Monografías de la Real Academia de Ciencias de Zaragoza **33**, 141–154, (2010).

Extending convergence results of Runge–Kutta methods for stiff semi linear initial value problems

M. Calvo[1], S. Gonzalez-Pinto[2] and J.I. Montijano[1]

[1] Departamento Matemática Aplicada
Universidad de Zaragoza, 50009-Zaragoza, Spain

[2] Departamento Análisis Matemático
Universidad de La Laguna, 38271-La Laguna, Spain

Abstract

In this paper some convergence results for Runge–Kutta methods applied to semi–linear variable coefficients differential systems with the stiffness contained in the linear part and under some assumptions on the relative variation of the jacobian matrix are derived. Previous results on this subject given by the authors in BIT 40, 4 (2000), pp. 611–634, are generalised. In particular, it is shown that some non B–stable methods such as those of the Lobatto IIIA family and some DIRK methods that have been used in practical problems are convergent of order greater or equal than the stage order for this kind of problems. Some numerical examples are presented to illustrate the theory.

1 Introduction

The concepts of B–stability and B–convergence introduced to study the behaviour of Runge–Kutta (RK) methods for the numerical integration of stiff IVPs

$$y'(t) = f(t, y(t)),\ y(t_0) = y_0 \in \mathbb{R}^m,\ t \in I := [t_0, t_0 + T], \tag{1}$$

where $f(t, y)$ satisfies a one-sided Lipschitz condition with respect to y, have provided in the last two decades a well established B–theory that allows us to identify those RK methods that are suitable for this class of stiff systems [8], [10].

However, as remarked by Alexander [1] there exist some globally well behaved stiff problems (i.e. IVPs whose solution is asymptotically stable) that possess strongly positive one sided Lipschitz constants and therefore do not fit into the B–theory. In particular, variable coefficient linear or semi linear problems with the stiffness contained in a non normal linear part are potential candidates to possess large one–sided Lipschitz constants (see [1, Sec. 1.1]).

Auzinger, Frank and Kirlinger [4] extended the B–theory to a class of stiff semi linear problems

$$y'(t) = f(t, y(t)) \equiv \widehat{J}(t)\, y(t) + g(t, y(t)), \quad t \in I, \quad y(t_0) = y_0 \in \mathbb{R}^m, \tag{2}$$

where $g(t, y)$ is Lipschitz continuous with respect to y and the logarithmic norm $\mu[S(t)\widehat{J}(t)S(t)^{-1}]$ is moderately sized for some smooth non singular matrix $S(t)$. These authors proved that any diagonally-stable and algebraically-stable RK method (A, b) with order p and stage order $q \leq p$, is B-convergent with order $\geq q$ for the class of semi linear problems (2). In the special case that $\widehat{J}(t) = J$ is a constant matrix, by requiring on the methods assumptions of linear-stability type, Burrage, Hundsdorfer and Verwer [5] obtained optimal convergence results with orders q (or $q + 1$), where q is the stage order of the method. Calvo, Montijano and Gonzalez-Pinto in [6], extended the convergence results of [5] to the class of semi-linear problems (2), with variation of $\widehat{J}(t)$ relatively bounded.

The purpose of this paper is to extend our results of [6] to the class of problems considered by Auzinger *et al.* [4], in the case that $J(t) = S(t)\widehat{J}(t)S(t)^{-1}$ varies on t in a relatively bounded form, when some smooth matrix $S(t)$ is considered. It will be seen that any A–stable RK method (A, b) with positive real part for each eigenvalue of A, is stable and convergent. Further, these properties also hold for some stiffly accurate A–stable methods having a first stage explicit, such as those of the Lobatto IIIA family and some DIRK methods, that have been used e.g. by Kennedy and Carpenter for convection–diffusion–reaction systems [11]. The paper is completed with some numerical experiments and comments about the convergence of some SDIRK methods. It must be also noticed that our results represent an improvement with regard to [4], in the case in which a bounded relative variation on $J(t)$ is assumed (see the assumption (H2) below). To keep the presentation within a reasonable length, we have omitted the proofs of the results, since they are based on an extension of those in [6]. An extended version of this paper that includes the proofs of the main results and more numerical experiments is given in [7].

2 Notations and basic assumptions

We assume that (2) possesses a unique smooth solution $y(t) = y(t; t_0, y_0)$, $t \in I$, in the sense that for a positive integer p as large as required $\|y^{(j)}(t)\| \leq M_j = \mathcal{O}(1)$, $j = 0, \ldots, p+1$, $t \in I$, and $f(t, y)$ has continuous partial derivatives up to order p in some cylinder $\mathcal{B}_\delta = \{(t,y);\ \|y - y(t)\| \le \delta, t \in I\}$ around the exact solution. The norm used is induced by some inner product with $\mu[\cdot]$ standing for the logarithmic norm associated to the induced norm and $\mathcal{O}(1)$ will mean any constant (or mapping) moderately sized independently of the stiffness.

We will consider semi-linear problems (2) under the following assumptions:

(H1) There exists a matrix $S(t) \in \mathbb{R}^{m,m}$ such that $J(t) := S(t)\widehat{J}(t)S(t)^{-1}$ satisfies $\mu[J(t)] \le 0$ and $S(t), S'(t), S^{-1}(t)$ are $\mathcal{O}(1)$ for $t \in I$.

(H2) There exist a constant $h^* > 0$ and a mapping $E_1(t, \Delta t)$, such that either

$$(i) \qquad J(t+\Delta t) - J(t) - \Delta t\ J(t)\ E_1(t,\Delta t) = \mathcal{O}(\Delta t),$$

or else

$$(ii) \qquad J(t+\Delta t) - J(t) - \Delta t\ E_1(t,\Delta t)\ J(t) = \mathcal{O}(\Delta t),$$

hold for all $t, t+\Delta t \in I$ with $|\Delta t| \le h^*$.

(H3) There exists a constant $\lambda_0 = \mathcal{O}(1)$ such that $\|g(t,y) - g(t,\tilde{y})\| \le \lambda_0 \|y - \tilde{y}\|$, for all $(t,y),\ (t,\tilde{y}) \in \mathcal{B}$.

The above assumptions imply that the stiffness of f is included in the linear term. Moreover the condition $\mu[J(t)] \le 0$ may be replaced by $\mu[J(t)] \le \nu = \mathcal{O}(1)$.

Observe that if $\widehat{J}(t)$ can be made diagonal by a smooth matrix $S(t)$ satisfying $\parallel S(t) \parallel \parallel S(t)^{-1} \parallel = \mathcal{O}(1)$, then the assumptions (H1)-(H2) are usually satisfied. This happens, for example, if the eigenvalues $\lambda_j(t)$ of $\widehat{J}(t)$ have a non–positive real part. The assumption (H2) is closely related to the relative Lipschitz condition used by Alexander in [1] as well as the assumption (a.1) introduced by van Dorsselaer and Spijker in [9] to study the convergence of Newton-type iterations in implicit RK methods. It has also been used by Calvo *et al.* [6] for the convergence analysis of Runge-Kutta methods by considering that the untransformed matrix $\widehat{J}(t)$ of (2) satisfies

$$\widehat{J}(t+\Delta t) - \widehat{J}(t) - \Delta t\ \widehat{J}(t)\ \widehat{E}_1(t,\Delta t) = \mathcal{O}(\Delta t),\ t \in I,\ \widehat{E}_1 = \mathcal{O}(1). \tag{3}$$

Assumptions similar to (H2) with $J(t)$ replaced by $\widehat{J}(t)$ have been also used by Strehmel and Weiner [14] and Schmitt [13] in connection with the analysis of stability and convergence of implicit Runge-Kutta methods and linearly implicit methods on time-dependant partial differential equations.

Remark 2.1 *For general non–linear problems (1), if $f(t,y)$ is analytic in a cylinder $\mathcal{B}_\delta$ around the exact solution $y = \varphi(t)$ of (1), then $f(t,y)$ can be written in the semilinear form (2) with $\hat{J}(t) = f_y(t,\varphi(t))$ and*

$$g(t,y) := f(t,\varphi(t)) - \hat{J}(t)\varphi(t) + \sum_{k\ge 2} \frac{1}{k!} f^{(k)}_{(t,\varphi(t))}(y - \varphi(t), \ldots^{(k}, y - \varphi(t)),$$

where $f^{(k)}_{(t,\varphi(t))}(u_1,\ldots,u_k)$ *denotes the* k*-Fréchet derivative of* f *with respect to* y *at* $(t,\varphi(t))$*. Now, if*

$$\|f^{(k)}_{(t,\varphi(t))}(u_1,\ldots,u_k)\| \le \lambda_k \|u_1\|\cdots\|u_k\|,\ t\in[0,T],\ k=2,3,\ldots,$$

with $\|\lambda_k\| = \mathcal{O}(1)$*, then the assumption (H3) is fulfilled.*

For the numerical solution of (2) we consider an s-stage RK method specified by the Butcher matrices (A,b), $A=(a_{ij})\in\mathbb{R}^{s,s}$, $b=(b_i)\in\mathbb{R}^s$ and the knot vector $c=(c_i)=Ae$, $e=(1,\ldots,1)^T\in\mathbb{R}^s$. The step from $(t_0,y_0)\longrightarrow(t_1=t_0+h, y_1=Y_{s+1})$ is defined by the equations

$$Y_i = y_0 + h\sum_{j=1}^{s} a_{ij}\left[(S_j^{-1}J_jS_j)Y_j + g(\tau_j,Y_j)\right], \qquad (i=1,\ldots,s+1) \tag{4}$$

with $c_{s+1}=1, a_{s+1,j}=b_j$ and $\tau_i=t_0+c_ih,\ S_i=S(\tau_i),\ J_i=J(\tau_i)$.

The stage order q of the RK method is defined as $\min\{p_j; j=1,\ldots,s+1\}$ where the positive integers p_j are given by

$$\varepsilon_j \equiv y(t_0+c_jh) - y(t_0) - h\sum_{k=1}^{s} a_{jk}\, y'(t_0+c_kh) = \mathcal{O}(h^{p_j+1}), \quad j=1,\ldots,s+1.$$

with $c_{s+1}=1, a_{s+1,j}=b_j$. Note that by the smoothness of $y(t)$, the residual errors ε_i satisfy $\|\varepsilon_i\|\le\mathcal{O}(h^{q+1})$.

To adapt the concepts of BSI-stability and BS-stability [8, Chaps. 5,7] to our class of problems, we consider the perturbed version of (4)

$$\widetilde{Y}_i = y_0 + h\sum_{j=1}^{s} a_{ij}\left[(S_j^{-1}J_jS_j)\widetilde{Y}_j + g(\tau_j,\widetilde{Y}_j)\right] + \eta_i, \quad (i=1,\ldots,s+1) \tag{5}$$

where $\eta_i\in\mathbb{R}^m$ are arbitrary perturbations.

In this situation, a RK method (A,b) is said to be BSI-stable if there exist two positive constants (independent of the stiffness) h^* and C_0 such that,

$$\max_{j=1,\ldots,s} \| Y_j - \widetilde{Y}_j \| \le C_0 \sum_{i=1}^{s}\|\eta_i\|, \quad \text{whenever } h\in(0,h^*].$$

Further, the RK method (A,b) will be called BS–stable if

$$\|\widetilde{y}_1 - y_1\| = \|\widetilde{Y}_{s+1} - Y_{s+1}\| \le C_1 \sum_{i=1}^{s+1}\|\eta_i\|, \quad C_1=\mathcal{O}(1).$$

We often require for the Runge-Kutta method (A,b) that each eigenvalue of its coefficient matrix A has a positive real part. This is equivalent to (see the assumption (M4) in [6])

$$(I-zA) \text{ is non singular for } \mathrm{Re}z\le 0 \text{ and } \sup_{\mathrm{Re}z\le 0}\|z(I-zA)^{-1}\|_2 < +\infty. \tag{6}$$

3 Main Results

In this section we state the main stability and convergence result whose proof has been omitted for the sake of brevity. These proofs are similar to the ones in previous paper of the authors [6], and can be seen in [7].

The first stability and convergence result is concerned with methods whose matrix A satisfies (6).

Theorem 3.1 *A Runge-Kutta method* (A, b) *satisfying (6),*

i) is BSI–stable and BS–stable for the class of stiff semi linear problems (2) under the assumptions (H1), (H2), (H3).

ii) If, moreover the method is A–stable, then it is convergent with order $\geq q$ *(the stage order).*

An important question is whether condition (6) is essential (apart from the A-stability) for convergence. We have found that there exist methods where A is a singular matrix with a special structure that are convergent for fixed step sizes or even on non uniform meshes, such that the number of given steps N multiplied by the maximum step-size is under some prefixed constant K. In particular, we have obtained positive convergence results for stiffly accurate methods, i.e., methods with $b^T = (a_{s1}, \ldots, a_{ss})$, with the first stage explicit and whose matrix A has the form

$$A = \begin{pmatrix} 0 & 0^T \\ a & \overline{A} \end{pmatrix} \in \mathbb{R}^{s,s}, \tag{7}$$

where the sub-matrix $\overline{A} \in \mathbb{R}^{(s-1),(s-1)}$ satisfies (6).

A convergence result for these stiffly accurate methods is given in the following:

Theorem 3.2 *Let* (A, b) *be a stiffly accurate RK method with a matrix* A *of type (7) and* $\overline{A}$ *satisfying (6), then*

i) If the method is A-stable, then it is convergent of order $\geq q$ *(stage order) on uniform meshes for the class of stiff semi linear problems (2) under the assumptions (H1), (H2), (H3).*

ii) In addition, it is also convergent with order $\geq q$ *on special non uniform meshes* $\{t_j\}_{j=0}^N$ *provided that* $N \max(t_j - t_{j-1}) \leq K$ *with* K *independent of the grid.*

Corollary 3.1 *The s-stage Lobatto IIIA method is convergent of order* $q \geq s$*, under the H-assumptions.*

Next result deals with AS– and ASI–stable methods. Recall (see e.g. [6]) that an s–stage RK method is said to be AS–stable (resp. ASI–stable) if $(I - zA)$ is a non singular matrix on Re $z \leq 0$ and

$$\sup_{\mathrm{Re} z \leq 0} \|zb^T(I - zA)^{-1}\|_2 < +\infty \quad \left(\text{resp. } \sup_{\mathrm{Re} z \leq 0} \|(I - zA)^{-1}\|_2 < +\infty \right). \tag{8}$$

For this case we have the weaker convergence (and internal stability) results,

Theorem 3.3 *An s-stage RK method (A, b) satisfying (8),*

i) is BSI–stable and BS–stable for the class of stiff semi linear problems (2) under the assumptions (H1), (H2-i), (H3), and (3).

ii) If, moreover the method is A–stable, then it is convergent with order not lesser than the stage order.

It must be remarked that if we replace in Theorem 3.2 the stiff-accuracy by the weaker assumption

$$b^T = d^T A, \quad \text{for some } d \neq e_j, \ j = 1, \ldots, s, \tag{9}$$

e_j denoting canonical vectors of $\mathbb{R}^s$, then the convergence statement in Theorem 3.2 does not necessarily hold as it will be clearly shown in the numerical experiments of the next section. However, under condition (9), the Theorem 3.3 holds true provided that the underlying method is also A-stable.

4 Numerical experiments

The aim of this section is to check the convergence behavior of several SDIRK methods on some variable coefficients stiff linear systems satisfying the H–assumptions. Two of these methods have been taken from the literature of stiff integrators and are not B–stable, but they satisfy the assumptions of either of the above theorems. The third one is a purposely chosen three–stage method with an A matrix of type (7) that is not stiffly accurate. They will be denoted by **SDIRKsX(p,ps)** where **s** is the number of stages, **X** is a reference to the author(s), **p** is the non stiff order and **ps** is the stage order. We have not included nonlinear Lipschitz continuous terms $g(t, y)$ in the problems presented here because they do not essentially modify the convergence behaviour of the methods. In our experiments we have used the Euclidean norm $\| \cdot \|_2$. The Runge-Kutta methods considered are:

- **SDIRK5HW(4,1)** is the five–stage SDIRK method given by Hairer–Wanner [10, p.100]. It has standard order four and stage–order one. It is stiffly accurate and L-stable. Since its coefficient matrix A satisfies (6), the method fulfils the assumptions of Theorems 3.1, 3.2 and 3.3.

- **SDIRK4A(3,2)** is the four–stage SDIRK method proposed by Alexander in [2, pp.2,6-8] and it is defined by

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ a_{31} & a_{32} & \gamma & 0 \\ b_1 & b_2 & b_3 & \gamma \end{pmatrix}, \qquad b = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \gamma \end{pmatrix}, \quad c = Ae,$$

  with $\gamma = 0.43586652...$ and $c_3 = 1/2+\gamma/4$. Since it was required to have standard order three and stage order two, the remainder coefficients are univocally determined. It is L–stable, stiffly accurate and it satisfies (8). Hence, it meets the assumptions in Theorems 3.2 and 3.3.

- **SDIRK3(3,2)** is the three–stage SDIRK method defined by the coefficient matrices

$$A = \begin{pmatrix} 0 & 0 & 0 \\ \gamma & \gamma & 0 \\ a_{31} & a_{32} & \gamma \end{pmatrix}, \qquad b = \frac{1}{88}\begin{pmatrix} 52 - 72\gamma \\ 25 + 50\gamma \\ 11 + 22\gamma \end{pmatrix}, \tag{10}$$

  with

$$a_{32} = \frac{2(1-2\sqrt{3})}{25}, \qquad a_{31} = c_3 - a_{32} - \gamma, \quad \gamma = \frac{3+\sqrt{3}}{6}, \; c_3 = \frac{4}{5}.$$

  This method is strongly A–stable ($R(\infty) = \gamma^{-1} - 2 = -0.732...$), has stage–order two and standard order three. It is not stiffly accurate but it satisfies (8), hence it fulfills the assumptions of Theorem 3.3, but not the ones of Theorem 3.2.

For the problems considered below, $\epsilon > 0$ is normally a small parameter, hence the stiffness of the problems, i.e. the Lipschitz condition on y for $f(t, y)$, is of order $\mathcal{O}(\epsilon^{-1})$.

Problem 1.- The two dimensional variable coefficient linear system

$$y'(t) = \widehat{J}(t)\,[y(t) - \varphi(t)] + \varphi'(t), \quad t \in [0, 4\pi], \quad y(0) = (1, 0)^T, \tag{11}$$

where $\varphi(t) = (\cos t, \sin t)^T$, $\widehat{J}(t) = S(t)^{-1}\Lambda S(t)$, $S(t) = PQ(t)$, and

$$\Lambda = \begin{pmatrix} -1 & 0 \\ 0 & -\epsilon^{-1} \end{pmatrix}, \; P = \begin{pmatrix} -1 & 0 \\ 1 & 1 \end{pmatrix}, \; Q(t) = \begin{pmatrix} 1 & \epsilon \\ e^{\sin t} & e^{\sin t} \end{pmatrix}. \tag{12}$$

The exact solution $y(t) = \varphi(t)$ is asymptotically stable and $\widehat{J}(t)$, that has as eigenvalues -1, and $-1/\varepsilon$, is highly non normal. In fact, its logarithmic norm behaves as ε^{-1} when $\varepsilon \to 0$. In this case $\widehat{J}(t)$ does not fulfill the assumptions of the B–theory. However, the assumption (H1), (H2-i) and (H3) are clearly satisfied.

It can be seen that for a matrix $\widehat{J}(t)$ of type $\widehat{J}(t) = S(t)^{-1} J S(t)$, with a non singular J, condition (3) holds if and only if $\| J^{-1} M J - M \|$ can be bounded independently of the stiffness, where $M = (\Delta t)^{-1}(S(t)^{-1} S(t+\Delta t) - I)$. In this case we have that

$$J^{-1} M J - M = \begin{pmatrix} 0 & -\nu(t) \\ -\nu(t) & 0 \end{pmatrix} \qquad \text{with} \quad \nu(t) = \frac{a(t+\Delta t) - a(t)}{a(t)\Delta t},$$

and therefore (3) is satisfied independently of ϵ.

Problem 2.- The variable coefficient linear problem similar to that one considered by Kreiss in [12],

$$y' = \widehat{J}(t) y \equiv S(t)^{-1}\, \Lambda\, S(t)\, y, \quad t \in [0, 4\pi], \tag{13}$$

where P and Λ are given by (12), $S(t) = P\Omega(t)$, and

$$\Omega(t) = \begin{pmatrix} \cos(t) & \sin(t) \\ -\sin(t) & \cos(t) \end{pmatrix}. \tag{14}$$

Since $y \to S(t) y$ transforms (13) into a constant coefficient linear system, the general solution of (13)-(14) can be written in the form

$$y(t) = S(t)^{-1} \begin{pmatrix} 1 & 1 \\ \lambda_+ & \lambda_- \end{pmatrix} \begin{pmatrix} C_+ e^{\lambda_+ t} \\ C_- e^{\lambda_- t} \end{pmatrix}, \tag{15}$$

with $\lambda_+ = -2\epsilon + \mathcal{O}(\epsilon^2)$ and $\lambda_- = -\epsilon^{-1} - 1 + \mathcal{O}(\epsilon)$. Moreover, all solutions tend quickly, after the initial transient layer, to the smooth stationary solution, which corresponds to the parameter $C_- = 0$. We have chosen $C_- = 0$ and $C_+ = 1$ for our numerical experiments. For the logarithmic norm, it can be shown that $\mu_2[\widehat{J}(t)] = \mathcal{O}(\epsilon^{-1}) \gg 1$. However, the assumptions (H1), (H2), (H3) are satisfied. In this case, the condition (3) is not accomplished.

Problem 3.- The linear system of partial differential equations of parabolic type,

$$\left. \begin{aligned} u_t &= a_{11}(t) u_{xx} + a_{12}(t) v_{xx} + r_1(x,t) \\ v_t &= a_{21}(t) u_{xx} + a_{22}(t) v_{xx} + r_2(x,t) \end{aligned} \right\} \tag{16}$$

where $u = u(x,t)$ and $v = v(x,t)$ represent the unknowns, the space variable x ranges in $[0,1]$, the functions $r_i(x,t)$, $(i = 1,2)$ are given by

$$r_1(x,t) = 2\cos t - \phi(x)\sin t, \quad r_2(x,t) = -(2\sin t + \phi(x)\cos t), \quad \phi(x) = x(1-x).$$

and $a_{ij}(t)$ are defined by

$$A(t) = \begin{pmatrix} a_{11}(t) & a_{12}(t) \\ a_{21}(t) & a_{22}(t) \end{pmatrix} \equiv \Omega(t)(-\Lambda)\Omega(t)^{-1}, \tag{17}$$

where $\Omega(t)$ and Λ are given by (14) and (12) respectively.

Taking as initial–boundary conditions

$$u(x,0)=\phi(x),\ v(x,0)=0, \qquad u(0,t)=u(1,t)=0,\ v(0,t)=v(1,t)=0,$$

the exact solution of (16) is $u(x,t)=\phi(x)\cos t, \quad v(x,t)=-\phi(x)\sin t$, independent of the parameter ϵ (only $\epsilon>0$ gives stable solutions).

Taking a uniform grid $x_j=j\Delta x, j=0,1,\dots,M+1$, in the spatial variable with $\Delta x=1/(M+1)$, and using second order centered differences, the semi discretization of (16) can be written as

$$\begin{aligned} u_j' &= a_{11}(t)\frac{u_{j-1}-2u_j+u_{j+1}}{(\Delta x)^2}+a_{12}(t)\frac{v_{j-1}-2v_j+v_{j+1}}{(\Delta x)^2}+r_1(x_j,t),\\ v_j' &= a_{21}(t)\frac{u_{j-1}-2u_j+u_{j+1}}{(\Delta x)^2}+a_{22}(t)\frac{v_{j-1}-2v_j+v_{j+1}}{(\Delta x)^2}+r_2(x_j,t),\\ & \qquad\qquad j=1,2,\dots,M, \end{aligned} \tag{18}$$

where $u_j(t)=u(x_j,t)$ and $v_j(t)=v(x_j,t)$, $j=0,\dots,M+1$. The initial conditions imply that $u_j(0)=\phi(x_j),\ v_j(0)=0,\ j=1,\dots,M$. It must be observed that the discrete exact solutions of (18) and (16) are the same.

By introducing the matrix

$$W=\frac{1}{(\Delta x)^2}\begin{pmatrix} -2 & 1 & & & \\ 1 & -2 & 1 & & \\ & & \ddots & \ddots & \\ & & 1 & -2 & 1 \\ & & & 1 & -2 \end{pmatrix}\in\mathbb{R}^{M,M},$$

and the vectors

$$U(t)=\begin{pmatrix} u_1(t)\\ \vdots \\ u_M(t)\end{pmatrix}, \qquad V(t)=\begin{pmatrix} v_1(t)\\ \vdots \\ v_M(t)\end{pmatrix}, \qquad y(t)=\begin{pmatrix} U(t)\\ V(t)\end{pmatrix},$$

$$\widetilde{R}_1(t)=\begin{pmatrix} r_1(x_1,t)\\ \vdots \\ r_1(x_M,t)\end{pmatrix}, \quad \widetilde{R}_2(t)=\begin{pmatrix} r_2(x_1,t)\\ \vdots \\ r_2(x_M,t)\end{pmatrix}, \quad g(t)=\begin{pmatrix} \widetilde{R}_1(t)\\ \widetilde{R}_2(t)\end{pmatrix},$$

we get the initial value problem

$$y'(t)=\widehat{J}(t)y+g(t), \qquad y(0)=(\phi(x_1),\dots,\phi(x_N),0,\dots,0)^T, \tag{19}$$

with $\widehat{J}(t)=A(t)\otimes W$.

Since $\widehat{J}(t) = (\Omega(t) \otimes I)\ (-\Lambda \otimes W)(\Omega(t)^{-1} \otimes I)$, the assumption (H1) is satisfied for $J(t) = S(t)\widehat{J}(t)S(t)^{-1} = -\Lambda \otimes W$ because it is a constant matrix that fulfills $\mu_2[J(t)] \simeq -\pi^2$. Further (H2) and (H3) are clearly accomplished.

For the smoothness of the exact solution $y(t)$, it is enough to consider the weighted Euclidean norm $\| \cdot \| = M^{-1/2} \| \cdot \|_2$. Thus, it follows that $\| y^{(j)}(t) \| \leq \max_{x\in[0,1]} |\phi(x)| = 1/4, \quad j = 0, 1, 2, \ldots, \quad t \in \mathbb{R}$. Observe that for any matrix B, both norms are identical, i.e., $\| B \| = \| B \|_2$. In conclusion, this problem fulfills the H–assumptions. For our numerical experiments we have selected $M = 20$, so that the dimension of the linear system is 40. In this case, unlike of the problems 1 and 2, the stiffness comes from two sources, from the small parameter ϵ and from the discretization in space.

For each problem and method we have carried out integrations for $t \in [0, 4\pi]$ with fixed step sizes $h = 4\pi/N$ for $N = 100, 200, 400, 800, 1600$. Note that, since in all problems the initial conditions have been taken on a stationary (smooth) solution, a fixed step size strategy can be used in the integrations. We have computed the global error at the end point $GE(N) = \|y(t_N) - y_N\|$ and also the numerical order of convergence p_N defined by

$$p_N = \frac{1}{\log(2)}\ \log\left(\frac{GE(N)}{GE(2N)}\right).$$

Concerning the method SDIRK5HW(4,1) observe that Th. 3.1 implies that it is convergent for all problems with order greater or equal that the stage order $q = 1$. This can be checked in Table 1, where the global errors $GE(N)$ and numerical orders p_N obtained for the Problems 1 and 2, have been displayed for several values of the parameter ϵ. Similar results, not included here, were encountered for Problem 3. From these results we follow:

- $GE(N) \to 0$ when $N \to \infty$ for a wide range of values of ϵ, i.e., the method is convergent also for the considered stiff problems.
- For the non–stiff case $\epsilon = \mathcal{O}(1)$, the computed numerical orders agree with the classical order of convergence $p = 4$ as expected.
- For small ϵ-values the observed orders range, in most cases, between the stage order and the classical order. Hence, the lower order of convergence can not be improved in general.
- In Problem 1, for fixed step-sizes, the $GE(N)$ decreases when $\epsilon \to 0$. For this problem, it can be shown that the local error tends to zero when $\epsilon \to 0$ and this property is also reflected in the global error behaviour.
- It must be also remarked that the first–order of convergence above cannot be explained from the ϵ–theory (see Corollary 3.10, pp. 402-403 in [10, Chap. VI.3]),

since from that theory global errors of size $\mathcal{O}(h^4 + \epsilon h)$ are expected, and this is not the case as it can be observed from the results in Table 1, when moving either on rows (fixed h) or on columns (fixed ϵ).

Table 1.— Method SDIRK5HW(4,1) for Problems 1 and 2

Problem 1

	$\epsilon = 0.5$		$\epsilon = 10^{-2}$		$\epsilon = 10^{-4}$		$\epsilon = 10^{-6}$		$\epsilon = 10^{-8}$	
N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N
50	3.1e-5		1.3e-3		1.8e-5		3.7e-6		3.9e-6	
100	2.4e-6	3.68	3.9e-4	1.78	1.0e-5	0.78	2.0e-7	4.26	2.8e-7	3.80
200	1.7e-7	3.85	7.2e-5	2.44	5.2e-6	0.98	3.7e-8	2.41	1.9e-8	3.93
400	1.1e-8	3.93	8.9e-6	3.02	2.6e-6	1.02	2.5e-8	0.54	1.0e-9	4.21
800	7.1e-10	3.96	8.4e-7	3.40	1.2e-6	1.04	1.3e-8	0.95	7.1e-11	3.82
1600	4.5e-11	3.98	6.7e-8	3.65	5.8e-7	1.09	6.5e-9	1.00	6.1e-11	0.21

Problem 2

50	4.9e-9		1.1e-3		2.4e-4		2.7e-4		2.7e-4	
100	3.3e-10	3.91	2.2e-4	2.37	1.0e-5	4.55	1.8e-5	3.88	1.8e-5	3.87
200	2.1e-11	3.95	2.7e-5	2.99	9.6e-7	3.41	1.2e-6	3.95	1.3e-6	3.82
400	1.3e-12	3.97	2.8e-6	3.26	4.8e-7	1.01	7.2e-8	4.03	9.8e-8	3.73
800	8.5e-14	3.99	2.8e-7	3.36	1.3e-7	1.85	4.0e-9	4.17	4.0e-8	1.28
1600	5.3e-15	3.99	2.4e-8	3.52	3.2e-8	2.05	5.7e-10	2.81	1.3e-8	1.61

Numerical experiments with the method SDIRK4A(3,2) for the three problems are presented in Table 2. Note that Theorem 3.1 can not be applied to this method but it meets the assumptions of Theorem 3.2. From the displayed results, apart from convergence behaviour for all problems, it can be observed that

- For the non–stiff case $\epsilon = \mathcal{O}(1)$, the computed numerical orders agree with the classical order of convergence $p = 3$.
- For small ϵ the order reduction is not observed in Problems 1 and 2. However, for Problem 3 numerical orders lower than 3 are found, therefore the stage order $q = 2$, seems to be the guaranteed order of convergence.
- In all problems, for a fixed stepsize h, the $GE(N)$ seem to be non-dependent on ϵ. This fact is explained because the local errors are practically independent on ϵ.

Table 2.— Method SDIRK4A(3,2) for Problems 1, 2 and 3

Problem 1

	$\epsilon = 0.5$		$\epsilon = 10^{-2}$		$\epsilon = 10^{-4}$		$\epsilon = 10^{-6}$		$\epsilon = 10^{-8}$	
N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N
50	2.7e-4		3.7e-4		4.0e-4		4.0e-4		4.0e-4	
100	4.4e-5	2.63	4.6e-5	2.98	5.3e-5	2.90	5.3e-5	2.90	5.3e-5	2.90
200	6.4e-6	2.76	5.5e-6	3.08	6.9e-6	2.95	6.9e-6	2.95	6.9e-6	2.95
400	8.8e-7	2.87	6.2e-7	3.16	8.7e-7	2.98	8.8e-7	2.97	8.8e-7	2.97
800	1.2e-7	2.93	6.7e-8	3.21	1.1e-7	3.00	1.1e-7	2.99	1.1e-7	2.99
1600	1.5e-8	2.97	7.3e-9	3.20	1.4e-8	3.01	1.4e-8	2.99	1.4e-8	2.99

Problem 2

50	5.6e-8		6.8e-3		9.1e-3		9.1e-3		9.1e-3	
100	7.7e-9	2.87	8.6e-4	2.99	1.2e-3	2.95	1.2e-3	2.96	1.2e-3	2.96
200	1.0e-9	2.92	1.0e-4	3.11	1.5e-4	2.98	1.5e-4	2.98	1.5e-4	2.98
400	1.3e-10	2.96	9.7e-6	3.36	1.9e-5	2.99	1.9e-5	2.99	1.9e-5	2.99
800	1.7e-11	2.98	6.5e-7	3.89	2.4e-6	2.99	2.4e-6	2.99	2.4e-6	2.99
1600	2.1e-12	2.99	1.0e-8	6.03	3.0e-7	2.99	3.0e-7	3.00	3.0e-7	3.00

Problem 3

50	2.0e-5		9.8e-7		1.5e-7		1.5e-7		1.5e-7	
100	3.3e-6	2.58	2.4e-7	2.06	2.8e-8	2.48	2.7e-8	2.49	2.8e-8	2.47
200	5.0e-7	2.70	5.5e-8	2.11	7.0e-9	1.98	6.2e-9	1.98	6.6e-9	2.07
400	7.2e-8	2.81	1.2e-8	2.19	1.1e-9	2.63	1.1e-9	2.63	9.2e-10	2.85
800	9.7e-9	2.89	2.5e-9	2.28	1.6e-10	2.80	1.6e-10	2.84	1.5e-10	2.57
1600	1.3e-9	2.94	4.7e-10	2.38	2.3e-11	2.81	2.2e-11	2.85	1.8e-10	-0.24

Finally, the method SDIRK3(3,2) meets the assumptions of Theorem 3.3 and therefore it is convergent with order greater or equal than the stage order ($q = 2$) for Problem 1. This fact agrees with the numerical results displayed in Table 3, where a third–order convergence is observed for that problem. However, SDIRK3(3,2) satisfies neither the assumptions of Theorem 3.1 nor those of Theorem 3.2, hence the convergence on Problems 2 and 3 is not guaranteed. In fact, the numerical experiments in Table 3 (Problem 2) show that the method is not B-convergent on the whole class of Problems 2 with $0 < \epsilon \leq 1$.

Similar results of not B-convergence were encountered for Problem 3 when considering small values for ϵ.

5 Conclusions

New theoretical results on stability and convergence for stiff semi–linear problems that extend previous results on the subject [4], [6] have been derived. The new results support the use of some SDIRK formulas [2], [10], [11] that have been designed taking into account their linear stability properties and efficient implementation in practical codes. Numerical experiments show that the assumptions on the theorems and stiff orders can not be improved for the class of problems under consideration.

Table 3.— Method SDIRK3(3,2) for Problems 1 and 2

Problem 1

	$\epsilon = 0.5$		$\epsilon = 10^{-2}$		$\epsilon = 10^{-4}$		$\epsilon = 10^{-6}$		$\epsilon = 10^{-8}$	
N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N	$GE(N)$	p_N
50	6.9e-4		1.1e-3		1.1e-3		1.1e-3		1.1e-3	
100	1.1e-4	2.64	1.5e-4	2.87	1.6e-4	2.84	1.6e-4	2.84	1.6e-4	2.84
200	1.7e-5	2.67	2.0e-5	2.97	2.1e-5	2.91	2.1e-5	2.91	2.1e-5	2.91
400	2.5e-6	2.80	2.4e-6	3.05	2.7e-6	2.95	2.7e-6	2.95	2.7e-6	2.95
800	3.4e-7	2.89	2.7e-7	3.12	3.4e-7	2.98	3.5e-7	2.98	3.5e-7	2.98
1600	4.4e-8	2.94	3.0e-8	3.17	4.3e-8	2.99	4.4e-8	2.99	4.3e-8	2.99

Problem 2

50	2.0e-7		3.3e-2		4.8e+65		1.9e+163	–	1.8e+261	–
100	2.7e-8	2.87	3.4e-3	3.25	7.7e+59	–	3.0e+251	–	1.4e+303	–
200	3.5e-9	2.92	3.4e-4	3.35	6.0e+10	–	4.9e+303	–	2.4e+304	–
400	4.6e-10	2.95	3.0e-5	3.49	8.2e-04	–	7.8e+303	–	1.8e+304	–
800	5.8e-11	2.97	2.0e-6	3.90	5.7e-05	3.84	5.0e+154	–	6.3e+303	–
1600	7.3e-12	2.99	2.8e-8	6.14	4.0e-06	3.83	3.0e-004	–	1.2e+303	–

Acknowledgements

This work was supported by projects MTM2007-67530-C02-01 and MTM2007-67530-C02-02.

References

[1] R. K. Alexander, *Stability of Runge–Kutta methods for stiff ordinary differential equations*, SIAM J. Numer. Anal., 31, 4 (1994), pp. 1147–1168.

[2] R. K. Alexander, *Design and implementation of DIRK integrators for stiff systems*, Appl. Numer. Math. 46 (2003), pp. 1-17.

[3] W. Auzinger, R. Frank and G. Kirlinger, *A note on convergence concepts for Stiff Problems*, Computing 44 (1990), pp. 197–208.

[4] W. Auzinger, R. Frank and G. Kirlinger, *An extension of B–convergence for Runge–Kutta methods*, Appl. Numer. Math., 9 (1992), pp. 91–109.

[5] K. Burrage, W. H. Hundsdorfer and J. G. Verwer, *A study of B-convergence of Runge-Kutta methods*, Computing 36 (1986), pp. 17–34.

[6] M. Calvo, J. I. Montijano and S. Gonzalez-Pinto, *Runge–Kutta methods for the numerical solution of stiff semilinear initial value problems*, BIT 40, 4 (2000), pp. 611–639.

[7] M. Calvo, J. I. Montijano and S. Gonzalez-Pinto, *On the convergence of Runge-Kutta methods for stiff semi linear initial value problems*, Technical report, Dept. de Matemática aplicada, Univ. de Zaragoza (2005), http://pcmap.unizar.es/numerico/reports.

[8] K. Dekker and J. G. Verwer, *Stability of Runge-Kutta methods for stiff nonlinear differential equations*, North Holland, Amsterdam, 1984.

[9] J. L. M. van Dorsselaer van and M. N. Spijker, *The error committed by stopping the Newton iteration in the numerical solution of stiff initial value problems*, IMA J. Num. Anal., 14 (1994), pp. 183–209.

[10] E. Hairer and G. Wanner, *Solving Ordinary Differential Equations II: Stiff and Differential–Algebraic Problems*, Springer Verlag, Berlin, 1996.

[11] C. A. Kennedy and M. H. Carpenter, *Additive Runge-Kutta schemes for convection-diffusion-reaction equations*, Appl. Numer. Math., 44, 1–2 (2003), pp. 139–181.

[12] H.O. Kreiss, *Difference methods for stiff ordinary differential equations*, SIAM J. Numer. Anal., 15 (1978), pp. 21–58.

[13] B. A. Schmitt, *Stability of implicit Runge-Kutta methods for nonlinear stiff differential equations*, BIT, 28, (1988), pp. 884–897.

[14] K. Strehmel and R. Weiner, *B-convergence results for linearly implicit one step methods*, BIT, 27, (1987), pp. 264–281.