SIAM J. SCI. COMPUT.
Vol. 31, No. 3, pp. 1926–1945

ON A CLASS OF UNIFORMLY ACCURATE IMEX RUNGE–KUTTA SCHEMES AND APPLICATIONS TO HYPERBOLIC SYSTEMS WITH RELAXATION*

SEBASTIANO BOSCARINO† AND GIOVANNI RUSSO†

Abstract. In this paper we consider hyperbolic systems with relaxation in which the relaxation time ε may vary from values of order one to very small values. When ε is very small, the relaxation term becomes very strong and highly stiff, and underresolved numerical schemes may produce spurious results. In such cases it is important to have schemes that work uniformly with respect to ε. IMplicit-EXplicit (IMEX) Runge–Kutta (R-K) schemes have been widely used for the time evolution of hyperbolic partial differential equations but the schemes existing in literature do not exhibit uniform accuracy with respect to the relaxation time. We develop new IMEX R-K schemes for hyperbolic systems with relaxation that present better uniform accuracy than the ones existing in the literature and in particular produce good behavior with high order accuracy in the asymptotic limit, i.e., when ε is very small. These schemes are obtained by imposing new additional order conditions to guarantee better accuracy over a wide range of the relaxation time. We propose the construction of new third-order IMEX R-K schemes of type CK [S. Boscarino, *SIAM J. Numer. Anal.*, 45 (2008), pp. 1600–1621]. In several test problems, these schemes, with a fixed spatial discretization, exhibit for all range of the relaxation time an almost uniform third-order accuracy.

Key words. Runge–Kutta methods, stiff problems, hyperbolic systems with relaxation, order conditions

AMS subject classifications. 65C20, 65L06

DOI. 10.1137/080713562

1. Introduction.

Several physical problems of great relevance for applications are described by stiff system in the form:

$$\partial_t U = F(U) + \frac{1}{\varepsilon} G(U), \tag{1}$$

where $U = U(t) \in R^M$, $F, G : R^M \to R^M$, and $\varepsilon > 0$ is the stiffness parameter.

The development of numerical schemes for systems (1) attracted a lot of attention in recent years. Systems of such form often arise from the discretization of partial differential equations, such as convection-diffusion equations and hyperbolic systems with relaxation. In this work we consider the latter case which in recent years has been a very active field of research, due to its great impact on applied sciences. In fact, relaxation is important in many physical situations, for example, it arises in discrete kinetic theory of rarefied gases, hydrodynamical models for semiconductors, linear and nonlinear waves, viscoelasticity, traffic flows, shallow water, etc. ([7], [16], [17], [18], [14], [6], [21], [8], [13]).

In one space dimension these problems can be described mathematically by the system

$$\partial_t u + \partial_x \mathcal{F}(u) = \frac{1}{\varepsilon} R(u), \quad u = u(x,t) \in R^N, \quad x \in R, \quad t \in R^+ \tag{2}$$

*Received by the editors January 16, 2008; accepted for publication (in revised form) November 10, 2008; published electronically March 27, 2009. This work was supported by Italian PRIN 2006 project Modelli cinetici e del continuo per il trasporto di particelle nei gas e nei semiconduttori: aspetti analitici e computazionali. Prot. 2006012132-003.
http://www.siam.org/journals/sisc/31-3/71356.html

†Department of Mathematics and Computer Science, University of Catania, viale A. Doria 6, 95125, Italy (boscarino@dmi.unict.it, russo@dmi.unict.it).

from which system (1) can be obtained by suitable finite difference discretization in space (methods of lines) [20]. System (2) is assumed to be hyperbolic; i.e., the Jacobian matrix $\mathcal{F}'(u)$ has real eigenvalues and admits a basis of eigenvectors for every $u \in R^N$. We will call system (2) the relaxation system and ε is called the relaxation time, which is small in many physical situations. Here we use the term relaxation in the sense of Whitham [21] and Liu ([8], [15]); i.e., the operator $R : R^N \to R^N$ is said relaxation operator, and consequently (2) defines a relaxation system, if there exists a constant $n \times N$ matrix Q, with $\operatorname{rank}(Q) = n < N$ such that $QR(u) = 0$ for $u \in R^N$. This yields n independent conserved quantities $v = Qu$. In addition, we assume that each such v uniquely determines a local equilibrium value $u = \mathcal{E}(v)$ satisfying $R(\mathcal{E}(v)) = 0$ and such that $Q\mathcal{E}(v) = v$ for all v. The image of $\mathcal{E}$ represents the manifold of local equilibria of R. Associated with Q are n conservation laws satisfied by every solution of (2) and that take the form $\partial_t(Qu) + \partial_x(Q\mathcal{F}) = 0$. These can be closed as a reduced system for $v = Qu$ if we take the local relaxation approximation for u, namely $u = \mathcal{E}(v)$, $\partial_t v + \partial_x H(v) = 0$, where H is defined by $H(v) = Q\mathcal{F}(\mathcal{E}(v))$. Typical examples of such systems are discrete velocity models in kinetic theory, such as the Broadwell model [14, 6], which in one space dimension has the structure of a semilinear hyperbolic system of three equations that degenerates to a quasilinear hyperbolic system with two equations, as $\varepsilon \to 0$.

The development of efficient numerical schemes for such systems is challenging, since in many applications the relaxation time varies from values of order one to very small values if compared to the time scale determined by the characteristic speeds of the system. In this second case the hyperbolic system with relaxation is said to be stiff.

Here we will focus on the development of schemes that work for all ranges of the relaxation time and we are interested in high order numerical schemes for the stiff relaxation system which are able to capture the correct physical behavior with high order accuracy. Underresolved numerical schemes may yield spurious numerical solutions that are unphysical and, in general for hyperbolic systems with stiff terms, high order schemes may also reduce to lower order when the mesh fails to resolve the small relaxation time.

In this article we will present some recent results on the development of high-order implicit-explicit (IMEX) Runge–Kutta (R-K) schemes suitable for time-dependent partial differential systems that work better than other IMEX R-K schemes existing in the literature. Classical high-order IMEX R-K schemes fail to maintain the high-order accuracy in time in the whole range of the relaxation time and in particular in the asymptotic limit, i.e., when the relaxation time is not temporally resolved. We introduce here a new class of IMEX R-K schemes that are able to handle the stiffness of the system (1) in the whole range of the relaxation time.

This work is motivated by the study of the global error for different types of IMEX R-K methods existing in literature [2]. This error analysis is accompanied by an order reduction phenomenon where the observed convergence rates of the methods drop the classical order of accuracy in time. The development of these schemes is aided by the knowledge of extra order conditions in addition to the classical ones [7, 16] which ensure accuracy in time at the various order in the stiffness parameter ε [3, 4]. We remark that in the classical literature IMEX R-K schemes do not satisfy these additional order conditions.

This approach allows the construction of new IMEX R-K schemes producing an error which is more uniform in the stiffness parameter. In particular, such conditions are used to improve some existing scheme (see also [2]). Our analysis is based on the

smoothness assumption of the solution and also applies to the stiff case when $\Delta t \gg \varepsilon$ (see [2], [3]).

The outline of the paper is as follows. In the next section we present the general structure of IMEX Runge–Kutta schemes. In section 3 we report the new order conditions up to third-order derived from [3]. Section 4 is devoted to the development of uniformly accurate IMEX R-K schemes. Important properties are proved and new assumptions are considered. In section 5 we perform several test problems to check the accuracy of new schemes for various values of the stiffness parameter and we discuss the result. Conclusions are drawn and work in progress is mentioned. Appendices are also included.

2. IMEX Runge–Kutta schemes and classification. An IMEX R-K scheme applied to (1) has the following form:

$$U_{n+1} = U_n + h\sum_{i=1}^{s} \tilde{b}_i F(t_n + \tilde{c}_i h, U^i) + h\sum_{i=1}^{s} b_i \frac{1}{\varepsilon} G(t_n + c_i h, U^i), \tag{3}$$

with internal stages given by

$$U^i = U_n + h\sum_{j=1}^{i-1} \tilde{a}_{ij} F(t_n + \tilde{c}_i h, U^i) + \sum_{j=1}^{i} a_{ij} \frac{1}{\varepsilon} G(t_n + c_i h, U^i). \tag{4}$$

The matrices $(\tilde{a}_{ij})$, with $\tilde{a}_{ij} = 0$ for $j \geq i$, and (a_{ij}) are $s \times s$ matrices such that the resulting method is explicit in F, and implicit in G. Using a diagonally implicit method ($a_{ij} = 0$, for $j > i$) for G gives a sufficient condition to guarantee that F is always evaluated explicitly; furthermore, this choice is usually preferred over full implicit schemes for efficiency reasons. The coefficient vectors $\tilde{c} = (\tilde{c}_1, \ldots, \tilde{c}_s)^T$, $\tilde{b} = (\tilde{b}_1, \ldots, \tilde{b}_s)^T$, $c = (c_1, \ldots, c_s)^T$, $b = (b_1, \ldots, b_s)^T$ complete the characterization of the schemes. They can be represented by a double *tableau* in the usual Butcher notation,

$$\begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} \qquad \begin{array}{c|c} c & A \\ \hline & b^T \end{array}.$$

We assume that the coefficients $\tilde{c}$ and c, used for the treatment of nonautonomous systems, are given by the relation

$$\tilde{c}_i = \sum_{j=1}^{i-1} \tilde{a}_{ij}, \quad c_i = \sum_{j=1}^{i} a_{ij}. \tag{5}$$

The great number of IMEX R-K methods presented in literature lead us to classify them in three different types characterized by the structure of the matrix $A = (a_{ij})_{i,j=1}^{s}$ of the implicit scheme [2].

DEFINITION 1. *We call an IMEX R-K scheme of type A, (see* [18]*), if the matrix* $A \in R^{s\times s}$ *is invertible.*

DEFINITION 2. *We call an IMEX R-K scheme of type CK, (see* [7]*), if matrix* $A \in R^{s\times s}$ *can be written as*

$$A = \begin{pmatrix} 0 & 0 \\ a & \hat{A} \end{pmatrix}$$

with the submatrix $\hat{A} \in R^{(s-1)\times(s-1)}$ *invertible.*

Remark 1. IMEX R-K schemes, called of type ARS (see [1]), are a special case of the type CK with the vector $a = 0$.

3. Additional order conditions. IMEX Runge–Kutta methods can be viewed as a particular class of partitioned Runge–Kutta methods. Therefore, their order conditions can be derived from the general theory of partitioned methods [12]. In particular, for a detailed account of the classical order conditions for the IMEX R-K schemes see, for example, [7] and [16].

In [3, 4] we derived additional order conditions that together with the classical ones ensure accuracy at the various order in the stiffness parameter ε.

In his paper [2] Boscarino studied the global error behaviour of some popular IMEX R-K schemes, showing a degradation of the order of accuracy for small values of ε.

By applying to IMEX R-K schemes a technique similar to the one used in [11] Chapt. VII in the case of implicit R-K schemes, additional order conditions are obtained by imposing that the numerical solution and the exact solution agree to various order in an expansion of ε. This analysis explains the degradation in the order of accuracy of existing IMEX R-K schemes, and allowed the construction of new schemes with better uniform accuracy in ε. For a detailed analysis we refer [3], [4].

Let us write explicitly the additional order conditions up to third order. Then, the index 1 and 2 order conditions for type A schemes read:

Index 1 Order Conditions

$$\sum_{i,j} b_i\omega_{ij}\tilde{c}_j = 1, \quad \textstyle\sum_{i,j} b_i\omega_{ij}\tilde{c}_j^2 = 1, \quad \sum_{i,j,k} b_i\omega_{ij}\tilde{a}_{jk}\tilde{c}_k = 1/2, \tag{6}$$

Index 2 Order Conditions

$$\begin{aligned} &\sum_{i,j,k} b_i\omega_{ij}\omega_{jk}\tilde{c}_k^2 = 2, && \textstyle\sum_{i,j,k} b_i\omega_{ij}\omega_{jk}c_k^2 = 2, \\ &\sum_{i,j,k} b_i\omega_{ij}\omega_{jk}\tilde{c}_k c_k = 2, && \textstyle\sum_{i,j,k,l} b_i\omega_{ij}\omega_{jk}\tilde{a}_{kl}\tilde{c}_l = 1, \\ &\sum_{i,j,k,l} b_i\omega_{ij}\omega_{jk}\tilde{a}_{kl}c_l = 1, \end{aligned} \tag{7}$$

where ω_{ij} are the elements of the inverse matrix of A. Observe that we deduced these order conditions for the type A; for the type CK (consequently for type ARS) we can rewrite the same ones replacing ω_{ij} with $\hat{w}_{ij}$ (the elements of the inverse of the matrix $\hat{A}$) with i, j, k from 2 to s.

Then, it may be advantageous to have schemes that satisfy these new order conditions. Although the goal of this paper is to present a class of numerical schemes that work with uniform accuracy in the whole range of the stiffness parameter ε, we will show that the use of these additional order conditions improves existing schemes.

In [2, 5], we introduced two IMEX R-K schemes, called MARK3(2)4L[2]SA and MARS(3,4,3), that are an improvement of the existing schemes, ARK3(2)4L[2]SA [7] and ARS(3,4,3) [1]. These schemes use the whole set of classical order conditions plus conditions (6) and show a good accuracy behavior with respect to the increasing stiffness in the limit of $\varepsilon = 0$ where the system becomes an index 1 differential algebraic system (see [2]). In section 5 we compare the different performances of these schemes.

4. Construction of a particular uniformly accurate IMEX R-K scheme. In [7], [16], [18], [17], [1], and [14] IMEX R-K schemes were introduced. As we saw in [2] and [4] none of them are suitable to preserve high-order accuracy uniformly with respect to the wide range of stiffness parameter ε when applied to stiff system (1).

Here we construct a third-order uniformly accurate IMEX Runge–Kutta scheme with respect to the stiffness parameter ε through several assumptions and additional order conditions. First of all, we require the following assumptions:

0. $b_i = \tilde{b}_i$, $c_i = \tilde{c}_i$ for $i = 1, \dots, s$,
1. $\sum_{j=1}^{i} a_{ij} c_j^{k-1} = c_i^k/k$ for $i = 2, \dots, s$ and $k = 1, 2$,
2. $\sum_{j=1}^{i-1} \tilde{a}_{ij} c_j^{k-1} = c_i^k/k$ for $i = 3, \dots, s$ and $k = 1, 2$,
3. $\sum_{i=1}^{s} b_i = 1, \quad \sum_{i=1}^{s} b_i c_i = 1/2, \quad \sum_{i=1}^{s} b_i c_i^2 = 1/3$,
4. $\tilde{b}_2 = b_2 = 0$.

1 and 2 are the simplifying assumptions $C(2)$ for the implicit and explicit part with $c_i = \tilde{c}_i$ for $i = 1, \dots, s$, which will be helpful in designing IMEX R-K schemes because they simplify the classical order conditions and conditions (6), (7) (see [11, 7]). The technical condition 4 is necessary because the assumption 2 cannot be satisfied for $i = 2$ in the explicit scheme. Furthermore, the classical order conditions for IMEX R-K schemes ([7], [16]) simplify a lot if assumption 0 is satisfied, so that the standard order p conditions for the explicit and implicit schemes are enough to guarantee the same order p for the IMEX R-K schemes, for $p \leq 3$. Notice that if assumptions 0 and 1 for $k = 1, 2$, are satisfied, then classical order conditions for the two R-K schemes imply that coupled order conditions are automatically satisfied for schemes up to third order.

Now, among the different types of IMEX schemes considered in [2], we choose to construct schemes of type CK because of their good properties. Concerning these type CK schemes, we consider schemes that are in the implicit part *stiffly accurate* ($a_{si} = b_i$, for $i = 1, \dots, s$) and *singly diagonally implicit* (SDIRK) with $a_{ii} = \gamma > 0$, for all $i = 2, \dots, s$ and $a_{11} = 0$. Stiffly accurate guarantees that an A-stable scheme is also L-stable (see [11]), and the SDIRK assumption here is used to simplified the analysis. Moreover, the implicit part of this scheme differs from the classical SDIRK one because $a_{ii} = 0$. In [7] these schemes are called *explicit* singly diagonally implicit R-K schemes (ESDIRK). A consequence to set $a_{11} = 0$ is that we have the possibility to guarantee a higher stage-order for ESDIRK scheme than the classical one for a SDIRK scheme, i.e., 1. For instance, a stage order of two for the scheme, i.e., $C(2)$ for every i-stage, is obtained by imposing the assumption 1 for $i = 2, .., s-1$ with $k = 2$. We remark that if the scheme is stiffly accurate condition 1 for $i = s$ and $k = 2$ is equivalent to the second condition in 3.

Moreover, the additional order conditions (6), (7) can be simplified using 2 and 1 written in the form

$$\sum_{j=2}^{i} \hat{\omega}_{ij} c_j^k = k c_i^{k-1} \quad \text{for} \quad i = 2, \dots, s \quad \text{and} \quad k = 1, 2, \tag{8}$$

where $\hat{\omega}_{i,j}$ are the elements of the inverse of $\hat{A}$.

Now, we have already combined the properties of the type CK IMEX scheme, order conditions (6), (7), and assumptions 0, 1, 2, 3, 4, but we have to require that other conditions are satisfied.

We first impose a stability condition on the implicit part of the scheme. Let us denote by $R(z)$ the stability function of the implicit part of the scheme. $R(z)$ is the numerical solution obtained after one time step Δt by applying a R-K scheme to the equation

$$y' = \lambda y, \quad y(0) = 1, \tag{9}$$

with $\lambda \in C$ and $z = \lambda \Delta t$.

We know that a standard SDIRK scheme with nonsingular A matrix that satisfies the condition $a_{si} = b_i$ for all i has $R(\infty) = 0$ and this makes an A-stable scheme L-stable [11].

However, for IMEX R-K schemes of type CK with $a_{11} = 0$, where the matrix A has the structure given in definition 2, the only stiffly accurate condition is not enough to guarantee that $R(\infty) = 0$, and, then an additional condition is required.

LEMMA 3. *If*

$$-e_s^T \hat{A}^{-1} a = \sum_{j\geq 2} \hat{\omega}_{sj} a_{j1} = 0, \tag{10}$$

then $R(\infty) = 0$, *where* $e_s = (0, \ldots, 0, 1)^T$ *and the elements* $\hat{\omega}_{ij}$ *are the elements of the inverse of* $\hat{A}$.

In order to obtain this result, we apply one step of the implicit part to (9), which yields for the internal stages

$$Y = \mathbf{1}_s + zAY \tag{11}$$

with $\mathbf{1}_s = (1, \ldots, 1)^T$. Now, if we set $Y = (Y_1, U^T)$, the expression (11) becomes

$$\begin{aligned} &Y_1 = 1, \\ &U = \mathbf{1}_{s-1} + za + z\hat{A}U, \end{aligned} \tag{12}$$

where $a = (a_{21}, \ldots, a_{s1})^T$. Therefore, we get

$$U = (I - z\hat{A})^{-1}(\mathbf{1}_{s-1} + za). \tag{13}$$

Now, if the scheme is stiffly accurate, the numerical solution is equal to the last internal stage of the scheme and, therefore, when $z \to \infty$, $R(\infty) = -e_s^T \hat{A}^{-1} a$, assumed $\hat{A}$ invertible. Then we can state that if the type CK scheme is stiffly accurate and the condition (10) is satisfied, we get $R(\infty) = 0$. However, thus is not sufficient for the L-stability because the implicit part of the type CK scheme should also be A-stable.

If we want this implicit part of the type CK schemes to be A-stable, we have to consider again the stability function $R(z)$. We know that the stability function $R(z)$ for a SDIRK scheme has the following form ([11], Sect. IV.6)

$$R(z) = \frac{P(z)}{Q(z)}, \tag{14}$$

where $Q(0) = 1$ and both $P(z)$ and $Q(z)$ are polynomials of degree at most s. For ESDIRK schemes with $a_{11} = 0$ and $a_{22} = \cdots = a_{ss} = \gamma$, we have $R(z) = \frac{P(z)}{(1-\gamma z)^{s-1}}$. In particular, $R(\infty) = 0$ implies that the degree of $P(z)$ has to be less than $s - 1$; consequently, the polynomial $P(z)$ has degree at most $s - 2$, and if the order of the scheme is at least $p \geq s - 2$, it follows from $(1 - z\gamma)^{s-1} e^z = P(z) + Cz^{s-1} + \cdots$ that the coefficients of $P(z)$ are uniquely determined in terms of γ and we have

$$P(z) = (-1)^{s-1} \sum_{j=0}^{s-2} L_{s-1}^{(s-1-j)}(\frac{1}{\gamma})(\gamma z)^j \tag{15}$$

with $C = (-1)^{s-1} L_{s-1}(1/\gamma)\gamma^{s-1}$, where

$$L_{s-1}(x) = \sum_{j=0}^{s-1} (-1)^j \binom{s-1}{j} \frac{x^j}{j!}, \tag{16}$$

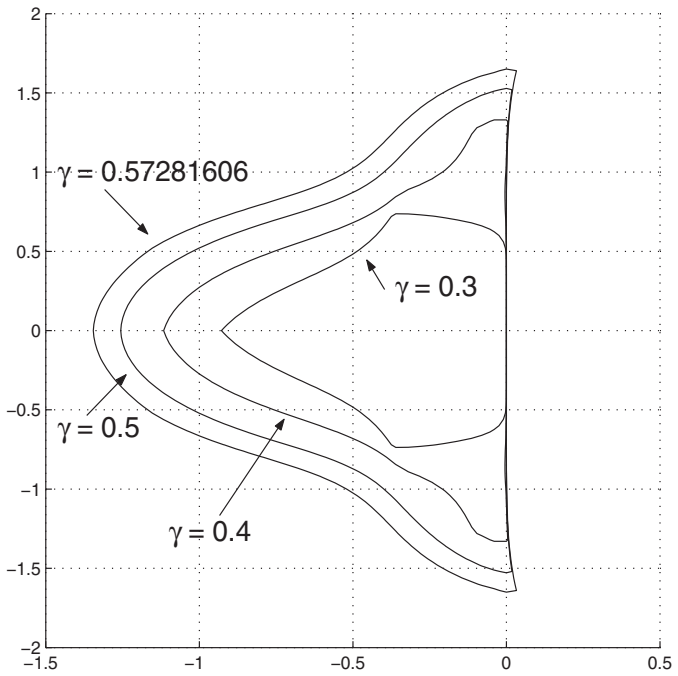

FIG. 1. *Stability domains for increasing values of* γ *in the interval* (25).

is the s-degree Laguerre polynomial and $L_s^{(k)}$ denotes the kth derivative (see [11], Sect. IV.6). Now, provided that $\gamma > 0$, it follows that the stability function (14) is analytic in C^- and we obtain A-stability if there is stability in the imaginary axis (I-stability, see [11]). It is equivalent to the fact that the polynomial

$$E(y) = |Q(iy)|^2 - |P(0, iy)|^2 \tag{17}$$

satisfies $E(y) \geq 0$ for all $y \in R$. Then, if $E(y) \geq 0$ for all y the implicit part of the type CK IMEX R-K scheme is A-stable, and we can state that the IMEX R-K scheme is L-stable.

We remark that it is straightforward to provide that the stability function $R(\tilde{z}, z)$ for a IMEX R-K scheme of type CK has the form

$$R(\tilde{z}, z) = \frac{P(\tilde{z}, z)}{(1 - \gamma z)^{s-1}}, \qquad P(\tilde{z}, z) = (-1)^{s-1} \sum_{j=0}^{s-2} L_{s-1}^{(s-1-j)}(\tilde{z}, \frac{1}{\gamma})(\gamma z)^j \tag{18}$$

with error constant $C(\tilde{z}) = (-1)^{s-1} L_{s-1}(\tilde{z}, \frac{1}{\gamma})(\gamma)^{s-1}$, where

$$L_{s-1}^{(k)}\left(\tilde{z}, \frac{1}{\gamma}\right) = \sum_{j=k}^{s-1} (-1)^j \binom{s-1}{j} \frac{(1/\gamma)^{j-k}}{(j-k)!} \tilde{P}^{(j-k)}(\tilde{z}), \tag{19}$$

and $\tilde{P}(\tilde{z}) = 1 + \tilde{z}/1! + \tilde{z}^2/2! + \cdots + \tilde{z}^s/s!$. $L_{s-1}^{(k)}$ denotes the kth derivative.

Several stability domains for IMEX R-K schemes of type CK are given in Figure 1 for different values of γ.

Finally, a careful consideration follows from the proof of the Theorem 6.2 in [2] about the type CK. In fact, we require that the formula

$$\sum_{j=2}^{s} b_i \hat{\omega}_{i,j} \hat{\omega}_{j,2} = 0, \tag{20}$$

for $i = 2, \dots, s$, is included among the assumptions in order to have at least second-order accuracy in time in the algebraic variable. This is due to the fact that without (20) the accuracy of the stiffness (or algebraic) variable drops to first order. Hence an extra care must be taken to properly obtain a third-order type CK IMEX R-K scheme. The derivation of condition (20) is reported in Appendix 1.

Remark 2. It is noteworthy that if the method is stiffly accurate the assumption $\sum_{i,j} b_i \hat{\omega}_{ij} \hat{\omega}_{j2} = 0$ is equivalent to $\hat{\omega}_{s2} = 0$ and in particular it follows from (10) that $\sum_{j \geq 3} \hat{\omega}_{sj} a_{j1} = 0$.

Now, in the construction of a third-order type CK IMEX R-K scheme, we first search schemes with four internal stages. In order to be able to design such a method we require to solve the system of equations from 1 to 4 with the condition (10) and the assumption (20). Unfortunately there are not third-order, four-stage schemes satisfying the previous assumptions. This is formally stated in the following.

PROPOSITION 1. *Consider IMEX Runge–Kutta scheme of type CK, stiffly accurate (i.e., $a_{si} = b_i$ for $i = 1, \dots, s$) in the implicit scheme, with $b_i = \tilde{b}_i$, $c_i = \tilde{c}_i$ for all i and $a_{11} = 0$. Then there exist no third-order, four-stage schemes satisfying assumptions 1–4 with condition* (10) *and assumption* (20).

Proof. By remark 2 for $s = 4$ we get $\hat{\omega}_{42} = 0$ and $\hat{\omega}_{43} a_{31} + \hat{\omega}_{44} b_1 = 0$. Then, by assumptions 1 with $k = 2$, 3, and 4 the resulting reduced system of equations can be explicitly written as

$$\frac{a_{32} b_3}{\gamma^3} = 0, \tag{21a}$$

$$\frac{b_3 a_{31}}{\gamma^2} + \frac{b_1}{\gamma} = 0, \tag{21b}$$

$$c_2 = 2\gamma, \tag{21c}$$

$$2a_{32}\gamma + \gamma c_3 = c_3^2/2, \tag{21d}$$

$$b_3 c_3 + \gamma = 1/2, \tag{21e}$$

$$b_3 c_3^2 + \gamma = 1/3 \tag{21f}$$

with $a_{21} = c_2 - \gamma$ and $a_{31} = c_3 - a_{32} - \gamma$ by 1 with $k = 1$, and $b_1 = 1 - b_3 - \gamma$. In order to state the proposition one can prove easily by trivial computational that it is possible to obtain several contradictions. There is a total of five coefficients to determine: γ, c_2, c_3, b_3, and a_{32}. We require that the conditions (21a), (21b) are satisfied simultaneously and as consequence from (21a) we consider $a_{32} = 0$ with $a_{31} = c_3 - \gamma$.

The resulting reduced system (21) can be computed as follows:

(a)

$$c_3 = \frac{1/3 - \gamma}{1/2 - \gamma}, \quad b_3 = \frac{1/2 - \gamma}{c_3} \tag{22}$$

from (21e) and (21f);

(b) by (21b) we compute γ and then b_1.

(c) Using (21d) we get $c_3 = 0$ or $c_3 = 2\gamma$ and from (a) we have $c_3 = (1/3 - \gamma)/(1/2 - \gamma)$.

Hence, if we substitute the values of γ computed above, we obtain contradictions about the values of c_3. Notice that if $c_3 = 0$, b_3 is not defined. Furthermore, it can be verified again by trivial computations that by choosing $b_3 = 0$ in (21a) one obtains other contradictions in (a). ∎

Then, a third-order type CK IMEX R-K scheme is designed in five stages, i.e., $s = 5$. Making use of conditions 0, 4, and 1, 2 with $k = 1$, the Butcher *tableau* of a IMEX R-K scheme of type CK becomes

$$\begin{array}{c|ccccc|ccccc}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
c_2 & c_2 & 0 & 0 & 0 & 0 & c_2-\gamma & \gamma & 0 & 0 & 0 \\
c_3 & c_3-\tilde{a}_{32} & \tilde{a}_{32} & 0 & 0 & 0 & c_3-a_{32}-\gamma & a_{32} & \gamma & 0 & 0 \\
c_4 & c_4-\tilde{a}_{42}-\tilde{a}_{43} & \tilde{a}_{42} & \tilde{a}_{43} & 0 & 0 & c_4-a_{42}-a_{43} & a_{42}-\gamma & a_{43} & \gamma & 0 \\
1 & b_1 & 0 & b_3 & b_4 & \gamma & b_1 & 0 & b_3 & b_4 & \gamma \\
\hline
 & b_1 & 0 & b_3 & b_4 & \gamma & b_1 & 0 & b_3 & b_4 & \gamma
\end{array}.$$

Then, there is a total of 16 coefficients to be determined: the 10 coefficients, c_2, c_3, c_4, a_{32}, a_{42}, a_{43}, b_1, b_2, b_3, b_4, γ for the implicit part and 6 coefficients $\tilde{a}_{32}$, $\tilde{a}_{42}$, $\tilde{a}_{43}$, $\tilde{a}_{52}$, $\tilde{a}_{53}$, $\tilde{a}_{54}$ for the explicit one.

In order to determine the 10 coefficients of the implicit scheme, we use the following conditions.

1. Assumption 3.
$$\begin{aligned} b_1 &= 1 - b_3 - b_4 - \gamma, \\ b_3c_3 + b_4c_4 + \gamma &= 1/2, \\ b_3c_3^2 + b_4c_4^2 + \gamma &= 1/3. \end{aligned} \tag{23a}$$
2. Assumption 2 for $i = 2, 3, 4$ with $k = 2$,
$$\begin{aligned} \gamma c_2 &= c_2^2/2, \\ a_{32}c_2 + \gamma c_3 &= c_3^2/2, \\ a_{42}c_2 + a_{43}c_3 + \gamma c_4 &= c_4^2/2. \end{aligned} \tag{23b}$$
3. By remark 2 condition (10) is $\sum_{j\geq 3} \hat{\omega}_{sj}a_{j1} = 0$, i.e.,
$$\frac{b_4a_{43}a_{31}}{\gamma^3} - \frac{b_3a_{31}}{\gamma^2} - \frac{b_4a_{41}}{\gamma^2} + \frac{b_1}{\gamma} = 0. \tag{23c}$$
with $a_{31} = c_3 - a_{32} - \gamma$ and $a_{41} = c_4 - a_{42} - a_{43} - \gamma$.
4. Analogously considering remark 2, additional condition (20) for $s = 5$ is $\hat{\omega}_{52} = 0$, i.e.,
$$\frac{b_3a_{32}}{\gamma^3} + \frac{b_4a_{42}}{\gamma^3} - \frac{b_4a_{43}a_{32}}{\gamma^4} = 0. \tag{23d}$$

Now, from the first and second formula in (23b) we get $c_2 = 2\gamma$ and $a_{32} = \frac{c_3(c_3-2\gamma)}{2c_2}$. In particular, we have $a_{21} = c_2 - \gamma = \gamma$. Using the third formula in (23b), the second and third one in (23a) and (23d), we compute b_3, b_4, a_{42}, and a_{43} as functions of c_3 and c_4 and b_1 by (23a).

Finally, by using the condition (23c) and substituting the quantities computed above b_1, a_{i1}, for $i = 3, 4$ and a_{43}, by algebraic manipulations, we obtain

$$c_3 = \frac{2(6\gamma^2 - 6\gamma + 1)}{3(2\gamma^2 - 4\gamma + 1)} \tag{24}$$

in function of γ. Then we have only two free parameters γ and c_4.

Now, in order to determine the optimal values for a good choice of the parameter γ, we, explicitly, give the formula for the polynomial (17), with $s = 5$. Then, we have with $s - 1 = 4$ and order $p \geq s - 2 = 3$,

$$E(y) = y^4(1/12 - 4\gamma/3 + 6\gamma^2 - 8\gamma^3 + 2\gamma^4) \\ + y^6(-1/36 + 2\gamma/3 - 6\gamma^2 + 76\gamma^3/3 - 52\gamma^4 + 48\gamma^5 - 12\gamma^6) + y^8\gamma^8.$$

Therefore, A-(and hence L-)stability means that all the coefficients of $E(y)$ must be nonnegative and, then, we obtain the following interval (see [11], Table 6.4, p. 98):

$$\gamma_1 \leq \gamma \leq \gamma_2, \tag{25}$$

where $\gamma_1 = 0.22364780\ldots$ and $\gamma_2 = 0.57281606\ldots$. Notice that for different values of γ in the interval, c_4 has been computed to minimize the fourth-order error constant ([10]).

We now turn our attention to the evaluation of the six coefficients of the explicit part of the scheme. Therefore, assumption 2 for $i = 3, 4, 5$, with $k = 2$ and the technical condition 4 give

$$\begin{aligned} &\tilde{a}_{32}c_2 = c_3^2/2, \\ &\tilde{a}_{42}c_2 + \tilde{a}_{43}c_3 = c_4^2/2, \\ &\tilde{a}_{52}c_2 + \tilde{a}_{53}c_3 + \tilde{a}_{54}c_4 = 1/2. \end{aligned} \tag{26}$$

Then, by (26) and $c_2 = 2\gamma$ we get

$$\begin{aligned} &\tilde{a}_{32} = c_3^2/(4\gamma), \\ &\tilde{a}_{43} = c_4^2/(2c_3) - 2\gamma\tilde{a}_{42}/c_3, \\ &\tilde{a}_{54} = 1/(2c_4) - (2\gamma\tilde{a}_{52}/c_4 + \tilde{a}_{53}c_3/c_4). \end{aligned} \tag{27}$$

Thus, we have only to determine the following coefficients: $\tilde{a}_{42}$, $\tilde{a}_{43}$, $\tilde{a}_{52}$, and $\tilde{a}_{53}$.

In order to compute these coefficients we make the following assumption:

$$\sum_{i,j,k} b_i\tilde{a}_{ij}\tilde{a}_{jk}c_k = 1/24, \tag{28}$$

which is one of the order conditions for the explicit scheme. When applied (28) to a linear system with constant coefficients, such condition provides fourth-order accuracy.

It is noteworthy that, considering the assumption 2 with $k = 2$, (28) implies $\sum_{i,j} b_i\tilde{a}_{ij}c_j^2 = 1/12$, if and only if $\sum_{i=3}^{s} b_i\tilde{a}_{i2} = 0$. Then, we here consider the following equation:

$$b_3\tilde{a}_{32} + b_4\tilde{a}_{42} + \gamma\tilde{a}_{52} = 0. \tag{29}$$

Hence if we choose $\tilde{a}_{42} = 0$ we can immediately compute $\tilde{a}_{52}$ by (29), $\tilde{a}_{43}$ by (27), and $\tilde{a}_{53}$ by $\sum_{i,j} b_i\tilde{a}_{ij}c_j^2 = 1/12$. Finally, we get $\tilde{a}_{31}$, $\tilde{a}_{41}$, $\tilde{a}_{51}$.

5. Test problems. In this section we investigate numerically the convergence rate for a wide range of parameter ε. To this aim we apply to several test problems the IMEX R-K scheme introduced in section 3 and the new schemes introduced in section 4.

Numerical convergence rate is calculated by the formula

$$p = log_2(E_{\Delta t_1}/E_{\Delta t_2}), \tag{30}$$

where $E_{\Delta t_1}$ and $E_{\Delta t_2}$ are the global errors computed with step Δt_1 and $\Delta t_2 = \Delta t_1/2$. The convergence rates for the different schemes are reported in the tables.

Before introducing some numerical examples, we start to suppose that the initial data of our test problems lie on a suitable manifold that allows smooth solutions even in the limit of infinite stiffness, and that the step size $\Delta t \gg \varepsilon$. In fact, arbitrary initial data introduce in the solution a fast transient. One possibility to overcome this difficulty is simply to ensure that the numerical scheme resolves the transient phase by taking time step $\Delta t \ll \varepsilon$ in the first few steps. Then, the following results are obtained assuming that the transient phase is over. A simple way to avoid the presence of the so-called initial layer is to suitably choose the initial data in such a way that the initial layer does not form. This is obtained by using the so-called *well-prepared* data, as outlined below.

Consider a linear system with stiff relaxation source term [19]

$$\partial_t u + \partial_x v = 0 \tag{31a}$$

$$\partial_t v + \partial_x u = (au - v)/\varepsilon, \tag{31b}$$

with $\varepsilon > 0$ and a constant. An expansion in ε can be performed directly at the level of the system (31), by using the so-called Chapman–Enskog expansion.

Consider a formal expansion of the solution to system (31) in the form

$$\begin{aligned} u &= u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \cdots \\ v &= v_0 + \varepsilon v_1 + \varepsilon^2 v_2 + \cdots \end{aligned} \tag{32}$$

and insert it in (31).

The leading term approximation to (31) is $v_0 = au_0$ and $\partial_t u_0 + a\partial_x u_0 = 0$. In fact, the second equation (31b), to order ε^{-1}, gives

$$v_0 = au_0,$$

and using this relation in equation (31a) to order ε^0 gives

$$\partial_t u_0 + a\partial_x u_0 = 0.$$

This equation is formally obtained as $\varepsilon \to 0$ and is called *relaxed* equation.

The first-order correction to the leading term approximation is obtained as follows. The equation (31b), to zero the order in ε, gives

$$\partial_t v_0 + \partial_x u_0 = au_1 - v_1,$$

which gives

$$v_1 = au_1 - (1 - a^2)\partial_x u_0.$$

The equation (31a), to first order, gives

$$\partial_t u_1 + a\partial_x u_1 = \varepsilon \partial_x((1 - a^2)\partial_x u_0).$$

By keeping first-order terms in the expansion (32) and neglecting second and higher order terms, one has

$$u = u_0 + \varepsilon u_1, \qquad v = v_0 + \varepsilon v_1$$

and obtain

$$\begin{aligned} &v = au - (1-a^2)\partial_x u, \\ &\partial_t u + a\partial_x u = \varepsilon \partial_x\big((1-a^2)\partial_x u\big). \end{aligned} \tag{33}$$

The second equation in (33) is a convection-diffusion equation with viscosity coefficient $\nu = \varepsilon(1-a^2)$. This equation is dissipative if $\nu \geq 0$, i.e., $|a| \leq 1$ (*subcharacteristc condition* of Liu [15] for (31)).

Motivated by this analysis we perform an accuracy test for the test problems considering *well-prepared* initial data. Then, concerning the test problem (31), we have considered a periodic smooth solution and well-prepared initial data by $u(x,0) = \sin(2\pi x)$ and $v(x,0) = v_0(x,0) + \epsilon v_1(x,0)$, where $v_0(x,0) = au(x,0)$ and $v_1(x,0) = (a^2-1)\partial_x u_0$. We set $a = 0.5$ and the final time is $t = 0.2$. The system is integrated for $x \in [0,2]$.

The second test problem is the *Broadwell model* equations:

$$\begin{aligned} &\partial_t \rho + \partial_x m = 0, \\ &\partial_t m + \partial_x z = 0, \\ &\partial_t z + \partial_x m = \tfrac{1}{\varepsilon}(\rho^2 + m^2 - 2\rho z). \end{aligned} \tag{34}$$

Here ε represents the mean free path of particles. The dynamical variables ρ and m are the density and the momentum, respectively, while z represents the flux of the momentum. As $\varepsilon \to 0$, z is given by a local Maxwellian distribution

$$z = z_E(\rho, m) = \frac{1}{2\rho}(\rho^2 + m^2), \tag{35}$$

and we are in the fluid dynamic limit, satisfying the equation

$$\begin{aligned} &\partial_t \rho + \partial_x(\rho u) = 0, \\ &\partial_t(\rho u) + \partial_x(\tfrac{1}{2\rho}(\rho^2 + m^2)) = 0, \end{aligned}$$

with $u = m/\rho$. We have considered a periodic smooth solution with initial *well-prepared* data given by

$$\begin{aligned} &\rho(x,0) = \rho_0(x),\ u(x,0) = u_0(x),\ m(x,0) = m_0(x) = \rho_0(x)u_0(x), \\ &z(x,0) = z_E(\rho_0(x), m_0(x)) + \varepsilon z_1(\rho_0(x), m_0(x)). \end{aligned}$$

where $\rho_0(x) = 1 + a_\rho \sin\left(\frac{2\pi x}{L}\right)$, $u_0(x) = \frac{1}{2} + a_u \sin\left(\frac{2\pi x}{L}\right)$, and

$$z_1(\rho_0, m_0) = \tfrac{-H(\rho_0, m_0)}{\rho_0}$$

with $H(\rho_0, m_0) = ((1 - \partial_\rho z_E + (\partial_m z_E)^2)\partial_x m_0 + (\partial_\rho z_E \partial_m z_E)\partial_x \rho_0)$.

Periodic boundary conditions are imposed. In our computations we used the parameters $a_\rho = 0.3$, $a_u = 0.1$, $L = 20$ and we integrate the equation for $t \in [0, 30]$.

Finally, we present other numerical results obtained concerning situation in which hyperbolic system with relaxation plays a major role in applications. In fact, we consider as third test problem that was used by Shi Jin in [13]:

$$\begin{aligned} &\partial_t h + \partial_x w = 0 \\ &\partial_t w + \partial_x(h + 0.5h^2) = -\tfrac{1}{\varepsilon}(w - 0.5h^2), \end{aligned} \tag{36}$$

with the *well-prepared* initial data given by $h(0,x) = 1 + 0.2\sin(8\pi x)$ and $w(0,x) = w_0 + \varepsilon w_1$ with $w_0 = f(h(0,x)) = 0.5h^2(0,x)$ and $w_1 = (f'(h(0,x)) - p'(h(0,x)))\partial_x h(0,x)$ where $p(h) = (h + 0.5h^2)$. The results have been obtained for fixed Δx.

TABLE 1
Convergence rate for v in L_∞-norm.

Schemes	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-5}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$
ARS	2.982	2.970	2.471	2.386	2.041	2.003	1.999
MARS	2.976	2.741	1.905	1.142	2.889	2.980	2.997
ARK3	3.012	2.833	2.443	2.108	2.012	2.002	2.002
MARK3	2.989	2.740	1.905	1.144	2.118	2.980	2.996

Our test problems are computed with coarse grids ($\Delta x, \Delta t \gg \varepsilon$) that do not resolve the small scales, and high accuracy in space is obtained by finite difference discretization with weighted essentially nonoscillatory (WENO) reconstruction [20], [17].

Discussion. Concerning our numerical experiments, we have constructed several IMEX R-K schemes of type CK considering different values of γ in interval (25) and we have obtained several schemes that qualitatively have shown a good accuracy behavior. Different values of the parameter γ in (25) with a reasonable large stability domain have been considered, and we report only the most relevant values for the different test problems that guaranteed a good behavior about the accuracy on the whole range of the parameters ε. These values are reported in the different tables presented in this article. In particular, we have chosen the values of $\gamma = \gamma_2$ and $\gamma = \tilde{\gamma}$ with $\tilde{\gamma} = 0.43586652150845$ because γ_2 guarantees the largest stability domain corresponding to the stability function (18), (see Figure 1), whereas $\gamma = \tilde{\gamma}$ is the same value used in the IMEX R-K schemes introduced in section 3. All these values are a good choice because the SDIRK scheme is L-stable. We report in the appendix only the values of the coefficients for the schemes with $\gamma = \gamma_2$ and $\gamma = \tilde{\gamma}$. In order to identify the schemes derived in this article we name this new scheme BHR(s, ρ, p), where the triplet (s, ρ, p) characterizes the number s of the stages of the implicit part, ρ the number of the stages of the explicit one, and p the order of the scheme. The coefficients of these schemes are displayed in the Appendix 2.

First we compare MARS(3,4,3) and MARK3(2)4L[2]SA schemes, with the ARS(3,4,3), [1], and ARK3(2)4L[2]SA, [7] (see Table 1). Tables show the corresponding convergence rate in L_∞-norm. Different norms give essentially the same results.

We have obtained an improvement for the convergence of the different stiff components. In fact, for the first problem we have increased the convergence rate for sufficiently stiff parameters ($\varepsilon < 10^{-4}$), namely when $\varepsilon \to 0$. These results show a third-order accuracy for small and large values of ε and note that for intermediate values of the parameter ε ($10^{-4} < \varepsilon < 10^{-2}$) we have a slight deterioration of the accuracy.

In a similar way, it is possible to improve the convergence of variable z in the Broadwell model (see Figures 2 and 3) and for the w variable in the shallow water.

Now we show that the new third order scheme have better uniform accuracy in ε. Investigating the numerical convergence rate of this scheme for a whole range of ε.

As it is evident, from Figures 4 and 5 and from Tables 2 and 3, the BHR(5,5,3) with $\gamma = \tilde{\gamma}$ verifies a third-order accuracy in the whole range of ε. Furthermore, we note that in a neighborhood of $\tilde{\gamma}$ (from $\gamma = 0.43$ to $\gamma = 0.45$) we have obtained several schemes with a good accuracy behavior (Tables 2, 3, and 4). Instead, BHR(5,5,3) schemes, with $\gamma = \gamma_2$ and $\gamma = 0.5$, show for intermediate values of ε a slight deterioration of the accuracy.

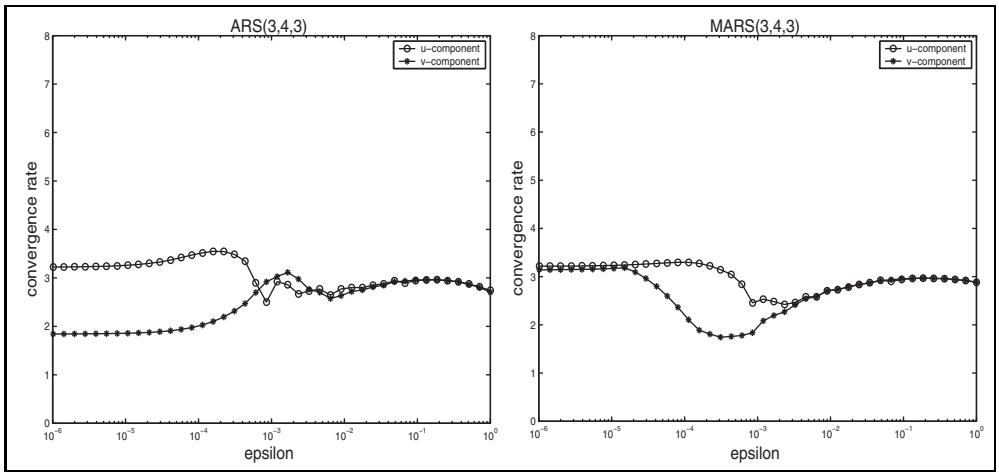

FIG. 2. *Convergence rate vs ε for the density ρ ($\circ$) (differential component) and the flux of the momentum z ($*$) (stiff component). On the left ARS(3,4,3) scheme, and on the right MARS(3,4,3) scheme.*

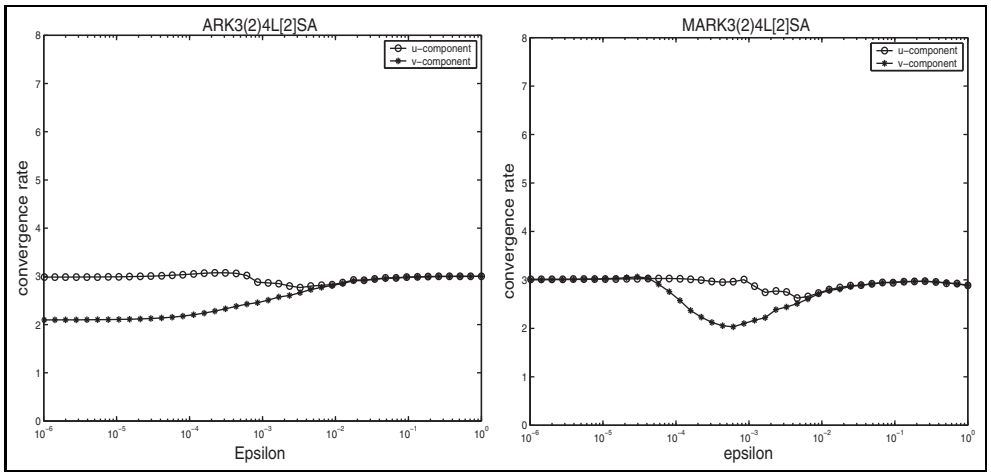

FIG. 3. *Convergence rate vs ε for the density ρ ($\circ$) (differential component) and the flux of the momentum z ($*$) (stiff component). On the left ARK3(2)4L[2]SA, and on the right ARK3(2)4L[2]SA scheme.*

In the rest of the section we show that the lack of accuracy of the schemes for intermediate values of the stiffness parameter is consistent with a theoretical analysis. In order to obtain these results in a general setting, we consider the singular perturbation problem

$$\begin{aligned} y' &= f(y,z), \\ \varepsilon z' &= g(y,z), \end{aligned} \tag{37}$$

with $0 < \varepsilon \ll 1$ where f and g are sufficiently differentiable. We are mainly interested in smooth solutions of (37) which are of the form

$$\begin{aligned} y(t) &= y_0(t) + \varepsilon y_1(t) + \varepsilon^2 y_2(t) + \cdots \\ z(t) &= z_0(t) + \varepsilon z_1(t) + \varepsilon^2 z_2(t) + \cdots, \end{aligned} \tag{38}$$

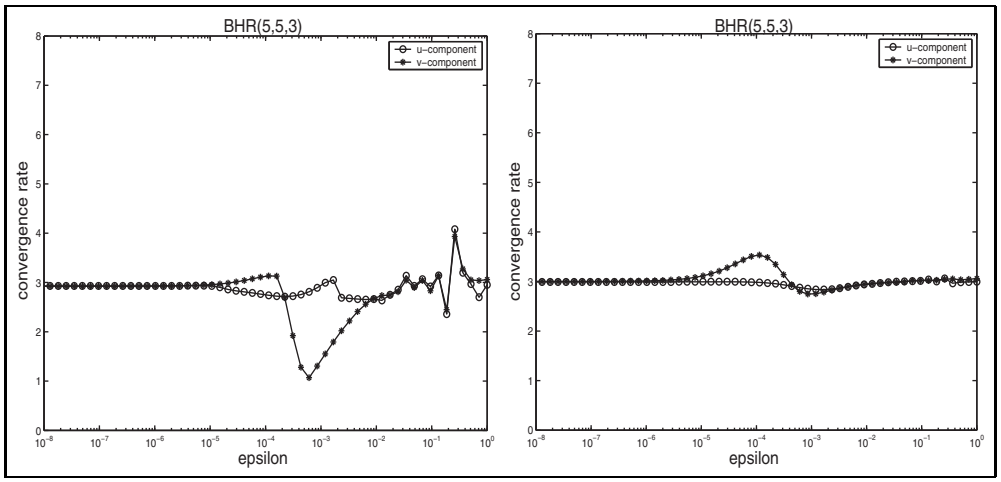

FIG. 4. *Convergence rate vs ε for the u-component ($\circ$), (differential component) and v-component ($*$), (stiff component). On the left BHR(5,5,3) scheme with $\gamma = \gamma_2$, and on the right BHR(5,5,3) scheme with $\gamma = \tilde{\gamma}$.*

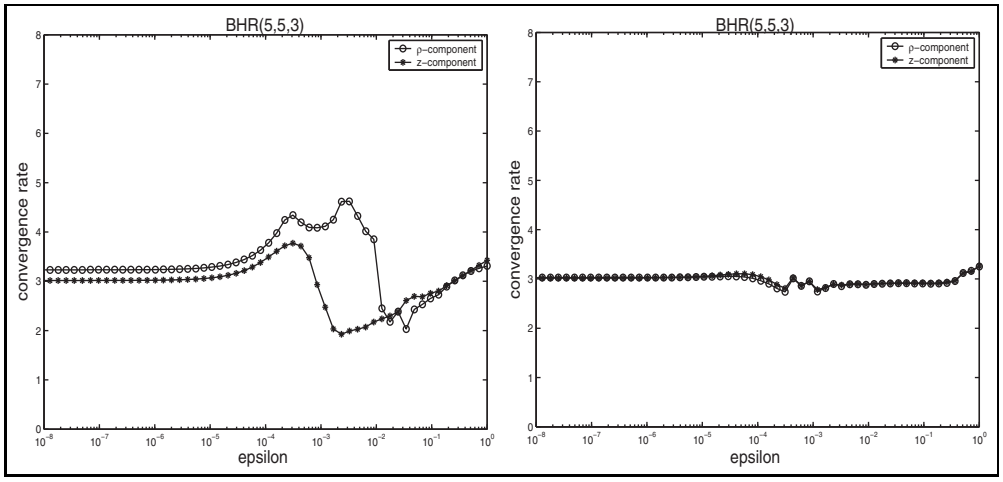

FIG. 5. *Convergence rate vs ε for the density ρ ($\circ$) (differential component) and the flux of the momentum z ($*$) (stiff component). On the left BHR(5,5,3) scheme with $\gamma = \gamma_2$, and on the right BHR(5,5,3) scheme with $\gamma = \tilde{\gamma}$.*

TABLE 2
Convergence rate for v in L_∞-norm for different values of γ.

γ	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$
BHR ($\gamma = 0.43$)	3.130	3.186	3.306	3.363	3.049	3.007	3.007
BHR ($\tilde{\gamma}$)	3.657	3.352	3.003	3.407	3.053	3.000	3.000
BHR ($\gamma = 0.44$)	3.112	3.101	3.132	3.371	3.051	3.008	3.007
BHR ($\gamma = 0.45$)	2.857	3.022	2.451	3.379	3.053	3.009	3.009
BHR ($\gamma = 0.5$)	3.257	2.882	1.596	3.370	3.062	3.009	3.008
BHR (γ_2)	3.070	2.678	1.405	3.096	3.010	3.000	3.000

TABLE 3
Convergence rate for z-component in L_∞-norm for different values of γ.

Schemes	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$
BHR ($\gamma = 0.43$)	3.055	2.916	2.670	2.776	3.218	3.019	3.017
BHR ($\gamma = 0.44$)	3.050	2.921	2.780	3.539	3.200	3.019	3.016
BHR ($\gamma = 0.45$)	2.896	2.975	2.520	2.911	3.172	2.986	2.984
BHR ($\gamma = 0.5$)	3.177	3.000	2.708	2.388	3.069	3.064	3.063

TABLE 4
Convergence rate for w in L_∞-norm for different values of γ.

Schemes	$\varepsilon = 1$	$\varepsilon = 10^{-1}$	$\varepsilon = 10^{-2}$	$\varepsilon = 10^{-3}$	$\varepsilon = 10^{-4}$	$\varepsilon = 10^{-6}$	$\varepsilon = 10^{-8}$
ARS	3.582	3.117	2.962	2.858	3.340	2.140	2.113
ARK3	2.963	3.013	2.982	2.860	2.482	2.060	2.044
BHR ($\tilde{\gamma}$)	3.028	3.026	2.998	3.072	3.525	3.210	3.185
BHR ($\gamma = 0.45$)	3.119	2.994	2.930	3.117	3.146	3.211	3.187
BHR ($\gamma = 0.5$)	3.186	3.024	2.970	2.730	2.070	3.036	3.012
BHR (γ_2)	3.545	3.317	2.930	2.542	2.165	3.071	3.050

where $y_i(t)$ and $z_i(t)$ are ε-independent functions. We suppose that the initial values lie on a suitable manifold that allows smooth solutions even in the limit of infinite stiffness. We remark that a sequence of differential algebraic systems arises in the study of problem (37). In [2], Boscarino showed that the problem (37) gives us the possibility of studying the dependence of the global error on the stiffness parameter ε, valid as $\varepsilon < \Delta t$. In fact, we may expand the global error in power of ε and show that the coefficients of the global error are the global errors of the IMEX R-K schemes applied to a differential algebraic system of different index.

Concerning (37), the global error satisfies

$$\begin{aligned} \Delta y_n &= \Delta y_n^0 + \varepsilon \Delta y_n^1 + \varepsilon^2 \Delta y_n^2 + \mathcal{O}(\varepsilon^3) \\ \Delta z_n &= \Delta z_n^0 + \varepsilon \Delta z_n^1 + \varepsilon^2 \Delta z_n^2 + \mathcal{O}(\varepsilon^3/\Delta t) \end{aligned} \tag{39}$$

with $\Delta y_n = y_n - y(t_n)$ and $\Delta z_n = z_n - z(t_n)$ where $\Delta y_n^0 = y_n^0 - y_0(t_n)$, $\Delta z_n^0 = z_n^0 - z_0(t_n), \dots$ are the global errors of the scheme applied to a differential-algebraic system of different index whereas $\mathcal{O}(\varepsilon^3)$, $\mathcal{O}(\varepsilon^3/\Delta t)$ are the estimates on the remainder.

By imposing additional order conditions obtained by matching the exact solution and the numerical solution at various orders in ε, we produce an error which is more uniform in ε. To do that, in [3], we imposed that the IMEX R-K scheme is of a given order p for the hierarchy of differential-algebraic systems (e.g., indexes 1 and 2).

In order to explain the lack of accuracy, we start observing that, for $\varepsilon > \Delta t$, classical analysis can be used to estimate the error. In fact, the classical global error is of order $\mathcal{O}(\Delta t/\varepsilon)^p$. As both estimates about the global error have to be satisfied, we can write, for example, for the z-component

$$|error| < min\left(C\left(\frac{\Delta t}{\varepsilon}\right)^p, \Delta z_n^0 + \varepsilon \Delta z_n^1 + \varepsilon^2 \Delta z_n^2 + \mathcal{O}(\varepsilon^3/\Delta t) \right) \tag{40}$$

(see Figure 6).

We put the order of the scheme $p = 3$. Now, the balance in the arguments of the right-hand side of equation (40) gives that:

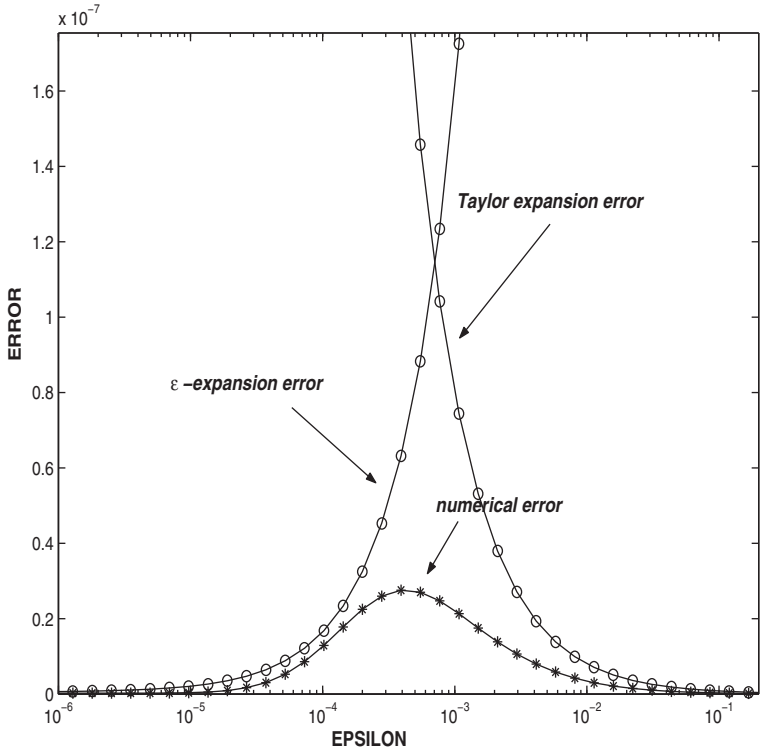

FIG. 6. *Numerical global error (∗) for z-component (stiff component) using BHR*(5,5,3) *IMEX R-K scheme and (∘) theoretical global error.*

1. about the algebraic variable z, the dominant term, in the worse case when $\Delta t \lesssim \varepsilon$, is $\mathcal{O}(\varepsilon^3/\Delta t)$ and then we have

$$C\left(\tfrac{\Delta t}{\varepsilon}\right)^3 \approx \mathcal{O}(\varepsilon^3/\Delta t), \tag{41}$$

which gives $\varepsilon \approx \Delta t^{\frac{2}{3}}$, so to get the maximum error $\mathcal{O}(\Delta t)$.

2. In a similar way, for the differential variable y, the dominant term is $\mathcal{O}(\varepsilon^3)$, which gives a maximum error $\mathcal{O}(\Delta t)^{\frac{3}{2}}$. These results state that the third-order schemes can degrade up to first order in the worst case, for example, for the z-component (such case is observed in Figures 3 and 4). Notice, however, that this is a lower bound for the convergence rate. In many case a better behavior is observed. For example, for BHR(5,5,3) scheme, in a neighborhood of $\gamma = \tilde{\gamma}$, the error is more uniform.

6. Conclusion. IMEX R-K schemes are introduced for application to hyperbolic systems with relaxation. Additional conditions up to the third order are provided to improve the accuracy of several schemes with respect to the stiffness parameter ε. For instance, MARK3(2)4L[2]SA and MARS(3,4,3) schemes are slightly more efficient in the limit of $\varepsilon = 0$ than the ARK3(2)4L[2]SA and ARS(3,4,3) ones. Construction of a third-order accurate IMEX R-K scheme is presented and numerical tests on several hyperbolic systems with relaxation term reveal better behaviors for this new scheme called BHR(5,5,3) than other ones presented in the literature. As it is evident from the figures and tables, the BHR(5,5,3) schemes with $\gamma = \gamma_2$ is third order accurate for

small and large values of ε, and there is a slight deterioration of the accuracy in the intermediate regime. This lack of the accuracy for intermediate values of the stiffness has been explained through a theoretical analysis.

A scalar stability analysis has also shown that the value of $\gamma = \gamma_2$ has the largest stability domain with respect to the other values in the interval (25). However, numerical stability is not the central issue here. We remark that a stability analysis for IMEX schemes applied to stiff systems has been performed by Higueras et al., see [9], using contractivity and monotonicity properties and concept of algebraic stability for additive R-K methods. Finally, concerning the asymptotic limit as $\varepsilon \to 0$, this new scheme becomes a good (high order) approximation of the equilibrium equations; i.e., it has the correct asymptotic limit.

Appendix 1. To derive the condition (20), here we just outline the main ideas. To this end, the condition (20) is an immediate consequence of Theorem 6.2 in [2]. We denoted the difference to the exact solution values by

$$\begin{array}{cc} \Delta y_n^\nu = y_n^\nu - y_\nu(t_n), & \Delta z_n^\nu = z_n^\nu - z_\nu(t_n), \\ \Delta Y_{ni}^\nu = Y_{ni}^\nu - y_\nu(t_n + \tilde{c}_i h), & \Delta Z_{ni}^\nu = Z_{ni}^\nu - Z_\nu(t_n + c_i h), \\ \Delta k_{ni}^\nu = k_{ni}^\nu - y_\nu^{'}(t_n + \tilde{c}_i h), & \Delta \ell_{ni}^\nu = \ell_{ni}^\nu - z_\nu^{'}(t_n + c_i h). \end{array} \tag{42}$$

$\nu = 0$ or $\nu = 1$ in (42) correspond, respectively, to the numerical solution of the index 1 or index 2 problem, (see [11, 2]).

Furthermore, from Theorem 5.2 in [2], we deduced that

$$\begin{array}{lll} \Delta y_n^0 = \mathcal{O}(h^{\tilde{q}+2} + h^p), & \Delta Y_{ni}^0 = \mathcal{O}(h^{\tilde{q}_i+1}), & \Delta k_{ni}^0 = \mathcal{O}(h^{\tilde{q}_i+1}), \\ \Delta z_n^0 = \mathcal{O}(h^{\tilde{q}+1} + h^p), & \Delta Z_{ni}^0 = \mathcal{O}(h^{\tilde{q}_i+1}), & \end{array} \tag{43}$$

with $\tilde{q} = \min\{\tilde{q}_s, \tilde{q}_i + 1, i = 2, \ldots, s-1\}$ where $\tilde{q}_i$ the stage order of the ith stage of the explicit part of the IMEX R-K scheme.

Moreover, we obtained that

$$\Delta \ell_{ni}^0 = \mathcal{O}(\|\Delta y_n^1\| + \|\Delta z_n^1\|) + h^{-1}\hat{\omega}_{i2}(\Delta Z_{n2}^0 - \Delta z_n^0) + \mathcal{O}(h^{\tilde{q}_i}) + \mathcal{O}(h^{q_i}), \tag{44}$$

with q_i the stage order of the ith stage of the implicit part, for $i = 2, \ldots, s$. Consequently, evaluated (see [2])

$$\Delta Z_{ni}^1 = C_1 \Delta \ell_{ni}^0 + C_2 \Delta Y_{ni}^1 + \mathcal{O}(h^{\tilde{q}_i+1}) \tag{45}$$

and

$$\Delta Y_{ni}^1 = \Delta y_n^1 + Ch(||\Delta y_n^1|| + ||\Delta z_n^1||) + \mathcal{O}(h^2), \tag{46}$$

substituting the formula (44) and (46) in the expression (45) we get

$$\begin{aligned} \Delta Z_{ni}^1 = \mathcal{O}(\|\Delta y_n^1\| + \|\Delta z_n^1\|) + Ch(\|\Delta y_n^1\| + \|\Delta z_n^1\|) \\ + C_1 h^{-1}\hat{\omega}_{i2}(\Delta Z_{n2}^0 - \Delta z_n^0) + \mathcal{O}(h^2). \end{aligned} \tag{47}$$

By the properties of the type CK IMEX scheme, (see [2]), we get the formula

$$\left\|\Delta z^1_{n+1}\right\| \le |\rho| \left\|\Delta z^1_n\right\| + \sum_{i,j} |b_i \hat{\omega}_{ij}| \left\|\Delta Z^1_{nj}\right\| + \mathcal{O}(h^{q_i+1})$$

with $|\rho| = |R(\infty)| < 1$. Then, substituting (47), it follows that

$$\left\|\Delta z^1_{n+1}\right\| \le \left\|\Delta y^1_n\right\| + (|\rho| + Ch) \left\|\Delta z^1_n\right\| + h^{-1} C_1 \sum_{i,j} |b_i \hat{\omega}_{ij} \hat{\omega}_{i2}| \left\|\Delta Z^0_{n2} - \Delta z^0_n\right\| + \mathcal{O}(h^2)$$

By (43) and by the fact that $\tilde{q}_2 = 1$, we have $\Delta Z^0_{n2} = \mathcal{O}(h^2)$ and $\Delta z^0_n = \mathcal{O}(h^2)$. Then, if (20) holds, we get

$$\left\|\Delta z^1_{n+1}\right\| \le \left\|\Delta y^1_n\right\| + (|\rho| + Ch) \left\|\Delta z^1_n\right\| + \mathcal{O}(h^2).$$

Analogously, a direct estimation of Δy^1_n is given and applying the Lemma 5.2 in [2] we have $\Delta y^1_n = \mathcal{O}(h^2)$ for the differential component and $\Delta z^1_n = \mathcal{O}(h^2)$ for the algebraic component. Again, we can conclude that if (20) is satisfied we have a second order accuracy in time for the algebraic variable otherwise it drops to first order and we get again the same estimates given in [2].

Appendix 2.

1. BHR(5,5,3) IMEX R-K method with $c_2 = 2\gamma$, $\gamma = \bar{\gamma} \approx 424782/974569$.

0	0	0	0	0	0
c_2	2γ	0	0	0	0
c_3	$\tilde{a}_{31}$	$\tilde{a}_{32}$	0	0	0
c_4	$\tilde{a}_{41}$	0	$\tilde{a}_{43}$	0	0
1	$\tilde{a}_{51}$	$\tilde{a}_{52}$	$\tilde{a}_{53}$	$\tilde{a}_{54}$	0
	b_1	0	b_2	b_3	γ

0	0	0	0	0	0
c_2	γ	γ	0	0	0
c_3	a_{31}	a_{32}	γ	0	0
c_4	a_{41}	a_{42}	a_{43}	γ	0
1	b_1	0	b_2	b_3	γ
	b_1	0	b_3	b_4	γ

$\tilde{a}_{31} = \gamma$,
$\tilde{a}_{32} = \tilde{a}_{31}$,
$\tilde{a}_{41} = -475883375220285986033264/594112726933437845704163$,
$\tilde{a}_{42} = 0$,
$\tilde{a}_{43} = 1866233449822026827708736/594112726933437845704163$,
$\tilde{a}_{51} = 628288458180731695856358816860913917376103082 47/176112910684412105319781630311686343715753056000$,
$\tilde{a}_{52} = -b_3$,
$\tilde{a}_{53} = 262315887293043739337088563996093207/297427554730376353252081786906492000$,
$\tilde{a}_{54} = -98761823189417658143812471708 7/2387733766020296931952690185 6000$,
$a_{31} = \tilde{a}_{31}$,
$a_{32} = -3173308231992731 3/455705377221960889379854647102$,
$a_{41} = -3012378541084922027361996761794919360516301377809610/$
$45123394056585269977907753045030512597955897345819349$,
$a_{42} = -628655892978071532942 68/102559673441610672305587327019095047$,
$a_{43} = 418769796920855299603146267001414900945214277000/212454360385257708555954598099874818603217167139$,
$b_1 = 4876985023367406786035 11/118115963692818592026020 8$,
$b_2 = 0$,
$b_3 = 3029877630811846226393001431379430 89/153535994420329331863918012936815650 0$,
$b_4 = -1052359283351006160729382188 63/228255445206466175657572719 8000$.
$c_3 = 902905985686/1035759735069$, $c_4 = 2684624/1147171$.

2. BHR(5,5,3) IMEX R-K method with $\gamma = \gamma_2 \approx 2051948/3582211$, $c_2 = 2\gamma$.

$\tilde{a}_{31} = 4734471154406558554524823578943 73/122630625634370615492007273557914 8$,
$\tilde{a}_{32} = 1292987660341318823230699787220 19/122630625634370615492007273557914 8$,
$\tilde{a}_{41} = 3749810521082814372451684 8/172642583546398006173766007$,
$\tilde{a}_{42} = 0$,
$\tilde{a}_{43} = 7628335974256148014080441 6/172642583546398006173766007$,
$\tilde{a}_{51} = (-3409975860212064612303539855622639333030782744869519)/$
$588670410236374513779238536111308431335187021647513 6$,
$\tilde{a}_{52} = (-2374163524338269788569417957340 73)/554681702576878342891447163499456$,
$\tilde{a}_{53} = 4298159710546228783638212411650783228275/216539851335209892458721148861040704620 8$,
$\tilde{a}_{54} = 6101865615855760853571922289749/272863973025878249803640374568448$,
$a_{31} = 2592522581696725239027084257846 9319755/439288776084324396892238867419171533622 8$,
$a_{32} = (-1720741747032619856470618958617 7)/122630625634370615492007273557914 8$,
$a_{41} = 1103202061574553405285863729195740268785131739395559693754/$
$98794577359372770706415224145904934590282646779257673058 37$,
$a_{42} = (-1037545205670589695665425562960873240 94)/459050363888246734833121482275319954529$,
$a_{43} = 3863207083069979654596872190377240608602701071947128/$
$192586902512876097652406833206114257457367626819505 51$,
$b_1 = (-2032971420760927701493589)/380171476565153841909974 16$,
$b_2 = 0$,
$b_3 = 2197602776651676983265261109643897073447/945067123279139583549933947379097184164$,
$b_4 = (-1281472151942603980706668262353 39)/6946848271068750338856295 2626424$.
$c_3 = 12015769930846/24446477850549$, $c_4 = 3532944/5360597$.

Acknowledgments. The authors wish to thank the anonymous referees for suggesting many improvements.

REFERENCES

[1] U. Ascher, S. Ruuth, and R. J. Spiteri, *Implicit-explicit Runge-Kutta methods for time dependent Partial Differential Equations*, Appl. Numer. Math., 25, (1997), pp. 151–167.
[2] S. Boscarino, *Error analysis of IMEX Runge-Kutta methods derided from differential algebraic systems*, SIAM J. Numer. Anal., 45 (2008), pp. 1600–1621.
[3] S. Boscarino, *On an accurate third order implicit-explicit Runge-Kutta method for stiff problems*, Appl. Numer. Math., DOI 10.1016/j.apnum.2008.10.003.
[4] S. Boscarino, *On the Uniform Accuracy of Implicit-Explicit Runge-Kutta Methods*, Ph.D. Thesis in Mathematics for the Technology, Department of Mathematics and Computer Science, University of Catania, Italy, 2005.
[5] S. Boscarino and G. Russo, *On the uniform accuracy of IMEX Runge-Kutta schemes and applications to hyperbolic systems with relaxation*, in Proceedings of SIMAI2006 VIII Convegno SIMAI Ragusa (Italy), May 2006. Published in Communications to SIMAI Conferences, 2006, DOI: 10.1685/CSC06028, ISSN 1827-9015, Vol. 2, 2007.
[6] R. E. Caflisch, S. Jin, and G. Russo, *Uniformly accurate schemes for hyperbolic systems with relaxation*, SIAM J. Numer. Anal., 34 (1997), pp. 246–281.
[7] M. H. Carpenter and C. A. Kennedy, *Additive Runge-Kutta schemes for convection-diffusion-reaction equations*, Appl. Numer. Math., 44 (2003), pp. 139–181.
[8] C. Q. Chen, C. D. Levermore, and T. P. Liu, *Hyperbolic conservation laws with relaxation terms and entropy*, Comm. Pure Appl. Math., 47 (1994), pp. 787–830.
[9] B. Garcia-Celayeta, I. Higueras, and T. Roldan, *Contractivity/monotonicity for additive Runge-Kutta methods: Inner product norms*, Appl. Numer. Math., 56 (2006), pp. 862–878.
[10] E. Hairer, S. P. Nørsett, and G. Wanner, *Solving Ordinary Differential Equation I: Non-stiff problems*. Springer Series in Comput. Mathematics, Vol. 8, Springer-Verlag, Berlin, 1987, second revised edition, 1993.
[11] E. Hairer and G. Wanner, *Solving Ordinary Differential Equation II: Stiff and Differential Algebraic Problems*. Springer Series in Comput. Mathematics, Vol. 14, Springer-Verlag, Berlin, 1991, second revised edition, 1996.
[12] E. Hairer, Ch. Lubich, and G. Wanner, *Geometric Numerical Integration: Structure-Preserving Algorithms for Ordinary differential Equations*, Springer series in Computational Mathematics, Vol. 31, Springer-Verlag, Berlin, 2002.
[13] S. Jin, *Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms*, J. Comput. Phys., 122 (1995), pp. 51–67.
[14] S. F. Liotta, V. Romano, and G. Russo, *Central schemes for balance laws of relaxation type*, SIAM J. Numer. Anal., 38 (2000), pp. 1337–1356.
[15] T. P. Liu, *Hyperbolic conservation laws with relaxation*, Comm. Math. Phys., 108 (1987) pp. 153–175.
[16] L. Pareschi and G. Russo, *Implicit-Explicit Runge-Kutta schemes for stiff systems of differential equations*, Recent trends in numerical analysis, Adv. Theory Comput. Math., 3, Nova Sci. Publ., Huntington, NY, 2001, pp. 241–251.
[17] L. Pareschi and G. Russo, *Implicit-Explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxations*, J. Sci. Comput., 25 (2005), pp. 129–155
[18] L Pareschi and G. Russo, *High order asymptotically strong-stability-preserving methods for hyperbolic systems with stiff relaxation*, Hyperbolic Problems: Theory, Numerics, Applications, Springer, Berlin, 2003, pp. 241–251.
[19] S. Qamar and G Warnecke, *A Space-Time Conservative Method for Hyperbolic System with Stiff and Non Stiff Terms*, Commun. Comput. Phys., 1, pp. 449–478.
[20] C. W. Shu, *Essentially Non Oscillatory and Weighted Essentially Non Oscillatory Schemes for Hyperbolic Conservation Laws*, in Advance Numerical Approximation of Nonlinear Hyperbolic Equations, Lecture Notes in Math. 1697, (2000).
[21] G. B. Whitham, *Linear and non-linear waves*, Wiley, New York, 1974.