1}} \cdots c_{i_1 i_0} \neq 0$, there follows: $z_{i_0} = 1$.

We have thus proved that $z_i = 1$ for all $i \in \{1, \ldots, s\}$, i.e. $Z = E$. □

6.2. **Completing the proof of Theorem 3.1.** In the present section, we will complete the proof of Theorem 3.1 by proving the inequalities $\varphi(\gamma) \leq \alpha$ and $\psi(\gamma) \leq \beta$, respectively, from the estimates $\max_k \|y_k^{[n]}\| \leq \alpha \cdot \|u_{n-1}\|$ and $\|u_n\| \leq \beta \cdot \|u_{n-1}\|$ occurring in (3.3). Our proof will make use of these estimates in situations where $\mathbb{V}$, $\|\cdot\|$ and F are specified by the following key Lemma 6.2. Our proof will also use Lemma 6.3, which gives a condition under which property (4.6), needed in Lemma 6.2, is present. Lemmas 6.2, 6.3 will also be used in the next section for completing the proof of Theorem 3.2.

We note that Lemma 6.2 is related to material in [20, proofs of Lemmas 4.5, 4.6], but the lemma does not follow directly from that paper.

Lemma 6.2. *Let the RKM be row-irreducible. Let positive τ_0, γ be given, such that* (4.6) *holds, and let $\theta \in \mathbb{R}$, $X \in \mathbb{R}^s$ be given, with*

$$\theta \geq 0, \quad 0 \leq X \leq |E - C\,E|\,\theta + |C|\,X. \tag{6.3}$$

Then there exist a seminorm $\|\cdot\|$ in $\mathbb{V} = \mathbb{R}^{s+2}$, a function $F : \mathbb{V} \to \mathbb{V}$ satisfying (1.5) *and vectors u_0, u_1, $y_k^{[1]} \in \mathbb{V}$ $(1 \leq k \leq s)$ satisfying* (1.2), *with $n = 1$, $\Delta t = \gamma \cdot \tau_0$,*

such that

$$[\|y_k^{[1]}\|] \;=\; |E - C\,E|\,\|u_0\| + |C|\,X, \tag{6.4a}$$

$$\|u_1\| \;=\; |1 - D\,E|\,\|u_0\| + |D|\,X, \quad \|u_0\| = \theta. \tag{6.4b}$$

Proof of Lemma 6.2. We put $\mathbb{V} = \mathbb{R}^{s+2}$, $n = 1$ and $\Delta t = \gamma \cdot \tau_0$. Below, in step 1, we shall determine u_0, u_1, $y_k^{[1]}$, $z_k^{[1]} \in \mathbb{V}$ satisfying (6.4) and (3.6). Next, in step 2, we shall derive (3.5), (1.5) from (3.6).

Step 1. We denote the components of X by x_k $(1 \le k \le s)$ and define the seminorm

$$\|v\| = \max\{|v_i| \,:\, 1 \le i \le s+1\} \quad (\text{for } v \in \mathbb{V} \text{ with components } v_i \ (1 \le i \le s+2)).$$

We write simply y_k, z_k instead of $y_k^{[1]}$, $z_k^{[1]}$ and define the components z_{ik} of $z_k \in \mathbb{V}$ $(1 \le k \le s)$, and $u_{i,\,0}$ of $u_0 \in \mathbb{V}$, as follows:

$$z_{ik} = \operatorname{sgn}(c_{ik})\,x_k \quad (1 \le i \le s), \quad z_{s+1,\,k} = \operatorname{sgn}(d_k)\,x_k, \quad z_{s+2,\,k} = \zeta_k,$$

$$u_{i,\,0} = \operatorname{sgn}\Big(1 - \sum_l c_{il}\Big) \cdot \theta \quad (1 \le i \le s), \quad u_{s+1,\,0} = \operatorname{sgn}(1 - D\,E) \cdot \theta, \quad u_{s+2,\,0} = 0.$$

Here c_{ik}, d_k denote the entries of C, D (cf. (3.1)), ζ_k will be specified in Step 2, and $\operatorname{sgn}(\xi) = 1$ (for $\xi \ge 0$), $\operatorname{sgn}(\xi) = -1$ (for $\xi < 0$).

We define $y^{[1]} = [y_k] \in \mathbb{V}^s$ and $u_1 \in \mathbb{V}$ by the equalities in (3.6) (with $n = 1$). It can be seen that, with these definitions, property (6.4) is present. Furthermore, combining the equality $[\|z_k\|] = X$ with (6.3), (6.4), we see that the inequalities in (3.6a) are fulfilled as well.

Step 2. To arrive at (3.5), (1.5), we define $f = [f_k] \in \mathbb{V}^s$ by $f = \frac{1}{\tau_0}\,(z^{[1]} - y^{[1]})$. Using (3.6), it can be seen that

$$y^{[1]} = \boldsymbol{E}\,u_0 + \Delta t\,\boldsymbol{A}\,f, \quad u_1 = u_0 + \Delta t\,\boldsymbol{B}\,f, \quad \|y_k + \tau_0\,f_k\| \le \|y_k\| \ \ (1 \le k \le s).$$

We denote the last component of the vector $y_k \in \mathbb{V}$ by η_k $(1 \le k \le s)$. Furthermore, we denote the vectors in $\mathbb{R}^s$, with components η_k, ζ_k, by η and ζ, respectively. From the equality in (3.6a), we have $\eta = C\,\zeta$. Because the rows of A are different from each other, the same holds for the rows of C. This allows us to choose $\zeta \in \mathbb{R}^s$ such that $\eta_j \ne \eta_k$ (for $j \ne k$). Hence $y_j \ne y_k$ (for $j \ne k$).

In view of the last property of the vectors y_k, we can define $F : \mathbb{V} \to \mathbb{V}$ by

$$F(v) = f_k \ \ (\text{for } v = y_k) \quad \text{and} \quad F(v) = 0 \ \ (\text{for } v \in \mathbb{V} \setminus \{y_1, \ldots, y_s\}).$$

With this F, we have (1.5) as well as (3.5) (with $n = 1$). □

Lemma 6.3. *For given $\gamma > 0$, assumption* (4.6) *is fulfilled, when γ_1 exists satisfying*

$$\tfrac{\gamma}{2} < \gamma_1 < \gamma, \quad I + \gamma_1\,A \text{ is invertible, and}$$
$$C_1 = \gamma_1\,A\,(I + \gamma_1\,A)^{-1} \text{ satisfies } spr(|C_1|) < 1.$$

Proof of Lemma 6.3. We have $I + \gamma\,A = (I + \gamma_1\,A)\,(I + H)$, with $H = \epsilon\,C_1$ and $\epsilon = \frac{\gamma - \gamma_1}{\gamma_1}$. Because $\operatorname{spr}(H) \le \operatorname{spr}(|H|) = |\epsilon|\operatorname{spr}(|C_1|) < 1$, we can conclude that the matrix $I + \gamma\,A$ is invertible. □

Proof of $\varphi(\gamma) \le \alpha$ for row-irreducible RKMs satisfying (3.3a). Assume property (3.3a) holds with $\alpha < \infty$. We shall prove $\varphi(\gamma) \le \alpha$, in two steps.

Step 1. Assume first (4.6). We shall prove $\mathrm{spr}(|C|) < 1$, by showing that the assumptions

$$Y \in \mathbb{R}^s, \;\; |C|\, Y = \lambda\, Y, \;\; Y \geq 0, \;\; \lambda \geq 1 \tag{6.5}$$

imply $Y = 0$. This implication proves that $\mathrm{spr}(|C|) < 1$, in view of the Perron-Frobenius theory (cf. e.g. [18, p. 503]).

Under the assumptions (6.5), the inequalities (6.3) hold with $\theta = 0$, $X = Y$, so that we can apply Lemma 6.2. It follows from (6.4), (3.3a) that $\|Y\|_\infty \leq \||C|\, Y\|_\infty = \max_k \|y_k^{[1]}\| \leq \alpha\, \|u_0\| = \alpha\, \theta = 0$, which proves $Y = 0$. Hence, $\mathrm{spr}(|C|) < 1$.

The inequalities (6.3) hold also with the choice $\theta = 1$, $X = (I - |C|)^{-1} |E - C\, E|$. Applying Lemma 6.2 in this situation, we conclude from (6.4), (3.3a) that $\varphi(\gamma) = \|X\|_\infty = \max_k \|y_k^{[1]}\| \leq \alpha\, \|u_0\| = \alpha$, which completes the proof of $\varphi(\gamma) \leq \alpha$ under assumption (4.6).

Step 2. Next, we shall prove (4.6). We choose any $\gamma_1 \in (\gamma/2,\, \gamma)$ for which $I + \gamma_1\, A$ is invertible. Because property (3.3a) holds for the given γ, it holds also with γ replaced by γ_1. By what we proved in Step 1, the matrix $C_1 = \gamma_1\, A\, (I + \gamma_1\, A)^{-1}$ has $\mathrm{spr}(|C_1|) < 1$. Applying Lemma 6.3, we arrive at (4.6). This completes the proof of $\varphi(\gamma) \leq \alpha$. □

Proof of $\psi(\gamma) \leq \beta$ *for irreducible RKMs satisfying* (3.3b). We assume (3.3b) with $\beta < \infty$, and shall prove $\psi(\gamma) \leq \beta$. By the same argument as used in the above proof of $\varphi(\gamma) \leq \alpha$ (Step 2), one sees that we can assume (4.6).

Under the assumptions (6.5), we shall again prove $Y = 0$.

Applying Lemma 6.2, with $\theta = 0$ and $X = Y$ similarly as above, we find from (6.4b), (3.3b): $|D|\, Y = \|u_1\| \leq \beta\, \|u_0\| = \beta\, \theta = 0$. Hence, the products of corresponding components of D and Y satisfy $d_i\, y_i = 0 \;\; (1 \leq i \leq s)$. It follows that $y_{i_0} = 0$ if $d_{i_0} \neq 0$.

On the other hand, if $d_{i_0} = 0$, then by Lemma 6.1 there exists $i_n, \ldots, i_1 \in \{1, \ldots, s\}$ with $d_{i_n}\, c_{i_n i_{n-1}} \cdots c_{i_1 i_0} \neq 0$. Hence, $d_{i_n} \neq 0$, so that $y_{i_n} = 0$. Because $0 = \lambda^n\, y_{i_n} = \sum |c_{i_n j_{n-1}}| \cdots |c_{j_1 j_0}|\, y_{j_0}$ (where the summation is over all $j_{n-1}, \ldots, j_1, j_0$) there follows: $|c_{i_n i_{n-1}}| \cdots |c_{i_1 i_0}|\, y_{i_0} = 0$. Thus we still have $y_{i_0} = 0$, so that $Y = 0$. By the Perron-Frobenius theory, we conclude that $\mathrm{spr}(|C|) < 1$.

Using Lemma 6.2, with the choice $\theta = 1$, $X = (I - |C|)^{-1} |E - C\, E|$, we find from (6.4b), (3.3b), that $\psi(\gamma) = \|u_1\| \leq \beta\, \|u_0\| = \beta$, which completes the proof of $\psi(\gamma) \leq \beta$. □

6.3. **Completing the proof of Theorem 3.2.** In this section, we consider an arbitrary row-irreducible RKM of general type (3.9). We shall denote by u_N the Runge-Kutta approximation, obtained after consecutive applications of formula (1.2) starting from u_0, and by u_n $(1 \leq n \leq N - 1)$ the corresponding intermediate approximations. In the following proof of the inequalities $\alpha_N \geq \varphi(\gamma)\, \psi(\gamma)^{N-1}$ and $\beta_N \geq \psi(\gamma)^N$, we assume $\alpha_N < \infty$, $\beta_N < \infty$ and (4.6). By the same argument as used in Section 6.2 (Step 2 of the proof of $\varphi(\gamma) \leq \alpha$), it can be seen that we lose no generality in making assumption (4.6).

In view of (3.5b), we have $u_n = u_0 + \Delta t \cdot \sum_{j=1}^{n} \boldsymbol{B}\,\boldsymbol{F}(y^{[j]})$ $(1 \le n \le N)$, so that by using (3.5a) there follows:

$$y^{[n]} \;=\; \boldsymbol{E}\,u_0 + \Delta t \cdot \big[\boldsymbol{A}\,\boldsymbol{F}(y^{[n]}) + \sum_{j=1}^{n-1} \boldsymbol{E}\,\boldsymbol{B}\,\boldsymbol{F}(y^{[j]})\big] \quad (1 \le n \le N), \tag{6.6a}$$

$$u_N \;=\; u_0 + \Delta t \cdot \sum_{j=1}^{N} \boldsymbol{B}\,\boldsymbol{F}(y^{[j]})\big). \tag{6.6b}$$

The formulas (6.6) can be viewed as describing one step of a (formal) RKM with $m = N \cdot s$ stages. Any vectors $y^{[n]} \in \mathbb{V}^s$, $u_N \in \mathbb{V}$ satisfy (6.6) if and only if they are generated by N consecutive applications of method (1.2), starting from u_0. Below, we shall prove the bounds $\alpha_N \ge \varphi(\gamma)\,\psi(\gamma)^{N-1}$ and $\beta_N \ge \psi(\gamma)^N$, by applying Lemma 6.2 and Theorem 3.1 to the RKM (6.6).

By T we denote the $m \times m$ coefficient matrix corresponding to the m internal stages (6.6a) of the m-stage RKM. One sees that T has a block Toeplitz structure, being made up of $s \times s$ blocks $T_{n,j} = T_{n-j+1}$. where $T_k = 0$ $(k \le 0)$, $T_1 = A$, $T_k = E\,B$ $(k > 1)$.

Because $I + \gamma\,A$ is invertible, the same holds for $I + \gamma\,T$. Furthermore, the row-irreducibility of method (1.2) combined with (3.9) implies that all rows of T are different from each other. Consequently, the RKM specified by (6.6) satisfies the assumptions required for an application of Lemma 6.2, with s, A, E, C, respectively, replaced by m, T, S and P, where S is the $m \times 1$ matrix with all entries equal to 1, and $P = \gamma\,T\,(I + \gamma\,T)^{-1}$. Note that P is a block Toeplitz matrix, made up of $s \times s$ blocks $P_{n,j} = P_{n-j+1}$ where $P_k = 0$ $(k \le 0)$, $P_1 = C$.

Proof of $\alpha_N \ge \varphi(\gamma)\,\psi(\gamma)^{N-1}$ for row-irreducible RKMs, of general type, satisfying (3.8a).

Step 1. To prove $\mathrm{spr}(|C|) < 1$, we assume (6.5). We shall show that $Y = 0$.

Let $X = [X_n] \in \mathbb{R}^m$, where the subvectors $X_n \in \mathbb{R}^s$ satisfy $X_n = 0$ $(1 \le n \le N-1)$ and $X_N = Y$. Because $|P|\,X = \lambda\,X \ge X \ge 0$, we have (6.3) with $\theta = 0$ and with $|C|$ replaced by $|P|$. Applying Lemma 6.2 (with this replacement) to RKM (6.6), we see via (6.4) that, for some vector space $\mathbb{V}$ with seminorm $\|\cdot\|$, some function F satisfying (1.5), and vectors u_0, $y^{[n]}$ satisfying (6.6a) with (1.6), we have $\|u_0\| = 0$ and $\eta = |P|\,X$, where $\eta \in \mathbb{R}^m$ is made up of subvectors $\eta_n \in \mathbb{R}^s$ $(1 \le n \le N)$ with components $\eta_{i,n} = \|y_i^{[n]}\|$ $(1 \le i \le s)$.

Property (3.8a) implies: $\alpha_N \cdot \|u_0\| \ge \|\eta_N\|_\infty = \|\lambda\,Y\|_\infty$. Hence, $Y = 0$, which (by the Perron-Frobenius theory) implies: $\mathrm{spr}(|C|) < 1$.

Step 2. Because $\mathrm{spr}(|C|) < 1$, we have $\mathrm{spr}(|P|) < 1$ as well. We apply Lemma 6.2 once more to the RKM specified by (6.6), but now with the choices $\theta = 1$ and $X = (I - |P|)^{-1}|S - P\,S|$.

It follows via (6.4) that $\eta = |S - P\,S| + |P|\,X = X$, where $\eta = [\eta_n] \in \mathbb{R}^m$ and the vectors $\eta_n \in \mathbb{R}^s$ $(1 \le n \le N)$ have components $\eta_{i,n} = \|y_i^{[n]}\|$ $(1 \le i \le s)$.

Property (3.8a) implies: $\alpha_N \ge \|\eta_N\|_\infty$. Writing $X = [X_n] \in \mathbb{R}^m$, with $X_n \in \mathbb{R}^s$, we have $\|\eta_N\|_\infty = \|X_N\|_\infty$, so that also $\alpha_N \ge \|X_N\|_\infty$. In step 3, below, we shall find that $\|X_N\|_\infty = \varphi(\gamma)\,\psi(\gamma)^{N-1}$, which will complete the proof.

Step 3. We write $R = [R_n] = S - P\,S$, with $R_n \in \mathbb{R}^s$ $(1 \le n \le N)$. Using $(I + \gamma\,T)\,[R_n] = S$, we obtain $(I + \gamma\,A)\,R_n + \gamma\,E\,B\,(R_1 + \cdots + R_{n-1}) = E$, which

implies that $(I+\gamma A)\,R_n = (I-E\,D)\,(I+\gamma A)\,R_{n-1}$ (for $n>1$), with D defined in (3.1). Because $(I+\gamma A)\,R_1 = E$, there follows: $R_n = (I+\gamma A)^{-1}\,(I-E\,D)^{n-1}\,E = (I+\gamma A)^{-1}\,E\,(1-D\,E)^{n-1}$ (for $n\geq 1$). Hence,

$$|R_n| = |E - C\,E|\,\rho^{n-1} \quad (1\leq n\leq N), \qquad \text{with } \rho = |1 - D\,E|.$$

Using $(I+\gamma T)\,P = \gamma T$, we obtain $(I+\gamma A)\,P_n + \gamma E\,B\,(P_1+\cdots+P_{n-1}) = \gamma E\,B$ $(n>1)$, which implies $(I+\gamma A)\,P_n = (I-E\,D)\,(I+\gamma A)\,P_{n-1}$ (for $n>2$). Because $(I+\gamma A)\,P_2 = E\,D$, there follows: $P_n = (I+\gamma A)^{-1}\,(I-E\,D)^{n-2}\,E\,D = (I+\gamma A)^{-1}\,E\,(1-D\,E)^{n-2}\,D$ (for $n\geq 2$). Hence,

$$|P_1| = |C|, \quad |P_n| = \rho^{n-2}\,|E - C\,E|\,|D| \quad (2\leq n\leq N).$$

Using $(I-|P|)\,[X_n] = |R|$, in combination with the above expressions for $|R_n|$ and $|P_n|$, we obtain $(I-|C|)\,X_n = \rho^{n-1}\,|E-C\,E|+|P_2|\,\big(\rho^{n-2}\,X_1+\cdots+X_{n-1}\big)$. We modify this relation, by multiplying it with ρ and replacing n by $n-1$. Subtracting this modified equality from the original one, we obtain $(I-|C|)\,(X_n-\rho\,X_{n-1}) = |P_2|\,X_{n-1}$, i.e. $X_n = \big(\rho\,I + (I-|C|)^{-1}\,|P_2|\big)\,X_{n-1}$ $(n\geq 2)$.

Repeated application of the last equality, in combination with the formulas $(I-|C|)^{-1}\,|P_2| = Z\,|D|$ and $X_1 = Z$ (where Z is defined in (3.1)), yields:

$$X_n = (\rho\,I + Z\,|D|)^{n-1}\,Z = Z\,(\rho + |D|\,Z)^{n-1} = Z\,\psi(\gamma)^{n-1} \ (1\leq n\leq N),$$

which implies $\|X_N\|_\infty = \varphi(\gamma)\,\psi(\gamma)^{N-1}$. □

Proof of $\beta_N \geq \psi(\gamma)^N$ for irreducible RKMs, of general type, satisfying (3.8b).

Step 1. We shall first prove that DJ-irreducibility of method (1.2) implies the same property for the formal RKM specified by (6.6). We denote by $\bar A = (\bar a_{k,l})$, $\bar B = (\bar b_l)$ the matrices which are related to (6.6) in the same manner as A, B are related to (1.2). (We could have stuck to the notation T instead of $\bar A$, but the latter is more convenient here.)

Suppose (6.6) would be DJ-reducible; i.e., $\bar b_l = 0$, $\bar a_{k,l} = 0$ (for all $k\in\bar{\mathcal{M}}$, $l\in\bar{\mathcal{N}}$), where $\bar{\mathcal{M}}\cup\bar{\mathcal{N}} = \{1,\ldots,m\}$, $\bar{\mathcal{M}}\cap\bar{\mathcal{N}} = \emptyset$, $\bar{\mathcal{N}}\neq\emptyset$. We write $j\equiv l$ when the difference $l-j$ is an integer multiple of s, and put

$$\mathcal{N} = \{j \,:\, 1\leq j\leq s, \ \text{ and } j\equiv l \text{ for some } l\in\bar{\mathcal{N}}\}, \quad \mathcal{M} = \{1,\ldots,s\}\setminus\mathcal{N}.$$

It can be seen that $\mathcal{N}$ is nonempty and $b_j = 0$ $(j\in\mathcal{N})$, $a_{ij} = 0$ $(i\in\mathcal{M},\ j\in\mathcal{N})$. This contradicts DJ-irreducibility of method (1.2), and thus proves DJ-irreducibility of method (6.6).

Step 2. Consider the vectors $y_i^{[N+1]}$ obtained after $N+1$ steps of method (1.2). By Theorem 3.1, Part (I), we have $\max_i\|y_i^{[N+1]}\| \leq \varphi(\gamma)\,\|u_N\|$, so that an application of (3.8b) yields: $\max_i\|y_i^{[N+1]}\| \leq \varphi(\gamma)\,\beta_N\,\|u_0\|$. Hence, (3.8a) holds (with $N+1$ replacing N, and $\alpha_{N+1} = \varphi(\gamma)\,\beta_N$). By Theorem 3.2, Part (I), there follows

$$\varphi(\gamma)\,\psi(\gamma)^N \leq \varphi(\gamma)\,\beta_N.$$

In Step 3, below, we shall prove that $\varphi(\gamma)$ is finite, which in combination with the above inequality will complete the proof of $\beta_N\geq\psi(\gamma)^N$.

Step 3. We denote by $\bar{\psi}(\gamma)$ the function which is related to the matrices $\bar{A}$, $\bar{B}$ in the same way as the function $\psi(\gamma)$ is related to A, B.

Because the RKM specified by (6.6) is row- and DJ-irreducible, we can apply Theorem 3.1, Part (II), to it, so as to conclude from (3.8b) that $\beta_N \geq \bar{\psi}(\gamma)$. By assumption, β_N is finite, so that the same holds for $\bar{\psi}(\gamma)$. Consequently, $\mathrm{spr}(|P|) < 1$, and therefore also $\mathrm{spr}(|C|) < 1$. It is evident from (3.2a) that $\varphi(\gamma)$ is finite. □

7. Conclusions

We have studied for s-stage Runge-Kutta methods (RKMs), specified by coefficients a_{ij}, b_j, the important properties of boundedness and monotonicity (also called strong-stability). The focus has been on stepsize-coefficients γ which are decisive for these properties.

The crucial question has been considered of whether stepsize-coefficients, relevant to boundedness, can be larger than optimal stepsize-coefficients for monotonicity. For irreducible RKMs we have found, to our surprise, that this may happen only within a class of very special RKMs. The class consists of the RKMs for which indices i, j exist with:

$$(a_{i,1}, \ldots, a_{i,s}) = (0, \ldots, 0) \quad \text{and} \quad (a_{j,1}, \ldots, a_{j,s}) = (b_1, \ldots, b_s).$$

As a result, any irreducible explicit RKM allows no positive stepsize-coefficient γ relevant to boundedness, as soon as it admits no positive γ for monotonicity.

As a by-product of studying the above question, we have found a new characterization of Kraaijevanger's coefficient and recovered the well-known fundamental relation between this coefficient and monotonicity, as stated in the existing literature. We think the derivation of this relation in the present paper is more transparent and shorter than the classical one. Moreover, it requires a simpler irreducibility condition than usually imposed in the literature.

As a further by-product, we have found extensions in various directions of the relation between monotonicity and Kraaijevanger's coefficient, mentioned above. Separate conclusions have been obtained for internal and external monotonicity, as well as a direct connection between these two properties. Moreover, new characterizations have been found of stepsize-coefficients for monotonicity or boundedness.

We have applied the theoretical findings of the paper to various well-known RKMs. Furthermore, a special instance has been examined of an irreducible RKM with no positive stepsize-coefficient γ for monotonicity, but with still a positive γ for boundedness. A systematic search for methods with this property may be a subject of future research.

References

[1] J.C. Butcher, *The numerical analysis of ordinary differential equations*, John Wiley, Chichester, UK, 1987. MR878564 (88d:65002)

[2] J.C. Butcher, *Numerical methods for ordinary differential equations*, John Wiley, Chichester, UK, 2003. MR1993957 (2004e:65069)

[3] G. Dahlquist, R. Jeltsch, *Reducibility and contractivity of Runge-Kutta methods revisited*, BIT **46** (2006), 567–587. MR2265575 (2007h:65068)

[4] L. Ferracina, M.N. Spijker, *Stepsize restrictions for the total-variation-diminishing property in general Runge-Kutta methods*, SIAM J. Numer. Anal. **42** (2004), 1073–1093. MR2113676 (2005k:65126)

[5] L. Ferracina, M.N. Spijker, *An extension and analysis of the Shu-Osher representation of Runge-Kutta methods*, Math. Comp. **74** (2005), 201–219. MR2085408 (2005g:65110)

[6] L. Ferracina, M.N. Spijker, *Stepsize restriction for total-variation-boundedness in general Runge-Kutta procedures*, Appl. Num. Math. **53** (2005), 265–279. MR2128526 (2005k:65127)
[7] L. Ferracina, M.N. Spijker, *Strong stability of singly-diagonally-implicit Runge-Kutta methods*, Appl. Num. Math. **58** (2008), 1675–1686. MR2458475 (2009i:65114)
[8] S. Gottlieb, *On high order strong stability preserving Runge-Kutta and multistep time discretizations*, Journ. Scientif. Computing **25** (2005), 105–128. MR2231945 (2007f:65029)
[9] S. Gottlieb, D.I. Ketcheson, C.-W. Shu, *High order strong stability preserving time discretizations*, Journ. Scientif. Computing **38** (2009), 251–289. MR2475652 (2010b:65161)
[10] S. Gottlieb, C.-W. Shu, *Total-variation-diminishing Runge-Kutta schemes*, Math. Comp. **67** (1998), 73–85. MR1443118 (98c:65122)
[11] S. Gottlieb, C.-W. Shu, E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Review **43** (2001), 89–112. MR1854647 (2002f:65132)
[12] E. Hairer, S.P. Nørsett, G. Wanner, *Solving ordinary differential equations. I. Nonstiff problems*, Springer-Verlag, Berlin, 1993. MR1227985 (94c:65005)
[13] E. Hairer, G. Wanner, *Solving ordinary differential equations. II. Stiff and differential-algebraic problems*, Springer-Verlag, Berlins, 1996. MR1439506 (97m:65007)
[14] A. Harten, *High resolution schemes for hyperbolic conservation laws*, J. Comput. Phys. **49** (1983), 357–393. MR701178 (84g:65115)
[15] I. Higueras, *On strong stability preserving time discretization methods*, Journ. Scientif. Computing **21** (2004), 193–223. MR2069949 (2005d:65112)
[16] I. Higueras, *Representations of Runge-Kutta methods and strong stability preserving methods*, SIAM J. Numer. Anal. **43** (2005), 924–948. MR2177549 (2006j:65184)
[17] I. Higueras, *Strong stability for additive Runge-Kutta methods*, SIAM J. Numer. Anal. **44** (2006), 1735–1758. MR2257125 (2008c:65164)
[18] R.A. Horn, C.R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1998. MR1084815 (91i:15001)
[19] Z. Horváth, *Positivity of Runge-Kutta and diagonally split Runge-Kutta methods*, Appl. Num. Math. **28** (1998), 309–326. MR1655167 (99i:65073)
[20] W. Hundsdorfer, A. Mozartova, M.N. Spijker, *Stepsize conditions for boundedness in numerical initial value problems*, SIAM J. Numer. Anal. 47 (2009), 3797–3819. MR2576521
[21] W. Hundsdorfer, J.G. Verwer, *Numerical solution of time-dependent advection-diffusion-reaction equations*, Springer-Verlag, Berlin, 2003. MR2002152 (2004g:65001)
[22] D.I. Ketcheson, *An algebraic characterization of strong stability preserving Runge-Kutta schemes*, Undergraduate Thesis, Brigham Young University, Provo, Utah, USA, 2004.
[23] D.I. Ketcheson, C.B. Macdonald, S. Gottlieb, *Optimal implicit strong stability preserving Runge-Kutta methods*, Appl. Num. Math. **59** (2009), 373–392. MR2484928 (2010a:65113)
[24] J.F.B.M. Kraaijevanger, *Contractivity of Runge-Kutta methods*, BIT **31** (1991), 482–528. MR1127488 (92i:65120)
[25] R.J. LeVeque, *Finite volume methods for hyperbolic problems*, Cambridge University Press, Cambridge, 2002. MR1925043 (2003h:65001)
[26] S.J. Ruuth, *Global optimization of explicit strong-stability-preserving Runge-Kutta methods*, Math. Comp. **75** (2006), 183–207. MR2176394 (2006k:65180)
[27] C.-W. Shu, *Total-variation-diminishing time discretizations*, SIAM J. Sci. Statist. Comput., **9** (1988), 1073–1084. MR963855 (90a:65196)
[28] C.-W. Shu, *A survey of strong stability preserving high-order time discretizations*, Collected lectures on the preservation of stability under discretization, D. Estep and S. Tavener eds., SIAM, Philadelphia, 2002, pp. 51–65. MR2026663
[29] C.-W. Shu, S. Osher, *Efficient implementation of essentially nonoscillatory shock-capturing schemes*, J. Comput. Phys. **77** (1988), 439–471. MR954915 (89g:65113)
[30] M.N. Spijker, *Contractivity in the numerical solution of initial value problems*, Numer. Math. **42** (1983), 271–290. MR723625 (85b:65067)
[31] M.N. Spijker, *Monotonicity and boundedness in implicit Runge-Kutta methods*, Numer. Math. **50** (1986), 97–109. MR864307 (88a:65076)

[32] M.N. Spijker, *Stepsize conditions for general monotonicity in numerical initial value problems*, SIAM J. Numer. Anal. **45** (2007), 1226–1245. MR2318810 (2008e:65199)

[33] R.J. Spiteri, S.J. Ruuth, *A new class of optimal high-order strong-stability-preserving time discretization methods*, SIAM J. Numer. Anal. **40** (2002), 469–491. MR1921666 (2003g:65083)

Center for Mathematics and Computer Sciences, P.O. Box 94079, NL-1090-GB Amsterdam, Nederland

E-mail address: willem.hundsdorfer@cwi.nl

Mathematical Institute, Leiden University, P.O. Box 9512, NL-2300-RA Leiden, Nederland

E-mail address: spijker@math.leidenuniv.nl