

Monotonicity for Runge–Kutta Methods: Inner Product Norms

Inmaculada Higuera¹

Received October 2, 2003; accepted (in revised form) February 25, 2004

An important class of ordinary differential systems is that whose solutions satisfy a monotonicity property for a given norm. For these problems, a natural requirement for the numerical solution is the reflection of this monotonicity property, perhaps under certain stepsize restriction. For Runge–Kutta methods, when the applied norm is an arbitrary one, the stepsize restrictions depend on the radius of absolute monotonicity. However for many problems, monotonicity holds for inner product norms and therefore it makes sense to restrict the analysis to this class of norms to obtain, if possible, less restrictive results. In this paper, we consider monotonicity issues for Runge–Kutta methods when the applied norm is an inner product norm.

KEY WORDS: strong stability; SSP; monotonicity; Runge–Kutta methods; coercive problems

AMS SUBJECT CLASSIFICATION: 65L06; 65M20; 65M12; 35L65

1. INTRODUCTION

Given initial value problems for ordinary differential systems (ODEs)

$$\begin{aligned} \frac{d}{dt}u(t) &= f(t, u(t)) \quad t \geq t_0, \\ u(t_0) &= u_0, \end{aligned} \tag{1.1}$$

an important class of problems widely studied in the literature is that whose solutions $u(t)$ satisfy a monotonicity property of the form

$$\|u(t)\| \leq \|u(t_0)\| \quad \text{for all } t \geq t_0 \tag{1.2}$$

¹Departamento de Matemática e Informática, Universidad Pública de Navarra 31006, Pamplona, Navarra, Spain. E-mail: higuera@unavarra.es

for a given norm $\|\cdot\|$.

When the ODE (1.1) is solved numerically, a natural requirement for the numerical solution is the reflection of the qualitative properties of the exact solution. Thus, if the solutions of the ODE satisfy (1.2), it is natural to require

$$\|u_{n+1}\| \leq \|u_n\| \quad \text{for all } n \geq 0 \quad (1.3)$$

on the numerical solution. Furthermore, when the numerical monotonicity property (1.3) holds, as the numerical solution depends on the stepsize h , a natural question is whether (1.3) holds for all stepsizes $h > 0$, obtaining thus unconditional monotonicity, or if (1.3) holds only with a stepsize restriction $h \leq H$, obtaining conditional monotonicity.

In this paper we consider initial value problems of the form (1.1). We assume that $t_0 \in \mathbb{R}$, $u_0 \in \mathbb{R}^m$ and

- (i) f is a continuous function from $\mathbb{R} \times \mathbb{R}^m$ to \mathbb{R}^m ;
- (ii) for each $t_0 \in \mathbb{R}$ and $u_0 \in \mathbb{R}^m$ the problem (1.1) has a unique solution $u : [t_0, \infty) \rightarrow \mathbb{R}^m$;
- (iii) $\|\cdot\|$ is an inner product norm on \mathbb{R}^m such that for any $t_0 \in \mathbb{R}$ and any solution $u(t)$ to (1.1) we have

$$\|u(t)\| \leq \|u(t_0)\| \quad \text{for all } t \geq t_0. \quad (1.4)$$

The class of all pairs $(f, \|\cdot\|)$ satisfying (i)–(iii) is denoted by \mathcal{F} . If $(f, \|\cdot\|) \in \mathcal{F}$ we will say that the ODE (1.1) is monotone with respect to $\|\cdot\|$.

Observe that we restrict the study to inner product norms. These are the applied norms for some important types of problems like semibounded and coercive ODEs ([9, 18]). Usually these ODEs arise from spatial discretizations of partial differential equations where the relevant problem norm is also an inner product norm [7].

Before proceeding, we have to impose some conditions on $(f, \|\cdot\|)$ to ensure that $(f, \|\cdot\|) \in \mathcal{F}$. The class of problems considered in this paper is $(f, \|\cdot\|)$ such that

$$\|f(t, y) + \rho y\| \leq \rho \|y\| \quad \text{for all } t \in \mathbb{R}, \quad \text{for all } y \in \mathbb{R}^m. \quad (1.5)$$

For $\rho > 0$, the solutions of the ODE (1.1) satisfy (1.4) [10]. As the applied norm is an inner product norm, inequality (1.5) can be expressed as

$$\operatorname{Re} \langle f(t, y), y \rangle \leq -\nu \|f(t, y)\|^2 \quad (1.6)$$

for $\nu = 1/(2\rho) > 0$. Problems of the form (1.6) with $f(t, y) = L(t)y$ are considered in [4, Sec. 3.2]. Recall that in the most general case, the expression ν in (1.6) may depend on t , i.e. $\nu = \nu(t)$.

In particular, if $f(t, y) = Ly$, with L a linear and independent of t operator, then inequality (1.6) is

$$\operatorname{Re} \langle Ly, y \rangle \leq -\nu \|Ly\|^2, \quad (1.7)$$

which is the class of problems considered in [9, Sec. 3.2] and [18, Sec. 3]. These problems are called in [4,9] coercive problems.

As it was pointed out in [8], from the numerical point of view, linear nonautonomous problems $f(t, y) = L(t)y$, and linear autonomous problems $f(t, y) = Ly$ are completely different. The nonautonomous case is closer to the nonlinear case than to the linear case, and therefore it should be considered within the nonlinear case.

Observe that if λ_i denotes an eigenvalue of L , inequality (1.7) implies

$$\operatorname{Re} \lambda_i \leq -\nu |\lambda_i|^2,$$

i.e., the eigenvalues of L are in $D(1/2\nu)$, where

$$D(r) = \{z \in \mathbb{C} : |z + r| \leq r\}, \quad (1.8)$$

denotes the disk of center $-r$ and radius r . Thus we are considering only dissipative problems with moderate stiffness where explicit methods may be used. In our analysis, dissipative problems with eigenvalues λ such that $\operatorname{Re} \lambda \ll 0$ or $\operatorname{Re} \lambda = 0$ and the imaginary parts are large, are not included.

A common class of one step methods to solve numerically (1.1) is the Runge–Kutta methods. An s -stage Runge–Kutta method is defined by an $s \times s$ real matrix \mathcal{A} and a real vector $b \in \mathbb{R}^s$; we will refer to it as (\mathcal{A}, b) . We denote by $c = \mathcal{A}e$, with $e = (1, \dots, 1)^t \in \mathbb{R}^s$. From u_n , the numerical approximation of the solution $u(t)$ at $t = t_n$, we obtain u_{n+1} , the numerical approximation of the solution at $t_{n+1} = t_n + h$ from

$$u_{n+1} = u_n + h \sum_{i=1}^s b_i f(t_n + c_i h, U_{ni}),$$

where the internal stages U_{ni} , $i = 1, \dots, s$ are computed from

$$U_{ni} = u_n + h \sum_{j=1}^s a_{ij} f(t_n + c_j h, U_{nj}), \quad i = 1, \dots, s.$$

In a Runge–Kutta method, the internal stages U_{ni} are approximations of the solution at $t_{ni} = t_n + c_i h$, and for many methods $c_i \geq 0$ for all i , giving $t_{ni} \geq t_n$. Therefore, if the exact solution is monotone (1.2), it also makes sense to impose monotonicity for the internal stages. We give the following definitions.

Definition 1.1. Consider an ODE (1.1) monotone with respect to the norm $\|\cdot\|$.

1. A Runge–Kutta method is called monotone for the stepsize h if the numerical solution satisfies

$$\|u_{n+1}\| \leq \|u_n\| \quad \text{for all } n \geq 0.$$

2. A Runge–Kutta method is called IS-monotone for the stepsize h if the internal stages U_{ni} satisfy

$$\|U_{ni}\| \leq \|u_n\|, \quad i = 1, \dots, s \quad \text{for all } n \geq 0.$$

For Runge–Kutta methods, when the applied norm is an arbitrary one, the stepsize restrictions to obtain monotonicity and *IS*-monotonicity depend on the radius of absolute monotonicity ([3, 8]). However for many problems, the monotonicity properties hold for inner product norms and therefore it makes sense to restrict the analysis to this class of norms to obtain, if possible, less restrictive results. The aim of this paper is to study monotonicity issues for Runge–Kutta methods in the particular case that the applied norm is an inner product norm.

Monotonicity of Runge–Kutta methods is also studied in [5, 6, 12, 13, 15, 17], (see [4, 14] for a review on this topic). In these references, high order monotone and *IS*-monotone Runge–Kutta methods are constructed by convex combinations of the forward Euler method. These schemes are known as strong-stability preserving methods (SSP). As it was pointed out in [8], the class of problems considered in the SSP context satisfy (1.5).

The rest of the paper is organized as follows. In Sec. 2, we consider nonlinear problems. Using the ideas in [2, Chap. 6], we give new results for monotonicity and *IS*-monotonicity. Some results obtained in [4] are recovered. Section 3 is devoted to linear problems. In this case, the results come out from the application of the Von Neumann’s theory of spectral sets [11, §153]. We give a rigorous justification of the needed stepsize restrictions, and review the results obtained in [19]. We apply this material to study a problem in [18] where the optimum stepsize restriction was not obtained.

2. MONOTONICITY FOR RK METHODS: NONLINEAR PROBLEMS

In this section we consider nonlinear ODEs such that (1.6) holds. We begin giving sufficient conditions to get monotonicity and *IS*-monotonicity.

In the next results, inequalities of the form $A \geq 0$, with A a real matrix, should be understood componentwise. When we say that a matrix A is nonnegative, we mean nonnegative definite.

Theorem 2.1. Consider an ODE such that (1.6) holds. Given a Runge–Kutta method with coefficients (\mathcal{A}, b) , we denote by

$$\begin{aligned} a_i^t &= (a_{i,1}, \dots, a_{i,s}), \\ \mathcal{A}_i &= \text{diag} (a_{i,1}, \dots, a_{i,s}), \\ M^{(i)} &= \mathcal{A}_i \mathcal{A} + \mathcal{A}^t \mathcal{A}_i - a_i a_i^t, \quad i = 1, \dots, s, \\ B &= \text{diag} (b_1, \dots, b_s), \\ M &= B \mathcal{A} + \mathcal{A}^t B - b b^t. \end{aligned}$$

1. Suppose that the matrix $\mathcal{A} \geq 0$ and there exists $r_i \neq 0$ such that the matrices

$$M^{(i)} + \frac{1}{r_i} \mathcal{A}_i \quad i = 1, \dots, s$$

are nonnegative, then the method is IS-monotone under the step-size restriction

$$\frac{h}{r} \leq 2\nu \tag{2.9}$$

with $r = \min\{r_1, \dots, r_s\}$.

2. Suppose that the matrix $B \geq 0$, and there exists $r \neq 0$ such that the matrix

$$M + \frac{1}{r} B$$

is nonnegative, then the method is monotone under the stepsize restriction

$$\frac{h}{r} \leq 2\nu. \tag{2.10}$$

Proof. The proof is similar to Theorem 6.2.2. in [2]. We denote by $w_j = h f(t_{nj}, U_{nj})$ and compute

$$\|U_{ni}\|^2 = \|u_n\|^2 + 2 \sum_{j=1}^s a_{ij} \text{Re}\langle U_{nj}, w_j \rangle - \sum_{k,j=1}^s m_{kj}^{(i)} \langle w_k, w_j \rangle$$

$$\begin{aligned} &\leq \|u_n\|^2 + 2 \sum_{j=1}^s a_{ij} \operatorname{Re}\langle U_{nj}, w_j \rangle + \frac{1}{r_i} \sum_{j=1}^s a_{ij} \langle w_j, w_j \rangle \\ &\leq \|u_n\|^2 + \sum_{j=1}^s \frac{a_{ij}}{h} \left(-2\nu + \frac{h}{r_i} \right) \|w_j\|^2, \end{aligned}$$

where we have used condition (1.6) and the nonnegativity of $M^{(i)} + (1/r_i)\mathcal{A}_i$. As $a_{ij} \geq 0$, we obtain that $\|U_{ni}\|^2 \leq \|u_n\|^2$, $i = 1, \dots, s$, provided that $-2\nu + (h/r_i) < 0$. Proceeding in a similar way we obtain that $\|u_{n+1}\|^2 \leq \|u_n\|^2$ under a similar stepsize restriction. \square

Remark 2.2. The algebraic conditions on the Runge–Kutta coefficients obtained in part 2 of Theorem 2.1 to get monotonicity are the same as the ones obtained for contractivity in [2, Chap. 6]. Observe that contractivity and monotonicity are different concepts. Contractivity deals with the growth of the difference of solutions whereas monotonicity deals with the growth of solutions themselves. Only for homogeneous linear problems both concepts are identical. However the kind of computations done to obtain monotonicity for Runge–Kutta methods do not differ much from the kind of computations done to obtain contractivity.

In the contractivity context, the difference of two solutions is considered, and this is the reason to impose the circle condition [10]

$$\|f(t, \tilde{y}) - f(t, y) + \rho(\tilde{y} - y)\| \leq \rho \|\tilde{y} - y\| \quad \text{for all } t \in \mathbb{R}, y, \tilde{y} \in \mathbb{R}^m$$

with $\rho > 0$, that when the applied norm is an inner product norm it becomes [2, Chap. 6]

$$\operatorname{Re} \langle f(t, y) - f(t, \tilde{y}), y - \tilde{y} \rangle \leq -\nu \|f(t, y) - f(t, \tilde{y})\|^2 \quad (2.11)$$

with $\nu > 0$, whereas when monotonicity is studied, as only one solution is considered, the natural condition to impose is (1.5) or (1.6). \square

Obviously, if a Runge–Kutta method is monotone (IS-monotone) for any $f(t, y)$, it is monotone (IS-monotone) for the linear scalar problem $f(t, y) = \lambda(t)y$. In this case, the numerical solution is given by

$$u_{n+1} = K(\xi)u_n,$$

with $K(\xi) = 1 + b^t \xi (I - \mathcal{A}\xi)^{-1} e$, $\xi = \operatorname{diag}(\xi_1, \dots, \xi_s)$ and $\xi_i = h\lambda(t_n + c_i h)$. Observe that $c_i = c_j$ gives $\xi_i = \xi_j$. The internal stages are given by

$$U_{ni} = K_i(\xi)u_n$$

with $K_i(\xi) = 1 + a_i^t \xi (I - \mathcal{A}\xi)^{-1} e$. We thus get that the method is monotone with stepsize restriction $h/r \leq 2\nu$ if and only if

$$|K(\xi)| \leq 1$$

for all $\xi = \text{diag}(\xi_1, \dots, \xi_s)$ such that $\xi_i = \xi_j$, whenever $c_i = c_j$, and such that $\xi_i \in D^s(r)$, with $D(r)$ the disk (1.8), and IS-monotone with stepsize restriction $h/r \leq 2\nu$ if and only if

$$|K_i(\xi)| \leq 1 \quad i = 1, \dots, s$$

for all ξ as above.

An important class of Runge–Kutta methods are the nonconfluent ones. A Runge–Kutta method is called nonconfluent if the coefficients $c_i \neq c_j$ for $i \neq j$. For nonconfluent Runge–Kutta methods we have the following result.

Proposition 2.3. Consider the linear scalar function $f(t, y) = \lambda(t)y$ satisfying (1.6) and a nonconfluent Runge–Kutta method.

1. [2, Lemma 6.2.3] If the method is monotone with stepsize restriction $h/r \leq 2\nu$, then the matrix $B \geq 0$ and the matrix $M + \frac{1}{r}B$ is nonnegative.
2. If the method is IS-monotone with stepsize restriction $h/r \leq 2\nu$, then the matrix $\mathcal{A} \geq 0$ and the matrices $M^{(j)} + (1/r)\mathcal{A}_j$, $j = 1, \dots, s$ are nonnegative.

Proof. The proof for the second part can be obtained following the ideas in the proof of Lemma 6.2.3 in [2]. □

We thus obtain the first main result in this section.

Theorem 2.4. For nonconfluent Runge–Kutta methods, the following statements are equivalent.

1. The matrices $B \geq 0$, $\mathcal{A} \geq 0$, and the matrices $M + 1/rB$ and $M^{(j)} + \frac{1}{r}\mathcal{A}_j$, $j = 1, 2, \dots, s$ are nonnegative;
2. for any function $f(t, y)$ satisfying (1.6), the method is monotone and IS-monotone with stepsize restriction $h/r \leq 2\nu$;
3. for any scalar function $f(t, y) = \lambda(t)y$ satisfying (1.6), the method is monotone and IS-monotone with stepsize restriction $h/r \leq 2\nu$.

Proof. Part (1) \Rightarrow (2) is Theorem 2.1. Part (2) \Rightarrow (3) is obvious. Part (3) \Rightarrow (1) is Proposition 2.3. □

Remark 2.5.

1. The fact that the method is nonconfluent has only been used to prove (3) \Rightarrow (1). Observe that (1) gives sufficient conditions for monotonicity and IS -monotonicity for any Runge–Kutta methods.
2. For $r = \infty$, we obtain in (1) that $B \geq 0$ and the matrix M is nonnegative, i.e. the concept of algebraic stability. It is known that for nonconfluent methods, BN -stability (i.e. contractivity for nonlinear problems), algebraic stability and AN -stability (i.e. contractivity for scalar linear time dependent problems) are equivalent concepts [2, Theorem 4.3.8]. However this equivalence is also valid for a wider class of methods: the S -irreducible methods [2, Corollary 4.5.7]. We conjecture that Theorem 2.4 is also valid if we consider S -irreducible methods instead of nonconfluent ones.
3. If the above conjecture is true, on the following, we can change the nonconfluence condition by the irreducibility one.

The radius of absolute monotonicity of a Runge–Kutta method [10] played an important role in the stepsize restrictions to get contractivity and monotonicity for Runge–Kutta methods ([3,10], see [8] for a review on the topic). We remember its definition.

Definition 2.6. An s -stage RK method with coefficients (\mathcal{A}, b) is said to be absolutely monotonic at a given point ξ if $I - \xi\mathcal{A}$ is nonsingular and

- (i) The stability function $\phi(\xi) = 1 + \xi b^t (I - \xi\mathcal{A})^{-1} e \geq 0$,
- (ii) $\mathcal{A}(\xi) = \mathcal{A}(I - \xi\mathcal{A})^{-1} \geq 0$,
- (iii) $b(\xi)^t = b^t (I - \xi\mathcal{A})^{-1} \geq 0$,
- (iv) $e(\xi) = (I - \xi\mathcal{A})^{-1} e \geq 0$,

where $e = (1, 1, \dots, 1) \in \mathbb{R}^s$ and the vector inequalities are understood component-wise. Further the method is said to be absolutely monotonic on a given set $\Omega \in \mathbb{R}$ if it is absolutely monotonic at each $\xi \in \Omega$. The radius of absolute monotonicity $R(\mathcal{A}, b)$ is defined by

$$R(\mathcal{A}, b) = \sup\{r | r \geq 0 \text{ and } (\mathcal{A}, b) \text{ is absolutely monotonic on } [-r, 0]\}.$$

If there is no $r > 0$ such that (\mathcal{A}, b) is absolutely monotonic on $[-r, 0]$ we set $R(\mathcal{A}, b) = 0$.

An algebraic criteria to obtain $R(\mathcal{A}, b) > 0$ is given in the following theorem. We remember that a method is irreducible if it is neither S -reducible nor DJ -reducible [2, Definitions 4.4.1 and 4.4.3].

Theorem 2.7. [10, Theorem 4.2] We denote by $\text{Inc}(F)$ the incidence matrix of the matrix F defined as $\text{Inc}(F) = (g_{ij})$, where $g_{ij} = 1$ if $f_{ij} \neq 0$ and $g_{ij} = 0$ if $f_{ij} = 0$. Then, for irreducible coefficient schemes (\mathcal{A}, b) we have $R(\mathcal{A}, b) > 0$ if and only if $\mathcal{A} \geq 0$, $b > 0$ and $\text{Inc}(\mathcal{A}^2) \leq \text{Inc}(\mathcal{A})$.

We state a technical Lemma.

Lemma 2.8. Consider a RK method with coefficients (\mathcal{A}, b) .

1. [2, Lemma 6.2.6] If $M + 1/rB$ is nonnegative for some $r \neq 0$ and $B \geq 0$ with B singular, then the RK method is reducible.
2. If $M^{(j)} + (1/r_j)\mathcal{A}_j$ is nonnegative for some $r_j \neq 0$, $j = 1, 2, \dots, s$, then the RK method satisfies the condition $\text{Inc} \mathcal{A}^2 \leq \text{Inc} \mathcal{A}$.

Proof. Part (2). We simply have to prove that $(\mathcal{A})_{ij} = 0$ implies $(\mathcal{A}^2)_{ij} = 0$. Assume that $a_{ij} = 0$. We fix the index i and construct the sets

$$S^i = \{j \mid a_{ij} = 0\}, \quad T^i = \{j \mid a_{ij} \neq 0\}.$$

Observe that $S^i \neq \emptyset$. We denote by e_i the vector in \mathbb{R}^s with all the components zero except the i th one. For $j \in S^i$ we have

$$e_j^t \left(M^{(i)} + \frac{1}{r_i} \mathcal{A}_i \right) e_j = 2a_{ij}a_{jj} - a_{ij}^2 + \frac{1}{r_i} a_{ij} = 0.$$

As $M^{(i)} + 1/r_i \mathcal{A}_i$ is nonnegative, it implies that $w^t (M^{(i)} + (1/r_i)\mathcal{A}_i) e_j = 0$ for any w [2, Lemma 4.5.1]. In particular for $w = e_k$ with $k \in T^i$,

$$0 = e_k^t \left(M^{(i)} + \frac{1}{r_i} \mathcal{A}_i \right) e_j = a_{ik} a_{kj}.$$

As $a_{ik} \neq 0$, we thus obtain that $a_{kj} = 0$ for $j \in S^i$, $k \in T^i$. We compute

$$(\mathcal{A}^2)_{ij} = \sum_{k=1}^s a_{ik} a_{kj} = \sum_{k \in S^i} a_{ik} a_{kj} + \sum_{k \in T^i} a_{ik} a_{kj} = 0$$

because for $k \in S^i$, the terms $a_{ik} = 0$, and for $k \in T^i$, the terms $a_{kj} = 0$. \square

We can now give the second main result in this section.

Theorem 2.9. Consider a nonconfluent irreducible RK method with coefficients (\mathcal{A}, b) . The following three statements are equivalent.

1. The method is monotone and IS -monotone for an arbitrary norm;

2. the method is monotone and *IS*-monotone for inner product norms;
3. $R(\mathcal{A}, b) > 0$.

Proof. The implication (1) \Rightarrow (2) is obvious. To prove that (2) \Rightarrow (3) we reason as follows. If the nonconfluent irreducible method is monotone and *IS*-monotone, by Theorem 2.4 and Lemma 2.8, we have that $B > 0$, $\mathcal{A} \geq 0$ and $\text{Inc } \mathcal{A}^2 \leq \text{Inc } \mathcal{A}$. By Theorem 2.7, we get that $R(\mathcal{A}, b) > 0$. The implication (3) \Rightarrow (1) follows Theorem 2.5 in [3] (see too Theorems 2.7 and 2.9 in [8]). \square

Remark 2.10. We conjecture that Theorem 2.9 is also valid for confluent irreducible methods.

Observe that Theorem 29 does not state anything on conditional or unconditional monotonicity and *IS*-monotonicity. In the following Sections, we go deeper in these concepts. Before proceeding, we state the following Lemmas that will be used to get stage order and order barriers.

Lemma 2.11. Consider a RK method with coefficients (\mathcal{A}, b) .

1. [10, Theorem 8.5] If $\mathcal{A} \geq 0$, then the stage order \tilde{p} is at most 2. Further, if $\tilde{p} = 2$, then \mathcal{A} has a zero row.
2. [10, Lemma 8.6] If the vector $b > 0$ and the order p satisfies $p \geq 2k + 1$ for some integer $k \geq 0$, then the stage order \tilde{p} is at least k .

We recall that for stiff problems, the stage order is a lower bound, sometimes a sharp one, for the order of convergence. Therefore RK methods with low stage order are not suitable for them [2].

2.1. Unconditional Monotonicity; Unconditional *IS*-monotonicity

2.1.1. Unconditional Monotonicity

If $r < 0$ or $r = \infty$ then the stepsize restriction (2.10) in Theorem 2.1 disappears obtaining a result on unconditional monotonicity. More precisely, we have unconditional monotonicity if the matrices B , and M are nonnegative. This is actually the definition of algebraically stable methods [2, Definition 4.2.1].

Recall that explicit methods are not algebraically stable. Recall too that irreducible algebraically stable methods must have $b > 0$ [2, Corollary 4.5.3].

There are many well known algebraically stable methods. For example the s -stage Gauss, Radau IA, Radau IIA and Lobatto IIC methods are algebraically stable and have orders $2s$, $2s - 1$, $2s - 1$ and $2s - 2$, respectively. Thus when we only consider monotonicity, there is not any order barrier.

2.1.2. Unconditional IS-monotonicity

Similarly, if $r < 0$ or $r = \infty$, then the stepsize restriction (2.9) in Theorem 2.1 disappears obtaining a result on unconditional IS-monotonicity. More precisely, we have unconditional IS-monotonicity if $\mathcal{A} \geq 0$, and the matrices $M^{(i)}$, $i = 1, \dots, s$ are nonnegative. Remember that as $\mathcal{A} \geq 0$, we have the stage order barrier $\tilde{p} \leq 2$ given by Lemma 2.8. In particular, as s -stage Gauss, Radau IA, Radau IIA and Lobatto IIC methods have stage orders s , $s - 1$, s and $s - 1$ respectively, for $s \geq 3$, in the case of Gauss and Radau IIA, and $s \geq 4$ in the case of Radau IA and Lobatto IIC we have not IS-monotonicity. A simple inspection of these methods [2, Sec. 3.3] gives that condition $\mathcal{A} \geq 0$ only holds for the one stage Gauss method, i.e. the midpoint rule, the one stage Radau IIA method, i.e. the backward Euler method, and the one stage Radau IA method

$$\begin{array}{c|c} 1 & 1 \\ \hline 2 & 2 \\ \hline & 1 \end{array}, \quad \begin{array}{c|c} 1 & 1 \\ \hline & 1 \end{array}, \quad \begin{array}{c|c} 0 & 0 \\ \hline & 0 \end{array}.$$

These methods are unconditionally IS-monotone and have orders 2, 1 and 1, respectively.

2.1.3. Unconditional Monotonicity and IS-monotonicity

Nonconfluent irreducible unconditional monotonic and IS-monotonic methods must have $\mathcal{A} \geq 0$, $B > 0$, and the matrices M and $M^{(j)}$, $j = 1, 2, \dots, s$ must be nonnegative. For example, the midpoint rule and the backward Euler are monotonic and IS-monotonic.

Observe that the midpoint rule is not unconditionally monotonic and IS-monotonic when any norm is used [8, Sec. 2.5]. It can be easily computed that for this method $R(\mathcal{A}, b) = 2$ and hence we have a stepsize restriction. We have the following stage order and order barriers.

Lemma 2.12.

1. If a RK method satisfies $\mathcal{A} \geq 0$ and $b > 0$, and the matrix M is nonnegative, then the stage order $\tilde{p} \leq 1$.

2. If a RK method satisfies $b > 0$, the matrices $M^{(j)}$, $j = 1, \dots, s$ are nonnegative and the stage order $\tilde{p} \leq 1$, then the order $p \leq 2$.

Proof. Part (1). From $\mathcal{A} \geq 0$, we have the stage order barrier $\tilde{p} \leq 2$ given by Lemma 2.11. Furthermore, if $\tilde{p} = 2$, then \mathcal{A} must have a zero row, e.g. the j th row. As the matrix M is nonnegative, we get $b_j = 0$. Hence $\tilde{p} \leq 1$.

Part (2). We left and right multiply the matrix $M^{(j)}$ by $e = (1, \dots, 1)^t$, and use the nonnegativity of $M^{(j)}$ to get

$$e^t M^{(j)} e = 2(a_{j1}, \dots, a_{js})c - c_j^2 \geq 0, \quad j = 1, \dots, s,$$

and hence

$$2\mathcal{A}c - c^2 \geq 0. \quad (2.12)$$

If the method had order 3, then $b^t(2\mathcal{A}c - c^2) = 0$. This equality together with expression (2.12) and the fact that $b > 0$ gives that $2\mathcal{A}c - c^2 = 0$, that contradicts that $\tilde{p} \leq 1$. Therefore the order is at most 2. \square

We have got the following order and stage order barriers.

Proposition 2.13. If an irreducible nonconfluent method is unconditionally monotonic and IS-monotonic, then its order p satisfies $p \leq 2$, and its stage order \tilde{p} satisfies $\tilde{p} \leq 1$.

Observe that there are unconditionally monotone and IS-monotone methods with order 2, e.g. the midpoint rule. We remark that in [8, Sec. 2.5], when any norm is considered, we had the order barrier $p \leq 1$.

2.2. Conditional Monotonicity; Conditional IS-monotonicity

2.2.1. Conditional Monotonicity

We require the matrix $B > 0$ and the matrix $M + (1/r)B$ to be nonnegative for some $r > 0$. As $B > 0$, by Lemma 2.11, a given order of the method implies certain stage order. Thus for explicit methods, as the stage order is $\tilde{p} \leq 1$, we get that for nonconfluent irreducible explicit monotone methods, the order p is $p \leq 4$. This result is similar to the one obtained for contractive RK methods [2, Corollary 6.2.8]. For implicit methods we do not get any order barrier.

Observe that condition $\mathcal{A} \geq 0$ is not needed for monotonicity. As we will see in the examples below, when inner product norms are used, we can have a conditional monotone method with $r > 0$ and negative elements in the matrix \mathcal{A} .

2.2.2. Conditional IS-monotonicity

In this case, we require $\mathcal{A} \geq 0$ and the matrices $M^{(j)} + (1/r)A_j$, $j = 1, 2, \dots, s$, to be nonnegative for some $r > 0$. As $\mathcal{A} \geq 0$, by Lemma 2.11, for nonconfluent irreducible IS-monotone methods, we have the stage order restriction $\tilde{p} \leq 2$.

2.2.3. Conditional Monotonicity and IS-monotonicity

In this case, as $B > 0$ and $\mathcal{A} \geq 0$, we have the restrictions given by Lemma 2.11. For implicit methods, the order $p \leq 6$, and for explicit ones, $p \leq 4$. Observe that we get the same order barriers as when any norm is used [8, Sec. 2.6]. A detailed lecture of [10] shows that the order barriers for any norm came from the conditions $\mathcal{A} \geq 0$ and $b > 0$ (see [10, Theorem 8.5, Lemma 8.6, and Corollary 8.7]).

Remember that if $r > 0$, the stepsize restriction (2.9) in Theorem 2.1 is

$$h \leq 2\nu r.$$

For the forward Euler method, the matrix $B = (1)$ is nonnegative and $M + (1/r)B = -1 + (1/r)$ is nonnegative for $r \leq 1$. Therefore, for the forward Euler method we obtain monotonicity and IS-monotonicity with stepsize restriction

$$h \leq \Delta t_{FE}$$

with $\Delta t_{FE} = 2\nu$. For any other RK method, the stepsize restriction $h \leq 2\nu r$ can be written as

$$h \leq r \Delta t_{FE}.$$

When an arbitrary norm is used, the stepsize restriction is given by [3, 8]

$$h \leq R(\mathcal{A}, b) \Delta t_{FE}.$$

If $R(\mathcal{A}, b) < r$, less restrictive results are obtained for inner product norms.

2.2.4. Computation of the CFL Coefficient: Examples

The next step is to get RK methods with the highest CFL coefficients for monotonicity and IS-monotonicity. With respect to monotonicity, the aim is to find the largest r such that $M + (1/r)B$ is nonnegative. Since $B > 0$, we can compute $B^{1/2}$ and thus the largest r is given by

$$r = -\lambda_{\min}^{-1} \left[B^{-1/2} M B^{-1/2} \right], \quad (2.13)$$

where $\lambda_{\min}[\cdot]$ denotes the smallest eigenvalue of the matrix in $[\cdot]$.

Concerning to IS-monotonicity, if the matrix $\mathcal{A} > 0$ we can compute in a similar way

$$r_j = -\lambda_{\min}^{-1} \left[\mathcal{A}_j^{-1/2} M \mathcal{A}_j^{-1/2} \right], \quad j = 1, 2, \dots, s. \tag{2.14}$$

Observe that for explicit methods, $a_i^t = (a_{i,1}, \dots, a_{i,i-1}, 0, \dots, 0)$. If we denote by $\mathcal{A}_i[i-1] = \text{diag} (a_{i,1}, \dots, a_{i,i-1})$ and $M^{(i)}[i-1]$ the matrix obtained by taking the first $i-1$ rows and columns in the matrix $M^{(i)}$, we get

$$M^{(i)} + \frac{1}{r} \mathcal{A}_i = \begin{pmatrix} M^{(i)}[i-1] + \frac{1}{r} \mathcal{A}_i[i-1] & 0 \\ 0 & 0_{s-i+1} \end{pmatrix}$$

Hence $M^{(i)} + (1/r)\mathcal{A}_i$ is nonnegative if and only if the matrix $M^{(i)}[i-1] + (1/r)\mathcal{A}_i[i-1]$ is nonnegative. We can thus use the above procedure (2.13) with the lower dimension matrices.

This way of computing r is quite simple for some given methods.

Example 1. We consider the optimal explicit s -stage order 1 methods (\mathcal{A}, b) with $a_{ij} = 1/s, 1 \leq j < i \leq s$, and $b_j = 1/s, 1 \leq j \leq s$ considered in [8, Sec. 2.7]. For these methods $B = (1/s)I_s, M = -(1/s^2)I_s$ and thus $B^{-1/2}MB^{-1/2} = -(1/s)I_s$. Therefore the method is monotone under step-size restriction $h \leq 2\nu s$. Furthermore, as

$$\mathcal{A}_j = \frac{1}{s} \begin{pmatrix} I_{j-1} & 0 \\ 0 & 0 \end{pmatrix}, \quad M^{(j)} = -\frac{1}{s^2} \begin{pmatrix} I_{j-1} & 0 \\ 0 & 0 \end{pmatrix},$$

we get

$$M^{(j)} + \frac{1}{r} \mathcal{A}_j = \left(-\frac{1}{s^2} + \frac{1}{sr} \right) \begin{pmatrix} I_{j-1} & 0 \\ 0 & 0 \end{pmatrix}$$

and thus $r_j = s, j = 2 \dots, s$. Therefore this method is IS-monotone under stepsize restriction $h \leq 2\nu s$. As $R(\mathcal{A}, b) = s$, the stepsize restriction is the same as for the general case. \square

Example 2. For the 3-stage order 3 method

$$\begin{array}{c|ccc}
 0 & 0 & & \\
 1 & 1 & 0 & \\
 \frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 \\
 \hline
 & \frac{1}{6} & \frac{1}{6} & \frac{2}{3}
 \end{array} \tag{2.15}$$

the eigenvalues of $B^{-1/2}MB^{-1/2}$ are

$$\lambda \left[B^{-1/2}MB^{-1/2} \right] = \left| -1, -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right|,$$

and thus the method is monotone under stepsize restriction $h \leq 2\nu$. For the internal stages,

$$\begin{aligned} \lambda \left[\mathcal{A}_2^{-1/2}[1]M^{(2)}[1]\mathcal{A}_2^{-1/2}[1] \right] &= \{-1\}, \\ \lambda \left[\mathcal{A}_3^{-1/2}[2]M^{(3)}[2]\mathcal{A}_3^{-1/2}[2] \right] &= \left| -1, \frac{1}{2} \right|, \end{aligned}$$

and thus $r_2 = 1, r_3 = 1$. Hence the method is IS-monotone under stepsize restriction $h \leq 2\nu$. This stepsize restrictions is the same as the one obtained for an arbitrary norm because $R(\mathcal{A}, b) = 1$. We remark that (2.15) is the 3-stage order 3 optimal method obtained when any norm is used [12, 8]. □

Example 3. For the 3-stage order 3 method

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

with $\beta = (-1 + \sqrt{97})/16$, the eigenvalues of $B^{-1/2}MB^{-1/2}$ are

$$\lambda \left[B^{-1/2}MB^{-1/2} \right] = \{-0.939062, -0.320286, 0.259349\},$$

and thus the method is monotone under stepsize restriction $h \leq 2\nu r$ with $r = 1/0.939062 \approx 1.06489$. For the internal stages,

$$\begin{aligned} \lambda \left[\mathcal{A}_2^{-1/2}[1]M^{(2)}[1]\mathcal{A}_2^{-1/2}[1] \right] &= \{-2/3\}, \\ \lambda \left[\mathcal{A}_3^{-1/2}[2]M^{(3)}[2]\mathcal{A}_3^{-1/2}[2] \right] &= \left| -1, \frac{1}{3} \right|, \end{aligned}$$

and thus $r_2 = 3/2, r_3 = 1$. Hence the method is IS-monotone under stepsize restriction $h \leq 2\nu r$, with $r = 1$.

The method is monotone and IS-monotone under stepsize restriction $h \leq 2\nu r$, with $r = 1$. For this method, $R(\mathcal{A}, b) = 1/4(-7 + \sqrt{97}) \approx 0.712214$. Therefore with inner products norms, we can use greater stepsizes. \square

Example 4. For the classical 4-stage order 4 method

$$\begin{array}{c|cccc}
 0 & 0 & & & \\
 \frac{1}{2} & \frac{1}{2} & 0 & & \\
 \frac{1}{2} & 0 & \frac{1}{2} & 0 & \\
 1 & 0 & 0 & 1 & 0 \\
 \hline
 & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6}
 \end{array}$$

the eigenvalues of the matrix $B^{-1/2}MB^{-1/2}$ are

$$\lambda \left[B^{-1/2}MB^{-1/2} \right] = \left| -1, -\frac{1}{2}, 0, \frac{1}{2} \right|,$$

and thus the method is monotone under stepsize restriction $h \leq 2\nu$. As condition $\text{Inc } \mathcal{A}^2 \leq \text{Inc } \mathcal{A}$ does not hold, the matrices $M^{(j)} + 1/r \mathcal{A}_j$ are not nonnegative, and hence we cannot ensure IS-monotonicity. Observe that this method is confluent. We remark that this method is not monotone and IS-monotone when any norm is considered (see [8, Sec.2]).

In [4, Proposition 3.3], this method is proved to be monotone under the same stepsize restriction, $h \leq 2\nu$, for the particular case $f(t, y) = L(t) y$. However the proof given in [4] is much more complicated than the proof given here. Furthermore, the result is also valid for nonlinear problems. \square

Example 5. [2, Example 6.3.3] For the 3-stage order 3 method

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

the eigenvalues of $B^{-1/2}MB^{-1/2}$ are

$$\lambda[B^{-1/2}MB^{-1/2}] = \{-2, 0, 1\},$$

and thus the method is monotone under stepsize restriction $h \leq 2 \nu r$, with $r = 0.5$. Observe that for this method the coefficient matrix \mathcal{A} has negative values and conditions for IS-monotonicity do not hold for any r . \square

Example 6. For the optimal 5-stage order 4 method given in [10] and [17] with $R(\mathcal{A}, b) = 1.508180$, when inner products norms are considered, we obtain that the method is monotone for $r = 1.834970$ and IS-monotone for $r = 1.775844$. Thus for inner product norms, the stepsize restriction is slightly milder. \square

3. MONOTONICITY FOR RK METHODS: LINEAR CONSTANT COEFFICIENT PROBLEMS

We consider now linear constant coefficient ODEs, i.e. $f(t, y) = Ly$ with L such that (1.7) holds,

$$\operatorname{Re} \langle Lu, u \rangle \leq -\nu \|Lu\|^2, \quad \text{with } \nu > 0.$$

Remember that this condition is the expression of inequality

$$\|L + \rho I\| \leq \rho$$

for inner product norms.

For linear problems, the numerical approximation with a RK method (\mathcal{A}, b) is given by

$$u_{n+1} = \phi(hL)u_n,$$

where $\phi(z) = 1 + zb^t(I - z\mathcal{A})^{-1}e$ is the stability function of the method. We immediately obtain that

$$\|y_{n+1}\| \leq \|y_n\| \quad \text{for } h \leq h_0$$

if and only if

$$\|\phi(hL)\| \leq 1 \quad \text{for } h \leq h_0 \tag{3.16}$$

holds. This kind of issues, namely, inequalities of the form $\|u(T)\| \leq 1$ where T is a linear transformation and $u(z)$ is a rational function, are studied in the Von Neumann's theory of spectral sets [11, Section 153] developed for linear transformations in Hilbert spaces. To state the main theorem in this theory, we need the definition of spectral set.

Definition 3.1. A set Z of points of the complex plane (completed by the point at infinity) will be called a spectral set of the linear transformation T if it is closed, and if for every rational function $u(z)$ satisfying in Z the inequality

$$\|u(z)\| \leq 1,$$

the transformation $u(T)$ exists and satisfies the inequality

$$\|u(T)\| \leq 1.$$

We can state the relevant part of the Von Neumann's Theorem [11, p. 442] for the class of problems we deal with in this section.

Theorem 3.2. A necessary and sufficient condition that the closed domain

$$|z - a| \leq r$$

be a spectral set of the linear transformation T is

$$\|T - aI\| \leq r.$$

We recall that in [11, p. 442] necessary and sufficient conditions so that the sets $|z - a| \geq r$ and $\operatorname{Re} z \geq 0$ be spectral sets are also given.

We can thus state the following theorem whose proof is immediate from Theorem 3.2.

Theorem 3.3. Assume that the linear transformation L satisfies the condition

$$\operatorname{Re} \langle Lu, u \rangle \leq -\nu \|Lu\|^2, \quad \nu > 0$$

and the stability function satisfies

$$|\phi(z)| \leq 1 \quad \text{for all } z \in D\left(\frac{1}{2\nu}\right),$$

then

$$\|\phi(L)\| \leq 1.$$

This is essentially Theorem 6.1 in [16].

Thus, given the method (A, b) , if we denote by \mathcal{R} the radius of the largest disk contained in the stability region

$$\mathcal{S} = \{z \in \mathbb{C} : |R(z)| \leq 1\},$$

we obtain that $\|\phi(hL)\| \leq 1$ under the stepsize restriction

$$h \leq 2\nu\mathcal{R}.$$

This radius \mathcal{R} is defined in [16] as the stability radius.

In particular, for the forward Euler method, it holds that $\mathcal{R} = 1$ and thus we obtain contractivity under the stepsize restriction

$$h \leq \Delta t_{FE}$$

with $\Delta t_{FE} = 2\nu$. Therefore the CFL coefficient given in terms of the stepsize restrictions of the forward Euler method is precisely \mathcal{R} .

In order to obtain these CFL coefficients, we simply have to compute the radius of the largest disk contained in the stability region \mathcal{R} . In [19] they are computed for s -stage methods with linear order p ; we show them in Table I.

We remark that for $s = 3, 4$, the s -stage linear order s optimal CFL coefficients are given in [1, Corollary 7.0.10]. We remark too that for these methods, the stepsize restriction given in [18, Proposition 1], namely $h \leq 2\nu$ is not optimal. It is optimal for an arbitrary norm (see [8, Table 3.1]).

Table I. Radius of the Maximum Disk Contained in the Stability Region

s	p									
	1	2	3	4	5	6	7	8	9	10
1	1									
2	2	1								
3	3	2	1.2564							
4	4	3	2.0731	1.3926						
5	5	4	2.9488	2.2295	1.6085					
6	6	5	3.8649	3.0598	2.3766	1.7767				
7	7	6		3.9037	3.1739	2.5554	1.9771			
8	8	7			3.9920	3.3232	2.7249	2.1568		
9	9	8				4.1059	3.4802	2.9086	2.3504	
10	10	9					4.2466	3.6461	3.0877	2.5348

Remark 3.4. As it is pointed out in [16, Sec. 5], the threshold factor of the RK method r , that dictates the stepsize restrictions when an arbitrary norm is used, satisfies $r \leq \mathcal{R}$.

We finish this section with a simple example showing the applicability of these results. It has been taken from [18, Sec. 5.1].

Example 7. We consider the advection equation

$$u_t(x, t) = a u_x(x, t) \quad -1 \leq x \leq 1, \quad a = \text{Constant} > 0$$

together with the zero boundary condition $u(x, t)|_{x=1} = 0$. We consider spatial equidistant points $x_j = -1 + j\Delta x$, $\Delta x := 2/N$, and one-sided differences for spatial differencing to obtain

$$\frac{d}{dt} u_N = L_N u_N, \quad L_N = \frac{a}{\Delta x} \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & -1 \end{pmatrix}.$$

It can be checked that condition (1.7) holds for $\nu = \Delta x/(2a)$. Thus for example for third order 3-stage explicit RK methods we have contractivity under the stepsize restriction

$$\Delta t \leq 1.2564 \frac{\Delta x}{a},$$

which coincides with stepsize restriction obtained with the von Neumann stability condition. Recall that with the analysis done in [18, Corollary 3], the CFL condition obtained is more restrictive, namely, $\Delta t \leq \Delta x/a$. \square

ACKNOWLEDGMENT

The author thanks the anonymous referees for their valuable remarks and comments on the paper. Research supported by the Ministerio de Ciencia y Tecnología, Project BFM2001-2188.

REFERENCES

1. Cavagna, C. (1998). Stability analysis in the numerical solution of convection-diffusion equations. Report May 2002, University of Leiden.

2. Dekker, K., and Verwer, J. G. (1984). *Stability of Runge–kutta methods for stiff nonlinear differential equations*, CWI Monographs, Amsterdam.
3. Ferracina, L., and Spijker, M. N. (2005). Step size restrictions for the total-variation-diminishing property in general Runge–Kutta methods. *Math. Comp.* **74**, 201–219.
4. Gottlieb, S., Shu, C. W., and Tadmor, E. (2001). Strong stability-preserving high order time discretization methods. *SIAM Rev.* **43**, 89–112.
5. Gottlieb, S., and Shu, C. W. (1998). Total variation diminishing Runge–Kutta schemes. *Math. Comp.* **67**, 73–85.
6. Gottlieb, S., and Gottlieb, L. J. (2003). Strong stability preserving properties of Runge–Kutta time discretization methods for linear constant coefficient operators. *J. Sci. Comput.* **18**, 83–109.
7. Gustafsson, B., Kreiss H. O., and Olinger, J. (1995). *Time Dependent Problems and Difference Methods*, John Wiley, New York.
8. Higuera, I. (2004). On strong stability preserving time discretization methods. To appear in *J. Sci. Comput.* **21**, 193–223.
9. Levy, D., and Tadmor, E. (1998). From semidiscrete to fully discrete: stability of Runge–Kutta schemes by the energy method. *SIAM Rev.* **40**, 40–73.
10. Kraaijevanger, J. F. B. M. (1991). Contractivity of Runge–Kutta methods. *BIT* **31**, 482–528.
11. Riesz, F., and Sz-Nagy, B. (1990). *Functional Analysis*, Dover, New York.
12. Ruuth, S. J., and Spiteri, R. J. (2002). Two barriers on strong stability preserving time discretization methods. *J. Sci. Comput.* **17**, 211–220.
13. Shu, C. W., and Osher, S. (1988). Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comput. Phys.* **77**, 439–471.
14. Shu, C. W. (2002). A survey of strong stability preserving high order time discretizations, In: D. Estep and S. Tavener, (eds). *Proceedings in Applied Mathematics* 109, SIAM, pp. 51–65
15. Spijker, M. N. (1983). Contractivity in the numerical solution of initial value problems. *Numer. Math.* **42**, 271–290.
16. Spijker, M. N. (1985). Step size restrictions for stability of one-step methods in the numerical solution of initial value problems. *Math. Comp.* **45**, 377–392.
17. Spiteri, R. J., and Ruuth, S. J. (2002). A new class of optimal high order strong stability preserving time discretization methods. *SIAM J. Numer. Anal.* **40**, 469–491.
18. Tadmor, E. (2002). From semidiscrete to fully discrete: stability of Runge-kutta schemes by the energy method II, Collected lectures on the preservation of stability under discretization. In: Estep, D. and Tavener, S. (eds). *Proceedings in Applied Mathematics* Vol. 109, SIAM, pp. 25–49
19. Van der Marel, R. P. (1990). Stability radius of polynomials occurring in the numerical solution of initial value problems. *BIT* **30**, 516–528.