View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Elsevier - Publisher Connector

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Computational and Applied Mathematics 189 (2006) 120–131

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

www.elsevier.com/locate/cam

Stability of Runge–Kutta–Nyström methods☆

I. Alonso-Mallo, B. Cano, M.J. Moreta*

Departamento de Matemática Aplicada, Universidad de Valladolid, C/ Dr. Mergelina s.n., 47011 Valladolid, Spain

Received 3 January 2005; received in revised form 10 January 2005

Abstract

In this paper, a general and detailed study of linear stability of Runge–Kutta–Nyström (RKN) methods is given. In the case that arbitrarily stiff problems are integrated, we establish a condition that RKN methods must satisfy so that a uniform bound for stability can be achieved. This condition is not satisfied by any method in the literature. Therefore, a stable method is constructed and some numerical comparisons are made.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Runge–Kutta–Nyström methods; Stability

1. Preliminaries

We are concerned with the study of the behaviour of the numerical solution obtained when integrating, with a Runge–Kutta–Nyström (RKN) method, the second-order ordinary differential system

$$\begin{aligned} &U''(t) = -B^2U(t) + f(t),\\ &U(0) = u_0,\\ &U'(0) = v_0, \end{aligned} \tag{1}$$

where B is a given symmetric positive definite matrix of order $m \geqslant 1$ and $U(t)$, $f(t)$, u_0, $v_0 \in \mathbb{C}^m$. These problems arise, for example, after the spatial discretization of second order in time partial differential equations. When the spatial discretization is refined, the value of m increases and the problem becomes *arbitrarily stiff*. A similar study for the numerical solution of first-order problems with one step and

☆ This research was supported by MTM 2004-08012 and JCYL VA103/04.
* Corresponding author. Tel.: +34 983 42 31 80; fax: +34 983 42 30 13.
E-mail addresses: isaias@mac.uva.es (I. Alonso-Mallo), bego@mac.uva.es (B. Cano), mjesus@mac.uva.es (M.J. Moreta).

0377-0427/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cam.2005.01.005

multistep numerical methods has been widely carried out in the literature (see for example [3,7], the recent paper [6] and the references therein).

The natural norm for the study of the well-posedness of (1) is the energy norm, given by

$$\|[U(t), U'(t)]^{\mathrm{T}}\|_B^2 = \|BU(t)\|_2^2 + \|U'(t)\|_2^2,$$

where $\|\cdot\|_2$ is the Euclidean norm in $\mathbb{R}^m$. The use of this norm will be crucial for the study of the stability of the numerical solution obtained with a RKN method, which is given by the Butcher array

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

with $c = [c_1, \ldots, c_s]^{\mathrm{T}}$, $\beta = [\beta_1, \ldots, \beta_s]^{\mathrm{T}}$, $b = [b_1, \ldots, b_s]^{\mathrm{T}}$ vectors of length s and $\mathscr{A} = [a_{ij}]$ an $s \times s$ matrix. We also denote $e = [1, \ldots, 1]^{\mathrm{T}}$, $c^j = [c_1^j, \ldots, c_s^j]^{\mathrm{T}}$ for $j \geqslant 1$.

We then solve (1) numerically with this RKN method using a time step size $k > 0$. If $n > 0$, we denote by u_n and v_n the numerical approximations to $U(t_n)$ and $U'(t_n)$ at $t_n = kn$. We obtain

$$[Bu_n, v_n]^{\mathrm{T}} = R(kB)^n [Bu_0, v_0]^{\mathrm{T}} + [f_n^1, f_n^2]^{\mathrm{T}}, \tag{2}$$

where f_n^1, f_n^2 depend on the source term $f(t)$, the coefficients of the RKN method and the matrix B. For $\theta \geqslant 0$, the matrix $R(\theta)$ is given by

$$R(\theta) = \begin{bmatrix} r_{11}(\theta) & r_{12}(\theta) \\ r_{21}(\theta) & r_{22}(\theta) \end{bmatrix} \tag{3}$$

with

$$\begin{aligned} &r_{11}(\theta) = 1 - \theta^2 \beta^{\mathrm{T}} (\mathscr{I} + \theta^2 \mathscr{A})^{-1} e, \quad r_{12}(\theta) = \theta[1 - \theta^2 \beta^{\mathrm{T}} (\mathscr{I} + \theta^2 \mathscr{A})^{-1} c], \\ &r_{21}(\theta) = -\theta b^{\mathrm{T}} (\mathscr{I} + \theta^2 \mathscr{A})^{-1} e, \quad r_{22}(\theta) = 1 - \theta^2 b^{\mathrm{T}} (\mathscr{I} + \theta^2 \mathscr{A})^{-1} c. \end{aligned} \tag{4}$$

From these formulas, we can define the matrix $R(kB)$ used in (2) by means of

$$\begin{aligned} r_{11}(kB) &= I - (\beta^{\mathrm{T}} \otimes k^2 B^2)(\mathscr{I} \otimes I + \mathscr{A} \otimes k^2 B^2)^{-1} (e \otimes I), \\ r_{12}(kB) &= kB[I - (\beta^{\mathrm{T}} \otimes k^2 B^2)(\mathscr{I} \otimes I + \mathscr{A} \otimes k^2 B^2)^{-1} (c \otimes I)], \\ r_{21}(kB) &= -(b^{\mathrm{T}} \otimes kB)(\mathscr{I} \otimes I + \mathscr{A} \otimes k^2 B^2)^{-1} (e \otimes I), \\ r_{22}(kB) &= I - (b^{\mathrm{T}} \otimes k^2 B^2)(\mathscr{I} \otimes I + \mathscr{A} \otimes k^2 B^2)^{-1} (c \otimes I), \end{aligned}$$

where I and $\mathscr{I}$ are, respectively, the $m \times m$ and $s \times s$ identity matrix, and $\otimes$ is the symbol for the Kronecker product of matrices.

Let us suppose that we approximate the initial values u_0 and v_0 with the values $\tilde{u}_0$ and $\tilde{v}_0$, and we denote by $\tilde{u}_n$ and $\tilde{v}_n$ the numerical solution obtained in this way. Since u_n, v_n and $\tilde{u}_n$, $\tilde{v}_n$ satisfy (2),

$$\|[\tilde{u}_n - u_n, \tilde{v}_n - v_n]^{\mathrm{T}}\|_B \leqslant \|R(kB)^n\|_2 \|[\tilde{u}_0 - u_0, \tilde{v}_0 - v_0]^{\mathrm{T}}\|_B. \tag{5}$$

From (5), the proof of stability when integrating an equation like (1) with a RKN method is related to the boundedness of the powers $\|R(kB)^n\|_2$ for $n \geqslant 0$. Since we assume that B is symmetric and positive definite, B is normal so that this well-known spectral result applies

$$\|R(kB)^n\|_2 \leqslant \sup_{\theta \in \sigma(kB)} \|R(\theta)^n\|_2, \tag{6}$$

where $\sigma(kB)$ is the spectrum of the matrix kB. (For the case when B is not normal, see [2] for a similar result to (6), but considering the numerical range instead of the spectrum.)

In the literature about stability for RKN methods, the boundedness of the second term in (6) is related to a certain interval of stability. However, the authors do not coincide in some definitions and nomenclature, although all of them have the same ideas in common. Therefore, we state in this paper the suitable definitions for our work. Let $\sigma(R(\theta))$ be the spectrum of (3) and $\rho(R(\theta))$ be its spectral radio. We remember that two squared matrices A and B of the same dimension are similar when there is a nonsingular matrix P such that $A = P^{-1}BP$ (see e.g. [5]). We also say that a matrix is simple when it is similar to a diagonal matrix.

Definition 1. The interval $\mathscr{C}=[0, \beta_{\text{stab}})$ is the interval of stability of the RKN method if $\beta_{\text{stab}} \in \mathbb{R}^+ \cup \{+\infty\}$ is the highest value such that

$$\mathscr{C} \subset \{\theta \in \mathbb{R}^+ \cup \{0\} / \rho(R(\theta)) \leqslant 1 \quad \textit{and} \quad R(\theta) \textit{ is simple when } \rho(R(\theta)) = 1\}.$$

We will say that the RKN method is R-stable if $\mathscr{C} = \mathbb{R}^+ \cup \{0\}$.

Definition 2. The interval $\mathscr{C}^* = [0, \beta_{\text{per}})$ is the interval of periodicity of the RKN method if $\beta_{\text{per}} \in \mathbb{R}^+ \cup \{+\infty\}$ is the highest value such that

$$\mathscr{C}^* \subset \{\theta \in \mathbb{R}^+ \cup \{0\} / R(\theta) \textit{ is simple and for all } \lambda \in \sigma(R(\theta)), |\lambda| = 1\}.$$

We will say that the RKN method is P-stable if $\mathscr{C}^* = \mathbb{R}^+ \cup \{0\}$.

Remark 3. The definitions of periodicity and stability intervals which can be found in the literature (see e.g. [1,4,9–13]) are given in terms of another stability matrix $\tilde{R}(kB)$ obtained from the recurrence relation

$$[u_n, kv_n]^{\mathrm{T}} = \widetilde{R}(kB)^n [u_0, kv_0]^{\mathrm{T}} + [\widetilde{f}_n^1, \widetilde{f}_n^2]^{\mathrm{T}}$$

which is similar to (2). It is straightforward to deduce that

$$\widetilde{R}(kB) = \begin{bmatrix} (kB)^{-1} & 0 \\ 0 & I \end{bmatrix} R(kB) \begin{bmatrix} kB & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} r_{11}(kB) & (kB)^{-1} r_{12}(kB) \\ kB r_{21}(kB) & r_{22}(kB) \end{bmatrix}.$$

Therefore, both matrices $R(kB)$ and $\widetilde{R}(kB)$ are similar and the previous definitions do not depend on the choice since both matrices have the same spectrum.

However, the norms $\|R(kB)^n\|_2$ and $\|\widetilde{R}(kB)^n\|_2$, that we need in the proofs of stability in Section 2, are very distinct. Notice that the use of the energy norm is more natural because it depends on the problem (1) but it is independent on the numerical time discretization and the energy norm is the suitable choice to study the well posedness of the differential problem (1).

Remark 4. The definitions of periodicity and stability intervals in the literature are slightly different to our previous definitions. The main difference is that we permit the case of 1 or -1 being double eigenvalues of $R(\theta)$ but requiring that $R(\theta)$ is a simple matrix. This condition is sufficient for the proof of the stability, i.e. the boundedness of $\|R(kB)^n\|_2$ for $n \geqslant 0$, in the following section. For example, our definition of R-stability is not equivalent to the one given in [11], but an R-stable method following [11] is R-stable according to our definition because [11] do not include the case of double eigenvalues on the

unit disk. The R-stability is nice if (1) is very stiff, for example, for spatial discretizations of some partial differential equations.

On the other hand, our definition of P-stability is not equivalent to the one given in [4,11], but a P-stable method according to our definition is P-stable following [4,11] because they include the case of $+1$ or -1 being double eigenvalues without $R(\pm 1)$ being simple. The concept of P-stability is suitable for very stiff problems in which we want the numerical solution not to be dissipative.

It is plain that, from our definitions, any P-stable method is also R-stable because $\mathscr{C}^* \subseteq \mathscr{C}$. A crucial practical difference between R-stability and P-stability is that, for P-stable methods, it is not possible that $R(\theta)^n \to 0$ as $n \to \infty$ for $\theta \in \mathscr{C}^*$ because $\rho(R(\theta)) = 1$. Therefore, for P-stable methods the initial errors do not diminish when the numerical approximation progresses in time.

It is natural to consider the previous definitions of stability because the asymptotic behaviour of $R(\theta)^n$ is governed by the spectral radio $\rho(R(\theta))$. When the matrix $R(\theta)$ is normal, $\rho(R(\theta)) = \|R(\theta)\|_2$ and, since $\sigma(B)$ is contained in a closed subinterval of $\mathbb{R}^+$, taking the time step size k small enough in order that $\sigma(kB) \subset \mathscr{C}$,

$$\|R(kB)^n\|_2 \leqslant \sup_{\theta \in \sigma(kB)} \|R(\theta)^n\|_2 \leqslant \sup_{\theta \in \mathscr{C}} \|R(\theta)^n\|_2 \leqslant \sup_{\theta \in \mathscr{C}} \|R(\theta)\|_2^n = \sup_{\theta \in \mathscr{C}} \rho(R(\theta))^n \leqslant 1$$

and we have stability. However, in the case of a nonnormal matrix $R(\theta)$, we have $\rho(R(\theta)) < \|R(\theta)\|_2$, and the difference between $\rho(R(\theta))$ and $\|R(\theta)\|_2$ can be *arbitrarily large* (we have found that $R(\theta)$ is always nonnormal for the RKN methods in the literature except for RKN derived from RK methods, see Remark 9 at the end of the following section).

The case of R- and P-stable methods is particularly interesting for very stiff problems, but it is necessary to bound uniformly the second term in (6) in an infinite interval. In fact, we have checked in the literature and we have not found any P- or R-stable RKN method satisfying a uniform bound of this second term. Therefore, we can deduce that actually there are no known stable RKN methods for very stiff problems. Our main contribution is a necessary and sufficient condition guaranteeing that the second term in (6) is uniformly bounded for $\theta \in \mathscr{C} = \mathbb{R}^+ \cup \{0\}$ and, therefore, the RKN method is stable, even when we apply it to an arbitrarily stiff problem. For this we impose a simple algebraic condition on the coefficients of the RKN method which can be used in practice to obtain a stable RKN method.

We now briefly give an outline of the rest of the paper. In Section 2, we state the results on stability. In the case of an infinite interval of stability, we obtain the necessary and sufficient condition in order to have a stable method. Since this condition is not satisfied by the methods in the literature, we construct in Section 3 a fourth-order SDIRKN method of four stages satisfying this condition. In Section 4 we present some numerical experiments showing the advantages of this method.

2. Theoretical results

In this section, we use the complex Schur decomposition [5] of the 2×2 matrix $R(\theta)$, given by (3). We know there exists a unitary basis change $Q(\theta)$ such that

$$R(\theta) = Q(\theta) \begin{pmatrix} \lambda_1(\theta) & \alpha(\theta) \\ 0 & \lambda_2(\theta) \end{pmatrix} Q^{\mathrm{H}}(\theta). \tag{7}$$

Obviously, here $\lambda_1(\theta)$ and $\lambda_2(\theta)$ are the eigenvalues of $R(\theta)$.

After introducing this notation, we can state the following theorem, which assures stability.

Theorem 5. *Let us assume the following hypotheses*:

(i) $|\alpha(\theta)| \leqslant K$ *for every* $\theta \in \mathscr{C}$ *for a certain constant* $K > 0$.
(ii) *There exists a value* $\bar{\theta} \in \mathbb{R}$ *such that* $R(\bar{\theta})$ *does not have double eigenvalues*.
Then, for every compact interval $\bar{I} \subset \mathscr{C}$ *there exists a constant* $C(\bar{I})$ *such that*

$$\|R^n(\theta)\|_2 \leqslant C(\bar{I}), \quad n \in \mathbb{N}, \quad \theta \in \bar{I} \subset \mathscr{C}.$$

Besides, a constant C *is valid for every* $\theta \in \mathscr{C}$ *if* $R(\beta_{\text{stab}})$ *is simple*.

Proof. Let us first assume that $\theta \in \mathscr{C}$ is such that $\lambda_1(\theta) = \lambda_2(\theta) = \lambda(\theta)$. Notice that the values of θ which lead to double eigenvalues of $R(\theta)$ are either all the real axis or just a finite number. This is due to the form of $R(\theta)$, which implies that the double eigenvalues come from the annihilation of the discriminant of a second-degree equation. This discriminant depends on the elements of $R(\theta)$ and therefore it is a certain rational expression. (This one vanishes when the numerator, which is a polynomial expression on θ, vanishes.) From here, using (ii), just for a finite number of values of θ, the eigenvalues are double. Because of this, in case $R(\theta)$ is not simple, as this implies $|\lambda(\theta)| < 1$, there exists a constant $N < 1$ such that $|\lambda(\theta)| < N < 1$ for every θ under this situation. Besides,

$$R(\theta)^n = Q(\theta) \begin{pmatrix} \lambda^n(\theta) & n\lambda(\theta)^{n-1}\alpha(\theta) \\ 0 & \lambda^n(\theta) \end{pmatrix} Q^{\mathrm{H}}(\theta) := Q(\theta) S_n(\theta) Q^{\mathrm{H}}(\theta).$$

Then, $\|R(\theta)^n\|_2 = \|S_n(\theta)\|_2 \leqslant \sqrt{\|S_n(\theta)\|_1 \|S_n(\theta)\|_\infty} = |\lambda(\theta)|^{n-1}[|\lambda(\theta)| + n|\alpha(\theta)|] \leqslant N^{n-1}[N + nK]$, which tends to zero when $n \to \infty$ and therefore is bounded. In case $R(\theta)$ is simple, $R(\theta) = \lambda(\theta) I$, so $\|R^n(\theta)\|_2 = |\lambda(\theta)|^n \leqslant 1$.

On the other hand, let us assume now that $\lambda_1(\theta) \neq \lambda_2(\theta)$. Let us use the notation

$$s_n(\theta) = \sum_{j=0}^{n-1} \lambda_1^j(\theta)\lambda_2^{n-1-j}(\theta) = \frac{\lambda_1^n(\theta) - \lambda_2^n(\theta)}{\lambda_1(\theta) - \lambda_2(\theta)}. \tag{8}$$

Then,

$$R(\theta)^n = Q(\theta) \begin{pmatrix} \lambda_1^n(\theta) & s_n(\theta)\alpha(\theta) \\ 0 & \lambda_2^n(\theta) \end{pmatrix} Q^{\mathrm{H}}(\theta)$$

and using that for $\theta \in \mathscr{C}$, $|\lambda_1(\theta)|$, $|\lambda_2(\theta)| \leqslant 1$, it happens that $\|R(\theta)^n\|_2 \leqslant 1 + |s_n(\theta)\alpha(\theta)|$. Therefore, to get the result, we just need to bound $|s_n(\theta)\alpha(\theta)|$ uniformly on $\theta \in \bar{I} \subset \mathscr{C}$ and $n \in \mathbb{N}$. As there is only a finite number of values of $\theta \in \mathbb{R}^+ \cup \{0\}$ $(\theta_1, ..., \theta_J)$ for which $\lambda_1(\theta) = \lambda_2(\theta)$, there exists a small value ε such that $|\lambda_1(\theta) - \lambda_2(\theta)| > \varepsilon$ except small subintervals of θ_j $(1 \leqslant j \leqslant J)$, which we will denote by I_j $(1 \leqslant j \leqslant J)$. Besides, these subintervals can be chosen such that, for $\theta \in \bigcup_{j=1}^{J} \{I_j \backslash \theta_j\}$, either $|\lambda_i(\theta)| < M < 1 (i = 1, 2)$ (which happens near double eigenvalues of modulus less than one) or the eigenvectors $p_i(\theta)$ $(i = 1, 2)$ of unit modulus corresponding to their diagonallization are far enough from each other, so that $|\langle p_1(\theta), p_2(\theta)\rangle| \leqslant L < 1$ (which happens near double eigenvalues of unit modulus because of continuity and the definition of $\mathscr{C}$).

In case $|\lambda_1(\theta) - \lambda_2(\theta)| > \varepsilon$, as $|\lambda_1^n(\theta) - \lambda_2^n(\theta)| \leqslant 2$, using (i) and (8),

$$|s_n(\theta)\alpha(\theta)| \leqslant \frac{2K}{\varepsilon}.$$

In case $|\lambda_i(\theta)| < M < 1$ $(i = 1, 2)$, using (i) and (8) again,

$$|s_n(\theta)\alpha(\theta)| \leqslant K n M^{n-1}.$$

As this tends to zero when $n \to \infty$, the searched bound is found.

In case $|\langle p_1(\theta), p_2(\theta)\rangle| \leqslant L < 1$, let us consider the QR factorization [5] of the change of basis $P(\theta) = [p_1(\theta), p_2(\theta)] = \bar{Q}(\theta)\bar{R}(\theta)$. Then,

$$R(\theta) = \bar{Q}(\theta)\bar{R}(\theta) \begin{pmatrix} \lambda_1(\theta) & 0 \\ 0 & \lambda_2(\theta) \end{pmatrix} \bar{R}(\theta)^{-1}\bar{Q}(\theta)^{-1},$$

where $\bar{R}(\theta) = (\bar{r}_{ij}(\theta))$ is an upper triangular matrix with positive diagonal elements and $\bar{Q}(\theta)$ is unitary. By making the calculations, it turns out that

$$R(\theta) = \bar{Q}(\theta) \begin{pmatrix} \lambda_1(\theta) & (\lambda_2(\theta) - \lambda_1(\theta)) \dfrac{\bar{r}_{12}(\theta)}{\bar{r}_{22}(\theta)} \\ 0 & \lambda_2(\theta) \end{pmatrix} \bar{Q}(\theta)^{-1}.$$

Comparing this with the Schur decomposition (7), it turns out that the central matrices in the right-hand sides are unitarily similar. Therefore,

$$|\alpha(\theta)| = |(\lambda_2(\theta) - \lambda_1(\theta))\bar{r}_{12}(\theta)/\bar{r}_{22}(\theta)|. \tag{9}$$

Now, by definition of QR factorization, taking into account that the eigenvectors $[p_1(\theta), p_2(\theta)]$ have been chosen of unit modulus, it happens that

$$|\bar{r}_{12}(\theta)| = |\langle p_2(\theta), p_1(\theta)\rangle|, \quad |\bar{r}_{22}(\theta)| = \sqrt{1 - |\langle p_2(\theta), p_1(\theta)\rangle|^2}.$$

Finally, as there exists a constant C_L such that $|x|/\sqrt{1 - |x|^2} \leqslant C_L$ if $|x| < L < 1$, and using again $|\lambda_1^n(\theta) - \lambda_2^n(\theta)| \leqslant 2$, (8) and (9),

$$|s_n(\theta)\alpha(\theta)| = |\lambda_1^n(\theta) - \lambda_2^n(\theta)| \frac{|\bar{r}_{12}(\theta)|}{|\bar{r}_{22}(\theta)|} \leqslant 2C_L.$$

Therefore, the result follows and it is obvious, from the proof, that $\bar{I}$ can be substituted by $\mathscr{C}$ when $R(\beta_{\text{stab}})$ is simple. □

When (1) is not stiff or it is moderately stiff, it is possible to consider RKN methods with a finite interval of stability, for example an explicit method [8]. In this case we can apply the following result.

Corollary 6. *When the method has a finite interval of stability, hypothesis* (ii) *of Theorem* 5 *is satisfied and* $\sigma(\mathscr{A}) \cap (-\infty, -1/\beta_{\text{stab}}^2] = \varnothing$, *stability follows whenever* $\sigma(kB) \subset \mathscr{C}$.

Proof. Because of the hypothesis on $\sigma(\mathscr{A})$, $(\mathscr{I} + \theta^2\mathscr{A})^{-1}$ is uniformly bounded on $\mathscr{C}$, and therefore, as this is finite, the same happens with the rational expressions $r_{ij}(\theta)$ (4). Then, $\|R(\theta)\|_2$ is uniformly bounded.

From here, considering the Schur decomposition (7), which leaves the norm invariant, hypothesis (i) of Theorem 5 is satisfied. As there always exists a compact interval $\bar{I}$ such that $\sigma(kB) \subset \bar{I} \subset \mathscr{C}$, the corollary follows. □

In many problems, (1) will be so stiff that infinite stability intervals will be required. In order to construct methods for which hypothesis (i) of the previous theorem is satisfied independently of the size of $\mathscr{C}$, we state the following.

Theorem 7. *Let us assume that the method is R-stable and* $\sigma(\mathscr{A}) \cap (-\infty, 0] = \emptyset$. *Then, the term* $\alpha(\theta)$ *in the Schur decomposition* (7) *is uniformly bounded for every* $\theta \in \mathscr{C}$ *if and only if the coefficients of the method satisfy*

$$\beta^{\mathrm{T}} \mathscr{A}^{-1} c = 1. \tag{10}$$

Proof. Notice that, as the method is *R*-stable, $\|R(\theta)\|_2$ is uniformly bounded on $\mathscr{C}$ if and only if $|\alpha(\theta)|$ also is (just take into account again that the Schur decomposition leaves the norm invariant). Notice also that, because of the assumption on $\mathscr{A}$, $r_{11}(\theta)$, $r_{21}(\theta)$ and $r_{22}(\theta)$ are rational expressions where the degree of the numerator is less than or equal to that of the denominator (see [12, Lemma 4.1]) and, besides, the denominator does not vanish on the whole interval. This implies that these expressions are uniformly bounded on the whole interval. As for $r_{12}(\theta)$, it behaves as $\theta(1 - \beta^{\mathrm{T}} \mathscr{A}^{-1} c)$ when $\theta \to \infty$. Therefore, the former does happen only when condition (10) is satisfied. □

Notice that, under condition (10), $R(\infty)$ is diagonal. As a result, in Theorem 5, a constant C can be chosen which is valid for every $\theta \in \mathscr{C}$. Then, from the previous theorems, the following result is true.

Corollary 8. *Under the assumptions of Theorem* 7, *condition* (10) *and hypothesis* (ii) *of Theorem* 5, *the following stability bound is true, where C is independent of the size of* $\sigma(kB)$,

$$\|R(kB)^n\|_2 \leqslant C, \quad n \in \mathbb{N}.$$

Besides, it is not possible to get this result if (10) *is not satisfied.*

Remark 9. In case the RKN method comes from a RK one (with stability function r)

$$R(kB) = r\begin{pmatrix} 0 & kB \\ -kB & 0 \end{pmatrix} := r(k\tilde{B}).$$

Since we suppose that B is symmetric and positive definite, $\tilde{B}$ is normal, and therefore, stability follows according to the theory for RK methods in the appropriate region of stability. Notice that, accordingly, stability also follows from the previous results taking into account that, for those methods, $\beta^{\mathrm{T}} = \bar{b}^{\mathrm{T}} \bar{\mathscr{A}}$, $\mathscr{A} = \bar{\mathscr{A}}^2$, where $\bar{b}$, $\bar{\mathscr{A}}$ are the coefficients of the RK method. Then,

$$\beta^{\mathrm{T}} \mathscr{A}^{-1} c = \bar{b}^{\mathrm{T}} \bar{\mathscr{A}} (\bar{\mathscr{A}})^{-2} c = \bar{b}^{\mathrm{T}} \bar{\mathscr{A}}^{-1} c = \bar{b}^{\mathrm{T}} e = 1.$$

Here we have used that RK methods are always constructed under the condition $\bar{\mathscr{A}} e = c$. The last equality is just due to consistency.

3. Construction of a stable RKN method for stiff problems

Since we do not know any RKN method in the literature satisfying (10), in this section we deal with the construction of a RKN method with this condition. We consider the case of SDIRKN methods which have all the diagonal elements of $\mathscr{A}$ equal to a number α. We concentrate on P-stable RKN methods because, as we have said in the preliminaries, the initial errors do not diminish with these methods when the computation progresses and therefore, stability is essential.

The construction of P- and R-stable SDIRKN methods is the subject of the Refs. [4,10–12]. The cases studied in [11] are third-order methods with $s = 2$ stages and fourth-order methods with $s = 3$ stages, assuming that the stage order of the method is 2, i.e. the condition $\mathscr{A}e = \frac{1}{2}c^2$ is satisfied. However, we have not obtained P-stable methods satisfying (10) with these assumptions (there exists an R-stable method for $s = 2$ stages).

Therefore, we consider the case of P-stable SDIRKN methods with $s = 4$ stages, the case studied in [4]. In this paper, the authors derive P-stable SDIRKN methods without imposing that the stage order is 2, but considering several conditions such as the method is symmetric, symplectic or dispersive of a certain order. We concentrate on the case of symplectic and symmetric fourth-order methods. With the previous notation, the coefficients satisfy the conditions of symplecticness

$$\beta_i = b_i(1 - c_i), \quad i = 1, 2, 3, 4,$$
$$a_{ij} = b_j(c_i - c_j), \quad i, j = 1, 2, 3, 4, \quad i > j$$

and the conditions of symmetry

$$b_1 = b_4, \quad b_2 = b_3, \quad c_1 + c_4 = 1, \quad c_2 + c_3 = 1.$$

Under these assumptions, the fourth-order conditions are

$$b^{\mathrm{T}}e = 1, \quad b^{\mathrm{T}}c^2 = \tfrac{1}{3}, \quad b^{\mathrm{T}}\mathscr{A}e = \tfrac{1}{6}.$$

As in [5], writing the nodes in the form

$$c_1 = \tfrac{1}{2} - \gamma, \quad c_2 = \tfrac{1}{2} - \lambda, \quad c_3 = \tfrac{1}{2} + \lambda, \quad c_4 = \tfrac{1}{2} + \gamma,$$

we obtain

$$b_1 = \frac{12\lambda^2 - 1}{24(\lambda^2 - \gamma^2)}, \quad b_2 = \frac{1 - 12\gamma^2}{24(\lambda^2 - \gamma^2)}, \quad \alpha = \frac{1}{6} - 4\gamma b_1 b_2 - 2\lambda b_2^2 - 2\gamma b_1^2.$$

Now, by imposing dispersion of order six, the free parameters γ and λ satisfy

$$\frac{1}{360} - \frac{\alpha}{6} + \alpha^2 = \frac{Q(\gamma, \lambda)}{6912(\gamma + \lambda)^3(\gamma - \lambda)} \tag{11}$$

with

$$Q(\gamma, \lambda) = (12\gamma^2 - 1)(12\lambda^2 - 1)(24\lambda\gamma^2 + 24\gamma\lambda^2 - 24\gamma\lambda - 12\lambda^2 + 2\gamma + 2\lambda - 1).$$

In [4], the authors select the values $\gamma = -0.45515766756706$ and $\lambda = 0.8$ in order to obtain a P-stable method such that the truncation error is small. For this method, we have $1 - \beta^{\mathrm{T}}\mathscr{A}^{-1}c \simeq 0.084728$, so (10) is not satisfied.

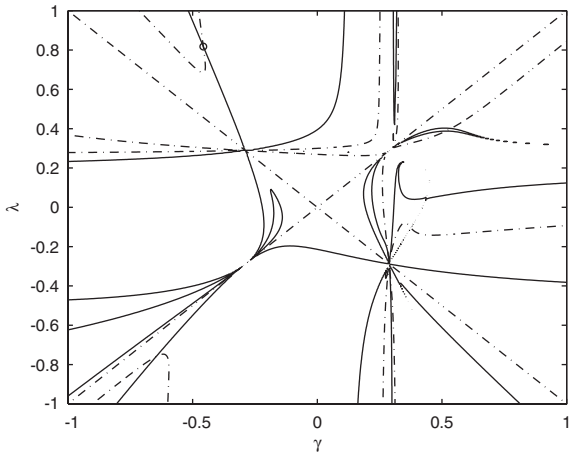

Fig. 1. Values of γ and λ for which the Eqs. (10) and (11), displayed respectively, as a continuous and a dash-dotted curve, are satisfied. The point marked with an ∘ is the point selected by us.

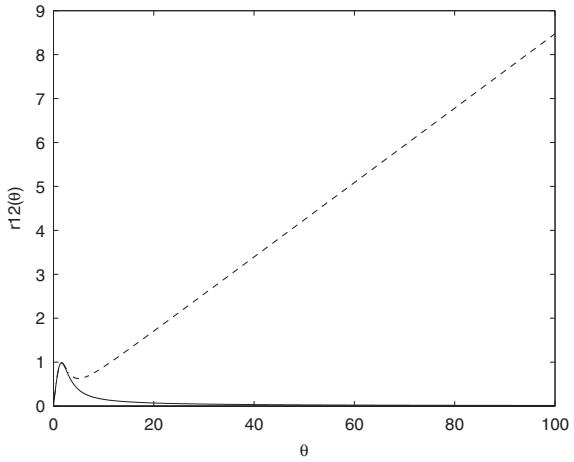

Fig. 2. Graphics of $r_{21}(\theta)$ for the stable method constructed in this paper (continuous line), and the method constructed by Franco et al. [4] (dashed line).

On the other hand, by imposing (10) we obtain another relation between the free parameters γ and λ and we have a system of two nonlinear equations in these unknowns. We show in Fig. 1 the curves of the values of γ and λ in the square $[-1, 1] \times [-1, 1]$ satisfying (10) and (11). There exist several points in the intersection of the curves (however, not all of them correspond to P-stable methods). In particular, we have checked that the point given by the values $\gamma = -0.4569794733108003$ and $\lambda = 0.8176615502464265$ is in this intersection and, moreover, the corresponding method is P-stable. Since these values are very close to the values selected in [4], the truncation error constants are also small. However, the behaviour of the element $r_{12}(\theta)$ of the matrix $R(\theta)$ is completely different for large values of θ because, as $\theta \to \infty$, $r_{12}(\theta) \to 0$ for our choice and $r_{12}(\theta) \to \infty$ for the choice in [4] (see Fig. 2).

Table 1
Errors for problem 1, with $k = 0.05$

Local error	Solution	Derivative	Energy norm
FGR(4,6)	$4.03917 - 05$	$2.12370 - 02$	2.99519
SRKN	$1.62527 - 08$	$2.25880 - 02$	$1.51999 - 02$
FGR(4,8)	$8.76246 - 05$	$3.15487 - 02$	6.49763
Global error for $T = 10$			
FGR(4,6)	$6.04291 - 05$	$5.23759 - 02$	5.90964
SRKN	$3.81809 - 08$	$5.71494 - 03$	$2.78350 - 03$
FGR(4,8)	$7.67099 - 04$	$3.75990 - 01$	75.0181

4. Numerical experiments

In this section, we show the numerical results obtained with the following methods with classical order 4.

- FGR(4,6): The SDIRKN method in [4] which is symplectic, symmetric, with dispersion of order six and with $1 - \beta^{\mathrm{T}}\mathscr{A}^{-1}c \simeq 0.084728$.
- SRKN: The SDIRKN method obtained in Section 3, which is stable, symplectic, symmetric, with dispersion of order six and with $1 - \beta^{\mathrm{T}}\mathscr{A}^{-1}c = 0$.
- FGR(4,8): The SDIRKN method in [4] which is symplectic, nonsymmetric, with dispersion of order eight and with $1 - \beta^{\mathrm{T}}\mathscr{A}^{-1}c \simeq 0.183651$.

All these methods are P-stable according to our definition and to the definitions of P-stability of Franco et al. and Sharp et al. [4,11] because the eigenvalues of $R(\theta)$ for $\theta > 0$ are always complex, distinct and of modulus 1. Therefore the periodicity and stability intervals for all of them is $\mathscr{C} = \mathscr{C}^* = [0, +\infty)$.

We have solved three problems, in order to include different situations that can appear in practice. When the solutions for all these problems are chosen such that they have no component in the stiff part, the results obtained with FGR(4,6) and SRKN are nearly the same. The difference comes when considering solutions of the form $y_1 + y_2$, where y_1 evolves slowly but y_2 is a small perturbation in the stiff component of the problem. We show the results obtained in some tables, which correspond to the local and global relative errors committed in the solution, its derivative and the relative error measured in the energy norm.

Problem 1 (*see Franco et al. [4]*).

$$y''(t) = -\frac{1}{2}\begin{bmatrix} \omega^2 + 1 & \omega^2 - 1 \\ \omega^2 - 1 & \omega^2 + 1 \end{bmatrix} y(t), \quad y(0) = y_0, \quad y'(0) = v_0.$$

In our experiments, we have chosen as solution $y(t) = y_1(t) + y_2(t)$, where $y_1(t) = [\cos(t) + \sin(t), -\cos(t) - \sin(t)]^{\mathrm{T}}$ and $y_2 = [10^{-7}(\cos(\omega t) + \sin(\omega t)), 10^{-7}(\cos(\omega t) + \sin(\omega t))]^{\mathrm{T}}$ with $\omega = 10^5$.

As we can see in Table 1, both in the local error as in the global error, the errors committed in the solution are much better with the SRKN method than with the other methods, and the same happens with the error measured in the energy norm, for which SRKN is clearly stable. The fact that there is less

Table 2
Errors for problem 2 with $k = 0.01$ and $h = 10^{-3}$

Local error	Solution	Derivative	Energy norm
FGR(4,6)	$4.87046 - 03$	$2.36457 - 01$	398.211
SRKN	$6.33085 - 04$	$7.09726 - 02$	$1.51180 - 01$
FGR(4,8)	$6.41754 - 03$	1.08744	862.960
Global error for $T = 1$			
FGR(4,6)	$1.07162 - 02$	$9.41193 - 01$	645.477
SRKN	$5.17065 - 03$	$2.19519 - 01$	$5.36328 - 01$
FGR(4,8)	$6.04834 - 02$	6.35906	6883.67

difference in the results for the derivative comes from the main influence of r_{12} on the solution, and not on the derivative; in fact, in the local error, r_{12} does not have influence at all. Notice also that FGR(4,8) gives worse results than FGR(4,6). This is because $1 - \beta^{\mathrm{T}} \mathscr{A}^{-1} c$ is bigger for the former than for the latter.

Problem 2. The second problem we have considered is the PDE

$$
\begin{aligned}
&u_{tt}(x,t) = -u_{xxxx}(x,t) + f(x,t), \quad x \in \Omega = [-1,1], \quad 0 \leqslant t \leqslant T < \infty,\\
&u(x,t) = g_1(t), \quad u_{xx}(x,t) = g_2(t), \quad x \in \partial\Omega = \{-1,1\}, \quad 0 \leqslant t \leqslant T < \infty,\\
&u(x,0) = u_0(x), \quad u_t(x,0) = v_0(x), \quad x \in \Omega
\end{aligned}
$$

with $f(x,t) = (24 - 4\pi^2(1-x^2)^2)\cos(2\pi t)$. In order to compare the methods, we have chosen as solution $u(x,t) = u_1(x,t) + u_2(x,t)$, with $u_1(x,t) = (1-x^2)^2 \cos(2\pi t)$ and $u_2(x,t) = 3 \cdot 10^{-6} \sin(10^6 t) \cos(10^3 x)$.

In this case, we have first made the space discretization of the fourth derivative, for which we have considered two successive approximations of the second-order derivative by the standard second-order difference method. As a result, we have obtained a problem like (1), which is arbitrarily stiff when the spatial discretization is refined.

From the results obtained in Table 2 for the errors corresponding to the space grid with diameter $h = 10^{-3}$, we draw the same conclusions as in Problem 1. We just want to point out here that the advantage of SRKN over FGR(4,6) and FGR(4,8) is stressed when h diminishes. This is logical since the problem becomes stiffer.

Problem 3. And finally, although the results in this paper just correspond to the linear case, we have wanted to corroborate numerically what happens in a nonlinear problem. The problem we have solved is (see [4])

$$
\begin{aligned}
&y''(t) = \begin{bmatrix} -1 & 0 \\ 0 & -\omega^2 \end{bmatrix} y(t) + \mu \begin{bmatrix} 1 \\ 1 \end{bmatrix} (\|y(t)\|_2^2 + \cos^2(t) - 1 + \varepsilon^2(\cos^2(\omega t) - 1)),\\
&y(0) = y_0, \quad y'(0) = v_0.
\end{aligned}
$$

The experiments have been made with $\omega = 10^5$, $\mu = 0.01$ and $\varepsilon = 10^{-7}$. We have chosen as solution $y(t) = y_1(t) + y_2(t)$, with $y_1(t) = [\sin(t), 0]^{\mathrm{T}}$ and $y_2(t) = [0, 10^{-7} \sin(\omega t)]^{\mathrm{T}}$.

Table 3
Errors for problem 3 with $k = 0.05$

Local error	Solution	Derivative	Energy norm
FGR(4,6)	$8.49606 - 04$	$1.02780 - 02$	4.24606
SRKN	$1.97868 - 06$	$1.15611 - 02$	$1.51975 - 02$
FGR(4,8)	$1.83925 - 03$	$2.00720 - 02$	9.19198
Global error for $T = 10$			
FGR(4,6)	$1.53412 - 04$	$1.42761 - 02$	8.34551
SRKN	$9.52624 - 08$	$7.14441 - 04$	$3.01700 - 03$
FGR(4,8)	$1.94859 - 03$	$1.28049 - 01$	106.002

The conclusions are the same as in Problems 1 and 2. In Table 3, we see that the local and global errors obtained with the SRKN method are much smaller than with the other methods, mainly in the solution and in the energy norm.

References

[1] M.M. Chawla, S.R. Sharma, Intervals of periodicity and absolute stability of explicit Nyström methods for $y'' = f(x, y)$, BIT 21 (1981) 455–464.
[2] M. Crouzeix, Numerical range and Hilbertian functional calculus, Institute of Mathematical Research of Rennes, Université de Rennes, preprint.
[3] J.L.M. van Dorsselaer, J.F.B.M. Kraaijevanger, M.N. Spijker, Linear stability analysis in the numerical solution of initial value problems, in: A. Iserles (Ed.), Acta Numerica, Cambridge University Press, Cambridge, 1993, , pp. 199–237.
[4] J.M. Franco, I. Gómez, L. Rández, Four-stage symplectic and P-stable SDIRKN methods with dispersion of high order, Numer. Algorithms 26 (2001) 347–363.
[5] G.H. Golub, C.F. Van Loan, Matrix Computations, second ed., The Johns Hopkins University Press, Baltimore and London, 1990.
[6] K.J. in 'T Hout, M.N. Spijker, Analysis of error growth via stability regions in numerical initial value problems, BIT 43 (2003) 363–385.
[7] D. Levy, E. Tadmor, From semidiscrete to fully discrete: stability of Runge–Kutta schemes by the energy method, SIAM J. Numer. Anal. 40 (1998) 40–73.
[8] I. Alonso-Mallo, B. Cano, M.J. Moreta, Order reduction and how to avoid it when explicit Runge–Kutta–Nyström methods are used to solve linear partial differential equations, J. Comput. Appl. Math. 176 (2005) 293–318.
[9] M.J. Moreta, Reducción de orden con métodos Runge–Kutta–Nyström, Degree Thesis, Department of Applied Mathematics, University of Valladolid, 2002.
[10] G. Papageorgiou, I.T. Famelis, C. Tsitouras, A P-stable singly diagonally implicit Runge–Kutta–Nyström method, Numer. Algorithms 17 (1998) 345–353.
[11] P.W. Sharp, J.M. Fine, K. Burrage, Two-stage and three-stage diagonally implicit Runge–Kutta–Nyström methods of order three and four, IMA J. Numer. Anal. 10 (1990) 489–504.
[12] P.J. Van Der Houwen, B.P. Sommeijer, Diagonally implicit Runge–Kutta–Nyström methods for oscillatory problems, SIAM J. Numer. Anal. 26 (1989) 414–429.
[13] P.J. Van Der Houwen, B.P. Sommeijer, Nguyen Huu Cong, Stability of collocation-based Runge–Kutta–Nyström methods, BIT 31 (1991) 469–481.