Available online at www.sciencedirect.com
ScienceDirect
Journal of Computational Physics 227 (2008) 2317–2343
JOURNAL OF COMPUTATIONAL PHYSICS
www.elsevier.com/locate/jcp

Numerical solution of under-resolved detonations

Luca Tosatto [1], Luigi Vigevano *

Dipartimento di Ingegneria Aerospaziale, Politecnico di Milano, Via La Masa, 34, 20158 Milano, Italy

Received 16 November 2006; received in revised form 22 June 2007; accepted 15 October 2007
Available online 24 October 2007

Abstract

A new fractional-step method is proposed for the numerical solution of high speed reacting flows, where the chemical time scales are often much smaller than the fluid dynamical time scales. When the problem is stiff, because of insufficient spatial/temporal resolution, a well-known spurious numerical phenomenon occurs in standard finite volume schemes: the incorrect calculation of the speed of propagation of discontinuities. The new method is first illustrated considering a one-dimensional scalar hyperbolic advection/reaction equation with stiff source term, which may be considered as a model problem to under-resolved detonations. During the reaction step, the proposed scheme replaces the cell average representation with a two-value reconstruction, which allows us to locate the discontinuity position inside the cell during the computation of the source term. This results in the correct propagation of discontinuities even in the stiff case. The method is proved to be second-order accurate for smooth solutions of scalar equations and is applied successfully to the solution of the one-dimensional reactive Euler equations for Chapman–Jouguet detonations.

Keywords: Stiff source terms; Shock capturing; Finite volume; Reactive Euler equations

1. Introduction

The appropriate modelling of high speed reacting flows needs to account for non-equilibrium gasdynamics: in the inviscid case this leads to a (possibly stiff) non-homogeneous system of hyperbolic equations – the so-called reactive Euler equations – where the source terms account for the modifications to the mixture composition due to chemical reactions. A wide range of kinetic reaction rates may be present, and the chemical time-scales are often orders of magnitude smaller than the typical relaxation time of fluid dynamics, leading to the stiffness of the problem.

Even if unconditional stability can be obtained with an operator splitting approach and an implicit treatment of the reactive term, spurious numerical phenomena may occur when dealing with discontinuous solu-

* Corresponding author.
E-mail addresses: luca.tosatto@yale.edu (L. Tosatto), luigi.vigevano@polimi.it (L. Vigevano).
[1] Present address: Department of Mechanical Engineering, Yale University, 15 Prospect Street, New Haven, CT 06520-8284, United States.

0021-9991/$ - see front matter
doi:10.1016/j.jcp.2007.10.011

tions: shock-capturing numerical methods may possibly predict wrong propagation speeds of the discontinuities if the source term is not resolved with appropriate spatial/temporal accuracy.

This numerical phenomenon was first observed by Colella et al. [1] who considered both the reactive Euler equations and a simplified 2×2 system obtained by coupling the inviscid Burgers equation with a single advection/reaction one. They derived analytically the conditions at which these spurious solutions may be produced, but failed to give an explanation for this numerical artifact. Starting from the slightly different context of hyperbolic equation systems with stiff relaxation, Pember [2] put in evidence the same behavior, observed also by Ben-Artzi when numerically integrating the reactive Euler equations by solving a generalized Riemann problem [3]. LeVeque and Yee [4] showed that a similar spurious propagation phenomenon can be observed even with scalar equations, by properly defining a model problem with a stiff source term. The analysis of such a simple scalar problem allowed LeVeque and Yee to argue that the propagation error is due to the introduction of some numerical viscosity in the solution of the convective terms, which smears the discontinuity front and produces fictitious non-equilibrium values that, in turn, erroneously activate the source term. A slightly different scalar problem featuring the same behavior was considered in [5].

The correct propagation speed of the reactive front may be restored by using a front-tracking approach, like the ghost fluid/level set method of Nguyen et al. [6], by resorting to a local grid/time step refinement [7,8], that eliminates the local stiffness of the problem, or by a combination of both [9]. However, a very thin resolution is what one wants to avoid if only the general behavior of the fluid dynamical system and not the detailed investigation of the chemical phenomena is to be reproduced. Chorin's random choice scheme [10,11] has been successfully used in [1,12] for the solution of under-resolved detonation waves: its good performance is due to the lack of numerical diffusion. The method is based on the exact solution of Riemann problems at randomly chosen locations within the computational cells and does not need to introduce any smearing effect near discontinuities, hence it can correctly handle a stiff advection/reaction equation. However, the introduction of some numerical viscosity is an essential feature of many of the presently used shock-capturing schemes. Therefore, several modifications to shock-capturing methods have been presented in the literature, in order to correctly reproduce discontinuous solutions even in the under-resolved case. In [13], Engquist and Sjogreen proposed a rather ad hoc temperature extrapolation method, which uses an extrapolated temperature value from outside the shock profile to activate the chemical source term. Different methods based on a modification of the ignition temperature derived from a more sound physical or analytical basis are those of Ton [14] and Berkenbosch [15]. Helzel et al. [16] presented a modified fractional step method for the solution of detonation waves in which the exact solution of Riemann problems is used to define the reactive portion of the cell, so as to limit the influence of the reactive term across discontinuities. The method is extended in two dimensions by accounting for transverse propagation. It is a very elegant approach to cure the very cause of the problem but it may be applied only in conjunction with an exact Riemann solver for the conservation equations. Bao and Jin [17,18] achieved correct average propagation speed of the discontinuity by adopting a random projection method, which consists of replacing the ignition temperature (or the unstable equilibrium value of the source term in the scalar case) by a uniformly distributed random variable. This method has however two drawbacks: (i) it assumes a priori a stiff source, which prevents its use for non-stiff problems and (ii) for multidimensional flows it requires a rather complex bookkeeping to apply locally the random projection. Recently, Kurganov [19] has proposed a very simple fix, termed Accurate Deterministic Projection (ADP) method, for both scalar equations and reactive Euler equations, that decouples the flow solution from that of the scalar variables dominated by the source terms. Again the assumption of an uniformly stiff behavior of the source-dominated equation prevents using the method in a more general context of stiff/non-stiff problems.

In the present work, we will first consider a simple linear scalar advection/reaction equation, with suitably selected source term and initial conditions, to devise a fractional-step method that can correctly reproduce the speed of the discontinuity in under-resolved calculations for stiff source conditions. The scalar model problem, although insufficient to reproduce the physics of rapid combustion, does feature the same numerical difficulties observed in reacting flow problems and allows to easily understand the very nature of them [4]. For this reason it has been considered as a first step in the developments of numerical methods for this class of problems also in [7,17,19].

The proposed fractional-step algorithm is termed the *MinMax* scheme, because it is based on a two-value variable reconstruction within each cell, where appropriate maximum and minimum values of the unknown

are considered. The resulting scheme is very general; it may be applied with no difficulties to both stiff and non-stiff problems and will be then extended to deal with the reactive Euler equations.

The paper is organized as follows. In Section 2 we define the selected one-dimensional, scalar, discontinuous model problem and describe a standard, first-order fractional-step/shock-capturing finite volume numerical method to solve it. The spurious solutions obtained for the stiff source conditions will be briefly discussed following the arguments of [4]. Section 3 presents the proposed scheme for the scalar case, starting from the variable reconstruction, followed by the treatment of the reaction and advection operators. The extrapolation procedure needed after the advection step is considered in detail, in order to guarantee that the resulting scheme preserves the accuracy of a standard finite volume method for the non-stiff case and correctly reproduces the propagation speed of a discontinuity for the stiff, under-resolved case. The section ends considering some results for stiff and non-stiff scalar test cases. Additional scalar model problems are presented in Section 4 to verify that the results obtained with a standard second-order finite volume method for a smooth problem are reproduced with the second-order version of the *MinMax* scheme. The latter is extended to the reactive Euler equations in Section 5, while some conclusions are drawn in Section 6.

2. The scalar problem

We consider the following simple model problem, first proposed in [4] and adopted, in a slightly modified nonlinear form, also in [17,19]:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = S(u) = v(1-u)(u-\beta)u \tag{1}$$

over $(x,t) \in (-\infty,+\infty) \times [0,+\infty)$, with initial conditions:

$$u(x,0) = \begin{cases} 1, & \text{if } x \leqslant x_d, \\ 0, & \text{if } x > x_d, \end{cases} \tag{2}$$

where x_d is the position of the initial discontinuity. In Eq. (1), v is a positive parameter that allows us to vary the stiffness of the problem and β is a second parameter $(0 < \beta < 1)$. Eq. (1) is an advection equation with constant propagation speed a and with a nonlinear source term $S(u)$ that becomes stiff for large v. The introduction of the more complex advection equation with variable propagation speed, as proposed in [7,17,19], does not modify the mechanism of generation of spurious wave speeds here investigated. Given the selected initial condition, it is easy to verify that the exact solution is

$$u(x,t) = \begin{cases} 1, & \text{if } x \leqslant x_d + at, \\ 0, & \text{if } x > x_d + at, \end{cases} \tag{3}$$

i.e. it is identical to the solution of the same equation without the source term and with the same initial data (2). The particular polynomial choice of the source term $S(u)$ allows for a simple interpretation of the solution. It may be noticed that the function $S(u)$ defines two stable equilibrium points, $u = 1$ and $u = 0$, and one unstable equilibrium point at $u = \beta$. Therefore, the solution $u(x,t)$ over the characteristic lines $x - at = \text{const}$ will tend to 1 or 0 depending on the initial conditions.

The problem (1),(2) here considered is very peculiar in both the form of the source term and the choice of the initial conditions, so that it cannot be considered a general example of an advection/reaction equation. It has been proposed by LeVeque and Yee [4] because it forces a standard finite volume solver to face the same difficulties encountered by the reactive Euler equations, as will be shown in this section. Some additional scalar problems will be considered later on in the paper.

To numerically solve problem (1)-(2), it is usual to adopt the so-called *fractional-step* approach. Considering a uniform space/time discretization like

$$\begin{aligned} x_i &= ih \quad i = 0,1,2,3,\ldots \\ t_n &= nk \quad n = 0,1,2,3,\ldots \end{aligned} \tag{4}$$

with $h = \Delta x$ and $k = \Delta t$, we approximate the exact solution at time t_{n+1} in terms of the solution $u(x, t_n)$ calculated at the last time level with

$$u(x, t_{n+1}) \approx \mathcal{R}^n \mathcal{A}^n u(x, t_n). \tag{5}$$

The operator $\mathcal{A}^n$ is defined as

$$\mathcal{A}^n u(x, t_n) = u^*(x, t_{n+1}), \tag{6}$$

where $u^*(x, t_{n+1})$ is the solution of the advection part of the problem on the time interval, that is

$$\begin{cases} \frac{\partial u^*}{\partial t} + a\frac{\partial u^*}{\partial x} = 0, & t_n \leqslant t \leqslant t_{n+1}, \\ u^*(x, t_n) = \tilde{u}^n(x), \end{cases} \tag{7}$$

where $\tilde{u}^n(x)$ is the initial condition that is inferred from some discrete representation of the solution at time t_n, for example a piecewise constant approximation of $u(x, t_n)$.

In a completely analogous way the operator $\mathcal{R}^n$ is defined as

$$\mathcal{R}^n u^*(x, t_{n+1}) = u^{**}(x, t_{n+1}), \tag{8}$$

where $u^{**}(x, t_{n+1})$ stands for the solution on a time step of the reaction problem:

$$\begin{cases} \frac{du^{**}}{dt} = S(u^{**}), & t_n \leqslant t \leqslant t_{n+1}, \\ u^{**}(x, t_n) = \tilde{u}^*(x), \end{cases} \tag{9}$$

where $\tilde{u}^*(x)$ is the initial condition that is provided by a suitable approximation of $u^*(x, t_{n+1}) = \mathcal{A}^n u(x, t_n)$.

While the fractional step approach (5) is known to be at best first-order accurate, second-order accuracy can be obtained using a three step procedure, the so-called *Strang-splitting* [20]. For sake of simplicity we restrict our attention here to first-order approximations, but the algorithm presented is proved to work also with higher order methods by immediate extension, as illustrated by the examples of second-order results given in Section 4; see also [21].

In this section, we solve the problem approaching both the reaction and advection operators by a standard finite volume method. We use 'finite volume' to describe any numerical method which is based on the partition of the space domain Ω into volumes Ω_i such that $\Omega_i \cap \Omega_j = \emptyset$ and $\bigcup_i \Omega_i = \Omega$. In this framework the numerical solution U_i^n at time t_n is considered to be an approximation of the average value of $u(x, t)$ on volume Ω_i:

$$U_i^n \approx \frac{1}{h} \int_{x_{i-1/2}}^{x_{i+1/2}} u(x, t_n) \mathrm{d}x, \tag{10}$$

where $x_{i-1/2}$ and $x_{i+1/2}$ are the coordinates of the cell boundaries.

To discretize the advection operator (6) we start by setting the initial condition in problem (7) $\tilde{u}^n(x)$ as a piecewise constant function, consistent with the vector of approximations U^n (see Fig. 1b):

$$\tilde{u}^n(x) = U_i^n \quad \text{for } x_{i-1/2} \leqslant x < x_{i+1/2}.$$

Starting from these initial data, the exact solution of the linear advection equation in (7) over the time interval k (Fig. 1c) is

$$u^*(x, t_{n+1}) = \tilde{u}^n(x - ak).$$

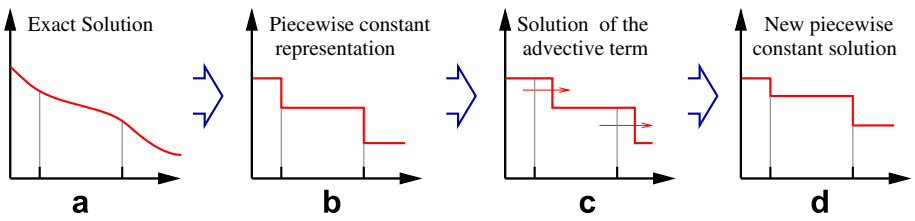

Fig. 1. Finite volume interpretation of the upwind method.

The discrete solution U_i^* is now defined by cell averaging $u^*(x, t_{n+1})$ (Fig. 1d):

$$U_i^* = \frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} u^*(x, t_{n+1})\mathrm{d}x.$$

The integral is immediately obtained since $u^*(x, t_{n+1})$ is piecewise constant. Whenever $0 < \frac{ak}{h} < 1$ we obtain

$$U_i^* = \left(1 - \frac{ak}{h}\right)U_i^n + \frac{ak}{h}U_{i-1}^n. \tag{11}$$

It is well known that this method is first-order accurate for smooth solutions, provided $0 < \frac{ak}{h} < 1$. Second-order accuracy can be achieved by using a piecewise linear reconstruction for the initial condition $\tilde{u}^n(x)$. It has been proved [4,21], however, that using higher order methods does not significantly change the behavior of the solution in the stiff case.

For the reaction operator (8), we still consider a piecewise constant initial condition $\tilde{u}^*(x)$ in problem (9):

$$\tilde{u}^*(x) = U_i^* \quad \text{for } x_{i-1/2} \leqslant x < x_{i+1/2}.$$

This leads to the solution of an ordinary differential problem in each cell. Using a linearized implicit Euler scheme we obtain

$$U_i^{n+1} = U_i^* + \frac{kS(U_i^*)}{1 - kS'(U_i^*)},$$

where S' is the Jacobian of the source term.

Summarizing, the numerical solution of Eq. (1) is calculated by the following standard finite volume algorithm:

$$\begin{aligned} U_i^* &= U_i^n - \frac{ak}{h}\left(U_i^n - U_{i-1}^n\right) \quad [\approx \mathcal{A}^k\tilde{u}^n(x)],\\ U_i^{n+1} &= U_i^* + \frac{kS(U_i^*)}{1 - kS'(U_i^*)} \quad [\approx \mathcal{R}^k\tilde{u}^*(x)]. \end{aligned} \tag{12}$$

From the expression of the source term in (1), it can be seen that the significant parameter to characterize the stiffness of the problem is kv.

The results computed with $a = 1$ (Fig. 2) show that in the non-stiff case (i.e. small values of kv) the problem is correctly solved, while in the stiff case the numerical solution overshoots and predicts a completely wrong propagation speed of the discontinuity. Similar results are obtained in [4] where different solvers are analyzed (see also [21]).

A detailed analysis of this wrong propagation phenomenon is reported in [4]; here we only recall some basic concepts that will be useful later. To pinpoint the source of the wrong propagation speed it is convenient to analyze with the help of Fig. 3 the first time step of algorithm (12).

- The numerical method initially applies the advection operator only. Starting from a discontinuous initial condition an intermediate state is introduced (see Fig. 3b). Such an intermediate state does not exist in the exact solution of the differential problem (which consists of a step between $u = 1$ and $u = 0$), but can anyhow be considered the most correct approximation to the differential problem as it respects the cell average of the exact solution.
- Although the intermediate state is a correct approximation for the advection problem, it generates a propagation error when the reaction solver is activated (see Fig. 3c). The intermediate state is a non-equilibrium condition for the source term and will hence be changed during the reaction step. If the problem is stiff (i.e. the relaxation time for the reaction problem is short compared to the time step used), then the intermediate state will be brought to the nearest stability condition, in the example of Fig. 3 to $u = 1$. This is why at the end of the integration step the discontinuity has moved a whole computational cell.

It can hence be inferred that when approaching the advection/reaction problems (1) and (2) with method (12) the observed wrong propagation speed of the discontinuity in the stiff case is due to the intermediate state

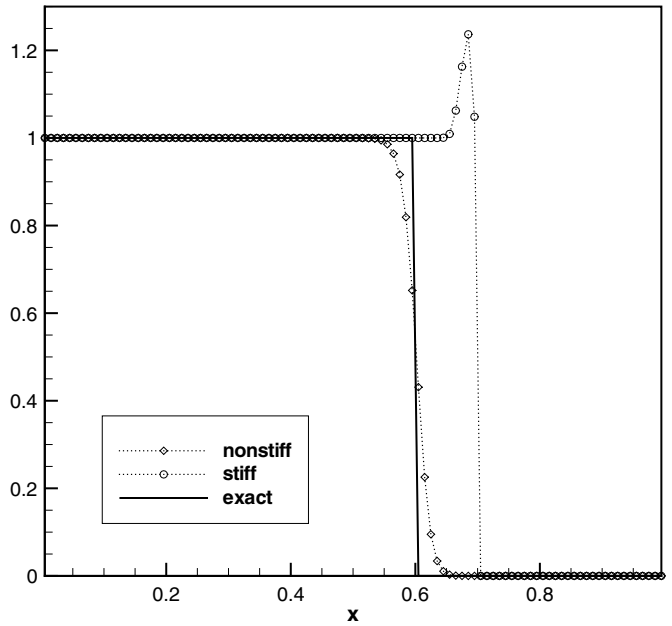

Fig. 2. Results obtained with the standard fractional step method (12) at $t = 0.3$ for problem (1)–(2) with $a = 1$, $\beta = 0.5$, in non-stiff ($kv = 0.15$) and stiff ($kv = 15$) cases. $h = 0.01$, $ak/h = 0.75$.

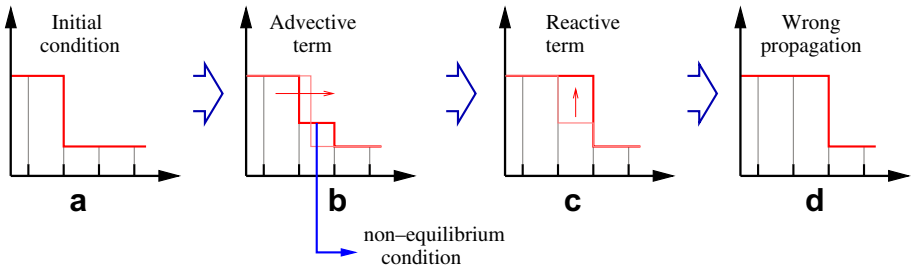

Fig. 3. Evolution of the numerical solution in one time step.

generated by the advection operator. This intermediate state constitutes a non-equilibrium condition for the source term and wrongly activates the source itself. This kind of behavior is not due to the particular solver considered, but it is typical of many finite volume solvers [22]: a conservative shock-capturing method needs to introduce some intermediate state to represent a discontinuity within a cell, but as soon as a non-equilibrium value is introduced, the source term mispropagates the discontinuity in attempt to reach an equilibrium condition.

3. The proposed method

A different treatment of the reaction operator is now introduced. Such a modified approach does not allow the intermediate condition to react and hence reproduces the correct propagation speed of the discontinuity.

The basic idea of the proposed method is to replace – during the reaction step – the representation of the unknown based on the cell average U_i^n with a more detailed one. We will identify the solution through three parameters: $\overline{U}_i^n$ and $\underline{U}_i^n$, which can be considered as an approximation of the maximum and minimum value of $u(x,t)$ in the cell, and U_i^n, which still represents the cell average value of the unknown, as defined in (10). This means that the present method needs three times more computational memory than a standard finite volume

approach. Nevertheless, it can still be considered more efficient, considering the very fine grid one needs to employ with the standard approach to correctly capture the shock velocity in the stiff case.

We begin by formalizing the method describing the reconstruction of the unknown inside the cell; then this reconstruction will be applied to the reaction and advection operators. It should be noted that for clarity of exposition the fractional-step approach is applied in this section as

$$u(x,t_{n+1}) \approx \mathcal{A}^n \mathcal{R}^n u(x,t_n), \tag{13}$$

i.e. in a reverse but strictly equivalent order with respect to the previous section. Consequently, the intermediate solution $u^*(x,t)$ represents hereafter the outcome of the *reaction* operator.

3.1. Cell reconstruction

We introduce the local coordinate ξ_i for the ith cell:

$$\xi_i(x) = \frac{x - x_{i-1/2}}{x_{i+1/2} - x_{i-1/2}}, \quad 0 \leqslant \xi_i(x) \leqslant 1 \tag{14}$$

and we define $u^n(\xi_i)$ as an approximation of $u(x,t)$ in the ith cell at time t_n. The function $u^n(\xi_i)$ must satisfy the following two constraints with respect to the triplet $\overline{U}_i^n$, $\underline{U}_i^n$ and U_i^n:

(1) Consistency with the cell average:

$$\int_0^1 u^n(\xi_i)\mathrm{d}\xi_i = U_i^n. \tag{15}$$

(2) $u^n(\xi_i)$ can assume only two discrete values, namely $\overline{U}_i^n$ and $\underline{U}_i^n$:

$$u^n(\xi_i) = \overline{U}_i^n \ \vee \ u^n(\xi_i) = \underline{U}_i^n \quad \forall \xi_i \in [0,1]. \tag{16}$$

The simplest example of $u^n(\xi_i)$ satisfying the conditions (15) and (16) is (see Fig. 4a):

$$u^n(\xi_i) = \begin{cases} \overline{U}_i^n, & \text{if } \xi_i \leqslant \gamma_i^n, \\ \underline{U}_i^n, & \text{if } \xi_i > \gamma_i^n, \end{cases} \tag{17}$$

where γ_i^n is the portion of cell occupied by the value $\overline{U}_i^n$, which can be determined from the numerical solution by a straightforward calculation of the average:

$$\gamma_i^n = \frac{U_i^n - \underline{U}_i^n}{\overline{U}_i^n - \underline{U}_i^n}. \tag{18}$$

We can observe that any two-value, piecewise constant reconstruction within the cell, characterized by the same value of γ_i^n, will be equivalent with respect to the reaction operator $\mathcal{R}^n$, which refers to the time-dependent ordinary differential equation in (9). For instance, a reconstruction that adopts the reversed order (Fig. 4b)

$$u^n(\xi_i) = \begin{cases} \underline{U}_i^n, & \text{if } \xi_i \leqslant 1 - \gamma_i^n, \\ \overline{U}_i^n, & \text{if } \xi_i > 1 - \gamma_i^n \end{cases} \tag{19}$$

is strictly equivalent to (17), as far as the reaction operator is concerned.

3.2. Reaction operator

Once the internal structure of the cell has been defined, it is possible to solve the reaction problem (9) in a consistent way with the two-value representation. Given the initial condition $u^n(\xi)$, the solution of the differential problem (9) can be obtained by a separate integration of the source term in each part of the cell:

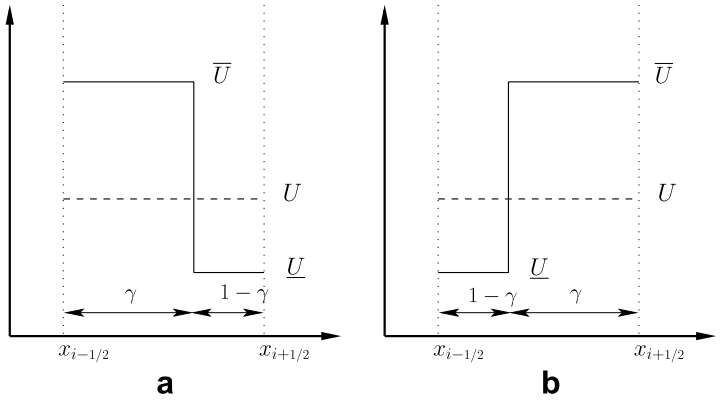

Fig. 4. Two-value reconstruction within one cell: (a) Eq. (17) and (b) Eq. (19).

$$\begin{aligned}\overline{U}_i^* &= \overline{\mathcal{R}}^n \overline{U}_i^n,\\ \underline{U}_i^* &= \overline{\mathcal{R}}^n \underline{U}_i^n,\\ \gamma_i^* &= \gamma_i^n,\end{aligned} \tag{20}$$

where $\overline{\mathcal{R}}^n$ is defined by the approximate solution of the differential problem (9) over the time step k. The fraction of cell γ_i occupied by the maximum value stands unchanged after solving the reaction operator. This allows us to compute the new average value after the integration of the source term as

$$U_i^* = \overline{U}_i^* \gamma_i^* + \underline{U}_i^* (1 - \gamma_i^*). \tag{21}$$

Remark 3.1. It can be noticed that the algorithm (20) is independent of the average value U_i^n; this means that, in presence of a discontinuity, the intermediate average states generated by the solution of the advection problem will not influence the solution of the reaction operator. This property allows the algorithm (20) to correctly handle the reaction term even in presence of numerical viscosity, and therefore makes it a suitable candidate for the formulation of a method for the solution of discontinuous advection–reaction equations with stiff source term.

3.3. Advection operator

The integration of the advection problem (7) under the assumption of the two-value reconstruction (16) is not as immediate as for the reaction term. A consistent solution would require the integration of the homogeneous hyperbolic problem with a discontinuous initial solution within the cell. However, in the attempt to obtain a flexible method (i.e. a numerical method that can independently use a variety of solvers for both the advection and reaction terms) we solve the hyperbolic part of the problem relying only on the average value, as in a standard finite volume scheme, as

$$U_i^{n+1} = \overline{\mathcal{A}}^n (U^*; i). \tag{22}$$

Here, with $\overline{\mathcal{A}}^n$ we intend a generic solver for homogeneous hyperbolic equations, while the notation $(U^*; i)$ means that the solution depends on node i as well as on other points on the computational grid. The algorithm (22) computes the updated average value but not the entire triplet necessary to identify the *MinMax* solution. As a consequence the values $\overline{U}^{n+1}$, $\underline{U}^{n+1}$ must be somehow extrapolated starting from the average value U^{n+1}.

This extrapolation clearly introduces an error in the solution. It will be shown however that, under appropriate conditions, the numerical scheme can be second-order accurate.

3.4. Extrapolation of the internal structure of the cell

The core of the *MinMax* algorithm is the definition of the functions employed to extrapolate the $\overline{U}^{n+1}$ and $\underline{U}^{n+1}$ values. For the ith cell, we impose these extrapolation functions, hereafter indicated with the notation M and m, to depend *only* on the average value U_i^{n+1} and on a certain number of values of $\overline{U}^*$ and $\underline{U}^*$ in the neighboring cells. We can henceforth write

$$\begin{aligned} \overline{U}_i^{n+1} &= M(\mathbf{U}^*, U_i^{n+1}), \\ \underline{U}_i^{n+1} &= m(\mathbf{U}^*, U_i^{n+1}), \end{aligned} \tag{23}$$

where $\mathbf{U}^*$ is a set containing the values of $\overline{U}^*$ and $\underline{U}^*$ in cells i, $i+1$ and $i-1$:

$$\mathbf{U}^* = \{\overline{U}_i^*, \overline{U}_{i+1}^*, \overline{U}_{i-1}^*, \underline{U}_i^*, \underline{U}_{i+1}^*, \underline{U}_{i-1}^*\}. \tag{24}$$

This choice of $\mathbf{U}^*$ is justified in the context of explicit methods: the stability constraint imposes that the characteristic lines can only reach the adjacent cells in a single time integration step. This means that the values $\overline{U}_i^{n+1}$ and $\underline{U}_i^{n+1}$ must be influenced only by the adjacent cells.

Now that the arguments of the extrapolation functions are properly defined, we ask the function M to satisfy the following two properties:

(1) *Convergence order*: The function M performs an extrapolation and hence introduces a further approximation on the solution. We wish the introduced error to be sufficiently small to allow second-order accuracy. It is possible to demonstrate (see Appendix) that if the value $\overline{U}^{n+1}$ uniformly converges to U^{n+1}, then a second-order method is obtained. We will hence ask the M function to satisfy:

$$\lim_{h\to 0} \frac{\max_i\left(\overline{U}_i^{n+1} - U_i^{n+1}\right)}{h} < C. \tag{25}$$

These requirements assure that, if the unknown $u(x,t)$ is smooth, the modified method behaves like a standard finite volume method, and preserves its convergence and accuracy characteristics.

(2) *Correct discontinuity propagation*: The second property imposes that M must be defined such as to avoid any spurious propagation phenomena in the stiff case. To grant this we require

$$\overline{U}_i^{n+1} = M(\mathbf{U}^*, U_i^{n+1}) \in \mathbf{U}^*, \tag{26}$$

that is the function M does not compute a new $\overline{U}$ value but just "chooses" it from the values available in $\mathbf{U}^*$.

Unlike the convergence condition, we are unable to give a demonstration to prove that (26) can assure a correct discontinuity propagation. Regardless, a large amount of numerical experiments seems to confirm this proposition.

An argument in favour of (26) can be obtained by considering the case in which the right, u_r, and left, u_l, values of $u(x,t)$ on each side of the discontinuity are equilibrium conditions for the reaction term, like in problem (1)–(2). Under these conditions we can state that the set $\mathbf{U}^*$ will assume, close to the discontinuity, only the two values u_r, u_l:

$$\mathbf{U}^* = \{u_r, u_l\},$$

since the discrete reaction operator (20) leaves unaltered the solution, $\mathcal{R}^n u_r = u_r$ and $\mathcal{R}^n u_l = u_l$. Therefore, the extrapolation procedure will select as minimum and maximum values of the unknown again the u_r, u_l values, independent of the average intermediate state generated after the advection step, thus allowing the correct propagation of the discontinuity.

Identical constraints are used to define the m function.

Remark 3.2. It must be noticed that requirement (26) is somehow opposite to (25): in the presence of a discontinuity the maximum and minimum values of the approximation are not influenced by the grid size but only by the strength of the discontinuity. Regardless, we can expect a good choice of M and m to respect both conditions where $u(x,t)$ is smooth but only condition (26) near a discontinuity.

The M function used here is

$$M(\mathbf{U}^*; U_i^{n+1}) = \min(U \in \mathbf{U}^* | U \geqslant U_i^{n+1}), \tag{27}$$

where $\mathbf{U}^*$ has been defined in (24) as the set containing the maximum and minimum values of the approximation in the ith cell and in the neighboring ones. The M function chooses the new value $\overline{U}_i^{n+1}$ within $\mathbf{U}^*$ as the closest value to U_i^{n+1} greater than the average value itself.

Analogously we define:

$$m(\mathbf{U}^*; U_i^{n+1}) = \max(U \in \mathbf{U}^* | U \leqslant U_i^{n+1}). \tag{28}$$

An intuitive representation of the M and m functions is given in Fig. 5.

Remark 3.3. It must be noticed that in definition (27) it is necessary that at least one element of U^* be greater than U^{n+1}, otherwise the function M will not produce any result. This implies that it is required to solve the advection problem with a TVD scheme [22]. A non-monotone method could lead to the paradoxical situation in which

$$U < U_i^{n+1} \quad \forall U \ \in \mathbf{U}^*.$$

In this condition the M function loses its meaning because the new maximum introduced by a spurious oscillation of the advection operator does not fit the maximum values obtained during the solution of the reaction operator. This difficulty is easily solved by including U_i^{n+1} itself in the set $\mathbf{U}^*$.

Remark 3.4. Many schemes proposed in the literature for both the scalar problem and the detonation wave problem can be considered *MinMax* methods, in which a different choice for the functions M and m is made. In these schemes the reconstruction step is not based on properties (25) and (26), but on less speculative and more physical considerations. For example, for problem (1)–(2), if we take

$$\underline{U}_i^{n+1} = m(\mathbf{U}^*; U_i^{n+1}) = 0,$$
$$\overline{U}_i^{n+1} = M(\mathbf{U}^*; U_i^{n+1}) = \frac{U_i^{n+1}}{\gamma_R},$$

where $0 \leqslant \gamma_R \leqslant 1$ is evaluated as the fraction of cell occupied by the shock in the solution of a Riemann problem, we obtain an analogue of the method of Helzel et al. [16] for the scalar equation, while if we set $\underline{U}_i^{n+1} = 0$ and compute $\overline{U}_i^{n+1}$ as the outcome of a random projection step [17] we obtain a scheme similar to the Bao–Jin random projection method.

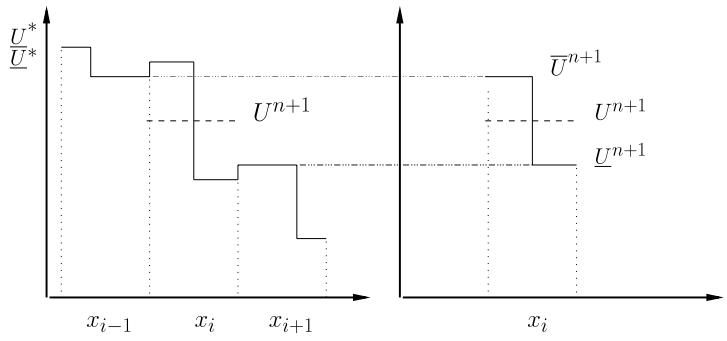

Fig. 5. An intuitive graphical representation of functions M and m.

3.5. The MinMax algorithm

It is useful at this point to summarize the structure of the modified fractional-step algorithm based on the two-value representation. The first-order *MinMax* algorithm follows four steps at each time integration:

(1) The reaction operator is solved in each computational cell using the procedure (20), as follows

$$\overline{U}_i^* = \overline{\mathcal{R}}^n \overline{U}_i^n,$$
$$\underline{U}_i^* = \overline{\mathcal{R}}^n \underline{U}_i^n,$$
$$\gamma_i^* = \gamma_i^n.$$

(2) After the reaction step, the cell average is computed using (21), so that a standard finite volume representation is obtained, namely

$$U_i^* = \overline{U}_i^* \gamma_i^* + \underline{U}_i^* (1 - \gamma_i^*).$$

(3) This average value is used as the initial condition for the computation of the advection step:

$$U_i^{n+1} = \overline{\mathcal{A}}^k (U^*; i).$$

(4) Now that the cell average at time t_{n+1} is known, the two-value structure of the solution is reconstructed via the functions M and m as

$$\overline{U}_i^{n+1} = M(\mathbf{U}^*, U_i^{n+1}),$$
$$\underline{U}_i^{n+1} = m(\mathbf{U}^*, U_i^{n+1})$$

and the variable γ_i^{n+1} is updated with (18) as

$$\gamma_i^{n+1} = \frac{U_i^{n+1} - \underline{U}_i^{n+1}}{\overline{U}_i^{n+1} - \underline{U}_i^{n+1}}.$$

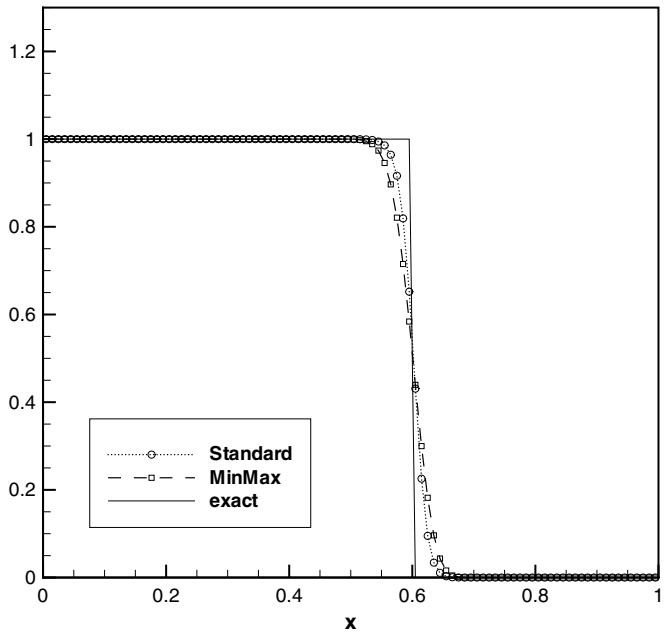

Fig. 6. Comparison of the standard fractional-step method (12) with the *MinMax* method, non-stiff case ($kv = 0.15$) at $t = 0.3$. $\beta = 0.5$, $h = 0.01$, $k/h = 0.75$.

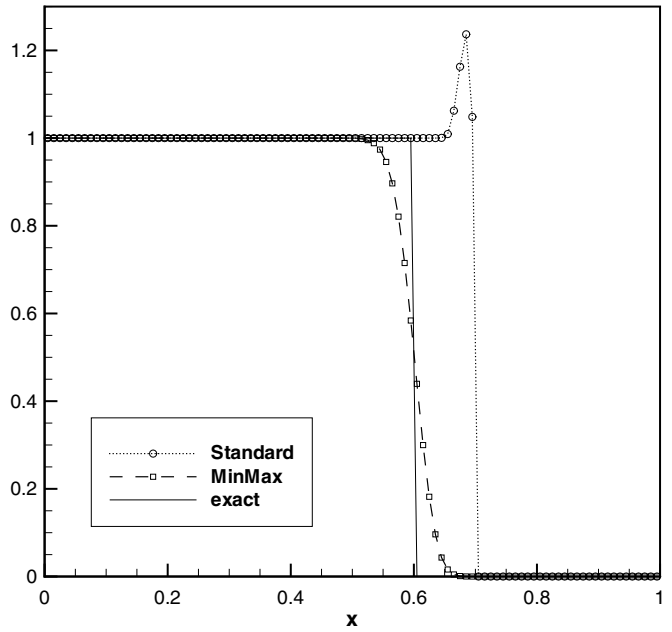

Fig. 7. Comparison of the standard fractional-step method (12) with the *MinMax* method, stiff case ($kv = 15$) at $t = 0.3$. $\beta = 0.5$, $h = 0.01$, $k/h = 0.75$.

3.6. Results

Figs. 6 and 7 present the results obtained with the standard-fractional step method described in Section 2 compared to our modified algorithm for the solution of problem (1)–(2) with $a = 1$. The two methods are almost equivalent in the non-stiff case, the *MinMax* scheme being slightly more dissipative. However, when the reaction term becomes dominant, only the modified method can correctly reproduce the position of the discontinuity.

4. Additional model problems

To verify that the *MinMax* scheme preserves the order of accuracy of a corresponding standard finite volume method for smooth solutions, we consider in this section two different scalar model problems, namely an initial value problem (IVP) in which a Gaussian pulse is damped by the source term, and an initial boundary value problem (IBVP) that allows for a steady-state solution. Both problems consider a scalar advection/reaction equation with linear source term:

$$\frac{\partial u}{\partial t} + a\frac{\partial u}{\partial x} = S(u) = -vu, \tag{29}$$

differing in the initial/boundary conditions.

The IVP reads: solve Eq. (29) over $(x,t) \in (-\infty, +\infty) \times [0, +\infty)$, with initial conditions

$$u(x,0) = u_0(x) = \exp\left(-\left(\frac{x - x_0}{\sigma}\right)^2\right), \tag{30}$$

where x_0 is the center of the initial pulse. This IVP is selected to examine the behavior of the scheme in computing a smooth, evolving solution. Since the source term is linear, it is easy to integrate the characteristic form of the equation to obtain the exact solution as

$$u(x,t) = u_0(x - at)\exp(-vt). \tag{31}$$

The IBVP reads: solve Eq. (29) over $(x,t) \in [0,+\infty) \times [0,+\infty)$, with initial and boundary conditions

$$\begin{cases} u(x,0) = 1, \\ u(0,t) = 1. \end{cases} \tag{32}$$

This second problem is chosen in an attempt of mimicking the reaction zone behind a strong detonation, where the combusting species are rapidly consumed by the chemical reactions. The exact solution in this case is as follows:

$$u(x,t) = \begin{cases} \exp(-vt), & \text{if } x - at > 0, \\ \exp(-vx), & \text{if } x - at \leqslant 0, \end{cases} \tag{33}$$

hence allowing for a steady-state solution for $at > 1$ in the domain $0 < x < 1$.

We solve the above problems with a second-order version of both the standard and the *MinMax* schemes. Strang-splitting [20] is used to reduce the splitting error. The advection operator is discretized in conservation form as

$$U_i^* = U_i^n - \frac{k}{h}\left(F_{i+\frac{1}{2}}^n - F_{i-\frac{1}{2}}^n\right), \tag{34}$$

where a numerical flux function corresponding to the second-order Lax–Wendroff scheme is selected:

$$F_{i+\frac{1}{2}} = a\frac{U_i^n + U_{i+1}^n}{2} - \frac{a^2 k}{2h}(U_{i+1}^n - U_i^n).$$

Finally, the reaction operator is solved with a linearized trapezoidal method as

$$U_i^{n+1} = U_i^* + \frac{\frac{k}{2}S(U_i^*)}{1 - \frac{k}{4}S'(U_i^*)}.$$

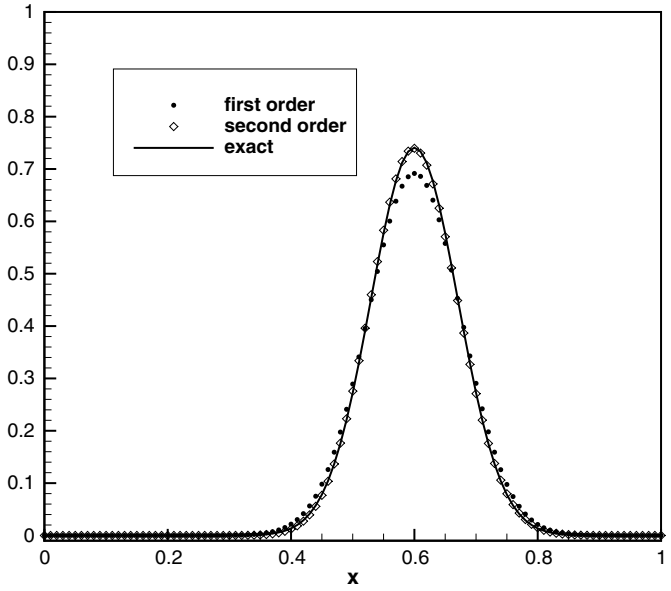

Fig. 8. Comparison of *MinMax* first and second-order results for the IVP: Gaussian pulse at $t = 0.3$; $h = 0.01$, $k/h = 0.75$.

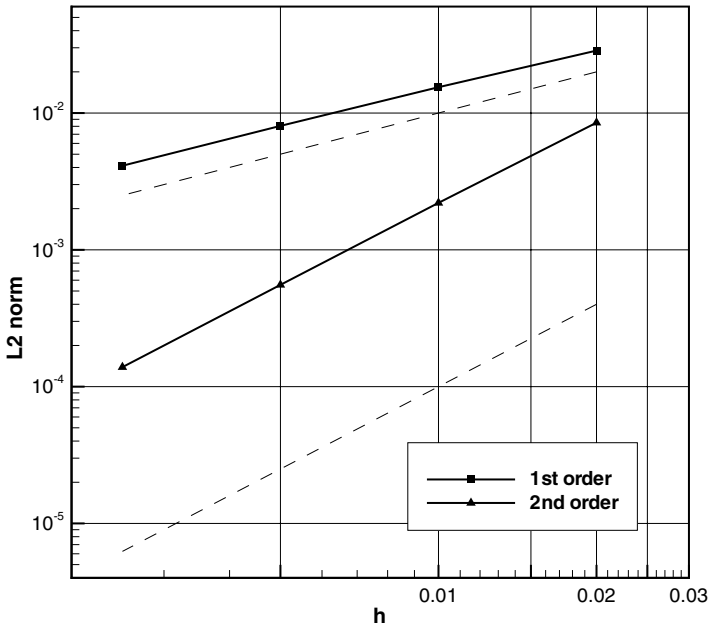

Fig. 9. L_2 norm of the results of Fig. 8. The dashed lines represent the h and h^2 error curves.

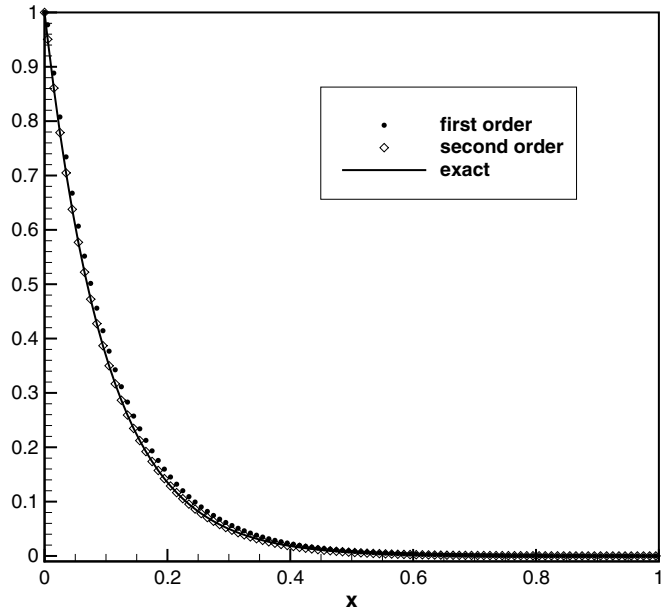

Fig. 10. Comparison of *MinMax* first and second-order results: IVBP at $t = 1.2$; $h = 0.01$, $k/h = 0.75$.

The structure of the *MinMax* algorithm is that described in Section 3.5, with the M and m extrapolation functions given by (27) and (28).

For the IVP, a Gaussian pulse with $\sigma = 0.1$, centered in $x_0 = 0.3$, is propagated from time $t = 0$ to time $t = 0.3$ with $a = 1$. The stiffness parameter is $v = 1$. Fig. 8 shows both first- and second-order *MinMax* solutions, that are virtually indistinguishable from those obtained with the standard finite volume scheme for this

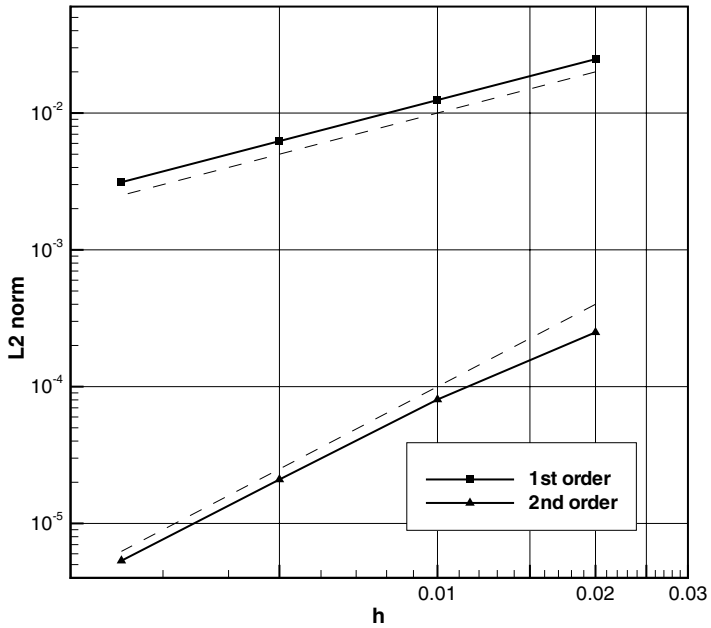

Fig. 11. L_2 norm of the results of Fig. 10. The dashed lines represent the h and h^2 error curves.

problem. The result of a convergence study is reported in Fig. 9, where the L_2 norm of the error is used. Second-order accuracy is obtained with the *MinMax* scheme, as for the standard method.

The IBVP (29)–(32) with $a = 1$ and $v = 10$ is integrated in the $0 < x < 1$ domain to time $t = 1.2$, so as to reach a steady-state condition in the considered domain. Fig. 10 shows the computed *MinMax* solutions, that again overlap the standard scheme results. The L_2 norm of the error, reported in Fig. 11, shows an effective second-order convergence.

We can conclude that the second-order *MinMax* algorithm behaves exactly as the corresponding standard fractional-step, second-order finite volume method for the smooth problems considered in this section.

5. The *MinMax* algorithm applied to the reactive Euler equations

We consider in this section the extension of the *MinMax* algorithm to the solution of the reactive Euler equations. Moving from a scalar equation to a system of partial differential equations poses some additional problems to the formulation of the algorithm that will be examined in the following sections.

5.1. The reactive Euler equations

The simplest description of a chemically reacting gas flow [23,16] assumes that the gas mixture is made only of two chemical species, respectively, burnt gas and unburnt gas. The unburnt gas is converted to burnt gas via a single irreversible reaction. The mixture state may be then represented by a single scalar variable, the mass fraction of the unburnt gas z. With the further assumption that gases in the mixture may be considered as ideal polytropic gases with equal ratio of specific heats γ and specific gas constant R, the reactive Euler equations in one dimension may be written as

$$\frac{\partial \mathbf{w}}{\partial t} + \frac{\partial \mathbf{f}(w)}{\partial x} = \mathbf{s}(w) \tag{35}$$

with

$$\mathbf{w} = \begin{Bmatrix} \rho \\ \rho u \\ E^{\mathrm{t}} \\ \rho z \end{Bmatrix}, \quad \mathbf{f} = \begin{Bmatrix} \rho u \\ \rho u^2 + P \\ u(E^{\mathrm{t}} + P) \\ \rho u z \end{Bmatrix}, \quad \mathbf{s} = \begin{Bmatrix} 0 \\ 0 \\ 0 \\ -K(T)\rho z \end{Bmatrix}, \tag{36}$$

where ρ is the mixture density, u the mixture velocity, $E^{\mathrm{t}} = \rho e + \frac{1}{2}\rho u^2$ the mixture total energy per unit volume, e the mixture specific internal energy, P the pressure, K the chemical reaction rate, and T represents the temperature. The two equations of state completing the model are

$$T = \frac{P}{\rho \mathsf{R}} \tag{37}$$

and

$$e = \frac{1}{\gamma - 1}\frac{P}{\rho} + q_0 z \tag{38}$$

where q_0 is the amount of heat per unit mass released in the chemical reaction.

The reaction rate of the irreversible chemical reaction, $K(T)$, is expressed in Arrhenius form as

$$K(T) = A \exp\left(\frac{-T_A}{T}\right), \tag{39}$$

where the pre-exponential coefficient A and the activation temperature T_A are empirical constants. When the reaction source term is stiff, however, the reaction rate may be simplified by adopting the so-called ignition temperature kinetic model, that is

$$K(T) = \begin{cases} v : T \geqslant T_{\mathrm{ign}}, \\ 0 : T < T_{\mathrm{ign}}, \end{cases} \tag{40}$$

where T_{ign} is a threshold ignition temperature and v represents the inverse of the characteristic time of the chemical reaction and determines the stiffness of the problem.

5.2. The standard fractional-step method

With the same fractional-step approach described for the scalar equation in Section 2 we approximate the exact solution on the grid defined in (4) as:

$$\mathbf{w}(x, t_{n+1}) \approx \mathcal{A}^n \mathcal{R}^n \mathbf{w}(x, t_n). \tag{41}$$

As before, the operator $\mathcal{R}^n$ is defined as

$$\mathcal{R}^n \mathbf{w}(x, t_n) = \mathbf{w}^*(x, t_{n+1}), \tag{42}$$

where $\mathbf{w}^*(x, t_{n+1})$ stands for the solution on a time step of the reaction problem:

$$\begin{cases} \frac{\mathrm{d}\mathbf{w}^*}{\mathrm{d}t} = \mathbf{s}(\mathbf{w}^n), \quad t_n \leqslant t \leqslant t_{n+1}, \\ \mathbf{w}^*(x, t_n) = \mathbf{w}_0^n(x) \end{cases} \tag{43}$$

with $\mathbf{w}_0^n(x)$ derived from a piecewise constant or linear approximation of $\mathbf{w}(x, t_n)$, while the operator $\mathcal{A}^n$ is defined as

$$\mathcal{A}^n \mathbf{w}^*(x, t_{n+1}) = \mathbf{w}^{**}(x, t_{n+1}), \tag{44}$$

where $\mathbf{w}^{**}(x, t_{n+1})$ is the solution of the advection part of the problem on the time interval, that is

$$\begin{cases} \frac{\partial \mathbf{w}^{**}}{\partial t} + \frac{\partial \mathbf{f}(\mathbf{w}^{**})}{\partial x} = 0, \quad t_n \leqslant t \leqslant t_{n+1}, \\ \mathbf{w}^{**}(x, t_{n+1}) = \mathbf{w}_0^*(x) \end{cases} \tag{45}$$

with $\mathbf{w}_0^*(x)$ inferred from $\mathbf{w}^*(x, t_{n+1})$. With a standard cell-centered formulation, the discrete values of the conservative variable vector, $\mathbf{W}_i^n$, approximate the cell averaging

$$\mathbf{W}_i^n = \frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} \mathbf{w}(x,t_n)\mathrm{d}x. \tag{46}$$

Considering the simple model (40) for the reaction constant, the reactive problem (43) has the following trivial analytical solution:

$$\begin{aligned}
&\rho_i^* = \rho_i^n,\\
&(\rho u)_i^* = (\rho u)_i^n,\\
&(E^{\mathrm{t}})_i^* = (E^{\mathrm{t}})_i^n,\\
&(\rho z)_i^* = \begin{cases} (\rho z)_i^n, & \text{if } T_i^n < T_{\mathrm{ign}},\\ (\rho z)_i^n \exp(-\nu k), & \text{if } T_i^n \geqslant T_{\mathrm{ign}}. \end{cases}
\end{aligned} \tag{47}$$

The advection problem is solved numerically with the popular Roe's scheme [24], extended to second-order accuracy with a flux limiting approach, blending upwind and Lax–Wendroff schemes with a Van Leer limiter [22] and using the entropy fix proposed by Harten and Hyman in [25].

The behavior of this standard method is tested on the example proposed in [16] for the same reactive Euler system considered here. The example consists of a Chapman–Jouguet detonation moving with constant unitary speed through the unburnt gas to the right of the domain. Setting $\gamma = 1.4$, $\mathsf{R} = 1$, $q_0 = 1$ and assigning the following *burnt* gas values:

$$\rho_b = \gamma, \quad u_b = 0, \quad P_b = 1, \quad z_b = 0$$

it is straightforward to compute the von Neumann state past the shock wave:

$$\rho_{\mathrm{vN}} = \frac{\gamma}{1-\delta}, \quad u_{\mathrm{vN}} = \delta, \quad P_{\mathrm{vN}} = 1 + \gamma\delta, \quad z_{\mathrm{vN}} = 1,$$

where $\delta = \sqrt{\frac{2(\gamma-1)}{\gamma+1}}$, and the *unburnt* gas values:

$$\rho_u = \frac{\gamma}{1+\delta}, \quad u_u = -\delta, \quad P_u = 1 - \gamma\delta, \quad z_u = 1$$

The resulting temperature of the unburnt gas, $T_u = 0.215995$, is only slightly lower than the assigned ignition temperature, $T_{ign} = 0.22$.

The initial condition for the computation consists of the cell integral values of the analytical solution, the front is placed at $x = -0.3$. The same cell-integrated detonation structure is used for comparison with the numerical experiments. It has to be noticed that a plot of the exact cell integral values will always smooth the peak von Neumann values after the shock, the amount of smoothing being strongly influenced by the width of the reaction zone, i.e. by the stiffness of the detonation itself, as can be observed in the presented results.

Fig. 12 shows the result achieved with the standard method for the non-stiff case ($k\nu = 0.1$): the peak values after the shock are quite well represented and the propagation speed of the detonation is properly predicted. At the intermediate value $k\nu = 1$ the numerical detonation speed is still correct, while the thinner reaction zone is less well resolved (Fig. 13). Finally, increasing the value of $k\nu$ to 10 brings us to the stiff case, where the numerical prediction (Fig. 14) shows a totally incorrect detonation propagating at the speed of one mesh cell per time step, followed by a non-reacting shock. This behavior of the numerical solution is similar to that presented and analyzed for the scalar case.

5.3. The MinMax method for a system of equations

To show how to extend the *MinMax* method to a system of hyperbolic equations like (35), we will first consider a very general approach in which the *MinMax* structure is applied to all the conservative variables. The resulting algorithm will be very similar to what is done in the scalar case, except for the necessity of *ordering*

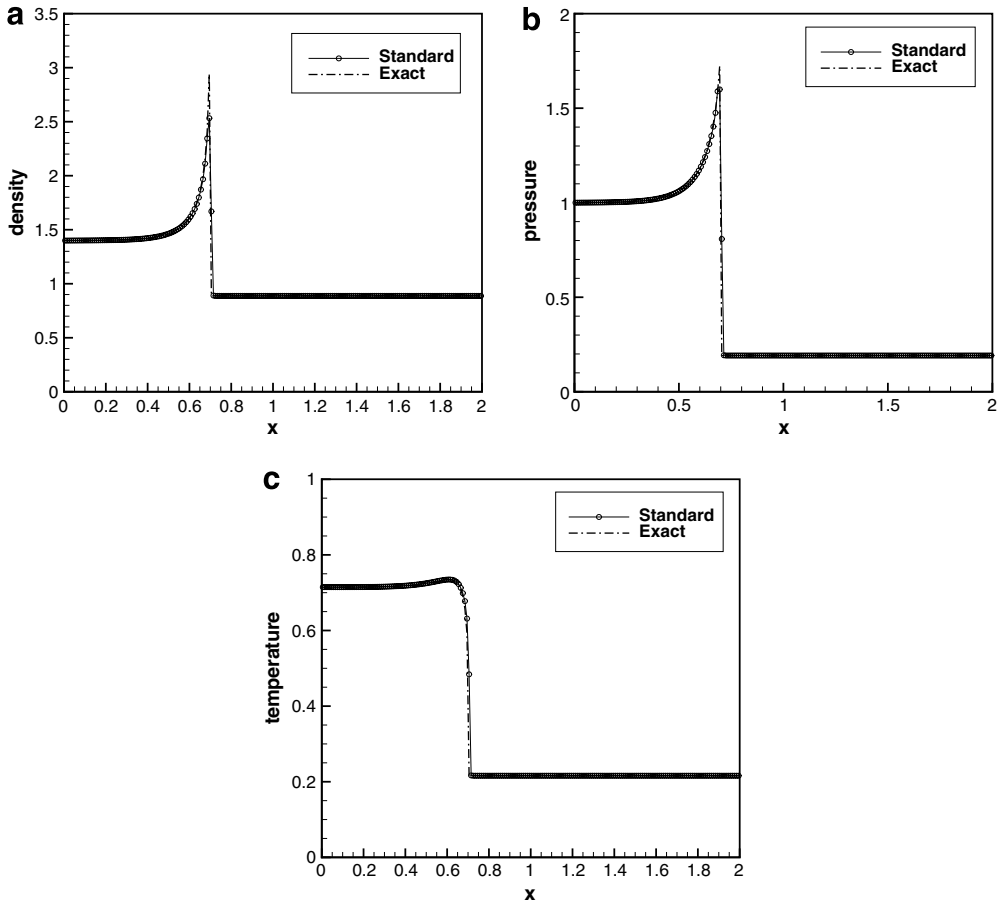

Fig. 12. Standard fractional-step method (Roe + VanLeer + HH2), $kv = 0.1$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density; (b) pressure; and (c) temperature.

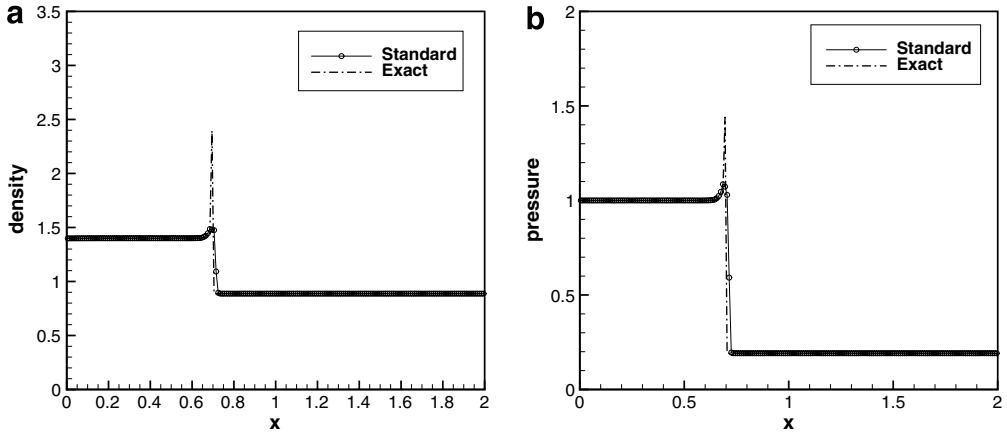

Fig. 13. Standard fractional-step method (Roe + VanLeer + HH2), $kv = 1$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

the variables inside the cell. This kind of approach will allow us to obtain the correct detonation structure for a one-dimensional problem, but its extension to multi-dimensional flow may not be straightforward.

We start by introducing two further vectors of unknowns, $\overline{\mathbf{W}}$ and $\underline{\mathbf{W}}$, that represent the maximum and minimum values of the conservative variables in every computational cell. Like in the scalar case, these further

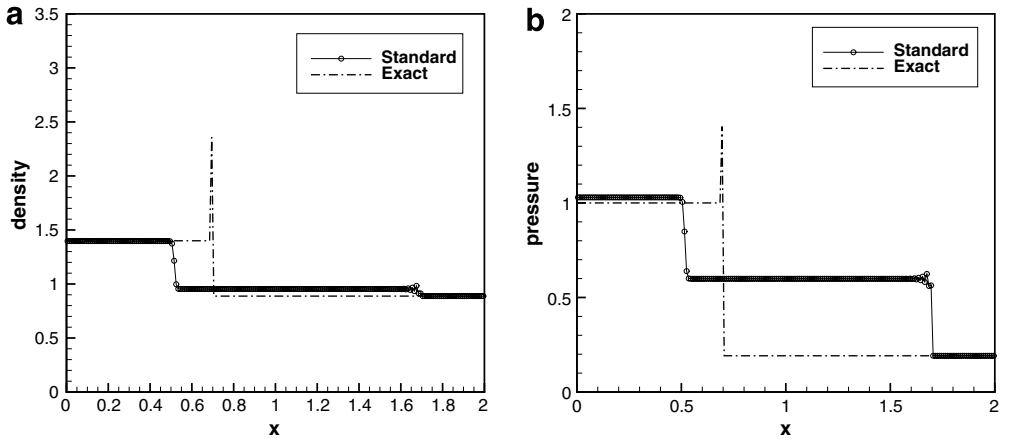

Fig. 14. Standard fractional-step method (Roe + VanLeer + HH2), $kv = 10$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

unknowns, along with the vector of the average values $\mathbf{W}$, allow to reconstruct the solution in such a way that a discontinuity for any of the variables may be present *inside* the cell.

Specifically, defining $\mathbf{w}^n(\xi_i)$ as a continuous approximation of the solution vector inside the cell i at time n (see Eqs. (14)–(16)), we impose

(1) Consistency with the average

$$\int_0^1 \mathbf{w}^n(\xi_i)\mathrm{d}\xi_i = \mathbf{W}_i^n.$$

(2) Each pth component of $\mathbf{w}^n(\xi_i)$ can assume only two values

$$w_p^n(\xi_i) = \overline{W}_{p,i}^n \ \vee \ w_p^n(\xi_i) = \underline{W}_{p,i}^n \quad \forall \xi_i \in [0,1],$$

where $\overline{W}_{p,i}^n$ and $\underline{W}_{p,i}^n$ are the pth components of $\overline{\mathbf{W}}_i^n$ and $\underline{\mathbf{W}}_i^n$.

The portion of cell occupied by the maximum value is also defined componentwise as

$$\gamma_p = \frac{W_p - \underline{W}_p}{\overline{W}_p - \underline{W}_p},$$

where the time level n and cell index i are not indicated for sake of clarity.

While solving the scalar problem, different choices of the function $u^n(\xi_i)$ are possible (see Eqs. (17)–(19)), all of which are equivalent to the solution procedure. This property no longer holds when solving a system of

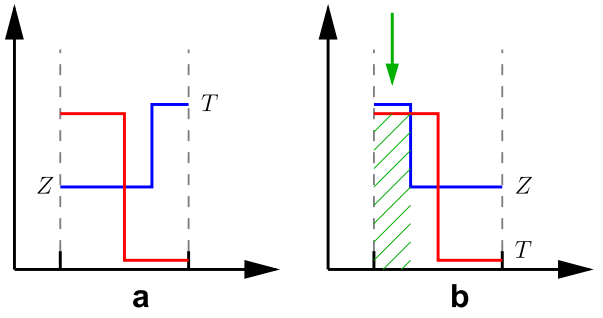

Fig. 15. Two possible ordering for the variables T (red) and Z (green) in the solution of a detonation wave. Only one of the two possibilities presents a region with both high temperature and reactive mixture that is necessary for combustion to occur (dashed region).

equations in which the *MinMax* reconstruction is applied to more than one variable. As a simple example, Fig. 15 shows two cell reconstructions for the variables temperature and mass fraction: only one of these possible combinations (Fig. 15b) presents a region with high temperature and high mixture fraction that is necessary to obtain combustion.

This non-uniqueness of the reconstruction requires an ordering of the maximum and minimum values within the cell. We choose to impose, componentwise, the maximum value of the solution in the left part of the cell if the spatial gradient of the solution is negative, or, conversely, in the right part of the cell if the gradient is positive, namely,

$$\begin{aligned} w_p^n(\xi_i) &= \begin{cases} \overline{W}_{p,i}^n & \xi_i \leqslant \gamma_{p,i}^n, \\ \underline{W}_{p,i}^n & \xi_i > \gamma_{p,i}^n, \end{cases} \quad \text{if } \left(W_{p,i+1}^n - W_{p,i-1}^n\right) \leqslant 0, \\ w_p^n(\xi_i) &= \begin{cases} \underline{W}_{p,i}^n & \xi_i \leqslant \left(1-\gamma_{p,i}^n\right), \\ \overline{W}_{p,i}^n & \xi_i > \left(1-\gamma_{p,i}^n\right), \end{cases} \quad \text{if } \left(W_{p,i+1}^n - W_{p,i-1}^n\right) > 0. \end{aligned} \tag{48}$$

Eq. (48) provides a unique way to define the solution, so that a structure like the one in Fig. 16 is obtained. This identifies a piecewise constant solution inside the cell, made of five regions. We will refer to the solution values in each region as $\mathbf{w}^{(q)}$, $q = 1, \ldots, 5$, and to the corresponding portion of cell occupied by the region as $\gamma^{(q)}$.

Once the internal structure of the cell has been reconstructed it is easy to solve the reaction operator by solving an ODE problem for each of the subparts:

$$\left\{ \frac{\mathrm{d}\mathbf{w}^*}{\mathrm{d}t} = \mathbf{s}(\mathbf{w}^*), \quad t_n \leqslant t \leqslant t_{n+1}, \mathbf{w}^*(x, t_n) = \mathbf{w}^{(q)} \right.$$

with $q = 1, \ldots, 5$.

From the solution of these differential problems the function $\mathbf{w}^*(\xi_i, t_{n+1})$ (i.e. the numerical solution inside the cell after the reactive step) is obtained. It is then possible to adjourn for each pth component of the solution the maximum, minimum and average values of the ith cell as

$$\begin{aligned} \overline{W}_{p,i}^* &= \max_{0 \leqslant \xi_i \leqslant 1} w_p^*(\xi_i, t_{n+1}), \\ \underline{W}_{p,i}^* &= \min_{0 \leqslant \xi_i \leqslant 1} w_p^*(\xi_i, t_{n+1}), \\ W_{p,i}^* &= \int_0^1 w_p^*(\xi_i, t_{n+1}) \mathrm{d}\xi_i. \end{aligned} \tag{49}$$

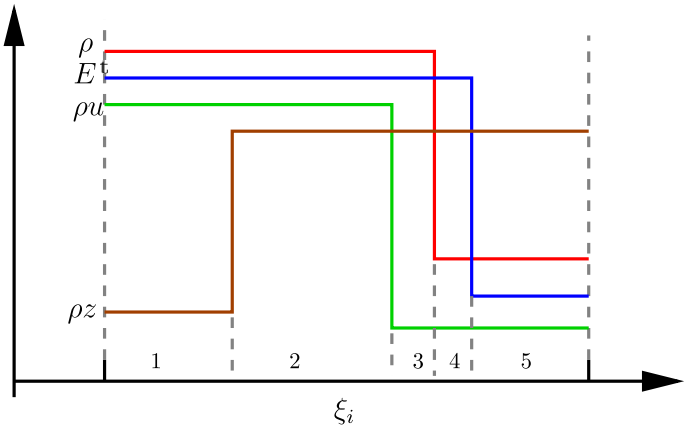

Fig. 16. The internal structure for the *MinMax* method in conservative variables identifies five regions in which the reaction occurs differently.

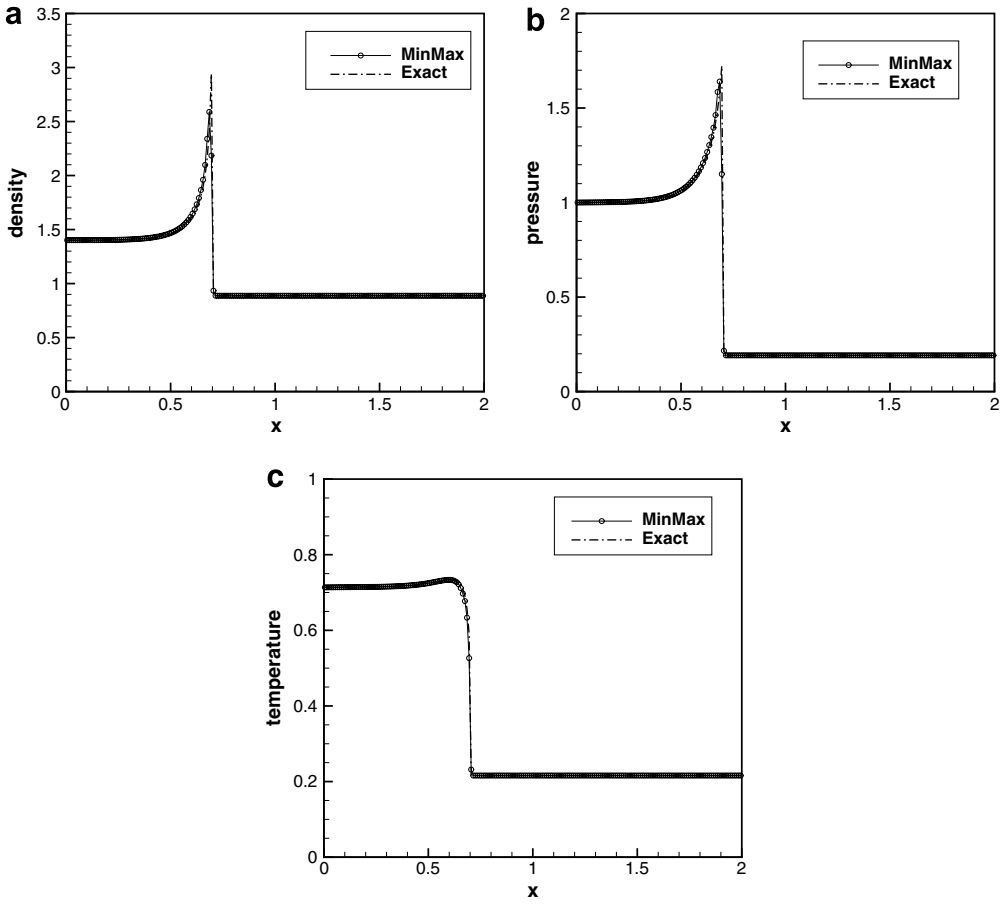

Fig. 17. *MinMax* method, $kv = 0.1$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density; (b) pressure and (c) temperature.

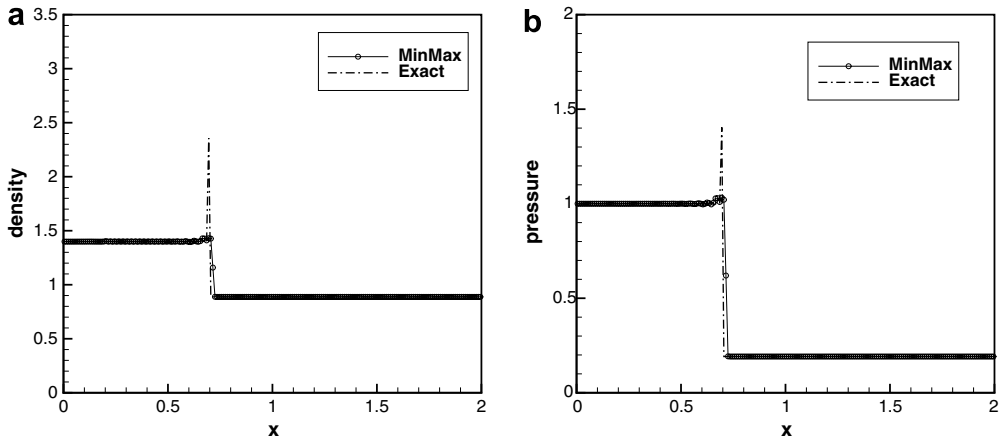

Fig. 18. *MinMax* method, $kv = 10$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

The advection problem that follows is solved with the same Roe's scheme, based only on the average values $\mathbf{W}^*$, as for the standard finite volume method. Finally, the updated maximum and minimum values are

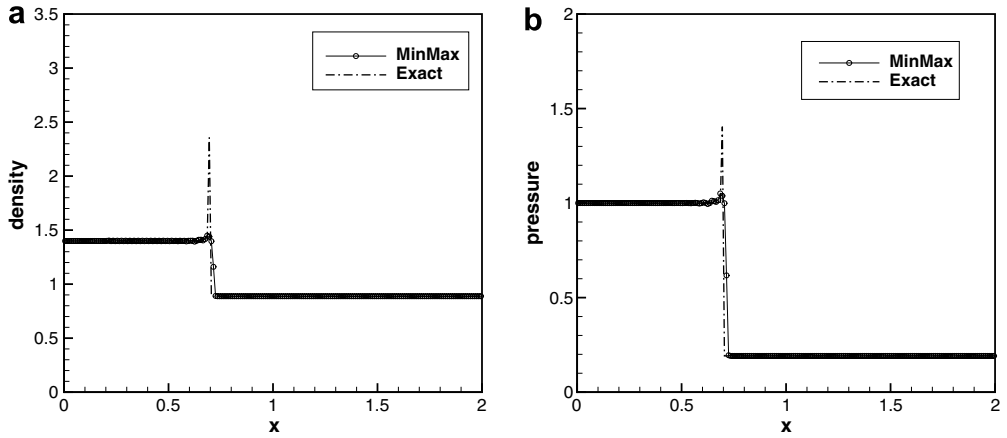

Fig. 19. *MinMax* method, $kv = 100$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

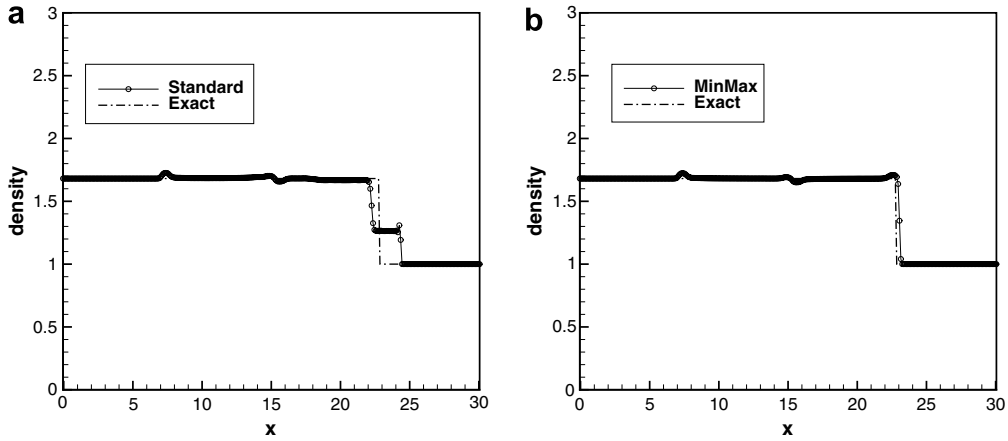

Fig. 20. Density for the stiff Arrhenius case at $t = 1.8$, with $h = 0.01$, $k = 0.005$. (a) Standard method. (b) *MinMax* method.

obtained by applying the extrapolation functions M and m (see Eqs. (27), (28)) to each component of the solution vector.

Figs. 17 and 18 show the results obtained by the *MinMax* method for the non-stiff and stiff conditions of the Chapman–Jouguet detonation described in the preceding section. They may be compared to Figs. 12 and 14 where the solutions obtained with the standard fractional-step method were reported. The *MinMax* solution retains the same accuracy as the standard method solution in describing the reaction zone for the non-stiff case. In the stiff case ($kv = 10$) the proposed scheme cannot resolve the reaction zone but allows us to obtain the correct propagation velocity of the detonation. In addition, increasing the value of kv to 100 (Fig. 19) does not prevent the *MinMax* solution from recovering the correct propagation speed.

The present method has also been tested with a second example, taken again from [16], in which the Arrhenius law (39) is considered. In this case the reaction problem is solved with an implicit trapezoidal method with subiterations. The unburnt gas state is given by

$$\rho_u = 1, \quad u_u = 0, \quad P_u = 1, \quad z_u = 1$$

with gas properties $\gamma = 1.4$, $\mathrm{R} = 1$, $q_0 = 25$. The detonation is initially located at $x = 10$ and travels with speed $u_{\mathrm{CJ}} = 7.12470242$. The activation temperature is set to the value $T_A = 25$, while the pre-exponential coefficient is given the value $A = 16418$ to impose a stiff problem on a grid with spacing $h = 0.01$. A comparison of the

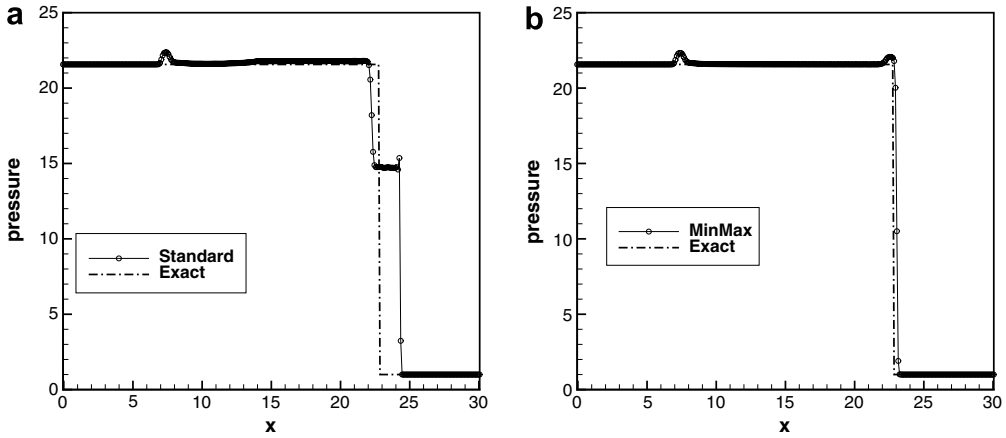

Fig. 21. Pressure for the stiff Arrhenius case at $t = 1.8$, with $h = 0.01$, $k = 0.005$: (a) standard method and (b) *MinMax* method.

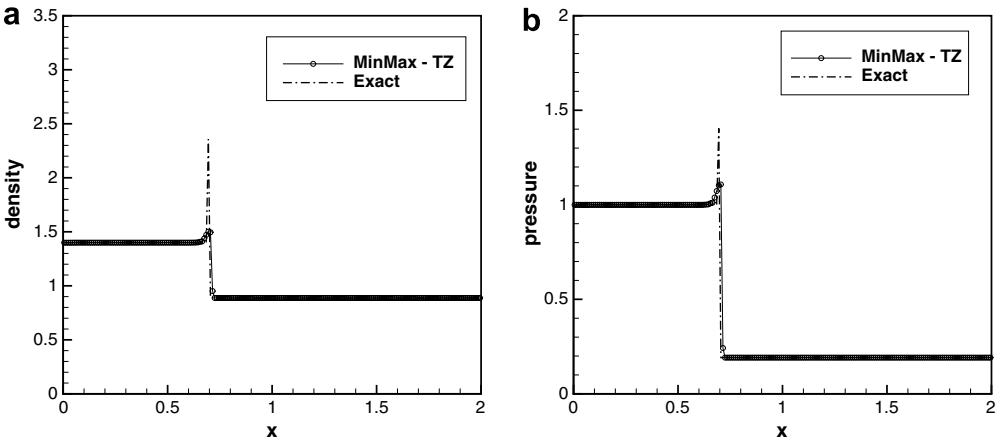

Fig. 22. *MinMax*–TZ method, $kv = 100$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

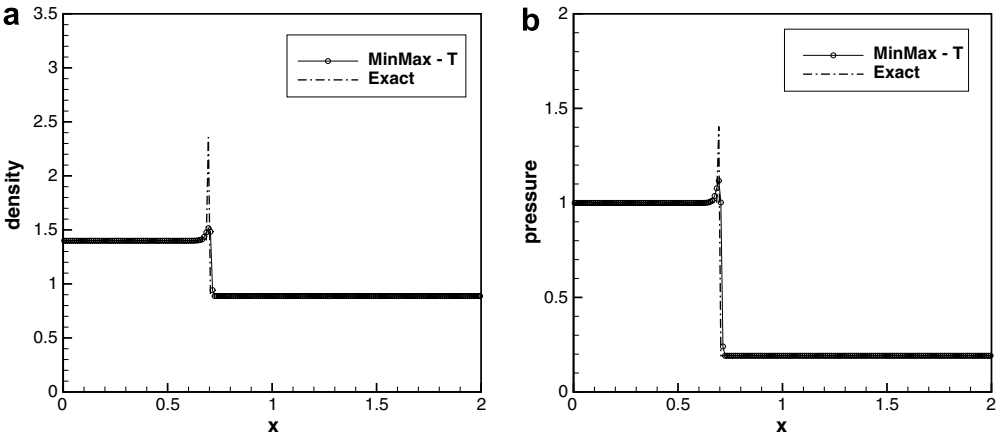

Fig. 23. MinMax–T method, $kv = 100$ at $t = 1$, with $h = 0.01$, $k = 0.005$: (a) density and (b) pressure.

standard and *MinMax* method results is presented in Figs. 20 and 21 and confirms the capability of the present approach to correctly approximate the detonation speed in under-resolved cases.

5.4. Simplified formulations of the MinMax method

In the previous section, it has been shown that the proposed method can be successfully applied to a system of equations, at the expense of an increase in memory requirements and computational cost. In fact, for the reactive Euler system considered in this work, the reaction step requires the separate solution of five ODE problems (43) in each cell. However, the most relevant difficulty of the *MinMax* algorithm, when applied to the whole vector of unknowns, is related to its multi-dimensional extension.

The spatial ordering, needed to select a unique *MinMax* reconstruction within the cell, appears in fact much more complicated for two- or three-dimensional problems. We hence look at possible simplifications of the method to reduce the number of unknowns for which the *MinMax* reconstruction is required. The expression of the source vector $\mathbf{s}(\mathbf{w})$ suggests immediately to consider a *MinMax* structure only for the variables temperature T and mass fraction per unit volume $Z = \rho z$. The resulting algorithm, labelled *MinMax–TZ*, follows the same steps described in the preceding section, with the difference that only the minimum and maximum values $\overline{T}$, $\underline{T}$ and $\overline{Z}$, $\underline{Z}$ are defined and ordered within the cell, giving rise to a three-region structure. The method is tested with the simplified rate law (40) and the unitary speed detonation described in Section 5.2. For the stiff case, the results achieved with the simplified *MinMax–TZ* algorithm are very similar to those of the original *MinMax* scheme (Fig. 22).

A further simplification is sought, in order to reduce the number of *MinMax* variables to one: a *MinMax–T* algorithm is proposed, in which only the temperature is considered to possess a *MinMax* structure. This hypothesis brings the method closer to those proposed in [14,15,17,18]: however, instead of modifying the ignition temperature, the present method enriches the representation of the temperature variable considering the $\overline{T}$ and $\underline{T}$ values within each cell. Also the *MinMax–T* algorithm yields very similar results to those of the other schemes for the stiff conditions (Fig. 23), thus suggesting a straightforward extension of the *MinMax* method to multi-dimensional problems.

6. Conclusions

A modification to standard fractional step, finite volume methods has been proposed to cure the problem exhibited by these methods when dealing with advection/reaction scalar equations and equation systems with stiff source terms. The calculation of an incorrect propagation speed of the discontinuities is avoided by resorting to a two-value reconstruction of the unknowns within each cell, to prevent the reaction operator from influencing the propagation speed itself. A key point of the proposed numerical procedure is the extrapolation of the two-value reconstruction after the advection step, in which the generation of new maxima or minima of the solution is avoided.

In the scalar case, the resulting algorithm, termed *MinMax*, has been proved to retain the formal accuracy of the related finite volume approach and presents three main advantages: (i) it is conceptually very simple; (ii) it may be used for both stiff and non-stiff problems; and (iii) it allows a straightforward extension to two-dimensional problems.

The *MinMax* approach may be extended to hyperbolic equation systems (i.e. the reactive Euler equations) by applying the same procedure to each of the scalar components of the vector of unknowns. Due to the number of variables possessing a *MinMax* structure, the algorithm gives rise to a complex distribution of the unknowns within the cell. However, for one-dimensional flows a simple ordering of the minimum and maximum values of the conservative variables in the cell allows us to achieve the correct numerical propagation speed for Chapman–Jouguet detonations.

Two simplified versions of the scheme have also been proposed, in which only a few of the variables possess a *MinMax* structure. Both the *MinMax–TZ* and *MinMax–T* algorithms yield results comparable to those of the original scheme for one-dimensional flows, thus representing the background for the future extension of the method to multi-dimensional problems.

Acknowledgements

The authors wish to thank Luigi Quartapelle, Alberto Guardone, Massimo Biava and Beth Anne Bennett for the useful discussions and suggestions during the writing of the manuscript. We wish also to thank the Editor and the Reviewers for their observations that allowed us to substantially improve the manuscript.

Appendix

The aim of this appendix is to demonstrate that the *MinMax* method can produce second-order accurate solutions of the advection/reaction Eq. (1). We begin by considering the following standard finite volume method:

$$\begin{aligned} W_i^* &= \overline{\mathcal{R}}^{\frac{k}{2}} W_i^n, \\ W_i^{**} &= \overline{\mathcal{A}}^k (W^*; i), \\ W_i^{n+1} &= \overline{\mathcal{R}}^{\frac{k}{2}} W_i^{**}, \end{aligned} \tag{50}$$

where the symbol W is used for the numerical solution of the standard finite volume method, to distinguish it from the *MinMax* solution defined below. The operator-splitting method (50) takes the name of *Strang-splitting* and it has been proved to be second-order accurate when the truncation error of the discrete reaction and advection operators $\overline{\mathcal{R}}$ and $\overline{\mathcal{A}}$ is at least second-order [20]. It has been extensively tested in [4,21] and in the stiff case shows exactly the same spurious propagation of the initial discontinuity as featured by the standard first-order method. We then consider the second-order version of the *MinMax* algorithm, applied to the same advection/reaction equation:

$$\begin{aligned} \overline{U}_i^* &= \overline{\mathcal{R}}^{\frac{k}{2}} \overline{U}_i^n, \\ \underline{U}_i^* &= \overline{\mathcal{R}}^{\frac{k}{2}} \underline{U}_i^n, \\ \gamma_i^* &= \gamma_i^n, \\ U_i^* &= \overline{U}_i^* \gamma_i^* + \underline{U}_i^* (1 - \gamma_i^*), \\ U_i^{**} &= \overline{\mathcal{A}}^k (U^*; i), \\ \mathbf{U}^* &= \{\overline{U}_i^*, \overline{U}_{i+1}^*, \overline{U}_{i-1}^*, \underline{U}_i^*, \underline{U}_{i+1}^*, \underline{U}_{i-1}^*\}, \\ \overline{U}_i^{**} &= M(\mathbf{U}^*, U_i^{**}), \\ \underline{U}_i^{**} &= m(\mathbf{U}^*, U_i^{**}), \\ \gamma_i^{**} &= \frac{U_i^{**} - \underline{U}_i^{**}}{\overline{U}_i^{**} - \underline{U}_i^{**}}, \\ \overline{U}_i^{n+1} &= \overline{\mathcal{R}}^{\frac{k}{2}} \overline{U}_i^{**}, \\ \underline{U}_i^{n+1} &= \overline{\mathcal{R}}^{\frac{k}{2}} \underline{U}_i^{**}, \\ \gamma_i^{n+1} &= \gamma_i^{**}, \\ U_i^{n+1} &= \overline{U}_i^{n+1} \gamma_i^{n+1} + \underline{U}_i^{n+1} (1 - \gamma_i^{n+1}). \end{aligned} \tag{51}$$

We will now demonstrate that $U^{n+1} = W^{n+1} + \mathrm{O}(h^2)$. Recalling that the Strang-splitting method is second-order accurate, this will also prove that the *MinMax* method is second-order accurate. The basic assumptions of this demonstration are: (i) the functions M and m satisfy condition (25); (ii) the initial conditions of the two methods are almost identical (i.e. $U^n = W^n + \mathrm{O}(h^2)$). We will also suppose that the exact solution is smooth in order to perform a Taylor expansion of the unknowns.

We begin by defining $D_{\mathcal{R}}$ as the difference between the two methods after the first reactive step:

$$D_{\mathcal{R}} = W_i^* - U_i^* = \overline{\mathcal{R}}^{\frac{k}{2}} W_i^n - \left(\gamma_i^* \overline{\mathcal{R}}^{\frac{k}{2}} \overline{U}_i^n) + (1-\gamma_i^*) \overline{\mathcal{R}}^{\frac{k}{2}} \underline{U}_i^n\right). \tag{52}$$

Introducing $\Delta\overline{U}$ and $\Delta\underline{U}$ as the difference between the average and the maximum (resp. minimum) values:

$$\Delta\overline{U} = \overline{U} - U,$$
$$\Delta\underline{U} = U - \underline{U}$$

and dropping suffix i for simplicity we can rewrite (52) as

$$D_{\mathcal{R}} = \overline{\mathcal{R}}^{\frac{k}{2}} W^n - \gamma^* \overline{\mathcal{R}}^{\frac{k}{2}} (U^n + \Delta\overline{U}^n) - (1-\gamma^*) \overline{\mathcal{R}}^{\frac{k}{2}} (U^n - \Delta\underline{U}^n). \tag{53}$$

Considering that the quantity γ is left unvaried by the reactive operator ($\gamma^* = \gamma^n$), we can also drop the superscripts n and $*$ as all the variables in (53) are evaluated at the same time level:

$$D_{\mathcal{R}} = \overline{\mathcal{R}}^{\frac{k}{2}} W - \gamma \overline{\mathcal{R}}^{\frac{k}{2}} (U + \Delta\overline{U}) - (1-\gamma) \overline{\mathcal{R}}^{\frac{k}{2}} (U - \Delta\underline{U}).$$

Assuming that the source term $S(u)$ in (1) is smooth, we may perform a first-order Taylor expansion of $\overline{\mathcal{R}}^{\frac{k}{2}}$ around U:

$$D_{\mathcal{R}} = \overline{\mathcal{R}}^{\frac{k}{2}} W - \gamma \left(\overline{\mathcal{R}}^{\frac{k}{2}} U + \left.\frac{\partial \overline{\mathcal{R}}^{\frac{k}{2}}}{\partial U}\right|_U \Delta\overline{U} + \mathrm{O}(\Delta\overline{U}^2) \right) - (1-\gamma) \left(\overline{\mathcal{R}}^{\frac{k}{2}} U - \left.\frac{\partial \overline{\mathcal{R}}^{\frac{k}{2}}}{\partial U}\right|_U \Delta\underline{U} + \mathrm{O}(\Delta\underline{U}^2) \right),$$

where $\frac{\partial \overline{\mathcal{R}}^{\frac{k}{2}}}{\partial U}$ is intended as the variation of the ODE solution with respect to a small perturbation of the initial condition.

If we now consider condition (25), then $\overline{U}$ and $\underline{U}$ uniformly tend to the average U, hence $\mathrm{O}(\Delta\underline{U}^2)$ and $\mathrm{O}(\Delta\overline{U}^2)$ may be substituted by $\mathrm{O}(h^2)$. We can write $D_{\mathcal{R}}$ as

$$D_{\mathcal{R}} = \overline{\mathcal{R}}^{\frac{k}{2}} W - \overline{\mathcal{R}}^{\frac{k}{2}} U + -\left.\frac{\partial \overline{\mathcal{R}}^{\frac{k}{2}}}{\partial U}\right|_U (\gamma\Delta\overline{U} - (1-\gamma)\Delta\underline{U}) + \mathrm{O}(h^2).$$

However, $\gamma\Delta\overline{U} - (1-\gamma)\Delta\underline{U}$ is null, as can be easily checked putting $\Delta\overline{U}$ and $\Delta\underline{U}$ in the definition of average operator (21). Furthermore, since $U^n = W^n + \mathrm{O}(h^2)$, we have

$$\overline{\mathcal{R}}^{\frac{k}{2}} U = \overline{\mathcal{R}}^{\frac{k}{2}} W + \left.\frac{\partial \overline{\mathcal{R}}^{\frac{k}{2}}}{\partial U}\right|_W \mathrm{O}(h^2),$$

so that the difference between the standard finite volume method and the *MinMax* method after the first reaction step is

$$D_{\mathcal{R}} = \mathrm{O}(h^2).$$

When solving the homogeneous part of the problem only the average is considered, so that we have

$$D_{\mathcal{A}} = W_i^{**} - U_i^{**} = \overline{\mathcal{A}}^k(W^*; i) - \overline{\mathcal{A}}^k(U^*; i) = \overline{\mathcal{A}}^k(W^*; i) - \overline{\mathcal{A}}^k(W^* + \mathrm{O}(h^2); i) = \mathrm{O}(h^2).$$

During the second reactive step we can estimate the difference between the two methods as previously described. We finally obtain

$$W_i^{n+1} - U_i^{n+1} = \mathrm{O}(h^2),$$

i.e. the *MinMax* approximation, applied to a smooth solution, behaves as the finite volume method, introducing an additional error of order $\mathrm{O}(h^2)$ that does not downgrade the formal accuracy of the scheme.

References

[1] P. Colella, A. Majda, V. Roytburd, Theoretical and numerical structure for reacting shock waves, SIAM J. Sci. Stat. Comput. 7 (1986) 1059–1079.

[2] R. Pember, Numerical methods for hyperbolic conservation laws with stiff relaxation, i. spurious solutions, SIAM J. Appl. Math. 53 (1993) 1293–1330.

[3] M. Ben-Artzi, The generalized Riemann problem for reactive flows, J. Comput. Phys. 81 (1989) 70–101.
[4] R. LeVeque, H. Yee, A study of numerical methods for hyperbolic conservation laws with stiff source terms, J. Comput. Phys. 86 (1990) 187–210.
[5] D. Griffiths, A. Stuart, H. Yee, Numerical wave propagation in an advection equation with a nonlinear source term, SIAM J. Numer. Anal. 29 (1992) 1244–1260.
[6] D. Nguyen, F. Gibou, R. Fedkiw, A fully conservative ghost fluid method & stiff detonation waves, in: Proceedings of the 12th International Detonation Symposium, S. Diego, CA, 2002.
[7] R. Jeltsch, P. Klingenstein, Error estimators for the position of discontinuities in hyperbolic conservation laws with source term which are solved using operator splitting, Comput. Vis. Sci. 1 (1999) 231–249.
[8] B. Bihari, D. Schwendeman, Multiresolution schemes for the reactive Euler equations, J. Comput. Phys. 154 (1999) 197–230.
[9] A. Bourlioux, A. Majda, V. Roytburd, Theoretical and numerical structure for unstable one-dimensional detonations, SIAM J. Appl. Math. 51 (1991) 303–343.
[10] A. Chorin, Random choice solution of hyperbolic systems, J. Comput. Phys. 22 (1976) 517–533.
[11] A. Chorin, Random choice methods with applications for reacting gas flows, J. Comput. Phys. 25 (1977) 253–272.
[12] A. Majda, V. Roytburd, Numerical study of the mechanisms for initiation of reacting shock waves, SIAM J. Sci. Stat. Comput. 11 (1990) 950–974.
[13] B. Engquist, B. Sjogreen, Robust Difference Approximations of Stiff Inviscid Detonation Waves, Technical Report, 1991, CAM 91-03, UCLA.
[14] V. Ton, Improved shock-capturing methods for multicomponent and reacting flows, J. Comput. Phys. 128 (1996) 237–253.
[15] A. Berkenbosch, E. Kaasschieter, R. Klein, Detonation capturing for stiff combustion chemistry, Combust. Theory Model. 2 (1998) 313–348.
[16] C. Helzel, R. LeVeque, G. Warneke, A modified fractional step method for the accurate approximation of detonation waves, SIAM J. Sci. Stat. Comput. 22 (1999) 1489–1510.
[17] W. Bao, S. Jin, The random projection method for hyperbolic conservation laws with stiff reaction terms, J. Comput. Phys. 163 (2000) 216–248.
[18] W. Bao, S. Jin, The random projection method for stiff detonation capturing, J. Sci. Comput. 23 (2001) 1000–1025.
[19] A. Kurganov, An accurate deterministic projection method for hyperbolic systems with stiff source terms, in: Proceedings of the 9th International Conference Hyperbolic Problems: Theory, Numerics, Applications, Springer, 2003.
[20] G. Strang, On the construction and comparison of difference schemes, SIAM J. Numer. Anal. 5 (1968) 506–517.
[21] L. Tosatto, Soluzione numerica di equazioni di advezione e reazione con termine di sorgente stiff: applicazione alle onde di detonazione, Technical Report, Master Thesis, Politecnico di Milano, 2005 (in Italian).
[22] R. LeVeque, Numerical Methods for Conservation Laws, Birkhäuser, 1992.
[23] E. Godlewski, P.-A. Raviart, Numerical Approximation of Hyperbolic Systems of Conservation Laws, Springer, 1996.
[24] P. Roe, Approximate Riemann solvers, parameter vectors and difference schemes, J. Comput. Phys. 43 (1981) 357–372.
[25] A. Harten, J.M. Hyman, Self adjusting grid methods for one-dimensional hyperbolic conservation laws, J. Comput. Phys. 50 (1983) 253–269.