High order Finite Volume Schemes for Balance Laws with Stiff Relaxation

High order Finite Volume Schemes for Balance Laws with Stiff Relaxation

S. Boscarino[a], G. Russo[a,*], M. Semplice[b]

[a] *Department of Mathematics and Computer Science, University of Catania, Catania, 95125, Italy*
[b] *Dipartimento di Matematica, Università di Torino, via C. Alberto 10, Torino, 10123, Italy*

Abstract

The paper deals with the construction and analysis of efficient high order finite volume shock capturing schemes for the numerical solution of hyperbolic systems with stiff relaxation. In standard high order finite volume schemes it is difficult to treat the average of the source implicitly, since the computation of such average couples neighboring cells, making implicit schemes extremely expensive. The main novelty of the paper is that the average of the source is split into the sum of the source evaluated at the cell average plus a correction term. The first term is treated implicitly, while the small correction is treated explicitly, using IMEX-Runge-Kutta methods, thus resulting in a very effective semi-implicit scheme. This approach allows the construction of effective high order schemes in space and time. An asymptotic analysis is performed for small values of the relaxation parameter, giving an indication on the structure of the IMEX schemes that have to be adopted for time discretization. Several numerical tests confirm the accuracy and efficiency of the approach.

Keywords: Finite Volume schemes, Stiff Problems, Runge-Kutta methods, Implicit-Explicit schemes, Hyperbolic systems, Relaxation.
AMS: 65M08, 65L04, 65L06, 35L45

*Corresponding author
Email addresses: boscarino@dmi.unict.it (S. Boscarino), russo@dmi.unict.it (G. Russo), matteo.semplice@unito.it (M. Semplice)

1. Introduction

The aim of the paper is to propose effective high order finite volume methods for the numerical solution of systems of hyperbolic equation with stiff relaxation of the form

$$u_t + f(u)_x = \frac{1}{\varepsilon} R(u), \tag{1}$$

where $(x,t) \in \mathbb{R} \times \mathbb{R}^+$, and $u(x,t) \in \mathbb{R}^M$.

Here we assume that the system may be *stiff*, i.e. that the parameter ε may be much smaller than 1.

Many physical systems are described by equations of the form (1), such as kinetic models near the fluid dynamic limit [14, 31], extended thermodynamics [1], gases with vibrational degrees of freedom [35], and many others.

Such systems are called *relaxation system* in the sense of Whitham [36] and Liu [28] if there exists a constant $m \times M$ matrix Q with rank $m < M$, such that $QR(u) = 0 \; \forall u \in \mathbb{R}^M$. This induces the existence of m independent conserved quantities $u_c = QR(u)$, which uniquely define each equilibrium u_{eq} such that $R(u_{eq}) = 0$: $u_{eq} = \mathcal{E}(u_c)$, with $R(\mathcal{E}(u_c)) = 0$. Multiplying system (1) by Q gives the following system of m conservation equations:

$$\partial_t u_c + \partial_x \left(Q f(u) \right) = 0 \tag{2}$$

In the limit $\varepsilon \to 0$, formally the variable u approaches the local equilibrium given by $u = u_{eq} = \mathcal{E}(u_c)$. Substituting this equation in the expression of $Qf(u)$ we obtain

$$\partial_t u_c + \partial_x \mathcal{F}(u_c) = 0 \tag{3}$$

where $\mathcal{F}(u_c) = QF(\mathcal{E}(u_c))$.

The convergence of the solution of the original system (1) to $\mathcal{E}(u_c)$, where u_c is the solution of the relaxed system (3) is guaranteed by the *sub characteristic condition* [15], that states that the characteristic velocities of the reduced system have to be bounded from above and from below by the characteristic velocities of the original system.

There is a vast literature on the numerical solution of systems of hyperbolic systems with stiff relaxation. Most shock capturing schemes for such problems are based on space discretization which can be finite volume, conservative finite difference (*a la* Shu, as in [34]) or discontinuous Galerkin [29], while time discretization is mainly based on implicit-explicit (IMEX)

schemes, which allow an explicit treatment of the flux and an implicit treatment of the source term (see for example [13], [30] and references therein).

One of the first attempts to construct a second order scheme for this class of problems was proposed in [12]. The space discretization was finite volume, and the time discretization had a structure that was later interpreted as an IMEX-Runge-Kutta (IMEX-RK) scheme. A similar approach was adopted in [27] in the context of central schemes.

The construction of shock-capturing finite-volume IMEX-RK schemes with spatial order higher than two requires:

1. high order reconstruction of point values from cell averages of the field variables from the cell averages,
2. computation of the numerical fluxes,
3. implicit treatment of the average of the source term.

The latter point is quite delicate for the following reason. If the source term is linear in the unknown vector field, than the cell average of the source and the source evaluated at the cell average coincide. In such cases the construction of high order finite volume scheme can be effectively performed by evaluating the source at cell averages. For non linear dependence of R on u, if we approximate the average of the source with the source term computed at the average of the solution, then we make an error which, for smooth solutions, is of second order in the spatial grid spacing Δx. Therefore this approximation, which was adopted in [12], would be sufficient for a second order scheme, but would prevent from reaching higher than second order accuracy.

On the other hand, in order to achieve third or higher accuracy, the source term can be integrated with a high order quadrature rule within each cell and a high order reconstruction procedure of point values of the solution at the quadrature nodes must be employed. Due to non linearity of high order non-oscillatory reconstructions, this in turn would make very expensive to treat implicitly the source term. There are special cases in which the implicit relaxation equation can be explicitly solved analytically. In such cases, one could express the implicit solution of the source in terms of quantities that can be explicitly computed, with no need of solution of non-linear equations or cell-coupling at the level of the source. Such an approach has been adopted, for example, by Banda et al. [4].

In general, however this is not possible, and one has to solve a large set of coupled non-linear equations to implicitly compute the source term. A work in this direction has been presented by Dumbser and collaborators [20].

In their paper the authors generalize the ADER-DG predictor to the case of stiff sources. They rely on the solution of a large sparse non-linear system, and give indications on how to start from a good guess in order to reduce the number of Newton iterations.

In other cases, third order accuracy is obtained by computing an implicit predictor on cell edges, and using it to construct quadrature formulas that allow high order accuracy of the average of the source term. This technique is presented by Balsara et al. in [3]. However, in spite of the accurate results, the analysis of the methods has still not been performed, and its justification is not yet fully understood.

For the above mentioned reasons, most high order schemes for hyperbolic systems with stiff relaxation have been based on finite difference space discretization both for the development of the schemes [30, 10], and for the analysis [6, 7]. In addition of being more naturally suited for implicit treatment of the source term, conservative finite difference schemes have the additional advantage, over finite volume schemes, that their high order implementation in several space dimensions is more efficient for uniform grid, because of the dimension-by-dimension reconstruction required.

However, finite volume schemes are much more flexible than finite difference ones, since they do not require uniform grid, and can be actually implemented on unstructured grids. Furthermore, recent developments of CWENO reconstructions [17, 18, 33, 32] make high order finite volume schemes almost as efficient as high order finite difference on regular grids, and allows efficient reconstruction also for unstructured grids [18] and multiple inner reconstruction points in the cells [17].

With this motivation in mind, the purpose of this paper is to provide effective tools for the construction of high order finite-volume IMEX Runge-Kutta schemes for hyperbolic systems with stiff relaxation, which do not suffer from the standard drawback of having to couple neighboring cells in the evaluation of the source term.

In fact, the reconstruction of u in a cell depends in a very nonlinear way not only on the cell average of that cell but also on the cell averages in a number of neighbours. Thus, treating the source term $\langle R(u(x))\rangle/\varepsilon$ implicitly would require the solution of a nonlinear system as large as the computational grid and with a degree of coupling that is larger and larger as the order of the scheme increases, due to the enlargement of the reconstruction stencil.

The idea of the proposed method is to adopt a penalization technique, [9, 11, 8, 19] on the average of the source $\langle R\rangle$ in each cell. Such average is

written as the sum of the source of the average plus a correction term:

$$\frac{1}{\varepsilon}\langle R(u)\rangle = \frac{1}{\varepsilon}R(\bar{u}) + \frac{1}{\varepsilon}\Delta R.$$

The term $R(\bar{u})$ can be treated implicitly, while the term $\Delta R = \langle R\rangle - R(\bar{u})$ is treated explicitly. Thus, the implicit part of the scheme does not depend on the reconstruction and requires only the solution of a set of independent equations.

The use of a penalization technique to split a source into a term which is easily invertible and a small correction which is computed explicitly has already been adopted in other contexts, such as kinetic equations [21], or hyperbolic systems with diffusive relaxation [9]. Note that the correction term is $O(\Delta x^2/\varepsilon)$, therefore it is small compared to the main source term, however it is not bounded as $\varepsilon \to 0$. As we shall see, this has some implication on the structure of IMEX schemes that have to be adopted if we want that the numerical scheme applied to (1) becomes a consistent scheme for the limit system (3) as $\varepsilon \to 0$.

The plan of the paper is the following. In the next section we describe how to construct the method. Section 3 is devoted to the analysis of the method. In particular, it will be shown that stability requires the use of *Globally Stiffly Accurate* IMEX Runge-Kutta (GSA IMEX-RK) methods, while third order accuracy will be maintained in the relaxed limit with no additional conditions. Section 4 is devoted to illustrate several numerical tests which emphasize the effectiveness of the scheme. Finally, in the last section we draw the conclusions.

2. Description of the method

In this paper we restrict to one dimensional problems. We expect the advantage of the approach in more space dimensions to be even higher, given that in this case a cell has several neighbors and a fully implicit treatment of the source term would be extremely expensive. Such an extension to more space dimensions should be performed in a straightforward manner, and will be considered in a forthcoming paper.

The computational interval is discretized in uniform cells, each cell Ω_j centered at x_j, and the mesh spacing is $\Delta x = x_{j+1/2} - x_{j-1/2}$, $j = 1, ..., N$, where N denotes the total number of cells.

We denote by $\bar{u}_j$ the cell average of u in the cell Ω_j:

$$\bar{u}_j = \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} u(t,x)dx.$$

To integrate system (1) we consider a standard finite volume discretization given as follows:

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{u}_j + \frac{1}{\Delta x}(F_{j+1/2} - F_{j-1/2}) = \frac{1}{\varepsilon}\left\langle R(u)\right\rangle_j \tag{4}$$

where the numerical flux $F_{j+1/2} = F(u^-_{j+1/2}, u^+_{j+1/2}) \approx f(u(x_{j+1/2},t))$ has to be defined in terms of the known cell-average numerical quantities $\bar{u}_j$. The numerical flux function that has been used throughout all the calculations performed for this paper is the local Lax-Friedrichs flux (also known as Rusanov flux),

$$F(u^-_{j+1/2}, u^+_{j+1/2}) = \frac{1}{2}\left[f(u^-_{j+1/2}) + f(u^+_{j+1/2}) - \alpha(u^+_{j+1/2} - u^-_{j+1/2})\right]$$

where $\alpha = \max_w |f'(w)|$, and the maximum is taken over the relevant range of w.

The averaged source term is defined by:

$$\begin{aligned}\left\langle R(u)\right\rangle_{\Omega_j} &= \frac{1}{\Delta x}\int_{x_{j-1/2}}^{x_{j+1/2}} R(u)dx = R\left(\int_{x_{j-1/2}}^{x_{j+1/2}} u(t,x)dx\right) \\ &\quad + \mathcal{O}(\Delta x^2) = R(\bar{u}_j) + \mathcal{O}(\Delta x^2).\end{aligned} \tag{5}$$

The second order error has the following expression:

$$\left\langle R(u)\right\rangle_{\Omega_j} - R(\bar{u}_j) = \frac{\Delta x^2}{24}\sum_{\alpha,\beta=1}^{M} u_x^\alpha u_x^\beta \partial^2_{\alpha\beta} R(\langle u_j\rangle) + O(\Delta x^3) \tag{6}$$

where α and β denote the components. Thus, in order to compute the cell average of the source term $R(u)$ at order higher than 2, one cannot use the approximation

$$\left\langle R(u)\right\rangle_{\Omega_j} \approx R(\bar{u}_j)$$

and must compute the average of the source term applying a quadrature rule with at least two nodes to a suitable reconstruction $u_j(x,t)$ of the numerical

solution within the cell Ω_j. In particular, in order to preserve the third order accuracy, the reconstruction in the j-th cell must depend at least on the cell averages of the same cell and of the first neighbours; in general one has

$$u_j(t,x) = p_j(x; \bar{u}_{j-p}(t), \bar{u}_j(t), \bar{u}_{j+q}(t)), \quad p, q \geq 1.$$

Note that the dependence of p_j on the cell averages must be nonlinear in order to ensure high order accuracy and essentially non-oscillatory properties for the scheme.

Applying the implicit-explicit Euler IMEX scheme to the above problem, treating the source implicitly, would lead to the following set of equations:

$$\forall j: \quad \frac{\bar{u}_j^{n+1} - \bar{u}_j^n}{\Delta t} + \frac{F_{j+1/2}(\bar{u}^n) - F_{j-1/2}(\bar{u}^n)}{\Delta x} = \frac{1}{\varepsilon} \left\langle R\left(u^{n+1}\right)\right\rangle_{\Omega_j} \tag{7}$$

where $\left\langle R\left(u^{n+1}\right)\right\rangle_{\Omega_j} \equiv \left\langle R\left(p_j\left(x; \bar{u}_{j-1}^{n+1}, \bar{u}_j^{n+1}, \bar{u}_{j+1}^{n+1}\right)\right)\right\rangle_{\Omega_j}$. Even in the case of a linear source term $R(u)$, the nonlinear dependence of the p_j on the cell averages would force us to solve the above equations as a system of coupled nonlinear equations, where the coupling is due to the stencil of the reconstruction procedure. In the third order case, the equation for $\bar{u}_j^{n+1}$ is coupled to the equations for $\bar{u}_{j\pm1}^{n+1}$, but this of course will get worse if the order (and thus the stencil) of the reconstruction procedure is increased.

Instead, we propose to rewrite system (4) as

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{u}_j + \underbrace{D_x f(u)_j}_{\text{EX}} = \frac{1}{\varepsilon}\left(\underbrace{R(\bar{u}_j)}_{\text{IM}} + \underbrace{\left\langle R(u)\right\rangle_{\Omega_j} - R(\bar{u}_j)}_{\text{EX}}\right), \tag{8}$$

where $D_x f$ denotes the flux difference at cell boundaries, and to treat implicitly only the (local) term $R(\bar{u}_j)$. In this way the implicit system to be solved for the source term is *de facto* a set of independent nonlinear equations, where the nonlinearity may originate from the nonlinearity of the source term only. The correction term, which contains nonlinearities coming from both the function $R(u)$ and from the nonlinearity of the reconstruction procedure, will be treated explicitly.

If the form (8) is employed instead of the standard form (1), the analogous of equation (7) reads

$$\begin{aligned} \forall j: \quad & \frac{\bar{u}_j^{n+1} - \bar{u}_j^n}{\Delta t} + \frac{F_{j+1/2}(\bar{u}^n) - F_{j-1/2}(\bar{u}^n)}{\Delta x} \\ = \quad & \frac{1}{\varepsilon} R\left(\bar{u}_j^{n+1}\right) + \frac{1}{\varepsilon}\left[\left\langle R\left(p_j\left(x; \bar{u}_{j-1}^n, \bar{u}_j^n, \bar{u}_{j+1}^n\right)\right)\right\rangle_{\Omega_j} - R\left(\bar{u}_j^n\right)\right] \end{aligned} \tag{9}$$

which is a set of independent equations, one for each cell Ω_j.

The difference between the right hand side of Eq. (7) and of Eq. (9) is

$$\left\langle R(u^{n+1})\right\rangle_{\Omega_j} - R(\bar{u}_j^{n+1}) - \left(\left\langle R(u^n)\right\rangle_{\Omega_j} - R(\bar{u}_j^n)\right)$$

which is $\mathcal{O}(\Delta t \Delta x^2)$, therefore this approach introduces an error of $\mathcal{O}(\Delta t \Delta x^2)$, thus allowing third order accuracy if the space and time accuracy of the scheme for system (8) is third order.

The aim of this paper is to generalize method (9) using high order IMEX schemes in order to obtain an overall high order method that avoids the solution of coupled non-linear systems like (7).

We briefly describe an s-stage IMEX-RK scheme applied to (8). An s-stage IMEX RK scheme can be represented with a double Butcher tableau

$$\text{Explicit}: \begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} \qquad \text{Implicit}: \begin{array}{c|c} c & A \\ \hline & b^T \end{array}. \tag{10}$$

Here the matrices $\tilde{A} = (\tilde{a}_{ij})$, $\tilde{a}_{ij} = 0$ for $j \geq i$, and $A = (a_{ij})$, $a_{ij} = 0$ $j > i$ are $s \times s$ matrices. We use diagonally implicit Runge-Kutta scheme (DIRK), [24], for the implicit part, so that we ensure that the flux and the correction term will be effectively treated explictly, (see [2], [13], [10], [7], [30]). The coefficients $\tilde{c}$ and c are given by the usual relation $\tilde{c}_i = \sum_{j=1}^{i-1} \tilde{a}_{ij}$, $c_i = \sum_{j=1}^{i} a_{ij}$ and vectors $\tilde{b} = (\tilde{b}_i)_{i=1\cdots s}$ and $b = (b_i)_{i=1\cdots s}$ provide the quadrature weights to combine the internal stages of the Runge-Kutta method.

We denote $F^{(i)}_{j+1/2}$ the numerical flux across the boundary between cells Ω_j and Ω_{j+1}, computed via the reconstruction of the cell averages $\bar{u}_j^{(i)}$ at stage i of the Runge-Kutta scheme. The stage values are given by

$$\begin{aligned} \bar{u}_k^{(i)} &= u_k^{*,i} + \frac{\Delta t}{\varepsilon} \sum_{j=1}^{i-1} a_{ij} R(\bar{u}_k^{(j)}) + \frac{\Delta t}{\varepsilon} a_{ii} R(\bar{u}_k^{(i)}), \quad \text{where} \\ u_k^{*,i} &= \bar{u}_k^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \left(D_x F_k^{(j)} - \frac{1}{\varepsilon} \left(\left\langle R(u^{(j)}(x)) \right\rangle_{\Omega_k} - R(\bar{u}_k^{(j)}) \right) \right) \end{aligned} \tag{11}$$

with

$$D_x F_k^{(j)} \equiv \frac{F^{(j)}_{k+1/2} - F^{(j)}_{k-1/2}}{\Delta x},$$

and the numerical solution by

$$\begin{aligned}
\bar{u}_k^{(n+1)} &= u_k^{*,n+1} + \Delta t \sum_{i=1}^{s} b_i R(\bar{u}_k^{(i)}), \quad \text{where} \\
u_k^{*,n+1} &= \bar{u}_k^n - \Delta t \sum_{i=1}^{s} \tilde{b}_i \left(D_x F_k^{(i)} - \frac{1}{\varepsilon} \left(\left\langle R(u^{(i)}(x)) \right\rangle_{\Omega_k} + R(\bar{u}_k^{(i)}) \right) \right).
\end{aligned} \tag{12}$$

2.1. Analysis of the scheme

Preliminary definitions. Some preliminary notions on IMEX RK schemes are necessary before we discuss our approach. First of all the double Butcher tableau must satisfy standard order conditions, see [13, 30] for details.

It is useful to characterize the different IMEX schemes we will consider in the sequel according to the structure of the DIRK method. Following [6] we have

Definition 1.

1. *We call an IMEX-RK method of* type A *(see [30]) if the matrix* $A \in \mathbb{R}^{s\times s}$ *is invertible, or equivalently* $a_{ii} \neq 0$, $i = 1, \ldots, s$.
2. *We call an IMEX-RK method of* type CK *(see [13]) if the matrix* A *can be written as*

$$A = \begin{pmatrix} 0 & 0 \\ \hat{a} & \hat{A} \end{pmatrix}, \tag{13}$$

with $\hat{a} = (a_{21}, \ldots, a_{s1})^T \in \mathbb{R}^{(s-1)}$ *and the submatrix* $\hat{A} \in \mathbb{R}^{(s-1)\times(s-1)}$ *is invertible, or equivalently* $a_{ii} \neq 0$, $i = 2, \ldots, s$. *In the special case* $\hat{a} = 0$, $b_1 = 0$ *the scheme is said to be of* type ARS *(see [2]) and the DIRK method is reducible to a method using* $s - 1$ *stages.*

We will also make use of the following representation of the matrix $\tilde{A}$ in the explicit Runge-Kutta method

$$\tilde{A} = \begin{pmatrix} 0 & 0 \\ \breve{a} & \breve{A} \end{pmatrix}, \tag{14}$$

where $\breve{a} = (\tilde{a}_{21}, \ldots, \tilde{a}_{s1})^T \in \mathbb{R}^{s-1}$ and $\breve{A} \in \mathbb{R}^{(s-1)\times(s-1)}$.

The following definition will be also useful to characterize the properties of the methods in the sequel.

Definition 2. *We call an IMEX-RK method* globally stiffly accurate (GSA)*, if*

$$a_{si} = b_i, \quad i = 1, \dots, s, \quad \tilde{a}_{si} = \tilde{b}_i, \quad i = 1, \dots, s-1, \tag{15}$$

Note that this definition implies that in this case the numerical solution coincides exactly with the last internal stage of the scheme. The GSA property were already considered in [9, 11]. In the appendix we list some example of such schemes.

2.2. Stability Analysis

The purpose of this section is to analyze the scheme as $\varepsilon \to 0$. We note that for small values of ε, even the correction term may be become stiff, therefore classical analysis of IMEX-RK schemes for hyperbolic systems with stiff relaxation ([30], [12], [10]) is not sufficient to guarantee stability. Here we show that stability for small values of ε is guaranteed provided the IMEX RK scheme is globally stiffly accurate (GSA).

First we start considering IMEX RK of type A for which the analysis is simpler. We write (11) in the compact form using the following notation: $\bar{U} = (\bar{u}_j^{(i)})$ is an $s \times N$ of vectors of dimension M, denoting the stage values between t^n and t^{n+1}, $(R(\bar{U}))_j^{(i)} := R(\bar{u}_j^{(i)})$, $(\Delta R(\bar{U}))_j^{(i)} := \left\langle R(\bar{U}^{(i)}) \right\rangle_j - R(\bar{u}_j^{(i)})$, $(D_x F(\bar{U}))_j^{(i)} := (F_{j+1/2}^{(i)} - F_{j-1/2}^{(i)})/\Delta x$, $e = (1, \cdots, 1)^T$ is the unit vector in $\mathbb{R}^s$.

Then we have:

$$\bar{U} = u^* + \frac{\Delta t}{\varepsilon} A R(\bar{U}) \tag{16}$$

with

$$u^* = \bar{u}^n e - \Delta t \tilde{A} \left(D_x F(\bar{U}) - \frac{1}{\varepsilon} \Delta R(\bar{U}) \right) \tag{17}$$

and for the numerical solution:

$$\bar{u}^{n+1} = \bar{u}^n - \Delta t \tilde{b}^T \left(D_x F(\bar{U}) - \frac{1}{\varepsilon} \Delta R(\bar{U}) \right) + b^T \frac{\Delta t}{\varepsilon} R(\bar{U})$$

By (16) we get

$$\frac{\Delta t}{\varepsilon} R(\bar{U}) = A^{-1}(\bar{U} - u^*)$$

and substituting it in the numerical solution we have

$$\bar{u}^{n+1} = \bar{u}^n - \Delta t \tilde{b}^T \left(D_x F(\bar{U}) - \frac{1}{\varepsilon} \Delta R(\bar{U}) \right) + b^T A^{-1}(\bar{U} - u^*).$$

By (17) and some algebraic manipulations we get

$$\begin{aligned}\bar{u}^{n+1} = (1 - b^T A^{-1} e)\bar{u}^n - \Delta t(\tilde{b}^T - b^T A^{-1}\tilde{A}) D_x F(\bar{U}) + \\ + \frac{\Delta t}{\varepsilon}(\tilde{b}^T - b^T A^{-1}\tilde{A})\Delta R(\bar{U}) + b^T A^{-1} U.\end{aligned} \tag{18}$$

In (18) consistency as $\varepsilon \to 0$ implies $\tilde{b}^T - b^T A^{-1}\tilde{A} = 0$. This condition is satisfied provided that the IMEX R-K is GSA. Indeed, if we assume that the scheme is GSA, we have $b^T A^{-1} = e_s^T = (0, ..., 0, 1)$, which implies

$$1 - b^T A^{-1} e = 0,$$

and

$$\tilde{b}^T - b^T A^{-1}\tilde{A} = \tilde{b}^T - e_s^T \tilde{A} = 0.$$

Note that if the scheme is not GSA, for small values of ε ($\varepsilon \to 0$), the numerical solution $\bar{u}^{n+1}$ blows up. In the numerical section we will show some tests that confirm this assertion.

On the other hand, under the assumption of GSA, in the limit $\varepsilon \to 0$, the numerical solution becomes

$$\bar{u}^{n+1} = e_s^T \bar{U},$$

i.e. the numerical solution coincides with the last internal stage of the method.

We now turn to IMEX RK schemes of type CK, which are very attractive because they allow some simplifying assumptions that make order conditions easier to deal with, thus allowing the construction of high order IMEX RK schemes, [23, 7, 10].

Following a technique similar to the one used in [9] in the case of the parabolic relaxation, one can extend the analysis to the case of IMEX CK schemes and prove the following:

Proposition 1. *If the IMEX RK scheme of type CK is GSA, then in the limit $\varepsilon \to 0$ one has $\bar{u}^{n+1} = \hat{e}_s^T \hat{\bar{U}}$.*

The proof of the statement is reported in the appendix. We conclude the subsection with a few remarks.

Remark 1. *Since the ARS type is a particular case of the type CK ($\breve{a} = \hat{a} = 0$, and $b_1 = \tilde{b}_1 = 0$) the previous condition guarantees the stability of ARS IMEX RK.*

Remark 2. *The requirement that the schemes are GSA, impose a lower bound on the number of the stages needed to reach order higher than two for the different types of IMEX RK. In particular, four stages are needed to obtain a second order GSA IMEX RK scheme of type A, [11, 9]. For this reason it is preferable to adopt CK (or ARS) type GSA IMEX RK schemes, which allow second order accuracy with three stages, and third order accuracy with four stages. The IMEX RK schemes tested in the paper are reported in the Appendix.*

Remark 3. *The IMEX RK schemes tested in this paper satisfy the condition $c_i = \tilde{c}_i$ for all i, which is often adopted to simplify the order conditions in designing standard IMEX schemes of type CK (or ARS), [10], and is satisfied by all CK (or ARS) GSA schemes existing in the literature, see for example [2, 13].*

2.3. Accuracy Analysis

In this section we prove that the scheme guarantees third order accuracy in space. We consider the simplest one dimensional prototype hyperbolic system with stiff source:

$$u_t + v_x = 0, \qquad v_t + u_x = -\frac{1}{\varepsilon}(v - f(u)), \tag{19}$$

where $f(u)$ is a given smooth function. It is easy to check that as $\varepsilon \to 0$, system (19) relaxes to the conservation law:

$$u_t + f(u)_x = 0. \tag{20}$$

A semi-discrete approximation for system (19) can be written as

$$\begin{aligned} &\frac{d\bar{u}_j}{dt} + D_x v_i = 0, \\ &\frac{d\bar{v}_j}{dt} + D_x u_i = -\frac{1}{\varepsilon} \left\langle (v - f(u)) \right\rangle_j . \end{aligned} \tag{21}$$

We apply the first order GSA IMEX RK scheme (39) to the system (21) and consider the limit case $\varepsilon \to 0$, thus obtaining:

$$\begin{aligned} \bar{v}_j^{n+1} &= \left\langle f(u) \right\rangle_i^{n+1}, \\ \bar{u}_j^{n+1} &= \bar{u}_j^n - \Delta t D_x \left\langle f(u) \right\rangle_j^n . \end{aligned}$$

Now we rewrite system (21) in the form (8), and apply the same time discretization (39), obtaining

$$\begin{aligned} \bar{u}_j^{n+1} &= \bar{u}_j^n - D_x v_i^n, \\ \bar{v}_j^{n+1} &= \bar{v}_j^n - D_x u_i^n + \frac{\Delta t}{\varepsilon} \Delta R(\bar{u}^n)_j + \frac{\Delta t}{\varepsilon} R(\bar{u}_j^{n+1}), \end{aligned} \tag{22}$$

where $\Delta R(\bar{u}^n)_j = (\langle R(u) \rangle_j^{n+1} - R(\bar{u}_j^{n+1})) = \mathcal{O}(\Delta x^2)$.

As $\varepsilon \to 0$, we obtain:

$$\begin{aligned} \bar{v}_j^{n+1} &= f(\bar{u}_j^{n+1}) + \langle f(u) \rangle_j^n - f(\bar{u}_j^n) \\ &= f(\bar{u}_j^{n+1}) + \langle f(u) \rangle_j^{n+1} - f(\bar{u}_j^{n+1}) - \frac{d}{dt} \left(\langle f(u) \rangle_j^{n+1} - f(\bar{u})_j^{n+1} \right) \Delta t \\ &\quad + \mathcal{O}(\Delta x^2 \Delta t^2) \\ &= \langle f(u) \rangle_j^{n+1} + \mathcal{O}(\Delta x^2 \Delta t) \end{aligned}$$

because $\langle f(u) \rangle - f(\bar{u}) = \mathcal{O}(\Delta x^2)$. At time t^n we have

$$\bar{v}_j^n = \langle f(u) \rangle_j^n + \mathcal{O}(\Delta x^2 \Delta t)$$

therefore the scheme for $\bar{u}$ becomes

$$\bar{u}_j^{n+1} = \bar{u}_j^n - \Delta t \left[D_x \langle f(u) \rangle_j^n + \mathcal{O}(\Delta x^2 \Delta t) \right]$$

indicating that the correction introduces an error which is second order accurate in space. Of course, one has to use an approximation of the average of the source, such as for example, Simpson's rule, which guarantees fourth order accuracy, introducing another error which depends on the accuracy of the reconstruction and on the quadrature formula adopted for the computation of the average. Using a parabolic reconstruction and Simpson's rule guarantees third order accuracy.

The generalization to high order IMEX-RK schemes requires a deeper analysis and is currently under investigation.

3. Numerical experiments

In this section we present some numerical tests concerning situations in which hyperbolic systems with stiff relaxation play a major role in applications in order to validate our method. Note that in some tests the relaxation term $R(u)$ is non-linear and, treating it implicitly, requires to solve a

non-linear algebraic system in each cell. An efficient approach is obtained by solving it iteratively with the Newton's method. All the numerical results are obtained using a class of existing third and second order IMEX RK schemes listed in the Appendix. For the third order accurate schemes, the spatial reconstruction is the third order accurate CWENO procedure first introduced in [26] as further developed in [18]. In particular the linear weights suggested in the latter paper were employed, namely $d_0 = 3/4$ for the central polynomial and $d_L = d_R = 1/8$ for the one-sided linear ones. The term $\langle R(u_j(x))\rangle$ is computed by the Simpson's quadrature rule applied in the cell Ω_j.

3.1. Euler Gas Dynamics with heat transfer

In agreement with what illustrated in Sect. 2, first we show the importance to have a GSA IMEX-RK scheme applied to (8). The GSA assumption for a IMEX-RK scheme guarantees that the numerical solution of system (8) in the limit $\varepsilon \to 0$ does not blow up, remains stable by using a classical hyperbolic CFL condition $CFL = \max_u |f'(u)|\Delta t/\Delta x < 1$ (using coarse grids ($\Delta t, \Delta x \geq \varepsilon$), and converges to the solution of the relaxed equation. Such a scheme is usually referred as an *underresolved numerical scheme* for the hyperbolic conservation laws with stiff source (1).

On the other hand, we expect that the non GSA (NGSA in short) IMEX RK scheme is not stable as underresolved numerical scheme and that forcing $\Delta t \sim \mathcal{O}(\varepsilon)$ (i.e. $CFL \sim \varepsilon/\Delta x$) is necessary in order to make the scheme stable.

In order to show this we consider the one dimensional Euler equations for gas dynamics coupled with a constant temperature bath via a simplified heat transfer rate equation [25]:

$$\begin{aligned} &(\rho)_t + (\rho u)_x = 0, \\ &(\rho u)_t + (\rho u^2 + p)_x = 0, \\ &(\rho E)_t + ((\rho E + p)u)_x = -K\rho(T - T_0). \end{aligned} \tag{23}$$

In the above equations ρ represents the gas density, u its velocity, $m = \rho u$ the momentum, $E = e + u^2/2$ the energy per unit mass, e the internal energy, T the temperature and p the pressure. We assume the gas to be a γ-law gas, i.e., $p = (\gamma - 1)\rho e$. We choose temperature units so that $T = e$ and K and T_0 are positive constants, where $K \gg 1$ is the heat transfer coefficient and T_0 is the temperature of the constant temperature bath. We compare GSA

and NGSA schemes by solving the Riemann problem with initial conditions

$$\begin{aligned} &\rho_l = 1, \quad m_l = 0, \quad E_l = 1, \qquad 0 < x < 0.5 \\ &\rho_r = 0.2, \quad m_r = 0, \quad E_r = 1, \qquad 0.5 \leq x < 1 \end{aligned} \tag{24}$$

with $K = 1/\varepsilon = 10^8$ (the stiffness of the system) and $T_0 = 1$. We integrate over $[0, 1]$ with $N = 200$ spatial cells and $CFL = 0.5$, reflecting boundary conditions, and we fix the final time at $t = 0.3$. A reference solution has been computed with 2000 spatial cells, using scheme (41). The numerical solutions are depicted in Fig. 1. We observe that the third order GSA IMEX-RK scheme (41) correctly captures the reference solution. On the other hand the NGSA IMEX-RK scheme of (43) blows up after a short time: the plot shows the computed solution at $t = 0.13$, when the simulation was stopped.

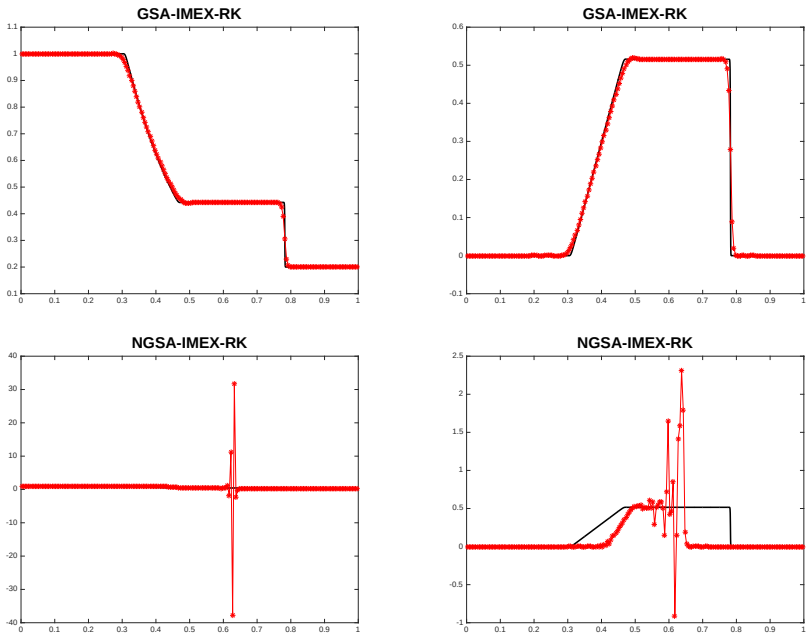

Figure 1: Euler gas dynamics with heat transfer. Comparison of the numerical solutions obtained with third order GSA-IMEX-RK and third order NGSA-IMEX-RK for $K = 1/\varepsilon = 10^8$. Density (left) and velocity (right) profiles. Black solid line is the reference solution.

3.2. Broadwell model.

The Broadwell model describes a two dimensional (three dimensional) gas composed of particles with four (six) discrete velocities with binary collision law and spatial variation in only one direction. For the two dimensional case, the evolution equations read [12]

$$\begin{aligned} &\partial_t \rho + \partial_x m = 0, \\ &\partial_t m + \partial_x z = 0, \\ &\partial_t z + \partial_x m = \frac{1}{2\varepsilon}(\rho^2 + m^2 - 2\rho z)\,, \end{aligned} \tag{25}$$

where the fluid dynamic moment variables are the density ρ, the momentum m and ε is the mean free path. As $\varepsilon \to 0$, equations (25) converges to

$$\begin{aligned} &\partial_t \rho + \partial_x (\rho v) = 0, \\ &\partial_t m + \partial_x \left(\frac{1}{2}(\rho + \rho v^2) \right) = 0, \\ &z = \frac{1}{2}(\rho + \rho v^2), \end{aligned} \tag{26}$$

with the velocity v defined by $v = m/\rho$. Equations (26) represent the fluid dynamic (Euler) limit of the Broadwell equation (25).

This system is an example of a semilinear hyperbolic system and it can be written in vector (conservative) form as (1) where:

$$u = (\rho, m, z)^T \qquad f(u) = (m, z, m)^T \qquad R(u) = (0, 0, \tfrac{1}{2}(\rho^2 + m^2 - 2\rho z))^T.$$

The correction introduced by our method is given by Eq. (6), which is a vector whose non zero component is the third one, given by

$$\Delta R_3(\bar{u}) = \frac{\Delta x^2}{24}(\rho_x^2 + m_x^2 - \rho_x\, z_x) + O(\Delta x^3).$$

This expression does not vanish, in general, as $\varepsilon \to 0$, thus causing a degradation of the accuracy to second order.

We consider again GSA IMEX-RK schemes to guarantee that our technique provides an *underresolved numerical method* for the limit equation in the case $\varepsilon \to 0$. In this section we confirm numerically our findings of Sect. 2, i.e. that our method is also able to increase the order of accuracy of the scheme. We show that by the following example.

Example 1, Accuracy test. We test numerically the convergence rate of the numerical schemes solving the Broadwell model for a smooth solution. The initial data as in [30, 27] are given by

$$[\rho, v, z](x, 0) = \left[1 + a_\rho \sin \frac{2\pi x}{L}, \frac{1}{2} + a_v \sin \frac{2\pi x}{L}, \frac{a_z}{2}\rho(x, 0)(1 + v(x, 0)^2)\right]$$

in a periodic domain of length $L = 20$ and with $a_\rho = 0.3$, $a_v = 0.1$, $a_z = 1.0$.

We first check the accuracy of a standard finite volume IMEX-RK (FV-IMEX-RK) scheme in which the source term is approximated by $R(\bar{u}_j)$. The final time was set to 10, CFL number 0.45, the number of points was chosen between 50 and 1600 and the solution with 3200 points obtained with the same scheme has been used as a reference for the computation of the error. The L^1-norm errors in each variable, together with the experimental convergence rates, are reported in the tables below. The numerical integration was carried out with third order GSA FV-IMEX-RK scheme (41). In Table 1 the numerical results show that for large values of N the method tends to be second order accurate for various values of ε.

A more accurate computation can be performed by using a fifth order CWENO reconstruction in space [17], and by choosing $\Delta t = \mathcal{O}(\Delta x^{4/3})$, so that the overall error scales like Δx^4. The number of points was chosen between 50 and 400 (Table 2). The second order accuracy of the method is now evident even for moderate values of N.

Finally we check the ability of our penalized technique to increase the order of accuracy of the scheme. The numerical integration was carried out with third order GSA IMEX-RK scheme (41) with CFL number 0.45. The L^1 errors in each variable, together with the experimental convergence rates, are reported in Table 3 for various values of ε. The table shows a clear third order accuracy for all the tested values of the relaxation parameter.

In Table 4 we show the convergence results obtained by using again a fifth order CWENO reconstruction and $\Delta t = \mathcal{O}(\Delta x^{4/3})$, adopting our penalization technique.

Fourth order accuracy is evident for $\varepsilon = 1$ (non-stiff regime) and $\varepsilon = 10^{-6}$ (stiff regime). On the other hand we observe a loss of order on the variables for moderate values of ε, which is typical for hyperbolic systems with relaxation [10].

Note that by using a non GSA scheme, a degradation of accuracy is observed in the stiff regime, even using the penalization technique (see Table 5).

Third order GSA FV-IMEX-RK, $\varepsilon = 10^{-6}$

N	error ρ	rate	error v	rate	error z	rate
50	5.18e-03		3.54e-03		3.57e-03	
100	6.16e-04	3.07	4.31e-04	3.04	4.20e-04	3.09
200	7.32e-05	3.07	5.59e-05	2.95	5.09e-05	3.04
400	8.99e-06	3.03	8.17e-06	2.77	6.83e-06	2.90
800	1.37e-06	2.72	1.42e-06	2.52	1.20e-06	2.51
1600	2.83e-07	2.27	2.78e-07	2.35	2.41e-07	2.31

Third order GSA FV-IMEX-RK-NP, $\varepsilon = 10^{-3}$

N	error ρ	rate	error v	rate	error z	rate
50	5.17e-03		3.54e-03		3.57e-03	
100	6.15e-04	3.07	4.31e-04	3.04	4.20e-04	3.09
200	7.31e-05	3.07	5.58e-05	2.95	5.09e-05	3.04
400	8.94e-06	3.03	8.10e-06	2.78	6.80e-06	2.90
800	1.30e-06	2.78	1.39e-06	2.55	1.17e-06	2.53
1600	2.59e-07	2.33	2.66e-07	2.38	2.30e-07	2.35

Third order GSA FV-IMEX-RK-NP, $\varepsilon = 1$

N	error ρ	rate	error v	rate	error z	rate
50	3.49e-03		3.08e-03		2.88e-03	
100	4.15e-04	3.07	3.68e-04	3.07	3.37e-04	3.10
200	4.95e-05	3.07	4.39e-05	3.07	3.88e-05	3.12
400	5.77e-06	3.10	5.44e-06	3.01	4.74e-06	3.03
800	7.72e-07	2.90	7.92e-07	2.78	7.14e-07	2.73
1600	1.39e-07	2.47	1.36e-07	2.54	1.30e-07	2.45

Table 1: Convergence test for the Broadwell model with classical FV scheme in which the source term is approximated as $R(\bar{u})$.

Example 2, Riemann Problems. Now we consider two Riemann test problems in order to validate our assertion concerning the importance to have GSA IMEX R-K scheme in our penalized technique in order to obtain an *underresolved numerical scheme* for small values of ε. We have considered two different initial data as in [30, 27]

$$\begin{aligned} &\rho_l = 2, \quad m_l = 1, \quad z_l = 1, \quad x < 2, \\ &\rho_l = 1, \quad m_r = 0.13962, \quad z_r = 1, \quad x > 2, \end{aligned} \tag{27}$$

Fourth order scaling GSA FV-IMEX-RK, $\varepsilon = 10^{-6}$

N	error ρ	rate	error v	rate	error z	rate
50	2.95e-04		2.52e-04		2.17e-04	
100	7.71e-05	1.93	6.43e-05	1.97	5.84e-05	1.89
200	1.95e-05	1.99	1.62e-05	1.99	1.48e-05	1.98
400	4.82e-06	2.01	4.00e-06	2.01	3.68e-06	2.01

Fourth order GSA FV-IMEX-RK, $\varepsilon = 10^{-3}$

N	error ρ	rate	error v	rate	error z	rate
50	2.94e-04		2.51e-04		2.16e-04	
100	7.69e-05	1.93	6.42e-05	1.97	5.80e-05	1.90
200	1.94e-05	1.99	1.61e-05	1.99	1.48e-05	1.98
400	4.79e-06	2.02	3.99e-06	2.02	3.67e-06	2.01

Fourth order scaling GSA FV-IMEX-RK, $\varepsilon = 1$

N	error ρ	rate	error v	rate	error z	rate
50	1.09e-04		9.52e-05		9.13e-05	
100	3.39e-05	1.69	3.10e-05	1.62	2.99e-05	1.61
200	8.87e-06	1.93	8.18e-06	1.92	7.95e-06	1.91
400	2.21e-06	2.00	2.05e-06	2.00	1.99e-06	2.00

Table 2: Convergence test for the Broadwell model with fifth order CWENO, $\Delta t = O(\Delta x^{4/3})$, and source approximated as $R(\bar{u})$. The degradation of accuracy due to the second order approximation of the source term is more evident.

$$\begin{aligned} &\rho_l = 2, \quad m_l = 0, \quad z_l = 1, \quad x < 0.5, \\ &\rho_l = 0.2, \quad m_r = 0, \quad z_r = 1, \quad x > 0.5, \end{aligned} \tag{28}$$

we integrate the equations for $t \in [0, 0.5]$ with $N = 200$, and $\varepsilon = 1.0, 0.02, 10^{-8}$. The results are shown in Fig. 2 and Fig. 3 with $CFL = 0.5$ and for the initial data (27) and (28).

In Figure 2 we report numerical results obtained only for some classical third order IMEX RK schemes listed in Appendix and that we referred to as NGSA-IMEX-RK (42) and (43), and GSA-IMEX-RK (41), respectively. Similar results are obtained by using other third order GSA and NGSA IMEX-RK scheme presented in the literature as in [2, 13, 7, 10, 5]. In Fig. 2, both NGSA-IMEX-RK and GSA-IMEX-RK give an accurate description of the solution for large values of $\varepsilon = 1$ and for moderate values $\varepsilon = 0.02$. Instead for small values $\varepsilon = 10^{-8}$, the NGSA-IMEX-RK blows up, and we stop the computation at time $t = 0.35$.

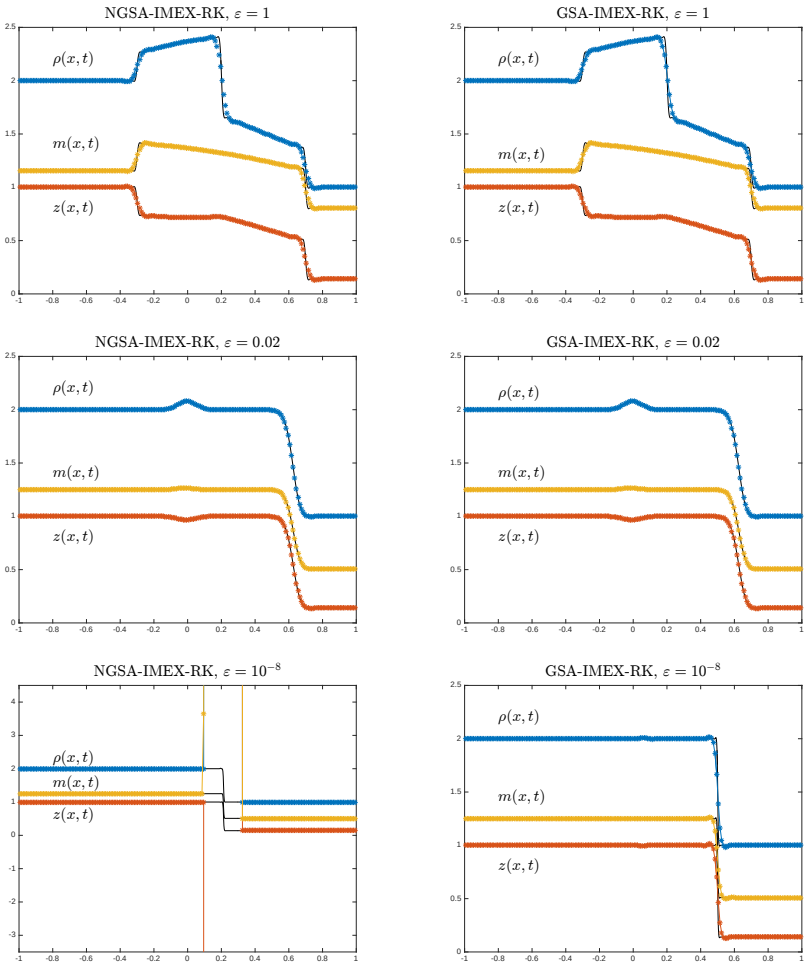

Figure 2: Broadwell model, example 2. Initial data (27). Comparison of the numerical solutions obtained with third order NGSA-IMEX-RK scheme (43) (left column), and third order GSA-IMEX-RK one (41), (right column), for different values of ε. The solid line is the reference solution. Final time is $t = 0.5$ for all panels except the lower left, which has been stopped at time $t = 0.35$.

GSA IMEX-RK, $\varepsilon = 1$

N	error ρ	rate	error v	rate	error z	rate
50	3.56e-03		3.14e-03		2.98e-03	
100	4.31e-04	3.05	3.82e-04	3.04	3.62e-04	3.04
200	5.32e-05	3.02	4.72e-05	3.02	4.49e-05	3.01
400	6.62e-06	3.01	5.87e-06	3.01	5.58e-06	3.01
800	8.15e-07	3.02	7.23e-07	3.02	6.88e-07	3.02

GSA IMEX-RK, $\varepsilon = 10^{-3}$

N	error ρ	rate	error v	rate	error z	rate
50	5.29e-03		3.50e-03		3.63e-03	
100	6.42e-04	3.04	4.16e-04	3.07	4.35e-04	3.06
200	7.92e-05	3.02	5.11e-05	3.03	5.34e-05	3.02
400	9.84e-06	3.01	6.32e-06	3.01	6.66e-06	3.00
800	1.22e-06	3.02	7.69e-07	3.04	8.43e-07	2.98

GSA IMEX-RK, $\varepsilon = 10^{-6}$

N	error ρ	rate	error v	rate	error z	rate
50	2.98e-03		2.04e-03		1.92e-03	
100	3.70e-04	3.01	2.57e-04	2.99	2.39e-04	3.01
200	4.59e-05	3.01	3.21e-05	3.00	2.98e-05	3.00
400	5.70e-06	3.01	4.01e-06	3.00	3.71e-06	3.01
800	6.97e-07	3.03	4.94e-07	3.02	4.56e-07	3.02

Table 3: Convergence test for the Broadwell model using our penalization technique and a fifth order CWENO reconstruction.

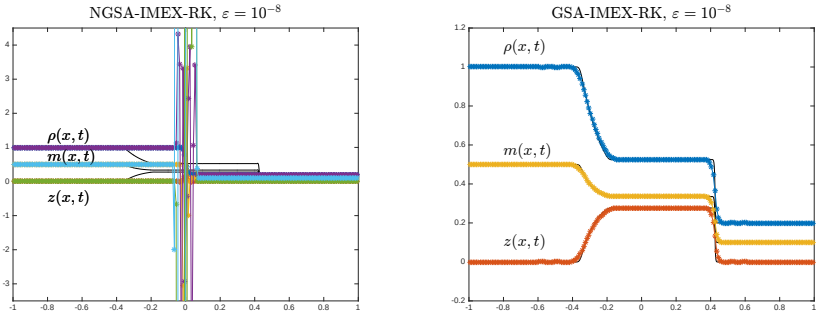

Figure 3: Broadwell model, example 2. Initial data (28) and $\varepsilon = 10^{-8}$. Numerical solutions obtained with third order NGSA-IMEX-RK scheme (42) at time $t = 0.01$ (left), and third order GSA-IMEX-RK (41) at time $t = 0.5$ (right). The solid line is the reference solution.

Fourth order scaling GSA IMEX-RK, $\varepsilon = 10^{-6}$

N	error ρ	rate	error v	rate	error z	rate
50	9.52e-05		7.24e-05		6.87e-05	
100	4.56e-06	4.39	3.78e-06	4.26	3.62e-06	4.25
200	2.40e-07	4.25	2.10e-07	4.17	2.02e-07	4.17
400	1.36e-08	4.14	1.22e-08	4.11	1.17e-08	4.11

Fourth order scaling GSA IMEX-RK, $\varepsilon = 10^{-3}$

N	error ρ	rate	error v	rate	error z	rate
50	9.52e-05		7.20e-05		6.87e-05	
100	4.54e-06	4.39	3.66e-06	4.30	3.93e-06	4.13
200	2.67e-07	4.09	2.10e-07	4.12	2.64e-07	3.89
400	5.21e-08	2.36	5.32e-08	1.98	3.79e-08	2.80

Fourth order scaling GSA IMEX-RK, $\varepsilon = 1$

N	error ρ	rate	error v	rate	error z	rate
50	7.56e-05		6.54e-05		6.69e-05	
100	3.96e-06	4.25	3.44e-06	4.25	3.52e-06	4.25
200	2.25e-07	4.14	1.94e-07	4.15	1.98e-07	4.15
400	1.34e-08	4.07	1.15e-08	4.08	1.18e-08	4.07

Table 4: Convergence test for the Broadwell model with penalized technique

Similarly in Fig. 3 we observe that NGSA-IMEX-RK schemes (43) blows up and (42) blows up after a short time $t = 0.01$. On the bottom the numerical solution is accurately computed by GSA-IMEX-RK (41) up to the final time $t = 0.5$. The "reference" solution is computed with $N = 2000$ points and GSA3-IMES-RK (41).

3.3. Some applications

Finally we present some numerical results obtained with GSA IMEX RK schemes concerning more general hyperbolic systems with stiff source term.

Euler equations with stiff friction. First we apply our method to the Euler equations of compressible gas dynamics with stiff friction [20]. This system reads as (1) with

$$u = \begin{pmatrix} \rho \\ \rho v \\ \rho E \end{pmatrix}, \quad f(u) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ (\rho E + p)v \end{pmatrix}, \quad R(u) = -\nu \begin{pmatrix} 0 \\ \rho v \\ \rho v^2 \end{pmatrix}. \quad (29)$$

NGSA IMEX-RK, $\varepsilon = 10^{-6}$

N	error ρ	rate	error v	rate	error z	rate
50	2.93e-03		1.99e-03		2.90e+00	
100	3.64e-04	3.01	2.50e-04	2.99	7.32e-02	5.31
200	4.51e-05	3.01	3.13e-05	3.00	1.82e-02	2.01
400	5.59e-06	3.01	3.91e-06	3.00	4.55e-03	2.00
800	6.83e-07	3.03	4.81e-07	3.02	1.15e-03	1.99

Table 5: Convergence rate obtained by a penalized method that adopts a third order non GSA IMEX-RK scheme.

The system is closed by an equation of state (EOS) of the form $p = p(u)$ and, using the ideal gas law, the equation of state reads as $p = (\gamma - 1)\rho(E - v^2/2)$ where $\gamma = 1.4$. For the numerical calculations we consider a computational domain $\Omega = [0, 1]$.

Two different initial conditions are considered for this model: the first one is a Riemann problem, and the second is a smooth one.

The initial conditions for the Riemann problem are the following:

$$u(x,0) = \begin{cases} (0.445, 0.3106, 8.9284) & \text{if } x \leq 0.5 \\ (0.5, 0.0, 1.4275) & \text{if } x > 0.5 \end{cases} \tag{30}$$

and for the stiffness parameter we take

$$\nu = \begin{cases} 0 & \text{if } x \leq 0.5, \\ 100 & \text{if } x > 0.5, \end{cases} \tag{31}$$

and

$$\nu = \begin{cases} 0 & \text{if } x \leq 0.5, \\ 10000 & \text{if } x > 0.5\,. \end{cases} \tag{32}$$

For all the following computations we set $CFL = 0.5$ and we solve the system up to $t = 0.1$. We consider second and third order GSA IMEX RK schemes given in the Appendix and we denoted as GSA2-IMEX-RK (40) and GSA3-IMES-RK (41). The results are depicted in Fig. 5.

The numerical results show the correct behavior in the stiff limit with respect to the reference solution computed with $N = 8000$ points and GSA3-IMES-RK (41).

By using a standard FV scheme with $R(\bar{u})$ with no correction, one obtains profiles that overlap with those shown in the figures, which shows at the

With correction			Without correction		
N	L_1-error	rate	N	L_1-error	rate
40	7.34e-03		40	6.82e-03	
80	5.57e-04	3.72	80	8.51e-04	3.00
160	3.59e-05	3.96	160	1.99e-04	2.09
320	2.00e-06	4.16	320	4.94e-05	2.01
640	8.14e-08	4.62	640	9.96e-06	2.31

Table 6: Full Euler with friction, smooth solution. Results for the density ρ.

same time that correction is quite small, and that the new schemes are able to capture shocks just as the classical ones.

For the smooth test we take an initial condition with constant data $\rho(x) = 1$, $v(x) = 1$ and $p(x) = 0.5$, and consider a variable friction coefficient of $\nu(x) = \hat{\nu}\,(\cos(4\pi x) - 1)$. The gas is slowed down considerably near the maxima of $\nu(x)$ and waves emerge from the flow. We computed numerical solutions for $\hat{\nu} = 100$ at final time $t = 0.05$, when the flow is still smooth. The numerical integration was carried out with third order GSA IMEX-RK scheme (41) at different spatial resolution, with CFL number 0.45 and periodic boundary conditions. The term treated implicitly in the ODE for the j-th cell is always $\nu(x_j)R(\bar{u}_j)$. We compare the cases in which we do not or we do employ our technique, which consists in adding to the explicit part of the scheme the difference $\langle \nu(x)R(u(x))\rangle_{\Omega_j} - \nu(x_j)R(\bar{u}_j)$. Notice that the explicit correction takes into account not only the proper approximation of the cell average of the source using the high order reconstruction $u(x)$, but also the fact that the friction coefficient is not constant. As in the case of the Broadwell model, we show the convergence results obtained by using again a fifth order CWENO reconstruction and $\Delta t = \mathcal{O}(\Delta x^{4/3})$. The solutions with 1280 points (with and without correction) have been used as references for the computation of the error. As it is clear from Table 6, the experimental order of convergence is approximately four only when our technique is proposed and is close to two otherwise. In Fig. 4 we report the error for the two methods as a function of the number of grid points. The degradation of accuracy for the method that does not use the correction is evident for small enough grid spacing.

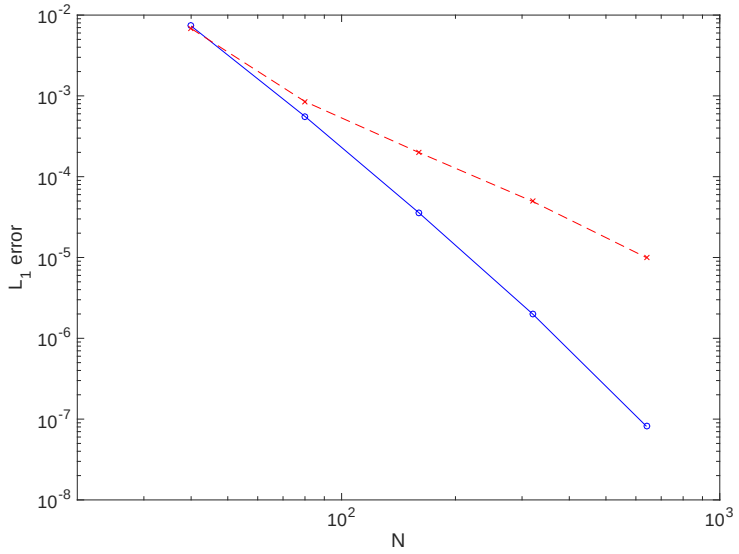

Figure 4: Full Euler with friction, smooth solution. L_1-error obtained with correction (blue continuous line) and without correction (red dashed in) for $N = 40, 80, 160, 320, 640$.

Reactive Euler-type equations. In this section we consider the set of reactive Euler-type equations derived from the kinetic theory of chemically reacting mixtures, [22, 16]. We consider a chemical reaction among four species of the form:

$$\mathcal{S}_1 + \mathcal{S}_2 \leftrightarrows \mathcal{S}_3 + \mathcal{S}_4.$$

We denote by m_i and E_i, with $i = 1, 2, 3, 4$, the particle masses and the energy levels corresponding to the chemical links. The total variation of internal energy of the reaction will be denoted by ΔE and is therefore given by $\Delta E = E_3 + E_4 - E_1 - E_2$.

This model is given by

$$\begin{aligned} &(\rho_i)_t + (\rho_i \mathbf{u})_{\mathbf{x}} = \mathcal{C}_i, \quad i = 1, ..., 4 \\ &(\rho \mathbf{u})_t + (\rho \mathbf{u} \otimes \mathbf{u} + p\mathbf{I})_{\mathbf{x}} = \mathbf{0}, \\ &(\mathcal{E})_t + ((\mathcal{E} + p)\mathbf{u})_{\mathbf{x}} = C_T, \end{aligned} \tag{33}$$

where $\mathcal{C}_i$, C_T are collision-like terms. The collision terms in the first equation are given by $\mathcal{C}_i = m_i \bar{\mathcal{C}}_i$, with

$$\bar{\mathcal{C}}_3 = \bar{\mathcal{C}}_4 = -\bar{\mathcal{C}}_1 = -\bar{\mathcal{C}}_2 = \mathcal{C}_{\text{chem}},$$

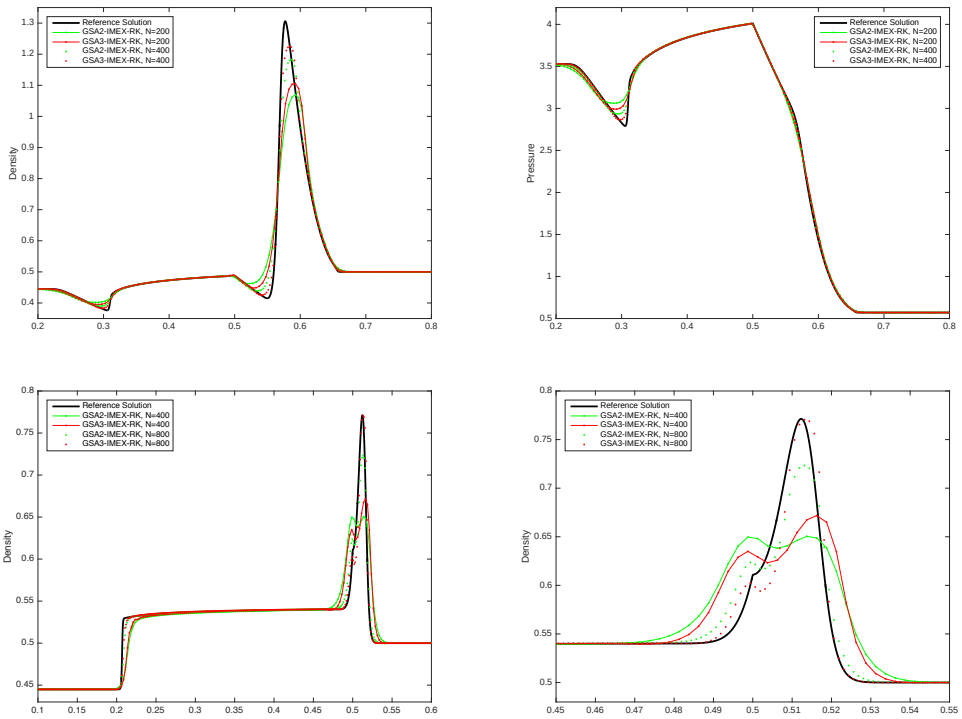

Figure 5: Reference solution (solid line) and numerical solutions at time $t = 0.1$ obtained using GSA2-IMEX-RK (40) and GSA3-IMEX-RK (41) for the Euler system with stiff friction. On top we plot the density and pressure computed at $N = 200$ and $N = 400$ with $\nu = 100$ for $x > 0.5$. Below on the left we plot the density computed with $N = 400$ and 800 with $\nu = 10000$ for $x > 0.5$. On the right we show a zoom of such density.

while in the energy equation $\mathcal{C}_T = -\Delta E\, \mathcal{C}_{\rm chem}$, where

$$\mathcal{C}_{\rm chem} = \frac{\gamma_T}{m_3 m_4} \left(\rho_1 \rho_2 \left(\frac{\mu_{34}}{\mu_{12}} \right) \exp(-\Delta E / K_B T) - \rho_3 \rho_4 \right).$$

Here the quantities μ denote the reduced masses: $\mu_{34} = \alpha_{34} m_4$, $\mu_{12} = \alpha_{12} m_2$, and $\alpha_{ij} = m_i/(m_i + m_j)$. Furthermore we have $\rho_i = m_i n_i$ for $i = 1, 2, 3, 4$, with n_i the number densities, $\rho = \sum_{i=1}^{4} \rho_i$ the total density, $\mathcal{E} = \rho E + \rho u^2/2$, T the kinetic temperature, K_B the Bolzmann constant and $\Delta E \geq 0$ (endothermic direct reaction).

Note that when $\gamma_T \to 0$ all the densities are just advected with the same velocity. Summing the first four equations, in this limit one obtains the well-known Euler equations of inviscid gas dynamics. On the other hand, for

larger γ_T, the stronger role played by the chemical reaction in the evolution makes system (33) stiff (see [22] for a detailed description of the mathematical model).

We perform a numerical simulation to test our technique when applied to the reactive Euler equation (33). As in the previous test, two different initial conditions are considered for this model: the first one is a Riemann problem, and the second is a smooth one.

First, we consider the equations in the interval $[0, 1]$ and final time $t = 0.07$ with the following initial data:

$$\begin{aligned} &\rho_1 = 1/10,\ \rho_2 = 2/10,\ \rho_3 = 3/10,\ \rho_4 = 4/10,\ u = 0,\ p = 5/3, \qquad x < 0.5 \\ &\rho_1 = 1/80,\ \rho_2 = 2/80,\ \rho_3 = 3/80,\ \rho_4 = 4/80,\ u = 0,\ p = 1/6, \qquad x > 0.5 \end{aligned} \tag{34}$$

with $\Delta E = 200$, $\gamma_T = 100$, $\Delta x = 1/200$. In this simulation, the chosen values for masses, appearing in the reactive term, are $m_1 = 58.5$, $m_2 = 18$, $m_3 = 40$, $m_4 = 36.5$.

Fig. 6 shows the behavior of the total density and velocity for type ARS GSA IMEX-RK schemes of different order. The schemes emploiyed are the first order (39), the second order (40) and the third order (41). It is worth noticing that our results capture the reference solution computed with 2000 cells and GSA3-IMEX-RK (41), and are in agreement with the results in [22]. Similarly to the Euler equations with stiff friction test, by using a standard FV scheme with $R(\bar{u})$ with no correction, one obtains profiles that overlap with those shown in Figure 6.

Finally, for the smooth test we take a $C^2([0, 1])$ initial condition with $\rho(x) = 1$, $v(x) = 0$ and

$$p(x) = \begin{cases} 1 + (1 + \cos((2\pi(x - 0.5))/0.2))^2/8 & \text{if } |\mathrm{x} - 0.5| \leq 0.1 \\ 1 & \text{otherwise} \end{cases}$$

The numerical integration was carried out with third order GSA IMEX-RK scheme (41) at different spatial resolution with CFL number 0.45 and final time $T = 0.2$. We report in Fig 7 the error for the two methods as a function of the number of grid points. The two methods give essentially the same result up to approximately $N = 160$, which is an indication that the correction term is quite small. A noticeable degradation of the order of accuracy for the method without the correction becomes evident at finer mesh.

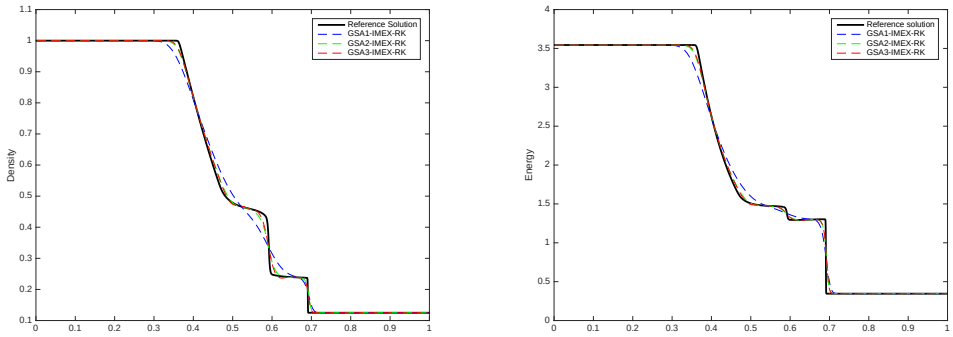

Figure 6: Reactive Euler-type equations, Riemann problem. Reference solution (solid line), total density (left) and energy (right) at time $t = 0.07$ obtained using GSA1-IMEX-RK, GSA2-IMEX-RK and GSA3-IMEX-RK, $\Delta E = 200$, $\gamma_T = 100$ and $N = 200$ number of cells.

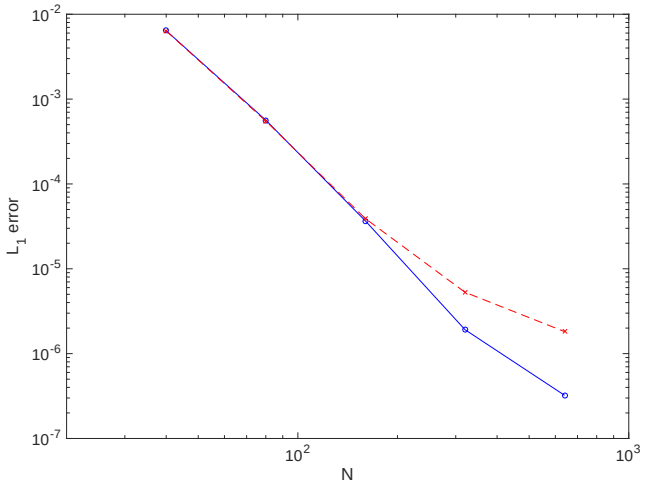

Figure 7: Reactive Euler-type equations, smooth solution. L_1-error obtained with correction (blue continuous line) and without correction (red dashed in) for $N = 40, 80, 160, 320, 640$.

4. Conclusions

In this paper we propose a new penalization technique to construct high order finite volume shock capturing schemes for hyperbolic systems with stiff relaxation.

The method is based on the idea of rewriting the average of the source in each cell as the sum of the source of the average plus a correction term: the first is treated implicitly, while the correction is treated explicitly. The main advantage of our technique over a high order fully implicit treatment of the source is to avoid coupling of neighboring cells at the level of the source term, thus dramatically improving efficiency over a fully implicit treatment of the source.

If, on the other hand, the source term is computed as a function of the average, a small error is introduced, which may have an influence if high accuracy is required.

We prove that in the stiff limit, i.e. $\varepsilon \to 0$, the scheme maintains the classical order of accuracy with no sign of order reduction. Globally stiffly accurate property for the IMEX R-K scheme is a fundamental assumption for very stiff systems, because it guarantees that the scheme is stable for small values of ε under a classical hyperbolic CFL condition that does not depend on ε. We have performed some numerical experiments that validate our method, and whose results are in agreement with the theoretical analysis: GSA IMEX-RK schemes appears to be stable and accurate even for extremely small values of ε, while lacking the GSA property results in code breaking, unless small scales are resolved. Furthermore, third order accuracy in space and time for smooth solution, and the capability of correctly capture shocks is verified numerically. A fifth order scheme in space has been also adopted, which better emphasizes the advantage of introducing the correction on the source term. Here we limit the construction and analysis to third order accuracy, both in space and time. Higher order accuracy would probably require additional conditions, which are currently under investigation.

5. Appendix.

Proof of Proposition 1

.

For the type CK, we have for the internal stages:

$$\begin{aligned} &\bar{U}_1 = \bar{u}^n, \\ &\hat{U} = \hat{U}^* + \frac{\Delta t}{\varepsilon} a R_1(\bar{U}) + \frac{\Delta t}{\varepsilon} \hat{A}\hat{R}(\bar{U}) \end{aligned} \tag{35}$$

with

$$\hat{U}^* = \bar{u}^n \hat{e} - \Delta t \left(\breve{a} \left(D_x F_1 - \frac{1}{\varepsilon} \Delta R_1 \right) + \breve{A} \left(D_x \hat{F} - \frac{1}{\varepsilon} \Delta \hat{R} \right) \right) \tag{36}$$

where we have used the notation in (14)

$$\bar{U} := \begin{pmatrix} U_1 \\ \hat{U} \end{pmatrix}, \; D_x F(\bar{U}) := \begin{pmatrix} D_x F_1 \\ D_x \hat{F} \end{pmatrix},$$

$$R(\bar{U}) := \begin{pmatrix} R_1 \\ \hat{R} \end{pmatrix}, \; \Delta R(\bar{U}) := \begin{pmatrix} \Delta R_1 \\ \Delta \hat{R} \end{pmatrix}$$

and $\hat{e} = (1, ..., 1)^T \in \mathcal{R}^{s-1}$, whereas for the numerical solution we have:

$$\begin{aligned} \bar{u}^{n+1} = & \bar{u}^n - \Delta t \left(\tilde{b}_1 \left(D_x F_1 - \frac{1}{\varepsilon} \Delta R_1 \right) + \breve{b}^T \left(D_x \hat{F} - \frac{1}{\varepsilon} \Delta \hat{R} \right) \right) \\ & + \frac{\Delta t}{\varepsilon} (b_1 R_1 + \hat{b}^T \hat{R}) \end{aligned}$$

where $\hat{b}^T = (b_2, \cdots, b_s)$ and $\breve{b}^T = (\tilde{b}_2, \cdots, \tilde{b}_s)$.

From (36) we get:

$$\frac{\Delta t}{\varepsilon} \hat{R}(\bar{U}) = \hat{A}^{-1} (\hat{U} - \hat{U}^*) - \hat{A}^{-1} a R_1,$$

and then for the numerical solution we have:

$$\bar{u}^{n+1} = u_n^* + R_1 (b_1 - \hat{b}^T \hat{A}^{-1} a) + \hat{b}^T \hat{A}^{-1} (\hat{U} - \hat{U}^*) \tag{37}$$

with

$$u_n^* = \bar{u}^n - \Delta t \, \tilde{b}_1 \, D_x F_1 + \frac{\Delta t}{\varepsilon} \tilde{b}_1 \Delta R_1 - \Delta t \, \breve{b}^T \, D_x \hat{F} + \frac{\Delta t}{\varepsilon} \, \breve{b}^T \, \Delta \hat{R}.$$

Substituting $\hat{U}^*$ in (37), we obtain:

$$\begin{aligned} \bar{u}^{n+1} = & u_n^* + R_1 (b_1 - \hat{b}^T \hat{A}^{-1} a) + \hat{b}^T \hat{A}^{-1} \hat{U} \\ & - \hat{b}^T \hat{A}^{-1} \left(\bar{u}^n \hat{e} - \Delta t \, \breve{a} \, D_x F_1 + \frac{\Delta t}{\varepsilon} \, \breve{a} \, \Delta R_1 - \Delta t \breve{A} D_x \hat{F} + \frac{\Delta t}{\varepsilon} \breve{A} \Delta \hat{R} \right). \end{aligned}$$

Then by some algebraic manipulation we obtain for the numerical solution

$$\begin{aligned}\bar{u}_{n+1} = &(1-\hat{b}^T\hat{A}^1\hat{e})\bar{u}^n + (b_1 - \hat{b}^T\hat{A}^{-1}\hat{a})R_1 + \hat{b}^T\hat{A}^{-1}\hat{U} \\ &- \Delta t\,(\tilde{b}_1 - \hat{b}^T\hat{A}^{-1}\breve{a})D_xF_1 - \Delta t\,(\breve{b}^T - \hat{b}^T\hat{A}^{-1}\breve{A})D_x\hat{F} \\ &+ \frac{\Delta t}{\varepsilon}(\tilde{b}_1 - \hat{b}^T\hat{A}^{-1}\breve{a})R_1 + \frac{\Delta t}{\varepsilon}(\breve{b}^T - \hat{b}^T\hat{A}^{-1}\breve{A})\Delta\hat{R}\end{aligned} \tag{38}$$

From the expression of the numerical solution it appears that requiring boundedness on $\bar{u}^{n+1}$ as $\varepsilon \to 0$, imposes that the terms: $\tilde{b}_1 - \hat{b}^T\hat{A}^{-1}\breve{a}$ and $\breve{b}^T - \hat{b}^T\hat{A}^{-1}\breve{A}$ have to vanish. Such terms are identically zero if the method is GSA. Note that GSA assumption also implies that the term $b_1 - \hat{b}^T\hat{A}^{-1}\hat{a}$ is zero.

We conclude the proof by using the following result:

Proposition 2. *If the IMEX R-K scheme of type CK is GSA then* $\breve{b}_1 - \hat{b}^T\hat{A}^{-1}\breve{A} = 0$, $\tilde{b}_1 - \hat{b}^T\hat{A}^{-1}\breve{a} = 0$ *and* $b_1 - \hat{b}^T\hat{A}^{-1}\hat{a} = 0$.

This proposition is easy to prove. In fact, by the GSA property (Definition 2) we have $e_s^T\tilde{A} = \tilde{b}^T$ and $e_s^TA = b^T$ and for a scheme of the type CK this reads $(\hat{e}_s^T\breve{a}, \hat{e}_s^T\breve{A}) = (\tilde{b}_1, \breve{b}^T)$, $(\hat{e}_s^T\hat{a}, \hat{e}_s^T\hat{A}) = (b_1, \hat{b}^T)$. The latter implies $\hat{b}^T\hat{A}^{-1} = \hat{e}_s^T$, therefore $\tilde{b}_1 - \hat{b}\hat{A}^{-1}\breve{a} = \tilde{b}_1 - \hat{e}_s^T\breve{a} = 0$, and $b_1 - \hat{b}^T\hat{A}^{-1}\hat{a} = b_1 - \hat{e}_s^T\hat{a} = 0$. Finally we have $\breve{b}^T - \hat{b}^T\hat{A}^{-1}\breve{A} = \breve{b}^T - \hat{e}_s^T\breve{A} = 0$.

As in the type A, under the assumption of GSA and proposition 2, from (38), the numerical solution becomes:

$$\bar{u}^{n+1} = \hat{e}_s^T\hat{\bar{U}}.$$

List of used schemes.

We list the IMEX RK schemes used in the numerical experiments.

1. First order GSA-IMEX-RK scheme of type ARS ([2]):

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 0 & 1 & 0 \\ \hline & 1 & 0 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 0 & 1 \\ \hline & 0 & 1 \end{array}. \tag{39}$$

Second order GSA-IMEX-RK scheme of type A:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & 1-\delta & \delta & 0 \\ \hline & 1-\delta & \delta & 0 \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}. \tag{40}$$

with $\gamma = 1 - 1/\sqrt{2}$, $\delta = 1 - 1/(2\gamma)$.

2. Third order GSA-IMEX-RK scheme of type ARS ([2]):

$$\begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 2/3 & 11/18 & 1/8 & 0 & 0 & 0 \\ 1/2 & 5/6 & -5/6 & 1/2 & 0 & 0 \\ 1 & 1/4 & 7/4 & 3/4 & -7/4 & 0 \\ \hline & 1/4 & 7/4 & 3/4 & -7/4 & 0 \end{array} \qquad \begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 2/3 & 0 & 1/6 & 1/2 & 0 & 0 \\ 1/2 & 0 & -1/2 & 1/2 & 1/2 & 0 \\ 1 & 0 & 3/2 & -3/2 & 1/2 & 1/2 \\ \hline & 0 & 3/2 & -3/2 & 1/2 & 1/2 \end{array} \tag{41}$$

3. Third order NGSA-IMEX-RK scheme of type ARS, ([2]):

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1084/2487 & 1084/2487 & 0 & 0 & 0 \\ 1181/1645 & 613/1908 & 332/837 & 0 & 0 \\ 1 & -880/8313 & 538/973 & 538/973 & 0 \\ \hline & 0 & 1849/1530 & -703/1091 & 1084/2487 \end{array} \tag{42}$$

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 1084/2487 & 0 & 1084/2487 & 0 & 0 \\ 1181/1645 & 0 & 464/1645 & 1084/2487 & 0 \\ 1 & 0 & 1.208496649 & -0.644363171 & 1084/2487 \\ \hline & 0 & 1849/1530 & -703/1091 & 1084/2487 \end{array} .$$

4. Third order NGSA-IMEX-RK scheme of type A, ([30]):

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/4 & 1/4 & 0 \\ \hline & 0 & 1/6 & 1/6 & 2/3 \end{array} \qquad \begin{array}{c|cccc} \alpha & \alpha & 0 & 0 & 0 \\ 0 & -\alpha & \alpha & 0 & 0 \\ 1 & 0 & 1-\alpha & \alpha & 0 \\ 1/2 & \beta & \eta & \delta & \alpha \\ \hline & 0 & 1/6 & 1/6 & 2/3 \end{array} . \tag{43}$$

with $\alpha = 0.24169426078821$, $\beta = 0.06042356519705$, $\eta = 0.12915286960590$, and $\delta = 1/2 - \beta - \eta - \alpha$.

Acknowledgments

The authors would like to thank Prof. Maria Groppi for useful discussion on the models of reacting Euler equations. The work has been partially supported by ITN-ETN Horizon 2020 Project *ModCompShock, Modeling and Computation on Shocks and Interfaces*, Project Reference 642768, by the National Group for Scientific Computing INdAM-GNCS project 2017: *Numerical methods for hyperbolic and kinetic equation and applications*, and by

the project F.I.R. *Charge transport in graphene and low dimensional systems*, University of Catania.

[1] A. M. Anile and S. Pennisi. Thermodynamic derivation of the hydrodynamical model for charge transport in semiconductors. *Phys. Rev. B*, 46:13186–13193, Nov 1992.

[2] Uri M Ascher, Steven J Ruuth, and Raymond J Spiteri. Implicit-Explicit Runge-Kutta methods for time-dependent partial differential equations. *Applied Numerical Mathematics*, 25(2):151–167, 1997.

[3] Dinshaw S Balsara, Takanobu Amano, Sudip Garain, and Jinho Kim. A high-order relativistic two-fluid electrodynamic scheme with consistent reconstruction of electromagnetic fields and a multidimensional Riemann solver for electromagnetism. *Journal of Computational Physics*, 318:169–200, 2016.

[4] Mapundi K Banda and Mohammed Seaid. Higher-order relaxation schemes for hyperbolic systems of conservation laws. *Journal of Numerical Mathematics jnma*, 13(3):171–196, 2005.

[5] S Boscarino and L Pareschi. On the asymptotic properties of IMEX Runge-Kutta schemes for hyperbolic balance laws. *Journal of Computational and Applied Mathematics*, 316, 2017.

[6] Sebastiano Boscarino. Error analysis of IMEX Runge-Kutta methods derived from differential-algebraic systems. *SIAM Journal on Numerical Analysis*, 45(4):1600–1621, 2007.

[7] Sebastiano Boscarino. On an accurate third order implicit-explicit Runge-Kutta method for stiff problems. *Applied Numerical Mathematics*, 59(7):1515–1528, 2009.

[8] Sebastiano Boscarino, Philippe G LeFloch, and Giovanni Russo. High-order asymptotic-preserving methods for fully nonlinear relaxation problems. *SIAM Journal on Scientific Computing*, 36(2):A377–A395, 2014.

[9] Sebastiano Boscarino, Lorenzo Pareschi, and Giovanni Russo. Implicit-Explicit Runge-Kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit. *SIAM Journal on Scientific Computing*, 35(1):A22–A51, 2013.

[10] Sebastiano Boscarino and Giovanni Russo. On a class of uniformly accurate IMEX Runge-Kutta schemes and applications to hyperbolic systems with relaxation. *SIAM Journal on Scientific Computing*, 31(3):1926–1945, 2009.

[11] Sebastiano Boscarino and Giovanni Russo. Flux-explicit IMEX Runge-Kutta schemes for hyperbolic to parabolic relaxation problems. *SIAM Journal on Numerical Analysis*, 51(1):163–190, 2013.

[12] Russel E Caflisch, Shi Jin, and Giovanni Russo. Uniformly accurate schemes for hyperbolic systems with relaxation. *SIAM Journal on Numerical Analysis*, 34(1):246–281, 1997.

[13] Christopher A. Kennedy Carpenter and Mark H. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Applied Numerical Mathematics*, 44:139–181, 2003.

[14] C. Cercignani. *The Boltzmann Equation and Its Applications*. Applied Mathematical Sciences. Springer New York, 2012.

[15] Gui Qiang Chen, C David Levermore, and Tai-Ping Liu. Hyperbolic conservation laws with stiff relaxation terms and entropy. *Communications on Pure and Applied Mathematics*, 47(6):787–830, 1994.

[16] Fiammetta Conforto, Maria Groppi, and Alessandra Jannelli. On shock solutions to balance equations for slow and fast chemical reaction. *Applied Mathematics and Computation*, 206(2):892–905, 2008.

[17] I. Cravero, G. Puppo, M. Semplice, and G. Visconti. CWENO: uniformly accurate reconstructions for balance laws. *Math. of Comp.*, in press. Preprint on arXiv at `https://arxiv.org/abs/1607.07319`.

[18] I. Cravero and M. Semplice. On the accuracy of WENO and CWENO reconstructions of third order on nonuniform meshes. *J. Sci. Comput.*, 2016.

[19] Giacomo Dimarco and Lorenzo Pareschi. Asymptotic preserving Implicit-Explicit Runge-Kutta methods for nonlinear kinetic equations. *SIAM Journal on Numerical Analysis*, 51(2):1064–1087, 2013.

[20] Michael Dumbser, Cedric Enaux, and Eleuterio F Toro. Finite volume schemes of very high order of accuracy for stiff hyperbolic balance laws. *Journal of Computational Physics*, 227(8):3971–4001, 2008.

[21] Francis Filbet and Shi Jin. A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources. *Journal of Computational Physics*, 229(20):7625–7648, 2010.

[22] Maria Groppi, Micol Pennacchio, et al. An IMEX finite volume scheme for reactive Euler equations arising from kinetic theory. *Communications in Mathematical Sciences*, 1(3):449–470, 2003.

[23] E. Hairer, S. P. Nørsett, and G. Wanner. *Solving Ordinary Differential Equations I (2nd Revised. Ed.): Nonstiff Problems*. Springer-Verlag New York, Inc., New York, NY, USA, 1993.

[24] E. Hairer and G. Wanner. *Solving Ordinary Differential Equations II. Stiff and Differential-Algebraic Problems. (2nd Revised. Ed.)*, volume 14 of *Springer Series in Comput. Mathematics*. Springer-Verlag New York, Inc., New York, NY, USA, 1996.

[25] Shi Jin. Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms. *Journal of Computational Physics*, 122(1):51–67, 1995.

[26] D. Levy, G. Puppo, and G. Russo. Central WENO schemes for hyperbolic systems of conservation laws. *M2AN Math. Model. Numer. Anal.*, 33(3):547–571, 1999.

[27] Salvatore Fabio Liotta, Vittorio Romano, and Giovanni Russo. Central schemes for balance laws of relaxation type. *SIAM Journal on Numerical Analysis*, 38(4):1337–1356, 2000.

[28] Tai-Ping Liu. Hyperbolic conservation laws with relaxation. *Communications on Mathematical Physics*, 108:153–175, 1987.

[29] Robert B. Lowrie and Jim E. Morel. *Discontinuous Galerkin for Hyperbolic Systems with Stiff Relaxation*, pages 385–390. Springer Berlin Heidelberg, Berlin, Heidelberg, 2000.

[30] Lorenzo Pareschi and Giovanni Russo. Implicit-Explicit Runge-Kutta schemes and applications to hyperbolic systems with relaxation. *Journal of Scientific computing*, 25(1-2):129–155, 2005.

[31] Tadeusz Platkowski and Reinhard Illner. Discrete velocity models of the Boltzmann equation: A survey on the mathematical aspects of the theory. *SIAM Review*, 30(2):213–255, 1988.

[32] G. Puppo and M. Semplice. Well-balanced high order schemes on non-uniform grids and entropy residuals. *J. Sci. Comput.*, 2016.

[33] M. Semplice, A. Coco, and G. Russo. Adaptive mesh refinement for hyperbolic systems based on third-order Compact WENO reconstruction. *J. Sci. Comput.*, 2016.

[34] Chi-Wang Shu and Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439 – 471, 1988.

[35] W.G. Vincenti and C.H. Krüger. *Introduction to Physical Gas Dynamics*. @Interscience tracts on physics and astronomy. John Wiley & Sons, 1965.

[36] G.B. Whitham. *Linear and Nonlinear Waves*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 2011.