XX Congreso de Ecuaciones Diferenciales y Aplicaciones
X Congreso de Matemática Aplicada
Sevilla, 24-28 septiembre 2007
(pp. 1–7)

Numerical positivity for Runge-Kutta methods

I. Higueras[1], T. Roldán[1]

[1] *Dpto. Ingeniería Matemática e Informática, Universidad Pública de Navarra, Edificio Las Encinas, E-31006 Pamplona. E-mails:* higueras@unavarra.es, teo@unavarra.es.

Palabras clave: Positivity, starting algorithms, Newton iteration, Runge-Kutta methods

Resumen

Over the last years, a great effort has been done to develop Runge-Kutta methods preserving qualitative properties of the exact solution like monotonicity of positivity. Some results are available in the literature that ensure these properties under certain stepsize restictions. However, these results are given for the exact numerical solution whereas in practice the numerical solution available is only an approximation of it. For example, when implicit Runge-Kutta methods are used, the numerical solution obtained comes out from the inexact numerical resolution of nonlinear systems.

The aim of this work is the study of the effective stepsize restrictions for positivity when implicit Runge-Kutta schemes are used. To achive this goal, we consider separately the problem of finding positive stage value predictors, and the analysis of the iterative scheme used, in this case Newton method.

1. Introduction

We consider IVPs for ordinary differential systems (ODEs) of the form

$$\begin{aligned} \frac{d}{dt}y(t) &= f(y(t))\,, \\ y(t_0) &= y_0\,. \end{aligned} \tag{1}$$

We assume that $f : \mathbb{R}^m \to \mathbb{R}^m$ is a sufficiently smooth function so that for each $t_0 \in \mathbb{R}$ and $y_0 \in \mathbb{R}^m$ the problem (1) has a unique solution $y : [t_0, \infty) \to \mathbb{R}^m$. We will denote this solution by $y(t; t_0, y_0)$.

In many situations the exact solution has a positivity property, i.e., if the initial condition $y_0 \geq 0$, then $y(t) \geq 0$ for all $t \geq t_0$, where the two vector inequalities should be

understood component-wise. This is the case for example in chemical reactions or models for population dynamics, where the variables represent densities, concentrations, or number of individuals. Much attention has been paid to problems like (1) with positivity properties (see [1], [12]). Positive ODEs are a particular case within the class of monotone problems [10]. The well known Kamke-Müller condition

$$x \leq y \quad \text{and} \quad x_i = y_i \quad \text{for some } i \text{ implies} \quad f_i(x) \leq f_i(y)\,, \tag{2}$$

ensures that

$$y_0 \leq \tilde{y}_0 \quad \text{implies} \quad y(t; t_0, y_0) \leq y(t; t_0, \tilde{y}_0)\,.$$

For example, for the problem considered in [14]

$$\begin{aligned} \dot{y}_1 &= -y_1 + y_3^3\,, \\ \dot{y}_2 &= -y_2 + y_1^3\,, \\ \dot{y}_3 &= -y_3 + y_2^3\,, \end{aligned} \tag{3}$$

we can apply criterion (2) for $x = 0$, obtaining that the solution $y(t) \geq 0$ whenever $y_0 \geq 0$. Similarly, criterion (2) for $y = e$, with $e = (1, \ldots, 1)^t$, gives that $y(t) \leq e$ whenever $y_0 \leq e$.

We assume that the ODE (1) is solved with a numerical method, e.g. a Runge-Kutta (RK) scheme $(\mathcal{A}, b^t)$. In this situation, if the exact solution to (1) has a qualitative property, it is desirable that the numerical method preserves this property too. RK methods preserving positivity have been studied in [9], [11], [12], [3]. Positivity is obtained under the stepsize restriction

$$h \leq \min\{\mathcal{R}(\mathcal{A}, b^t)\, \tau_0, H\}\,, \tag{4}$$

where $\mathcal{R}(\mathcal{A}, b^t)$ is the radius of absolute monotonicity of the RK method (see e.g. [13, 2, 6, 7] for a definition), τ_0 is the stepsize restriction for positivity for the explicit Euler method, and H is the stepsize restriction for solvability of the nonlinear systems. Stepsize restriction (4) is also valid to obtain $y \leq e$.

For example, for the trapezoidal rule

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array} \tag{5}$$

one gets $\mathcal{R}(\mathcal{A}, b^t) = 2$. For the second order 2-stage DIRK method ([2]),

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

it can computed that $\mathcal{R}(\mathcal{A}, b^t) = 4$.

The computation of the parameter τ_0 should be obtained for each problem. For example, for system (3), explicit Euler method gives

$$\begin{aligned} y_{1,n+1} &= y_{1,n} + h(-y_{1,n} + y_{3,n}^3)\,, \\ y_{2,n+1} &= y_{2,n} + h(-y_{2,n} + y_{1,n}^3)\,, \\ y_{3,n+1} &= y_{3,n} + h(-y_{3,n} + y_{2,n}^3)\,. \end{aligned}$$

Therefore, as for $h \leq 1$, we obtain $y_{n+1} \geq 0$, we can set $\tau_0 = 1$. Similarly, for $h \leq 1$, we obtain $y_{n+1} \leq 1$, and hence, we can also consider $\tau_0 = 1$.

Consequently, for problem (3), we can ensure positivity for $h \leq \min\{2, H\}$ when the trapezoidal rule is used, and for $h \leq \min\{4, H\}$, when the SDIRK (6) method is used.

However, these results are given for the exact numerical solution whereas often the numerical solution available is only an approximation of it. For example, when implicit Runge-Kutta (IRK) methods are used and f is a nonlinear function, the numerical solution obtained comes out from the inexact numerical resolution of nonlinear systems. This numerical solution is inexact because usually these systems are solved with an iterative scheme (e.g. Newton method), and iterations are done until a stopping criterion is fulfilled.

The aim of this work is the study of the effective stepsize restrictions for positivity when implicit RK schemes are used. To achieve this goal, we consider separately the problem of finding positive stage value predictors (section 2), and the analysis of the iterative scheme used, in this case Newton method (section 3). To show the difficulties that may occur, we consider some concrete methods applied to problem (3). Some conclusions are given in section 4.

2. Stage value predictors

Given the ODE (1), if the function f is non linear and the RK method is implicit, nonlinear systems must be solved in order to obtain the internal stages Y_{n+1}. To solve these non linear systems, usually an iterative scheme is used. In this case, an initial approximation $Y_{n+1}^{(0)}$ to Y_{n+1} is required. The values $Y_{n+1}^{(0)}$, known in the literature as stage value predictors or initializers ([16, 15, 4, 8]), should be as accurate as possible, because, otherwise, the number of iterations in each step may be too high or, even worse, the convergence may fail.

In this section we consider initializers built with the information from the previous step. We are going to assume that we have just given a step from t_{n-1} to t_n with stepsize h, and we have already computed the numerical solution y_n as well as the internal stages Y_n, and we are about to give another step from t_n to t_{n+1} with stepsize $r\,h$. In this process, we need to solve a nonlinear system to compute Y_{n+1}. The predictors studied in [8] were of the form

$$Y_{n+1}^{(0)} = b \otimes y_{n-1} + (B \otimes I)\, Y_n\,, \tag{7}$$

were the coefficients of matrix B and vector b are determined by imposing some order conditions. Our aim is to study whether it is possible to construct b and B such that $Y_{n+1}^{(0)} \geq 0$ provided that $y_{n-1} \geq 0$ and $Y_n \geq 0$. As y_{n-1} and Y_n may be any positive vector,

a sufficient and necessary condition to obtain positive initializers is $b \geq 0$ and $B \geq 0$. In order to obtain that $Y_{n+1}^{(0)} \leq e$ provided that $y_{n-1} \leq e$ and $Y_n \leq e$, we require not only $b \geq 0$ and $B \geq 0$ but also the condition $b + B\,e = e$. This equality is the consistency condition imposed in [8] for the initializers.

For any method, the trivial predictor

$$Y_{n+1}^{(0)} = e \otimes y_n \tag{8}$$

satisfies trivially $0 \leq Y_{n+1}^{(0)} \leq e$ provided that $0 \leq y_n \leq e$. However, its order is zero. The study done in [8, 5] shows that there is a reduction of the number of iterations and convergence problems can be avoid when high order stage value predictors are used.

2.1. Example 1

First we consider the method (6). The analysis done in [8] allow us to construct a family of order one predictors, namely

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \qquad B = \begin{pmatrix} \frac{1}{2}(-3b_1 - r - 1) & \frac{1}{2}(b_1 + r + 3) \\ \frac{1}{2}(-3b_2 - 3r - 1) & \frac{1}{2}(b_2 + 3r + 3) \end{pmatrix},$$

where r is the step ratio and $b = (b_1, b_2)$ is a vector of free parameters. When one analyzes positivity, the results are discouraging as it is not possible to impose positivity on the coefficients of b and B for $r > 0$. This fact can be easily seen from the first column of B. Consequently, for this method, the best positive initializer is the order zero family of the form (7) with $B\,e + b = e$, $b \geq 0$, $B \geq 0$.

2.2. Example 2

We consider now the trapezoidal rule (5). Following [8], we construct the family of order one predictors given by

$$b = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \qquad B = \begin{pmatrix} -b_1 & 1 \\ -b_2 - r & r + 1 \end{pmatrix}.$$

Observe again that positivity for the coefficients of b and B is not possible for $r > 0$. Consequently, for this method, the best positive initializer is the order zero family of the form (7) with $B\,e + b = e$, $b \geq 0$, $B \geq 0$.

3. Newton iterations

As we have pointed out above, the theoretical stepsize restriction for positivity is given by (4). However, in practice, the internal stages obtained are an approximation of the exact ones. To analyze the problems that may occur, it is enough to consider a simple problem like (3) and methods like (5) or (6). It can be checked that the nonlinear system can be solved for $h \leq H$ with $H \geq \mathcal{R}(\mathcal{A}, b^t)$, and therefore, the stepsize restriction for positivity given by (4) is

$$h \leq \mathcal{R}(\mathcal{A}, b^t)\,.$$

Let's see what happens when the Newton method is used as iterative scheme.

For the trapezoidal rule, the non linear system that must be solved is of the form

$$\begin{cases} Y_1 = y_n\,, \\ Y_2 = y_n + h\,\dfrac{1}{2}\,f(Y_1) + h\,\dfrac{1}{2}\,f(Y_2)\,. \end{cases}$$

For

$$F(Y) = Y - y_n - h\,\frac{1}{2}\,f(Y_1) - h\,\frac{1}{2}\,f(Y_2)\,,$$

the first Newton iteration for the second stage Y_2 reads

$$\begin{cases} F'(Y_2^{(0)})\,\Delta Y_2^{(0)} = -F(Y_2^{(0)})\,, \\ \qquad\qquad Y_2^{(1)} = Y_2^{(0)} + \Delta Y_2^{(0)}\,. \end{cases} \tag{9}$$

As there is no confusion, we drop the index in Y_2 and in the following we will use Y. We assume that we use the trivial initializer $Y^{(0)} = e \otimes y_n$. After solving the linear system in (9), we obtain $Y^{(1)}$, whose first component is given by

$$Y_1^{(1)} = \frac{(2-h)\,(2+h)^2 y_1 - 9h^3 y_1^3 y_2^2 y_3^2 + 4(1-h)\,h\,y_3^2\big(2\,y_3 + h\,(3y_2^3 + y_3)\big)}{8 + 12\,h + 6h^2 + h^3(1 - 27 y_1^2\,y_2^2\,y_3^2)}\,. \tag{10}$$

An exhaustive numerical search gives us that for $0 \le h \le 1$ and $0 \le y \le e$, then $0 \le Y_1^{(1)} \le e$. The symmetry of the problem permits to ensure that the other components $Y_2^{(1)}$ and $Y_3^{(1)}$ also satisfy $0 \le Y_i^{(1)} \le e$, $i = 2, 3$, for $0 \le h \le 1$. Furthermore, for $h > 1$, for some values of $y_i \in (0, 1)$, $i = 1, 2, 3$, some components of the vector $Y^{(1)}$ are either negative or greater than 1. Consequently, for the first Newton iteration, the effective stepsize restriction for positivity is not $h \le 2$ but $h \le 1$.

Now we consider the Runge-Kutta method (6) and the same problem (3). For this method the equations for the internal stages are

$$\begin{cases} Y_1 = y_n + h\,\dfrac{1}{4}\,f(Y_1)\,, \\ Y_2 = y_n + h\,\dfrac{1}{2}\,f(Y_1) + h\,\dfrac{1}{4}\,f(Y_2)\,. \end{cases}$$

Now both stages are implicit. For

$$F(Y) = Y_1 - y_n - h\,\frac{1}{4}\,f(Y_1)\,,$$

a Newton iteration for the first equation gives

$$\begin{cases} F'(Y_1^{(0)})\,\Delta Y_1^{(0)} = -F(Y_1^{(0)})\,, \\ \qquad\qquad Y_1^{(1)} = Y_1^{(0)} + \Delta Y_1^{(0)}\,. \end{cases}$$

Again, we drop the index in Y_1, and on the following we will use Y. We assume that we use the trivial predictor $Y^{(0)} = e \otimes y_n$. After solving the above linear system, we obtain an expression analogous to (10) for the first component of $Y^{(1)}$

$$Y_1^{(1)} = \frac{-2\left(9h^3y_2^2y_3^2y_1^3 - 2(h+4)^2y_1 + (h-2)hy_3^2\left(4y_3 + h\left(3y_2^3 + \mathrm{y3}\right)\right)\right)}{\left(1 - 27y_1^2y_2^2y_3^2\right)h^3 + 12h^2 + 48h + 64},$$

We have repeated an exhaustive numerical search for $0 \leq h \leq 2$ and $0 \leq y \leq e$, obtaining $0 \leq Y_1^{(1)} \leq e$. The symmetry of the problem permits to ensure that the other components $Y_2^{(1)}$ and $Y_3^{(1)}$ also satisfy $0 \leq Y_i^{(1)} \leq e$, $i = 2, 3$ for $0 \leq h \leq 2$. Similar results are obtained for the second internal stage.

4. Conclusions

In this paper, we have studied the effective stepsize restrictions for positivity when implicit RK schemes are used. To show the difficulties that may occur, we have considered some concrete methods applied to a simple problem.

In relation to the problem of finding positive stage value predictors, the results obtained are discouraging. Methods with large radius of absolute monotonicity only allow the trivial predictor. An open question is wether the order restriction obtained for positive stage value predictors for these two methods considered is valid for any implicit method.

With regard to the analysis of the Newton method, our example shows that the theoretical stepsize restriction is halved in practice. Hence the gain in size of the radius of absolute monotonicity for implicit RK schemes may be lost in the resolution of the nonlinear systems. A deeper analysis of positivity for implicit schemes is required.

Agradecimientos

This research has been supported by the Ministerio de Educación y Ciencia, Project MTM2005-03894.

Referencias

[1] J. Bruggeman, H. Burchar, B. Kooi, B. Sommeijer, *A second-order, unconditionally positive, mass-conserving integration scheme for biochemical systems*. Applied Numerical Mathematics 57 (2007), 36–58.

[2] L. Ferracina, M. N. Spijker, *Computing monotonicity preserving Runge-Kutta methods*. Submitted.

[3] A. Gerisch, R. Weiner, *On the positivity of low order explicit Runge-Kutta schemes applied in splitting methods*. Comput. Math. Appl. 45 (2003), 53–67.

[4] S. González-Pinto, J. I. Montijano, S. Pérez Rodríguez, *On the starting algorithms for fully implicit Runge-Kutta methods*. BIT, 40 (2000), 685–714.

[5] I. Gómez, I. Higueras, T. Roldán, *Starting algorithms for low stage order RKN methods*. Journal of Computational and Applied Mathematics, 140 (2002), 345–367.

[6] I. Higueras, *On strong stability preserving time discretization methods*. J. Sci. Comput. 21 (2004) 193–223.

[7] I. Higueras, *Representations of Runge-Kutta methods and strong stability preserving methods*, *SIAM J. Numer. Anal., 43* (2005) 924-948.

[8] I. Higueras, T. Roldán, *Starting algorithms for some DIRK methods*, Numerical Algorithms, 23 (2000) 357-369.

[9] I. Higueras, T. Roldán, *Positivity for Runge-Kutta and additive Runge-Kutta methods*. In preparation.

[10] M. W. Hirsh, H. Smith, *Monotone dynamical systems*. Handbook of Differential Equations, Ordinary Differential Equations (second volume), eds. A. Canada, P. Drabek, A. Fonda, Elsevier (2005), 239–357.

[11] Z. Horváth, *Positivity of RK and diagonally split RK methods*. Appl. Numer. Math. 28 (1998), 309–326.

[12] W. Hundsdorfer, J. Verwer, *Numerical solution of time-dependent Advection-Diffusion-Reaction equations*. Springer Series in Computational Mathematics, 2003.

[13] J. Kraaijevanger. *Contractivity of Runge-Kutta methods*. BIT 43 (2003), 571–586.

[14] P. E. Kloeden, J. Schropp, *Runge-Kutta methods for monotone differential and delay equations*. BIT 31 (1991), 482–528.

[15] M. P. Laburta, *Starting algorithms for IRK methods*. Comput. Appl. Math. 83 (1997), 269-288.

[16] J. Sand. *Methods for starting iterations schemes for implicit Ruge-Kutta formulae*. Computational ordinary differential equations (London, 1989), 115–126, Inst. Math. Appl. Conf. Ser. New Ser., 39, Oxford Univ. Press, New York, 1992.