



HAL
open science

a posteriori stabilized sixth-order finite volume scheme for one-dimensional steady-state hyperbolic equations

Stéphane Clain, Raphaël Loubère, Gaspar J Machado

► **To cite this version:**

Stéphane Clain, Raphaël Loubère, Gaspar J Machado. a posteriori stabilized sixth-order finite volume scheme for one-dimensional steady-state hyperbolic equations. 2016. hal-01418899

HAL Id: hal-01418899

<https://hal.science/hal-01418899>

Preprint submitted on 20 Dec 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

a posteriori stabilized sixth-order finite volume scheme for one-dimensional steady-state hyperbolic equations

Stéphane Clain · Raphaël Loubère · Gaspar J. Machado*

Abstract We propose a new family of finite volume high-accurate numerical schemes devoted to solve one-dimensional steady-state hyperbolic systems. High-accuracy (up to the sixth-order presently) is achieved thanks to polynomial reconstructions while stability is provided with an *a posteriori* MOOD method which control the cell polynomial degree for eliminating non-physical oscillations in the vicinity of discontinuities. Such a procedure demands the determination of a chain detector to discriminate between troubled and valid cells, a cascade of polynomial degrees to be successively tested when oscillations are detected, and a parachute scheme corresponding to the last, viscous, and robust scheme of the cascade. Experimented on linear, Burgers', and Euler equations, we demonstrate that the schemes manage to retrieve smooth solutions with optimal order of accuracy but also irregular solutions without spurious oscillations.

Contents

1	Introduction	1
2	Framework	3
3	The MOOD method	6
4	Advection problem	8
5	The Burgers' equation	11
6	The Euler system	18
7	Conclusion and perspectives	29

1 Introduction

Numerical approximations of the steady-state Euler system date back to the early seventies with the NASA Ames group [39]. The development of implicit time algorithm [8] with high-order finite difference methods [9] led to the first simulations for two- and three-dimensional complex geometries calculated on Illiac IV Computer [42,41,40]. The steady-state was achieved as the limit stage of the non-stationary problem using fictitious time step (time marching method) [47]. The main

Stéphane Clain and Gaspar J. Machado
Centre of Mathematics, University of Minho, Campus of Azurém, 4800-058 Guimarães, Portugal
E-mail: clain@math.uminho.pt, E-mail: gjm@math.uminho.pt

Raphaël Loubère
CNRS and Institut de Mathématiques de Bordeaux (IMB), Université de Bordeaux, Talence, France
E-mail: raphael.loubere@u-bordeaux.fr

difficulties are, on one hand, to achieve an accurate approximation where the solution is smooth enough, and, on the other hand, to produce a robust solution without non-physical oscillations.

Most of the high-order technologies developed for the non-stationary case were adapted to the steady-state case with the time marching algorithm using an explicit scheme [25] or an implicit formulation [50,30] equipped with a Newton-Krylov iterative procedure [47,45,37]. Taking the time step to infinity leads to a Newton-Krylov method for the steady-state problem which represents an alternative to the time marching method [27]. Second-order finite volume approximations with MUSCL limiter [43–46] on structured [49,25] or unstructured meshes [15,5,24,31] revealed to be the standard approach in the nineties and have been widely developed. Unfortunately, the accuracy remains below the optimal value two due to an excessive limiting even for smooth solution where limiting should not act to provide the maximum order. Moreover, higher-order methods are appealing because they may dramatically reduce the computational time since they provide an excellent approximation with coarse meshes for smooth solutions. Of course, shocks and vortices still require an adequate local mesh to catch all the structures but high-order schemes prove to be efficient in reducing the numerical diffusion. For instance the k -exact reconstruction [3,4,34] increases the method accuracy using quadratic or cubic polynomial approximations [32,35,36]. Nevertheless, traditional TVD (Total Variation Diminishing) limiting procedures drastically reduce the order of accuracy despite the construction of alternative limiters [45,33] to enhance the quality of the solution.

More recently, Weighted Essentially Non-Oscillatory (WENO) technology [1,48,26,20] delivers more accurate approximations while preserving the robustness for the numerical solution and it enables to achieve higher-orders than traditional TVD methods. The technique was to a later time adapted for the steady-state situation [53,28]. The main difficulties lie on the non-differentiability of the numerical scheme leading to the creation of slight oscillations nearby shock discontinuities. More adapted WENO schemes then have been developed [52,23,21] to overcome the problem of convergence toward the steady-state solution. We also mention some interesting developments for the steady-state case based on the Discontinuous Galerkin method [6,7,38], the Spectral Volume technique [10,18], and the Residual Distribution schemes [12,2,22].

All the technologies listed above employ an *a priori* limiting procedures to reduce the accuracy or the degree of the polynomial reconstructions in the presence of non-physical oscillations. Indeed, the reconstructed values are modified before feeding the numerical flux for the subsequent update stage. Consequently, unexpected over-limiting usually happens since the correction is carried out with the spirit to eliminate any potential risks. Moreover, the limiting procedure is only based on the numerical regularity of the approximation and rarely on the physics of the problem. For example, the positivity preserving is a consequence of the TVD restriction on the reconstruction but is not directly imposed as a restriction *per se*. The Multidimensional Optimal Order Method (MOOD) has been designed in [13,16,17] on different paradigms. It is an *a posteriori* method since the modification of the polynomial reconstruction is performed after calculating a candidate solution. The main advantage is that we only modify the solution where it is necessary leading to an improvement of the accuracy with respect to other methods. Another point is the capacity of the method to directly take into account the physics of the problem (positivity, entropy production, etc.). The MOOD paradigm has been developed for non-stationary hyperbolic system [51,29,14] and we aim at demonstrating that the methodology is adapted to steady-state situations. The key point is the introduction of an additional unknown vector which represents the maximum admissible polynomial degree on each cell. We seek for the numerical solution (density, the velocity, and the pressure for Euler system for instance) as well as the cell polynomial degree that produce the best approximation exempt from non-physical oscillations. In this work we only tackle the one-dimensional situation in order to focus on the MOOD algorithm and prove the ability of the method to provide high-order approximations for classical hyperbolic equation or system of

equations. We also skip the question of the iterative solver efficiency since most of the implicit or explicit techniques used in the WENO, DG, or SV methods do apply in our context.

This document is divided in seven sections. After the introduction, we present in section 2 the framework (notation, mesh, schemes) while section 3 is devoted to the MOOD methodology. We start by considering the simple scalar linear equations in section 4, then the non-linear Burgers' equation case in section 5, and dedicate section 6 to the Euler system. Numerical experiments involving regular and irregular solutions are presented for each equation or system of equations. We end the document with the conclusions of this work and present some perspectives.

2 Framework

Let $\Omega =]x_L, x_R[$ be the computational domain discretized into a regular mesh \mathcal{T}_h constituted of cells $K_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ for $i = 1, \dots, I$ (see Fig. 1) of constant length $h = (x_R - x_L)/I$ with center $x_i = \frac{1}{2}(x_{i+\frac{1}{2}} + x_{i-\frac{1}{2}})$. The boundaries of Ω are denoted by $x_{\frac{1}{2}} = x_L$ and $x_{I+\frac{1}{2}} = x_R$.

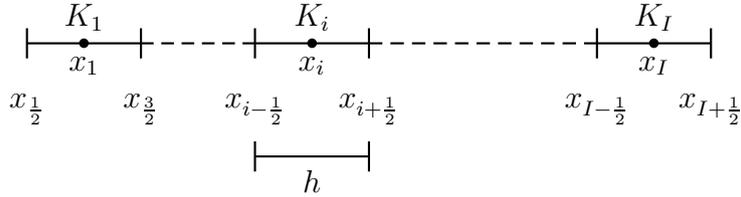


Fig. 1 Notation: point/interface, mesh, and cells.

For any bounded function ϕ defined on Ω , ϕ_i stands for an approximation of the mean value ϕ_i^{ex} of function ϕ over cell K_i , that is

$$\phi_i \approx \phi_i^{\text{ex}} = \frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \phi(x) dx, \quad i = 1, \dots, I,$$

while vector $\Phi = (\phi_i)_{i=1, \dots, I} \in \mathbb{R}^I$ gathers the values of the approximation.

2.1 Polynomial reconstruction

Let K_i be a cell and $d \in \mathbb{N}$. The stencil S_i is composed of the n_i closest neighbor cells (excluding cell K_i) such that $n_i > d$. The polynomial of degree d associated to cell K_i and approximation ϕ_i is defined as

$$\phi_i(x; d) = \phi_i + \sum_{k=1}^d R_{i,k} ((x - x_i)^k - X_{i,k}),$$

with $\mathbf{R}_i = (R_{i,k})_{k=1, \dots, d}$ the vector that gathers the unknown polynomial coefficients. We set $X_{i,k} = \frac{1}{h} \int_{K_i} (x - x_i)^k dx$ to achieve the conservative property

$$\frac{1}{h} \int_{K_i} \phi_i(x; d) dx = \phi_i. \tag{1}$$

For a given stencil S_i , we consider the quadratic functional

$$E_i(\mathbf{R}_i) = \sum_{j \in S_i} \left(\frac{1}{h} \int_{K_j} \phi_i(x; d) dx - \phi_j \right)^2.$$

We denote by $\widehat{\mathbf{R}}_i = (\widehat{R}_{i,k})_{k=1,\dots,d}$ the unique vector which minimizes the quadratic functional and set $\widehat{\phi}_i(x; d)$ the associated polynomial that corresponds to the best approximation in the least squares sense of the data of the stencil

$$\widehat{\phi}_i(x; d) = \phi_i + \sum_{k=1}^d \widehat{R}_{i,k} ((x - x_i)^k - X_{i,k}).$$

Such a reconstruction satisfies the conservation property (1).

Remark 1 We have adopted the following notations: ϕ_i represents an approximation of ϕ over cell K_i while $\phi_i(x; d)$ is a generic polynomial function of degree d associated to K_i which satisfies the conservation property (1).

2.2 Generic finite volume scheme

We consider the steady state one-dimensional equation with source term

$$\frac{d\mathbf{F}(\phi)}{dx} = \mathbf{S}, \quad \text{in } \Omega, \quad (2)$$

where $\mathbf{F}(\phi)$ stands for the physical flux, \mathbf{S} represents a regular source term, and Dirichlet boundary conditions ϕ_D are prescribed on the boundaries.

Integration of (2) over a cell K_i gives

$$\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \frac{d\mathbf{F}(\phi(x))}{dx} dx = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{S}(x) dx,$$

and integration by parts yields

$$\left(\mathbf{F}(\phi(x_{i+\frac{1}{2}})) - \mathbf{F}(\phi(x_{i-\frac{1}{2}})) \right) - h \left(\frac{1}{h} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathbf{S}(x) dx \right) = 0.$$

We substitute the exact relation with the numerical scheme

$$\left(\mathcal{F}(\phi_{i+\frac{1}{2},-}, \phi_{i+\frac{1}{2},+}; x_{i+\frac{1}{2}}) - \mathcal{F}(\phi_{i-\frac{1}{2},-}, \phi_{i-\frac{1}{2},+}; x_{i-\frac{1}{2}}) \right) - h \mathcal{S}_i = 0, \quad (3)$$

where $\mathcal{F}(\phi_{i+\frac{1}{2},-}, \phi_{i+\frac{1}{2},+}; x_{i+\frac{1}{2}})$ is the two-point numerical flux evaluated at point $x_{i+\frac{1}{2}}$, scalar values $\phi_{i+\frac{1}{2},-}$ and $\phi_{i+\frac{1}{2},+}$ stand for approximations on the left and right sides of interface $x_{i+\frac{1}{2}}$ respectively, and \mathcal{S}_i is an approximation of the source term computed over the cell with Gaussian quadrature formula. Boundary conditions are taken into account for the first and last cell equations taking $\phi_{\frac{1}{2},-} = \phi_D(x_L)$ and $\phi_{I+\frac{1}{2},+} = \phi_D(x_R)$.

2.3 Cell polynomial degree and residual formulation

Polynomial reconstruction is an efficient tool to provide accurate approximations. Nonetheless, one has to pay attention to the local regularity and choose the appropriate degree for the representation. For a cell K_i , we denote by d_i its associated degree and $\mathcal{M} = (d_i)_{i=1,\dots,I}$ stands for the Cell Polynomial Degree (CPD) map, that is, the vector which indicates the degree of the reconstruction for each cell. Moreover, we denote by

$$d_{i+\frac{1}{2}} = \min(d_i, d_{i+1})$$

the effective polynomial degree we use to compute approximations on interface $x_{i+\frac{1}{2}}$ and set

$$\phi_{i+\frac{1}{2},-} = \widehat{\phi}_i(x_{i+\frac{1}{2}}; d_{i+\frac{1}{2}}), \quad \phi_{i+\frac{1}{2},+} = \widehat{\phi}_{i+1}(x_{i+\frac{1}{2}}; d_{i+\frac{1}{2}}).$$

We introduce the residual $\mathcal{G}_i(\Phi, \mathcal{M})$ at cell K_i that is given by

$$\mathcal{G}_i(\Phi, \mathcal{M}) = \left(\mathcal{F}(\phi_{i+\frac{1}{2},-}, \phi_{i+\frac{1}{2},+}; x_{i+\frac{1}{2}}) - \mathcal{F}(\phi_{i-\frac{1}{2},-}, \phi_{i-\frac{1}{2},+}; x_{i-\frac{1}{2}}) \right) - h \mathcal{S}_i, \quad (4)$$

and the global residual $\mathcal{G}(\Phi, \mathcal{M}) \in \mathbb{R}^I$ gathers the residual expressions $\mathcal{G}_i(\Phi, \mathcal{M})$ for all cell i .

For a given CPD map \mathcal{M} , we introduce the residual operator

$$\Phi \in \mathbb{R}^I \rightarrow \mathcal{G}(\Phi, \mathcal{M}) \in \mathbb{R}^I$$

and seek a solution $\Phi^{\mathcal{M}}$ for the system of equations $\mathcal{G}(\Phi, \mathcal{M}) = 0$ using an iterative solver and given an initial guess.

We also notice that the case where $d_i = 0$ corresponds to the first-order approximation. On the other hand, if we are dealing with smooth solutions, let us choose $d_i = d$ with d given by the user and solving $\mathcal{G}(\Phi, \mathcal{M}) = 0$ provides a $d + 1$ th-order approximation. The point is to deal with solutions involving discontinuities. In this case the CPD map can not be constant and requires a low degree where the solution is rough and high degree where the solution is smooth. Since we do not know the regularity of the solution, \mathcal{M} also turns to be an unknown of the problem and one search a solution (Φ^*, \mathcal{M}^*) such that $\mathcal{G}(\Phi^*, \mathcal{M}^*) = 0$.

2.4 Steady-state iterative solvers

Even dealing with linear equations, the problem is nonlinear due to the choice of \mathcal{M} . Therefore iterative procedures have to be designed to produce a succession of approximations (Φ^k, \mathcal{M}^k) that converge to a satisfying solution (Φ^*, \mathcal{M}^*) . Furthermore, if the system is nonlinear (Burgers equation, Euler system), an extra inner loop is required to solve the nonlinear problem for a fixed CPD map.

One way to proceed requires a solver which is problem dependent $\Phi \rightarrow \mathcal{G}(\Phi, \mathcal{M})$ such that, being given a CPD map \mathcal{M}^k and an initial guess Φ^k , the (non)linear ((N)L) solver computes Φ^{k+1} solution of $\mathcal{G}(\Phi, \mathcal{M}^k) = 0$. Then a new CPD map \mathcal{M}^{k+1} is provided on the base of the data of Φ^{k+1} (evaluation of \mathcal{M}^{k+1} will be explained in the next section) and the iterative process goes one until we reach to a limit CPD map, *i.e.* $\mathcal{M}^k = \mathcal{M}^{k+1}$ since in that case, we have $\Phi^k = \Phi^{k+1}$ and the steady-state approximation is obtained.

A second strategy consists in reformulating equation (2) by introducing a fake time dependency

$$\frac{\partial \tilde{\phi}}{\partial t} + \frac{\partial \mathbf{F}(\tilde{\phi})}{\partial x} = \mathbf{S} \quad (5)$$

with prescribed initial and boundary conditions. The time parameter t is devoted to tend to infinity providing the steady state solution as $\phi = \lim_{t \rightarrow \infty} \tilde{\phi}(\cdot, t)$. Equation (5) is solved using an explicit high accurate finite volume code as presented in [17] and we tag the solver as “Time Marching solver” (TM).

3 The MOOD method

The dynamical determination of the CPD map \mathcal{M} is of paramount importance to ensure stability, robustness, and accuracy where appropriate. In the case of a discontinuous solution, high-order schemes produce numerical instabilities resulting into non-physical oscillations in the vicinity of the shocks. Consequently, in the presence of a discontinuous solution or a steep gradient, the cell polynomial degree d_i must drop to zero to avoid spurious oscillations whereas the maximal user-defined degree d^{\max} can be employed in the zones where the solution is regular enough, ensuring *de facto* a local high accuracy of the approximate solution.

3.1 MOOD loop

We dynamical determine the CPD map using a *a posteriori* MOOD-like approach developed in [13, 16, 17]. Namely, for a given stage k and its associated map \mathcal{M}^k , a candidate solution Φ^k is computed. We then perform a detection procedure to identify which cells present numerical artifacts that would demand more dissipation. For those cells only we reduce the polynomial degree following a cascade, *i.e.* a decreasing sequence of polynomial degrees, keeping the good cell polynomial degrees alike. This determines the updated CPD map \mathcal{M}^{k+1} which is further employed to compute the new candidate solution Φ^{k+1} . We depict in Fig. 2 the MOOD loop embedding the solver that provides solution Φ^k for a fixed CPD map \mathcal{M}^k .

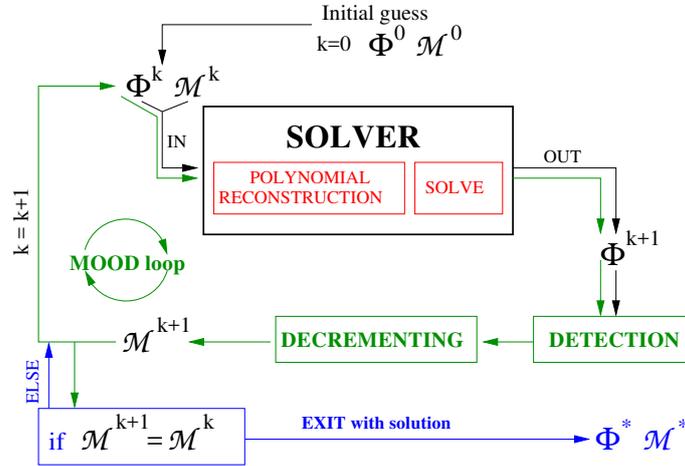


Fig. 2 Sketch of the MOOD loop. Starting from an initial guess Φ^0 and an initial CPD map \mathcal{M}^0 ($k = 0$), the solver furnishes a candidate solution Φ^{k+1} which is tested (detection) and subsequent decrementing of the polynomial degree map may occur giving a new CPD map \mathcal{M}^{k+1} . If so, or if convergence is not reached then another iterate is performed.

3.2 Detection chain

The detection chain is the procedure by which the algorithm analyses a candidate solution. For each cell K_i , we apply the chain to assess if the value is considered as acceptable (with respect to the neighbor cells). If not, the solution needs to be recomputed with more dissipation locally by reducing the degree of the polynomial reconstruction associated to K_i . The detection chain is composed of a succession of elementary detector procedures that focus on a specific potential problem of the solution. We list in the following the elementary detectors used in the applications.

3.2.1 Basic detectors

The basic detectors are designed to check if the candidate solution is physically admissible and to prevent the appearance of oscillations characterized by the creation of local extrema.

Physical Admissible Detector (PAD) The candidate solution satisfies the PAD condition on cell K_i if ϕ_i^k is a physical state associated with the solved system of PDEs. Such a condition is crucial for the Euler system and it consists of the positivity of the density and pressure values.

Extrema Detector (ED) One states that the candidate solution does not present an extremum at cell K_i if

$$\min(\phi_{i-1}^k, \phi_{i+1}^k) \leq \phi_i^k \leq \max(\phi_{i-1}^k, \phi_{i+1}^k).$$

When such a condition is violated, we have detected a potential oscillation of the discrete candidate solution.

3.2.2 Relaxation detectors

New extrema may derive from local oscillations associated to the Gibbs phenomenon, but may also correspond to smooth physical extrema. Another false detection situation arises when the solution is locally constant and extremely small perturbations due to the real number IEEE format may be interpreted as oscillations. Relaxation detectors aim at eliminating non-critical situations that have been falsely detected by ED. To this end, we compute an approximation of the local curvature setting

$$C_i(\Phi) = \widehat{\phi}_i''(x; 2), \quad i = 1, \dots, I.$$

We then define the following local curvature indicators

$$\chi_{i,m} = \min(C_{i-1}, C_i, C_{i+1}), \quad \chi_{i,M} = \max(C_{i-1}, C_i, C_{i+1}), \quad i = 2, \dots, I - 1,$$

where we omit dependency on the vector Φ for the sake of simplicity. Notice that one can have $|\chi_{i,m}| > |\chi_{i,M}|$.

Plateau Detector (PD) Let ε_{PD} be a given tolerance parameter. Then, d_i is not modified if

$$\max(|\chi_{i,m}|, |\chi_{i,M}|) \leq \varepsilon_{PD}.$$

Such condition means that the local curvature is so small that the numerical solution is locally linear and therefore the degree of the polynomial reconstruction should not be altered.

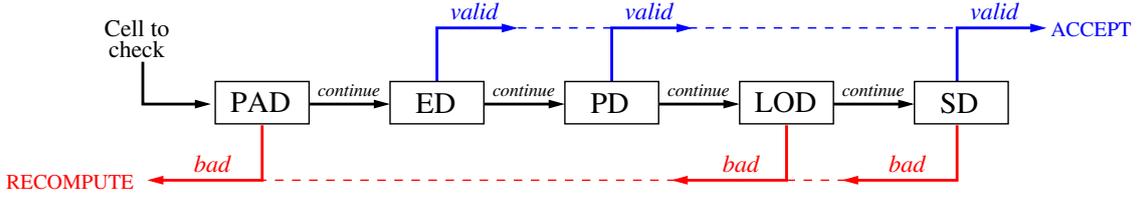


Fig. 3 Chain of detectors used to check if a cell value is valid and accepted, or, if it is discarded and needs to be recomputed after reduction of the local cell polynomial degree. Each of the top/bottom arrow permits to exit the chain prematurely. PAD: Physical Admissible Detector, ED: Extrema Detector, PD: Plateau Detector, LOD: Local Oscillation Detector, SD: Smoothness Detector.

Local Oscillation Detector (LOD) We must reduce the cell polynomial degree d_i if one has

$$\chi_{i,m}\chi_{i,M} \leq 0,$$

for any new extrema. This condition detects a local oscillation due to the variation of the curvature signs.

Smoothness Detector (SD) This detector determines if the minimum and the maximum curvatures are close enough with respect to the threshold parameter, meaning that the numerical solution is considered locally smooth. Let ε_{SD} be a given tolerance parameter. The numerical solution is considered locally smooth if

$$1 \geq \frac{\min(|\chi_{i,m}|, |\chi_{i,M}|)}{\max(|\chi_{i,m}|, |\chi_{i,M}|)} \geq 1 - \varepsilon_{SD}.$$

If that is the case we do not modify the polynomial degree, otherwise the degree should be reduced.

In Fig. 3 we present a sketch of the chain of detectors proposed in this work. Note that some detectors may be removed depending on the system of equations (for instance the PAD may not make sense in the case of linear advection of a scalar). Note also that the PD permits to validate and accept a cell which is on a “plateau” and not wasting time trying to compute and compare zero valued curvatures in the subsequent detectors of the chain. The top/bottom arrows permit to exit the chain sooner if possible. In this work we fix $\varepsilon_{PD} = h$ and $\varepsilon_{SD} = 1/2$.

3.3 The cascade

The last ingredient of a MOOD approach is to choose the cascade of polynomial degrees that the algorithm will successively try. A costly but systematic cascade consists in decrementing the polynomial degree by one from d^{\max} to 0. A less expensive choice consists in ignoring some intermediate stages. For example, $d^{\max} = 5 \rightarrow 2 \rightarrow 0$ is a less computational demanding cascade since we ignore the intermediate stages 4, 3, and 1.

In this work we mainly use the cascade $d^{\max} = 5 \rightarrow 2 \rightarrow (1 \rightarrow)0$ or $d^{\max} = 1 \rightarrow 0$.

4 Advection problem

We first consider the scalar linear convective steady-state problem

$$\frac{d}{dx} (u(x)\phi(x)) = S(x), \quad x \in \Omega, \quad (6)$$

where $u(x)$ is a given velocity function. We also prescribe boundary function ϕ_D , which is just relevant at the boundary interfaces with inflow velocities. Since u or S may present some irregularities, the solution is not necessarily smooth enough to perform all the polynomial reconstructions with d^{\max} hence the CPD map will be adapted with respect to the numerical approximation.

4.1 Numerical flux

The numerical flux $\mathcal{F}(\phi_{i+\frac{1}{2},-}, \phi_{i+\frac{1}{2},+}; x_{i+\frac{1}{2}})$ is the classical upwind one. For a given CPD map \mathcal{M} and noticing that the problem $\mathcal{G}(\Phi, \mathcal{M}) = 0$ is affine, we use a GMRES procedure to nullify the residual and tagged this first strategy as “Linear solver” (L).

Next, the chain detector associated to this linear equation follows the chain depicted in Fig. 4 where all detectors have been defined in the previous section. We consider only two cascades:

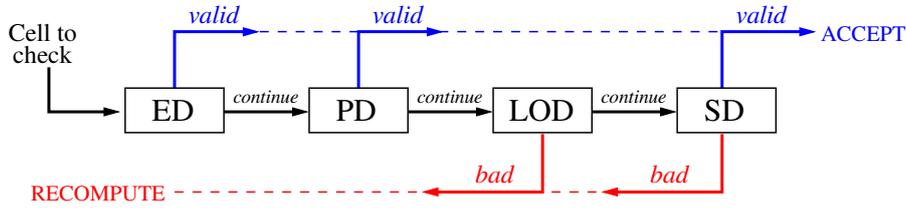


Fig. 4 Chain of detectors used for Linear and Burgers’ equation to check if a cell value is valid and accepted, or, if it is discarded and need to be recomputed after reduction of the local cell polynomial degree. ED: Extrema Detector, PD: Plateau Detector, LOD: Local Oscillation Detector, SD: Smoothness Detector.

$d^{\max} = 1 \rightarrow 0$, or $d^{\max} = 5 \rightarrow 2 \rightarrow 1 \rightarrow 0$, that is second-order and sixth-order accurate schemes. The first-order accurate scheme with $d^{\max} = 0$ is also employed for comparison purposes.

4.2 Numerical experiments

The L^1 and L^∞ errors calculated for the whole mesh are given by

$$E_1 = \sum_{i=1}^I |\phi_i - \phi_i^{\text{ex}}| h, \quad E_\infty = \max_{i=1}^I |\phi_i - \phi_i^{\text{ex}}|. \quad (7)$$

Note that the previous norms explicitly depend on the numerical solution Φ and mesh size I , as $E_1(\Phi, I)$ and $E_\infty(\Phi, I)$. Next we define the rate of convergence between two solutions/meshes (Φ_k, I_k) , for $k = 1, 2$ where $I_1 < I_2$ as

$$\mathcal{O}_\alpha \equiv \mathcal{O}_\alpha((\Phi_1, I_1); (\Phi_2, I_2)) = \frac{|\log E_\alpha(\Phi_1, I_1)/E_\alpha(\Phi_2, I_2)|}{|\log I_1/I_2|}, \quad \alpha = 1, \infty. \quad (8)$$

In order to derive an exact solution, we systematically choose $\phi(x) = 1/u(x)$, where $u(x)$ is non-zero on the considered computational domain $\bar{\Omega} = [0; 1]$. This leads to $u(x)\phi(x) = 1$, hence $(u(x)\phi(x))' = 0$, for all x in Ω and $S(x) = 0$. Consequently, an exact solution associated to a given velocity field u is easily accessible. Moreover boundary conditions are given by $\phi_{\frac{1}{2},-} = 1/u(0)$ and $\phi_{I+\frac{1}{2},+} = 1/u(1)$, where u is known.

We experiment two algorithms: the linear (L) and the time-marching (TM) solvers which both employ a MOOD like approach to stabilize the computation. The methodology of testing is based on the following test cases with the previous manufactured solutions.

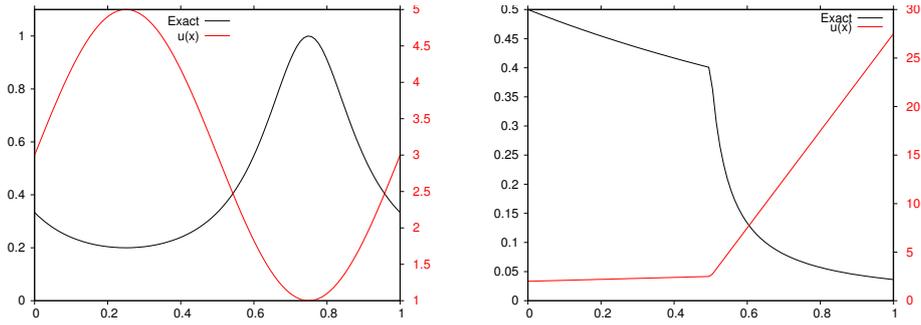


Fig. 5 Exact solutions (black) and associate velocities (red) for the linear equation on $\bar{\Omega} = [0; 1]$ — $\phi(x) = 1/u(x)$ — Left: regular solution $u(x) = 2 \sin(2\pi x) + 3$ — Right: irregular solution $u(x) = x + 2$, if $x < 1/2$ or $u(x) = 50(x - 1/2) + 5/2$, otherwise.

Regular solution (see Fig. 5-left). With this test, one tries to retrieve a regular steady-state solution of (6) with high accuracy. The unlimited version of the schemes are used to determine the maximal possible accuracy, then the MOOD-stabilized version of the schemes are employed and it is shown that optimal order of accuracy is attained. In other words we expect that no bad cells should be detected, leading *de facto* to a strict equivalence between unlimited and *a posteriori* MOOD schemes.

Irregular solution (see Fig. 5-right). An irregular solution is simulated to prove the oscillatory character of the unlimited schemes, which, however is cured by the MOOD approach leading to a stable, yet more accurate, solution than the one provided by the first-order scheme.

4.2.1 Regular solution

Let us consider the case where $u(x) = 2 \sin(2\pi x) + 3$ (see Fig. 5-left for a plot of the solution and the velocity) and the Linear and Time Marching schemes employ polynomial reconstructions \mathbb{P}_0 , \mathbb{P}_1 , and \mathbb{P}_5 . That is to say first-, second-, and sixth-order accurate schemes. The boundary conditions are set to $1/3$ at $x = 0$ and an outflow condition at $x = 1$. In Table 1 we present E_1 the errors in L_1 and E_∞ the L_∞ norms, the rates of convergence \mathcal{O}_1 and \mathcal{O}_∞ for L and TM schemes. Successively refined meshes made of $I = 70$ to 200 cells are considered. We also report the number of bad cells detected by the chain detector during the whole simulation (column labeled with the tag “ \mathcal{B} ”). From the numbers in this table we observe that the optimal convergence rates are systematically attained. Moreover no bad cells are detected meaning that the *a posteriori* MOOD scheme would produce exactly the same results. As a consequence, because no bad cells were detected, a *a posteriori* MOOD L or TM scheme also reaches the optimal order of accuracy associated to its d^{\max} for this smooth test case.

4.2.2 Irregular solution

In this second test we consider velocity

$$u(x) = \begin{cases} x + 2, & \text{if } x \in [0; 1/2], \\ 50(x - 1/2) + 5/2, & \text{if } x \in [1/2; 1], \end{cases}$$

leading to the solution depicted in Fig. 5 on right panel. This solution is continuous but not regular at location $x = 1/2$. The boundary conditions are set to $1/2$ at $x = 0$ and an outflow condition at

Table 1 Errors and convergence rates for a regular solution of linear equation for the Linear (L) and Time-Marching (TM) schemes. \mathcal{B} represents the number of bad cells detected by MOOD loop.

	I	Linear scheme					Time Marching scheme				
		E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}	E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}
\mathbb{P}_0	70	1.1E-02	—	2.8E-02	—	0	1.1E-02	—	2.8E-02	—	0
	80	1.0E-02	1.0	2.4E-02	1.0	0	1.0E-02	1.0	2.4E-02	1.0	0
	90	8.9E-03	1.0	2.2E-02	1.0	0	8.9E-03	1.0	2.2E-02	1.0	0
	100	8.0E-03	1.0	2.0E-02	1.0	0	8.0E-03	1.0	2.0E-02	1.0	0
	200	4.0E-03	1.0	9.8E-03	1.0	0	4.0E-03	1.0	9.8E-03	1.0	0
\mathbb{P}_1	70	2.9E-04	—	1.4E-03	—	0	2.9E-04	—	1.4E-03	—	0
	80	2.2E-04	2.1	1.1E-03	2.0	0	2.2E-04	2.1	1.1E-03	2.0	0
	90	1.7E-04	2.1	8.2E-04	2.1	0	1.7E-04	2.1	8.2E-04	2.1	0
	100	1.4E-04	2.1	6.7E-04	2.0	0	1.4E-04	2.1	6.7E-04	2.0	0
	200	3.4E-05	2.0	1.7E-04	2.0	0	3.4E-05	2.0	1.7E-04	2.0	0
\mathbb{P}_5	70	3.7E-07	—	3.2E-06	—	0	3.6E-07	—	3.1E-06	—	0
	80	1.7E-07	6.0	1.4E-06	6.1	0	1.6E-07	5.8	1.4E-06	6.2	0
	90	8.2E-08	6.1	7.4E-07	5.6	0	8.2E-08	5.9	7.0E-07	5.7	0
	100	4.3E-08	6.0	3.9E-07	6.1	0	4.5E-08	5.7	3.6E-07	6.2	0
	200	6.9E-10	6.0	6.4E-09	5.9	0	8.1E-10	5.8	5.6E-09	6.0	0

$x = 1$. The first set of simulations considers unlimited schemes and the results are presented in top panel of Fig. 6. Note that the L and TM solvers are producing the very same solutions, therefore we skip TM results (only the number of iterations, and, consequently the CPU time is larger for the TM scheme).

From the figures we deduce that the schemes are capturing the correct solution but with spurious oscillations for the high accurate schemes. On the same figure, on bottom panels, we plot the *a posteriori* MOOD limited solutions when the following cascades are employed: \mathbb{P}_0 (left panel, red color), $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ (middle panel, blue color) and $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ (right panel, green color). We can observe that the numerical solutions are now exempt from spurious oscillations in the vicinity of the irregularity. There, cells are updated with lower order schemes employing either \mathbb{P}_1 or \mathbb{P}_0 reconstructions. Note that the color of the cell corresponds to its cell polynomial degree. Consequently we observe that for the cascade $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ two cells are updated with first order of accuracy, while, for the nominally sixth-order accurate scheme two extra cells are updated with \mathbb{P}_1 reconstructions. As such the convergence rate can not exceed a first order.

Nonetheless, it may still be of interest to employ high accurate schemes. Indeed, in Fig. 7, we present the logscale of the cell error for the first-, second-, and sixth-order accurate L schemes and three meshes $I = 80, 240$, and 320 cells. (The results for the TM scheme are equivalent.) As expected the schemes behave alike close to the discontinuity. There, the high accurate schemes are not more accurate than first order as expected. Contrarily, in the smooth parts of the flow, the high accurate schemes perform with an appreciable high accuracy. Note that, even if the discontinuity is located on one point, its effect on the numerical error spreads along a cone, the width of which is larger with the nominal accuracy of the scheme. Away from this cone the error produced by a high accurate scheme dramatically drops towards machine precision.

5 The Burgers' equation

Burger's equation is an attractive benchmark equation since we are dealing with a very simple scalar non-linear model which may provide discontinuous solutions. The equation writes

$$\frac{1}{2} \frac{d\phi^2}{dx} = S(x, \phi), \quad x \in \Omega, \tag{9}$$

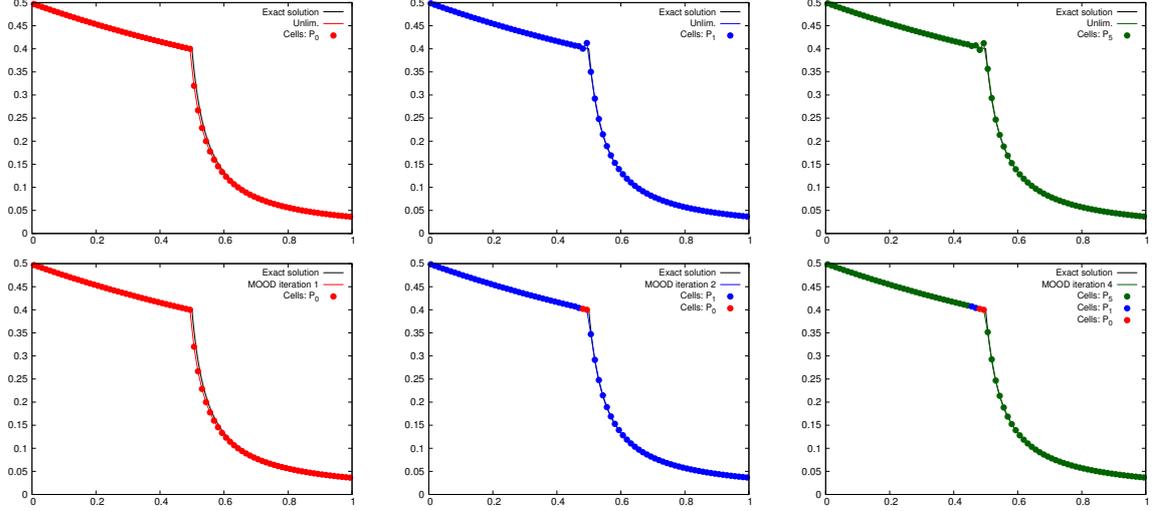


Fig. 6 Numerical solutions for linear equation and $I = 80$ cells — Top: unlimited schemes with reconstruction degrees 0 (left), 1 (middle), and 5 (right) — Bottom: *a posteriori* MOOD limited schemes with cascade \mathbb{P}_0 (left), $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ (middle), and $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ (right). (L and TM schemes provide similar pictures so only L solver's results are shown.)

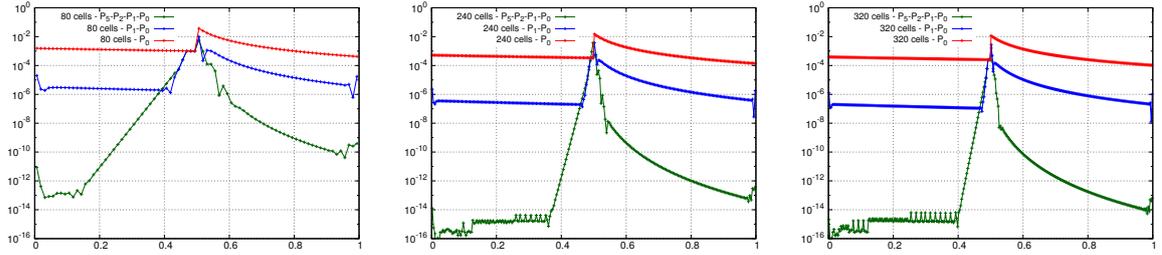


Fig. 7 Numerical cell errors in logscale for linear equation and $I = 80, 240$ and 320 cells from left to right — *a posteriori* MOOD limited schemes with cascade \mathbb{P}_0 (red), $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ (blue) and $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ (green). (L and TM schemes provide similar pictures so only L solver's results are shown.)

where $S(x, \phi)$ is a nonlinear source depending on the position and the solution.

5.1 Numerical flux

The numerical flux $\mathcal{F}(\phi_{i+\frac{1}{2},-}, \phi_{i+\frac{1}{2},+}; x_{i+\frac{1}{2}})$ is the classical Rusanov one and the chain detector associated to the Burgers' equation follows the chain depicted in Fig. 4 for the linear case. For Burgers equation we consider only two cascades: $d^{\max} = 1 \rightarrow 0$, $d^{\max} = 5 \rightarrow 2 \rightarrow 1 \rightarrow 0$, that is second-order and sixth-order accurate schemes. Moreover the first-order accurate scheme is carried out for comparison purposes.

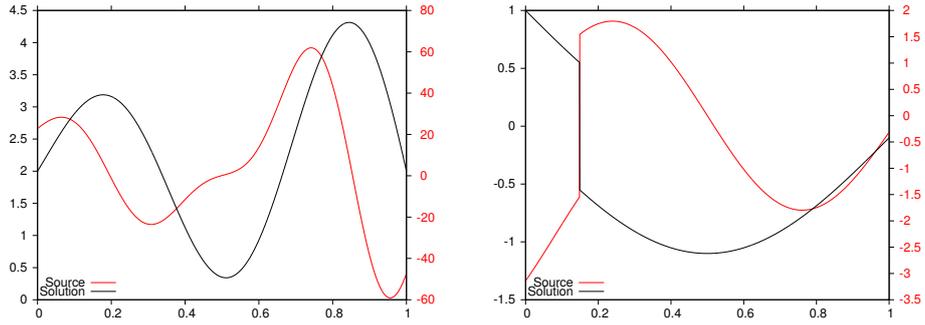


Fig. 8 Exact solutions (black) and source term (red) for each of the proposed problem for Burgers' equation — Left: regular solution — Right: irregular solution.

5.2 Numerical experiments

In this section we again experiment with the two algorithms, the nonlinear (NL) and the explicit time-marching (TM) solvers which both employ a MOOD like approach to stabilize the computation. The methodology of testing is based on simulating a regular solution (see Fig. 8-left) and an irregular one (see Fig. 8-right) following the experiments carried out for the linear equation. We adopt the same error norms and rate of convergence as for the linear equation, refer to (7)-(8).

5.2.1 Regular solution

In order to test the implementation of the numerical method we manufacture an exact solution. Let us consider the regular function

$$\phi(x) = \sin(3\pi x) \exp x + 2 \tag{10}$$

and formally compute the source term

$$S(x) = \frac{1}{2} \frac{d\phi^2}{dx^2} = (\sin(3\pi x) \exp x + 2) (3\pi \cos(3\pi x) \exp x + \sin(3\pi x)). \tag{11}$$

We set the domain to be $\overline{\Omega} = [0, 1]$ and compute the boundary conditions according to the exact solution, namely $\phi_{\frac{1}{2},-} = \phi(0)$ and $\phi_{I+\frac{1}{2},+} = \phi(1)$.

In Table 2 we present E_1 the errors in L_1 and E_∞ the L_∞ norms, and the rates of convergence \mathcal{O}_1 and \mathcal{O}_∞ for the NL and TM schemes employing \mathbb{P}_0 , \mathbb{P}_1 , and \mathbb{P}_5 polynomial reconstructions. Successively refined meshes made of 70, 80, 90, and 100 cells are considered. We also report the number of bad cells detected by the chain detector during the whole simulation with value \mathcal{B} the number of bad cells detected by the MOOD loop.

From the table we can conclude that both solvers provide the optimal expected order of accuracy. The NL solver converges in about 10 outer iterations for all simulation with a residual at machine tolerance. The TM scheme demands from 4000 to 5000 iterations to converge at machine tolerance for \mathbb{P}_0 and \mathbb{P}_1 polynomial degrees due to its explicit nature while the \mathbb{P}_5 reconstructions demands about 8000 to 10000 iterations to reduce the residual around 5×10^{-5} , hence explaining the non optimal order of convergence. We illustrate this behavior in Fig. 9 where we plot on the bottom panels the evolution of the residuals as a function of the nonlinear solver iterations (NL scheme) or as a function of the time cycles (TM scheme). From this plot we can observe that the evolution of the residual for the TM scheme presents a plateau at 0.1 during few thousands of cycles before dropping to another plateau at $\sim 10^{-6}$. For all other simulations, that is with

Table 2 Errors and convergence rates for a regular solution of Burgers' equation for the Nonlinear (NL) and Time-Marching (TM) schemes. \mathcal{B} represents the number of bad cells detected by MOOD loop.

	I	Nonlinear scheme					Time Marching scheme				
		E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}	E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}
\mathbb{P}_0	70	7.8E-02	—	1.9E-01	—	0	7.8E-02	—	1.9E-01	—	0
	80	6.8E-02	1.1	1.6E-01	1.0	0	6.8E-02	1.1	1.6E-01	1.0	0
	90	6.0E-02	1.0	1.5E-01	1.0	0	6.0E-02	1.0	1.5E-01	1.0	0
	100	5.4E-02	1.0	1.3E-01	1.0	0	5.4E-02	1.0	1.3E-01	1.0	0
\mathbb{P}_1	70	1.8E-03	—	7.4E-03	—	0	1.8E-03	—	7.4E-03	—	0
	80	1.4E-03	2.0	5.4E-03	2.4	0	1.4E-03	2.0	5.4E-03	2.4	0
	90	1.1E-03	2.0	4.1E-03	2.3	0	1.1E-03	2.0	4.1E-03	2.3	0
	100	8.6E-04	2.0	3.2E-03	2.3	0	8.6E-04	2.0	3.2E-03	2.3	0
\mathbb{P}_5	70	1.1E-07	—	8.8E-07	—	0	1.8E-07	—	6.2E-07	—	0
	80	4.9E-08	6.0	3.8E-07	6.2	0	1.0E-07	4.3	3.6E-07	4.0	0
	90	2.4E-08	6.1	1.8E-07	6.2	0	5.7E-08	4.8	2.1E-07	4.7	0
	100	1.2E-08	6.2	1.1E-07	5.2	0	3.3E-08	5.1	1.2E-07	5.0	0

different $d^{\max} = 0, 1$, the same plateau at 0.1 is systematically observed, but the dropping reaches machine precision.

On top panel of Fig. 9 we plot the solution provided by both schemes for 100 cells and \mathbb{P}_5 reconstructions. As expected, even if machine precision has not been reached for the TM scheme, the solution is visually indistinguishable from the NL converged one. Note that the solution is visually improved when the mesh is refined and/or the maximal polynomial degree d^{\max} increases (not shown). Also we have observed that the MOOD detector does not flag any cell during the whole simulation, $\mathcal{B} = 0$ systematically in Table 2, proving that a limiting procedure is not needed. As such MOOD limited TM or NL schemes provide exactly the same results as the unlimited ones. Consequently we do not show these tables and curves.

We have also tested other regular solutions such as polynomials, exponentials, etc., and the same conclusions hold. Consequently we also omit those results.

5.2.2 Irregular solution

Next, following [11] for instance, we will consider a discontinuous solution. Considering the source term $S(x) = -\pi \cos(\pi x)\phi(x)$ and the boundary conditions $\phi_{\frac{1}{2},-} = 1$ and $\phi_{I+\frac{1}{2},+} = -0.1$, we can determine the exact solution with shock

$$\phi(x) = \begin{cases} 1 - \sin(\pi x), & \text{if } x \in [0; x_s], \\ -0.1 - \sin(\pi x), & \text{if } x \in [x_s; 1], \end{cases} \quad (12)$$

where x_s is the location of the shock. Two shocks are possible — $x_s = 0.1486$ and $x_s = 0.8514$ — but only the first shock is stable for small perturbation. Exact solution is imposed at the boundaries of the domain. We plot in Fig. 8 the solutions (black line) along with the source term (red line).

Unlimited solvers. Let us first observe the behavior of unlimited TM and NL solvers. We represent in Fig. 10 the solutions produced by the NL and TM solvers when $d^{\max} = 0, 1$, and 5 and $I = 70$ cells are considered. We observe that all simulations are capable of capturing the shock wave but only the schemes without reconstructions (\mathbb{P}_0) produce monotone solutions. Any high accurate unlimited scheme, $\mathbb{P}_1, \mathbb{P}_5$ for TM or NL solvers does produce parasitical oscillations in the vicinity of the shock as expected. The same behavior is observed when the mesh is refined up to $I = 100$ cells. This behavior is confirmed by the number of bad cells detected which is of the order 5 for

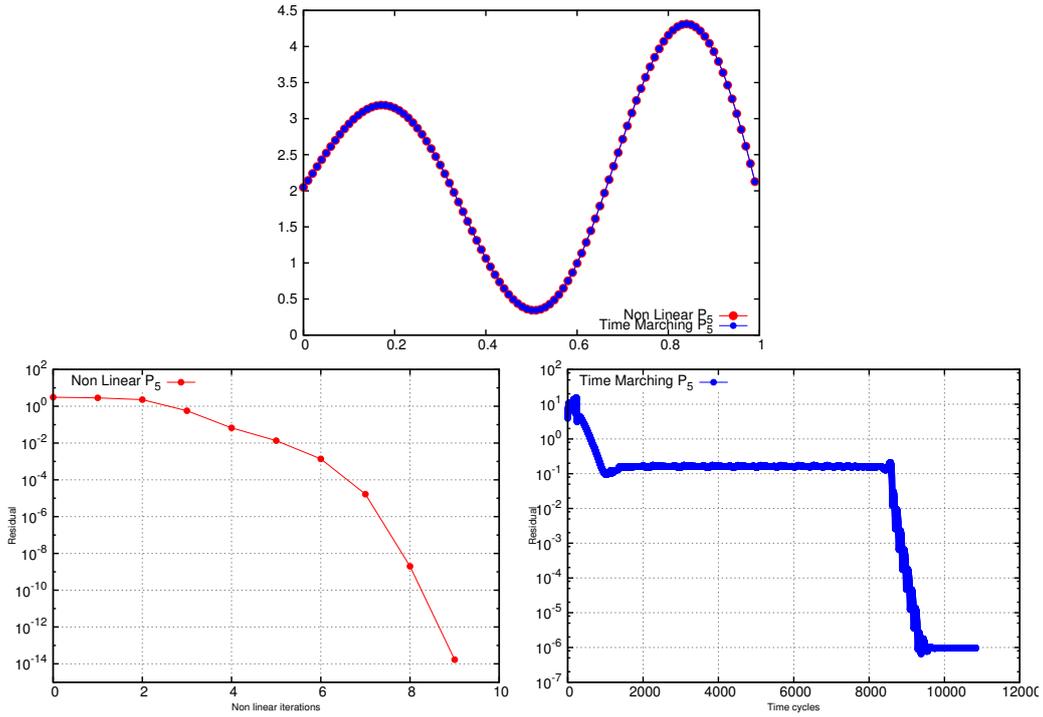


Fig. 9 Simulation of regular solution $\phi(x) = \sin(3\pi x) \exp x + 2$ of Burgers equation for the Nonlinear (NL, circle sybols) and Time-Marching (TM, cross symbols) schemes — 100 cells and \mathbb{P}_5 polynomial reconstructions — Top: numerical solutions. Bottom: evolution of the residuals as a function of the nonlinear solver iterations (NL scheme) or as a function of the time cycles (TM scheme).

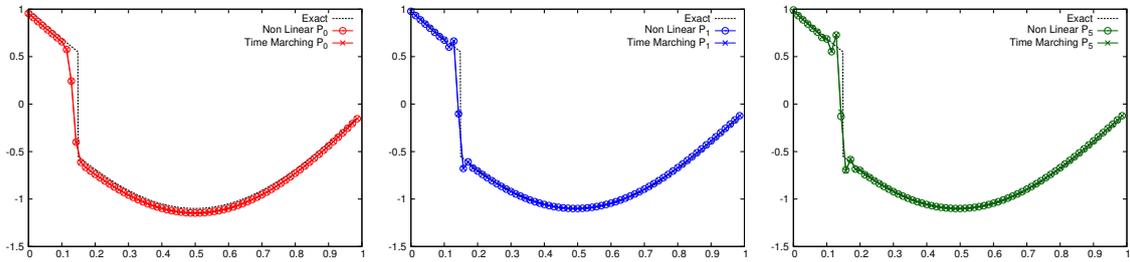


Fig. 10 Simulation of irregular solution of Burgers equation for the Nonlinear (NL, red) and Time-Marching (TM, blue) unlimited schemes — 70 cells — From left-to-right: solutions for \mathbb{P}_0 , \mathbb{P}_1 , and \mathbb{P}_5 polynomial reconstructions and exact solution (dashed line).

\mathbb{P}_1 and \mathbb{P}_5 simulations, and, of course 0 for \mathbb{P}_0 simulation. All simulations have converged with machine precision residual apart from the TM \mathbb{P}_5 simulations for which the residual saturates at about 10^{-4} .

Both solvers, NL and TM, are performing alike in terms of accuracy and, from now on we will only use the NL solver to pursue our test campaign.

2nd order limited MOOD NL solver. In this paragraph we only test the 2nd order limited MOOD NL solver, that is, with the cascade $d^{\max} = 1 \rightarrow 0$. The MOOD loop does produce several candidate

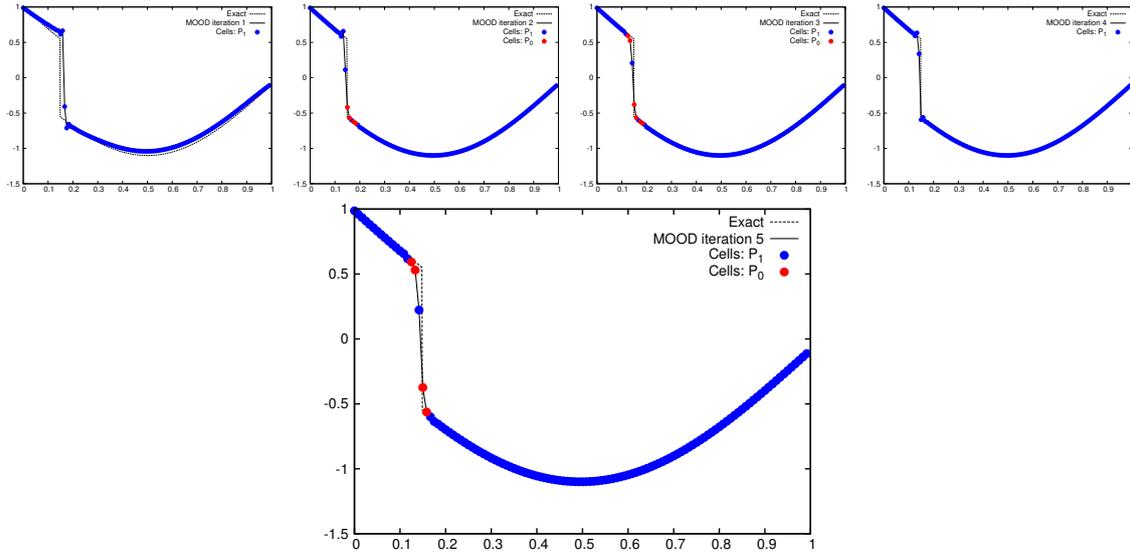


Fig. 11 Simulation of irregular solution of Burgers equation for the Nonlinear (NL) scheme with *a posteriori* MOOD limiting — 120 cells and $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ polynomial reconstructions — Top: numerical solutions for the first 4 iterations. Bottom: last iteration and valid solution.

solutions, each at convergence of the residual and an associated cell polynomial degree map that we plot in Fig. 11. As observed, for $I = 120$ cells, five MOOD iterations are needed to produce a valid solution in terms of the detection criteria. In red we emphasize the bad cells, updated with a first order accurate scheme, while in blue the cells updated with maximal 2nd order accuracy.

6th order limited MOOD NL solver. Next we test the 6th order limited MOOD NL solver with the cascade $d^{\max} = 5 \rightarrow 2 \rightarrow 0$. The MOOD loop produces several candidate solutions. Each candidate solution is obtained at convergence of the residual, and, has an associated cell polynomial degree map that we plot in Fig. 12 (colours). For $I = 120$ cells, ten MOOD iterations are needed to produce a valid solution in terms of the detection criteria. In red we emphasize the bad cells, updated with a first order accurate scheme, while in purple the cells updated with maximal 3rd order accuracy and in green the cells updated with the 6th order scheme. We present nine iterations in the figure, the final solution being plotted on the larger bottom panel. We observe that the bad cells are located in the vicinity of the shock wave as expected. About five bad cells for the final iterations need limiting. Note that the first MOOD iteration produces a solution which is not correct, but, when used as initial guess for the second iteration, it permits to capture the correct solution in the smooth parts. Then the other iterations fix the shock region.

Mesh refinement for the limited MOOD NL solver. Next, in Fig. 13, top panels, we present the behavior of the 6th order scheme when the mesh is refined (80, 120, 240, and 320 cells). The number of MOOD iterations is respectively 8, 10, 10, and 13. From this figure we can clearly observe that no spurious oscillations are present, while they were produced in previous iterations, and the number of corrected cells in the vicinity of the shock is about 4. For comparison purposes we have added on bottom panels the results obtained with a different cascade adding \mathbb{P}_1 reconstructions, *i.e.* $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$. The results are almost equivalent even if some cells are updated with \mathbb{P}_1 reconstructions. The number of iterations are respectively 5, 12, 15, and 17.

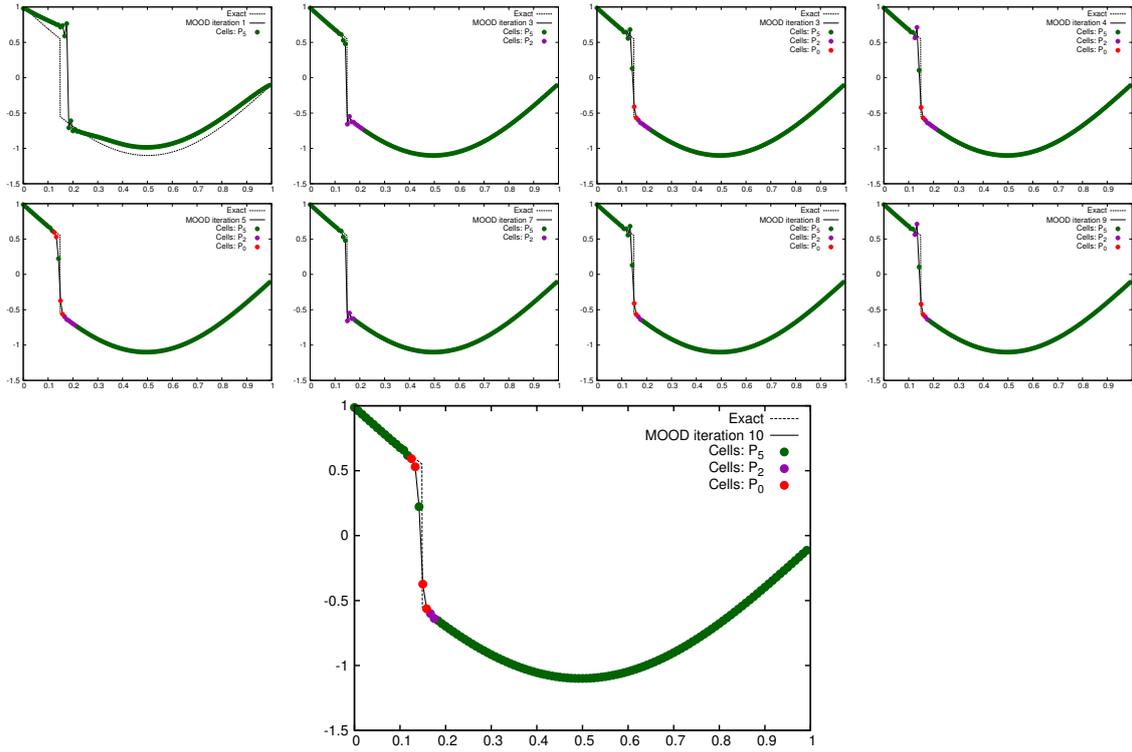


Fig. 12 Simulation of irregular solution of Burgers equation for the Nonlinear (NL) scheme with *a posteriori* MOOD limiting — 120 cells and $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$ polynomial reconstructions — Top: numerical solutions for 8 iterations. Bottom: last iteration and valid solution.

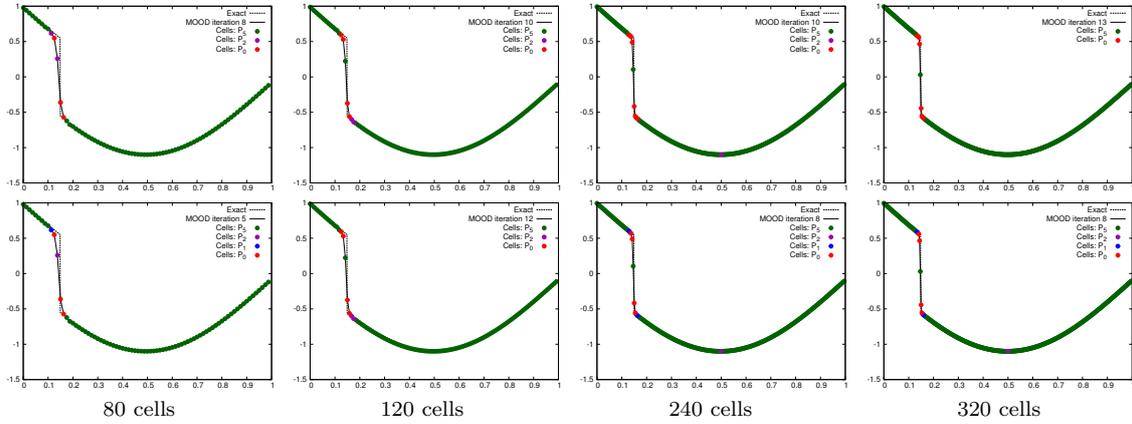


Fig. 13 Simulation of irregular solution of Burgers equation for the Nonlinear (NL) scheme with *a posteriori* MOOD limiting — Top: $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$ polynomial reconstructions — Bottom: $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ polynomial reconstructions — From left to right: last and valid solutions for 80, 120, 240, and 320 cells.

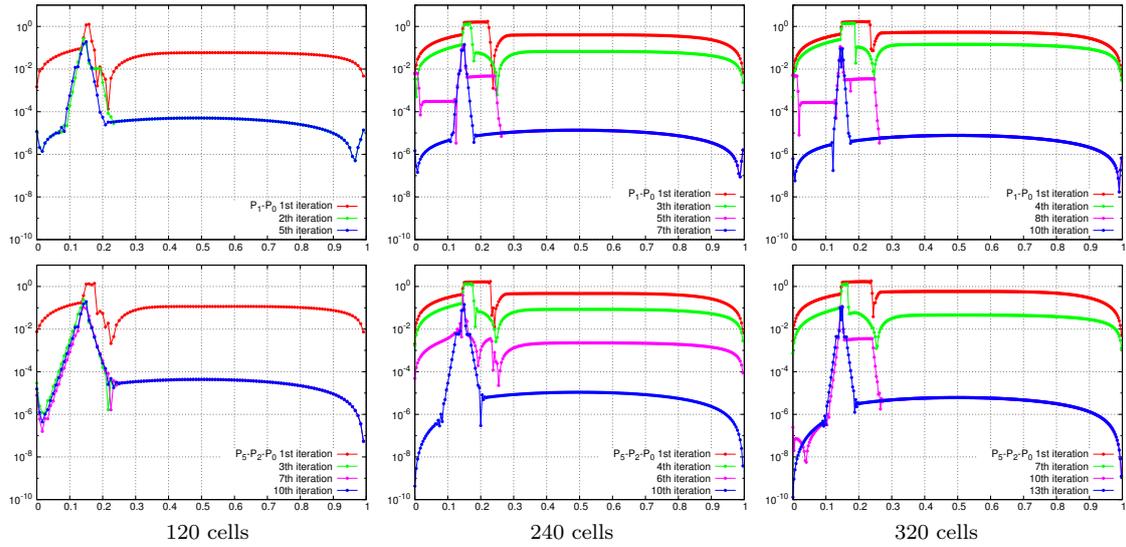


Fig. 14 Simulation of irregular solution of Burgers equation for the Nonlinear (NL) scheme with *a posteriori* MOOD limiting — Top: cascade $\mathbb{P}_1 \rightarrow \mathbb{P}_0$, Bottom: cascade $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$ — From left to right: logscale of the errors for 120, 240, and 320 cells for the first iteration (red), intermediate iterations (green/purple), and the last iteration (blue).

Last, we display in Fig. 14 the absolute value of the difference between the exact and the approximate solution for each cell for three of the previous meshes in logscale and for cascades $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ and $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$. Due to the discontinuous nature of the solution we can not expect more than first-order accuracy. Nonetheless we observe that the errors visually decrease with the iteration number and the mesh size. The maximal error is appropriately located in the vicinity of the shock wave, but the smooth parts of the flow are genuinely better approximated for the last iterations.

The *a posteriori* MOOD limiting is preserving the non-oscillatory character of the numerical solution for the highest order schemes. We observe that the procedure demands several iterations to reach a steady cell polynomial degree map. Nevertheless, once found, this map ensures the correctness of the solution by having the cell polynomial degrees reduced in the vicinity of the shock wave but nowhere else. We remind that the MOOD limiting *de facto* produces a (very) good guess. A very last step of the algorithm consists in using this guess and resetting the cell polynomial degree map to its maximal value.

In the case of the time-marching solver the MOOD procedure has more difficulties to converge as a “blinking” phenomena occurs due to the non-differential character of the method with respect to the CPD map (we jump from one degree to another one). Some pairs of cells alternatively drop then increase their degree so that the cell polynomial degree map can never attain a steady situation which is mandatory for convergence. This drawback could be solved (unsatisfactory) by setting a maximal number of iterations after which the map is kept fixed. Doing so the scheme converges at the price of fixing such a sensitive parameter.

6 The Euler system

We seek $\mathbf{V} \equiv \mathbf{V}(x) = (\rho, u, p)^T \equiv (\rho(x), u(x), p(x))^T$, the vector of state variables density, velocity, and pressure, solution of the steady-state one-dimensional Euler equations with source terms given

by

$$\frac{d\mathbf{F}(\mathbf{V})}{dx} = \mathbf{S}, \quad \text{in } \Omega, \quad (13)$$

where the flux vector is given by $\mathbf{F}(\mathbf{V}) = (\rho u, \rho u^2 + p, u(E + p))^T$ and the regular source vector term is given by $\mathbf{S} \equiv \mathbf{S}(x) = (\frac{dD}{dx}, \frac{dF}{dx}, \frac{dH}{dx})^T \equiv (\frac{dD}{dx}(x), \frac{dF}{dx}(x), \frac{dH}{dx}(x))^T$. Total energy E , specific internal energy e , and sound-speed a are expressed as

$$E = \frac{1}{2}\rho u^2 + \rho e, \quad e = \frac{p}{\rho(\gamma - 1)}, \quad a = \sqrt{\frac{\gamma p}{\rho}}, \quad (14)$$

respectively, where $\gamma > 1$ is the ratio of specific heat. System (13) is completed with Dirichlet boundary conditions

$$\mathbf{V}(x_L) = \mathbf{V}_L \equiv (\rho_L, u_L, p_L)^T \quad \text{and} \quad \mathbf{V}(x_R) = \mathbf{V}_R \equiv (\rho_R, u_R, p_R)^T,$$

depending on the flow regime at x_L and x_R . On the other hand, $\mathbf{U} = (\rho, \rho u, E)^T$ stands for the conservative variables vector.

We deduce the algebraic system to be solved

$$\rho u = D, \quad (15)$$

$$\rho u^2 + p = F, \quad (16)$$

$$u(E + p) = H, \quad (17)$$

$$\frac{1}{2}\rho u^2 + \frac{p}{\gamma - 1} = E. \quad (18)$$

Notice that D , F , and H depend on additional constants we shall fix with the boundary conditions.

6.1 Solution for the Euler system

We aim to derive exact solutions for system (15)-(18) that we shall employ as manufactured solutions in the numerical section. First of all, functions F and H have to respect some restrictions to provide a physical solution. From (15) and (16), we deduce that $p = F - \frac{D^2}{\rho}$. The pressure and density non-negativity yields that

$$F \geq 0, \quad \rho \geq \frac{D^2}{F}. \quad (19)$$

On the other hand from relations (17) and (14), we deduce that

$$p = \frac{\gamma - 1}{\gamma} \left(\frac{H}{u} - \frac{1}{2}\rho u^2 \right).$$

The non-negativity of the pressure and (15) implies

$$p = \frac{\gamma - 1}{\gamma} \left(\frac{H}{u} - \frac{1}{2}\rho u^2 \right) \geq 0 \implies Hu \geq \frac{1}{2}\rho u^4 \implies DH \geq \frac{1}{2}\frac{D^4}{\rho^2}. \quad (20)$$

6.1.1 Supercritical and subcritical solutions

Eliminating p and u from system (15)-(18) provides an implicit relation between D , F , H , and ρ given by

$$F = \frac{\gamma + 1}{2\gamma} \frac{D^2}{\rho} + \frac{\gamma - 1}{\gamma} \frac{\rho H}{D}, \quad (21)$$

such that ρ is a positive solution.

Equation (21) is rewritten as a quadratic polynomial in ρ of the form

$$(2(\gamma - 1)H)\rho^2 - (2\gamma DF)\rho + (\gamma + 1)D^3 = 0. \quad (22)$$

Existence of solution demands that

$$4D^2(\gamma^2 F^2 - 2(\gamma - 1)(\gamma + 1)DH) \geq 0 \implies \frac{DH}{F^2} \leq \frac{\gamma^2}{2(\gamma^2 - 1)} \quad (23)$$

leading to an additional restriction. From restrictions (19), (20), and (23) we define the admissible space

$$\Lambda = \{(D, F, H), F \geq 0, DH \geq 0, \gamma^2 F^2 - 2(\gamma^2 - 1)DH \geq 0\}. \quad (24)$$

Assuming $(D, F, H) \in \Lambda$, the supersonic solution $\hat{\rho}_{\text{sup}}(D, F, H)$ and the subsonic solution $\hat{\rho}_{\text{sub}}(D, F, H)$ correspond to the solution of (22) and are given by

$$\hat{\rho}_{\text{sup}}(D, F, H) = \frac{\gamma F - \sqrt{\gamma^2 F^2 - 2(\gamma^2 - 1)DH}}{2(\gamma - 1)\frac{H}{D}},$$

and

$$\hat{\rho}_{\text{sub}}(D, F, H) = \frac{\gamma F + \sqrt{\gamma^2 F^2 - 2(\gamma^2 - 1)DH}}{2(\gamma - 1)\frac{H}{D}},$$

respectively.

Remark 2 Solutions of the steady-state problem are constituted of branches of the supersonic and subsonic solutions with possible jump transitions. Consequently, since boundary conditions on the left and right sides have to satisfy the same relations (15)-(18), they also correspond to subcritical or subcritical states.

Remark 3 We also deduce an important relation between the total energy and the density. Indeed, eliminating ρu^2 between (16) and (18) and mutipling (17) by ρ , we obtain

$$\frac{F - p}{2} + \frac{p}{\gamma - 1} = E, \quad E + p = \frac{H}{D}\rho.$$

Eliminating the pressure between the two last equations gives an affine relation between the total energy and the density

$$E = \frac{\gamma - 1}{\gamma + 1}F + \frac{3 - \gamma}{\gamma + 1}\frac{H}{D}\rho. \quad (25)$$

6.1.2 Shock condition

We now analyse the condition to provide an admissible entropic steady-state shock.

Theorem 4 Assume that a steady-state solution admits an entropic stationary genuinely nonlinear shock at point x_c . Then there only exist two admissible situations:

- (a) the solution is supercritical on the left and subcritical on the right with $D > 0$ and $H > 0$;
- (b) the solution is subcritical on the left and supercritical on the right with $D < 0$ and $H < 0$.

Proof. We present the proof for the case $D > 0$ and $H > 0$, *i.e.* the velocity is positive. A steady-state genuinely nonlinear shock requires that the flow regime on the left and right sides of x_c are different. We state that a subcritical state on the left and supercritical state on the right is not possible. Indeed, a supercritical flow on the right side $u > 0$ means that three simple waves go out of the shock while only two simple waves enter on the left side (u and $u + c$). Consequently the shock is not entropic in the Lax sense. On the contrary, if the supercritical state lies on the left side, three simple waves enter in the shock while only two waves go out in accordance to the entropy principle. The entropic steady-state shock is then associated to the simple wave $u - c$. ■

One difficulty when dealing with steady-state solution is that the shock can take place at any point x_c of the domain. The transition location is not contained in system (15)-(18) and requires an additional restriction. To this end, we consider the steady-state situation as the limit case of the non-stationary problem solution $\tilde{\mathbf{U}} = (\tilde{\rho}, \tilde{\rho u}, \tilde{E})^T$ in conservative variables with initial conditions ρ^0 , $\rho^0 u^0$, and E^0 . We prescribe the same Dirichlet boundary conditions as in the steady-state situation.

Theorem 5 Assume that density $\tilde{\rho}(x, t)$, the mass flow $\tilde{\rho u}(x, t)$, and the total energy \tilde{E} are continuous at the boundary for $t > 0$, that is, there are no jumps between the boundary conditions and the solution at x_L and x_R (as a limit from the inner domain) for ρ , ρu , and E . Then for any time $t > 0$, we have

$$\int_{x_L}^{x_R} \tilde{\phi}(x, t) dx = \int_{x_L}^{x_R} \phi^0(x) dx, \quad \tilde{\phi} = \tilde{\rho}, \tilde{\rho u}, \tilde{E}.$$

Moreover, if the solution converges to a steady-state solution in time, denoted by ρ , u , and E , then we have

$$\int_{x_L}^{x_R} \rho^0(x) dx = \int_{x_L}^{x_R} \rho(x) dx, \quad \int_{x_L}^{x_R} \rho^0(x) u^0(x) dx = \int_{x_L}^{x_R} D(x) dx, \quad \int_{x_L}^{x_R} E^0(x) dx = \int_{x_L}^{x_R} E(x) dx.$$

Proof. Integrating the non-stationary mass conservation equation on the whole domain provides

$$\int_{x_L}^{x_R} \left(\frac{\partial \tilde{\rho}}{\partial t} + \frac{\partial(\tilde{\rho u})}{\partial x} \right) dx = \int_{x_L}^{x_R} \frac{dD}{dx} dx.$$

Integration by part yields

$$\frac{d}{dt} \int_{x_L}^{x_R} \tilde{\rho} dx + (\tilde{\rho u})|_{x=x_R} - (\tilde{\rho u})|_{x=x_L} = D(x_R) - D(x_L).$$

Since the boundary conditions satisfy the algebraic relation (15) with no discontinuity, we deduce that $(\tilde{\rho u})|_{x=x_L} = D(x_L)$, $(\tilde{\rho u})|_{x=x_R} = D(x_R)$ and so

$$\frac{d}{dt} \int_{x_L}^{x_R} \tilde{\rho} dx = 0.$$

Integration in time over the interval $[0, t]$ gives

$$\int_{x_L}^{x_R} \bar{\rho}(x, t) dx = \int_{x_L}^{x_R} \rho^0(x) dx.$$

Similar operations provide the two other relations.

Assuming that the unstationary solution converges to a limit steady-state solution, we get

$$\begin{aligned} \int_{x_L}^{x_R} \rho(x) dx &= \int_{x_L}^{x_R} \rho^0(x) dx, \\ \int_{x_L}^{x_R} \rho(x)u(x) dx &= \int_{x_L}^{x_R} \rho^0(x)u^0(x) dx = \int_{x_L}^{x_R} D(x) dx, \\ \int_{x_L}^{x_R} E(x) dx &= \int_{x_L}^{x_R} E^0(x) dx. \end{aligned} \quad \blacksquare$$

Remark 6 The second relation is automatically satisfied if one assumes that for $\rho^0 > 0$ given, we define u^0 such that $\rho^0 u^0 = D$. In the following, we assume that the velocity u^0 satisfies relation (15). In the same way, we assume that E^0 is compute with the relation (25) and provides

$$E^0 = \frac{\gamma - 1}{\gamma + 1} F + \frac{3 - \gamma}{\gamma + 1} \frac{H}{D} \rho^0. \quad (26)$$

Corollary 7 Assume that the steady-state solution is the limit of the unstationary problem with a unique discontinuity at point x_c . Moreover, assume that the initial density ρ^0 satisfies the condition

$$\int_{x_L}^{x_R} \rho_{\text{sup}}(x) dx \leq \int_{x_L}^{x_R} \rho^0(x) dx \leq \int_{x_L}^{x_R} \rho_{\text{sub}}(x) dx. \quad (27)$$

Assume at last that there exists $C \in \mathbb{R}$ such that on the whole domain

$$\frac{H(x)}{D(x)} = C. \quad (28)$$

Then we have only two situations.

(a) Case $D > 0, H > 0$.

The steady-state solution is constituted of the supercritical solution in $[x_L, x_c[$ and the subcritical solution on $]x_c, x_R]$ with a shock located at the unique point x_c such that

$$\int_{x_L}^{x_c} \rho_{\text{sup}}(x) dx + \int_{x_c}^{x_R} \rho_{\text{sub}}(x) dx = \int_{x_L}^{x_R} \rho^0(x) dx. \quad (29)$$

(b) Case $D < 0, H < 0$.

The steady-state solution is constituted of the subcritical solution in $[x_L, x_c[$ and the supercritical solution on $]x_c, x_R]$ with a shock located at the unique point x_c such that

$$\int_{x_L}^{x_c} \rho_{\text{sub}}(x) dx + \int_{x_c}^{x_R} \rho_{\text{sup}}(x) dx = \int_{x_L}^{x_R} \rho^0(x) dx. \quad (30)$$

Proof. We only consider the first case. Steady-state solutions have to satisfy relation (21), hence are composed of subsonic and supersonic branches we join with discontinuities. Under the assumptions given in Theorem 5, mass conservation reads

$$\int_{x_L}^{x_R} \rho(x) \, dx = \int_{x_L}^{x_R} \rho^0(x) \, dx.$$

From Theorem 4, the entropic shock requires that the supercritical solution lies on the left side of the shock situated at point x_c . Hence $\rho = \rho_{\text{sup}}$ for $x < x_c$ while $\rho = \rho_{\text{sub}}$ for $x > x_c$ leading to relation (29). Let

$$A_{\text{sup}} = \int_{x_L}^{x_R} \rho_{\text{sup}}(x) \, dx, \quad A_{\text{sub}} = \int_{x_L}^{x_R} \rho_{\text{sub}}(x) \, dx.$$

Since $\rho_{\text{sup}} < \rho_{\text{sub}}$, function

$$x_c \rightarrow \int_{x_L}^{x_c} \rho_{\text{sup}}(x) \, dx + \int_{x_c}^{x_R} \rho_{\text{sub}}(x) \, dx \in [A_{\text{sup}}, A_{\text{sub}}]$$

is strictly increasing with respect to x_c . Consequently condition (27) is mandatory to provide existence and uniqueness of x_c such that (29) holds. Relation (30) is obtained in a similar way.

The last point we have to check is that we must respect the additional constraint for the total energy. Since the initial total energy is defined by (26), we have, using relation (25)

$$\begin{aligned} \int_{x_L}^{x_R} E(x) \, dx &= \int_{x_L}^{x_R} \left(\frac{\gamma-1}{\gamma+1} F + \frac{3-\gamma}{\gamma+1} \frac{H}{D} \rho \right) dx \\ &= \frac{\gamma-1}{\gamma+1} \int_{x_L}^{x_R} F + \frac{3-\gamma}{\gamma+1} C \int_{x_L}^{x_R} \rho \, dx \\ &= \frac{\gamma-1}{\gamma+1} \int_{x_L}^{x_R} F + \frac{3-\gamma}{\gamma+1} C \int_{x_L}^{x_R} \rho^0 \, dx \\ &= \int_{x_L}^{x_R} E^0(x) \, dx. \end{aligned}$$

Consequently, the additional constraint is automatically fulfilled. ■

Remark 8 Notice that the corollary requires that $\frac{H}{D}$ is constant otherwise, the compatibility does not hold. In that case, continuity of the density and total energy at the boundary no longer holds.

Corollary 7 suggests the additional condition to fix the shock position by imposing the total mass. Therefore, the system (13) will be augmented with the relation

$$\int_{x_L}^{x_R} \rho(x) \, dx = A, \tag{31}$$

for a given $A \in [A_{\text{sup}}, A_{\text{sub}}]$.

Remark 9 Since the boundary conditions have to correspond to one of the two states (subsonic or supersonic), we have 4 possible situations. Attending to Theorem 4 and Corollary 7, and assuming $D > 0$ and $H > 0$, we have

- (a) supersonic states on both boundary points — the solution for this case will be the regular function ρ_{sup} ;
- (b) subsonic states on both boundary points — the solution for this case will be the regular function ρ_{sub} ;

- (c) supersonic state on the left boundary and subsonic state on the right boundary point — one shock sharing the supersonic solution situated on the left and the subsonic one on the right;
- (d) subsonic state on the left boundary point and supersonic state on the right boundary point — the case is impossible since no entropic shock transition is possible from a subcritical to a supercritical branch.

6.2 High accurate numerical scheme

The high-accurate scheme is derived from the one proposed for Burgers' equation. Namely, one has to define the numerical flux to plug into the residual formulation. Let us denote by $\mathbf{U} = (U_i)_{i=1,\dots,I}$ the $3 \times I$ matrix containing the approximations of the mean-values for ρ , ρu , and E while \mathbf{U}^k stands for the data associated to stage k . The numerical flux $\mathcal{F}(\mathbf{U}_-, \mathbf{U}_+; x)$ is the HLL one proposed in [19] where \mathbf{U}_- and \mathbf{U}_+ represent the left and right states at interface x .

For a given CPD map \mathcal{M}^k , we introduce the nonlinear operator

$$\mathbf{u} \in \mathbb{R}^{3 \times I} \rightarrow \mathcal{G}(\mathbf{u}, \mathcal{M}^k) \in \mathbb{R}^{3 \times I}$$

and we seek for \mathbf{u}^{k+1} such that $\mathcal{G}(\mathbf{u}^{k+1}, \mathcal{M}^k) = 0$. Again we consider either a nonlinear solver (NL) and the time-marching (TM) technique to solve the set of equations. Notice that for the NL case, an additional condition is considered to fix the shock position when dealing with two different regimes on the left and right boundaries. For the TM method, the initial condition for density will fix the position of the shock.

The chain detector associated to the Euler equation follows the chain depicted in Fig. 15. Here the PAD box checks the positivity of density and pressure. The MOOD procedure is adapted to this

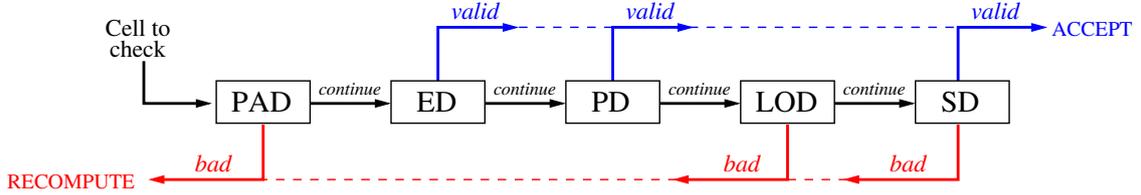


Fig. 15 Chain of detectors used for Euler system of equations to check if a cell value is valid and accepted, or, if it is discarded and need to be recomputed after reduction of the local cell polynomial degree. PAD: Physical Admissible Detector, ED: Extrema Detector, PD: Plateau Detector, LOD: Local Oscillation Detector, SD: Smoothness Detector.

system of equations. Because we consider only one polynomial degree per cell for each component of \mathbf{u} *i.e.* density, mass flow and total energy, we determine only one polynomial degree map by only testing the density variable for the ED, LOD, and SD detection criteria. Consistently with Burgers' equation experiments we consider only two cascades: $d^{\max} = 1 \rightarrow 0$ or $d^{\max} = 5 \rightarrow 2 \rightarrow 1 \rightarrow 0$. The MOOD loop is escaped if $\mathcal{M}^{k+1} = \mathcal{M}^k$.

6.3 Numerical experiments

To assess the validity of the method, we consider the following data. The domain is given by $x_L = 0$, $x_R = 1$ and we take $D(x) = 1$, $F(x) = 0.027x + 0.6137$, and $H(x) = 0.375$. The first two

conditions for the admissible space (24) are trivially satisfied. The third condition, $\frac{DH}{F^2} \leq \frac{\gamma^2}{2(\gamma^2-1)}$ is also satisfied since

$$\max_{x \in [0,1]} \frac{DH}{F^2} = \max_{x \in [0,1]} \frac{0.375}{(0.027x + 0.6137)^2} = 0.9957 \leq \frac{\gamma^2}{2(\gamma^2-1)} = 1.0208.$$

The expressions for the supersonic solution and the subsonic solution are given by

$$\rho_{\text{sup}}(x) = 0.126x - 3.3333\sqrt{1.96(0.027x + 0.6137)^2 - 0.72} + 2.8639, \quad (32)$$

$$\rho_{\text{sub}}(x) = 0.126x + 3.3333\sqrt{1.96(0.027x + 0.6137)^2 - 0.72} + 2.8639, \quad (33)$$

respectively and depicted in Fig. 16.

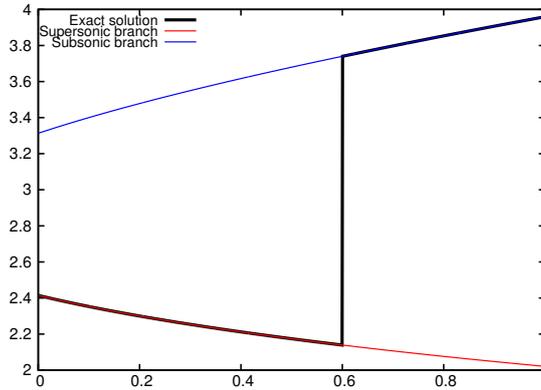


Fig. 16 Plots of the supersonic and subsonic branches ρ_{sup} (red) and ρ_{sub} (blue) from (32), and the exact solution of the irregular density (shock located at $x = 0.6$) employed in this paper (black).

As mentioned in Remark 9, only three situations have to be considered. We have two regular situations if considering supersonic or subsonic states on both boundary points. The case where we prescribe subsonic state at the left boundary point and supersonic state at the right does not provide an entropic solution and must be discarded. At last, prescribing supersonic state on the left boundary point and subsonic state on the right boundary produces a solution with a shock wave located at x_c . We adopt the same error norms and rate of convergence as for the linear and Burgers equations, refer to (7)-(8), and only monitor the error for the density variable.

6.3.1 Regular case

We first try to retrieve a regular steady-state solution of (13) with high accuracy. The unlimited versions of the schemes are used by canceled the MOOD stage to determine the maximal possible accuracy. Then the MOOD-stabilized version of the scheme is employed and it is shown that optimal order of accuracy is attained. We consider the regular supersonic/supersonic case which corresponds to the bottom-red curve of Fig. 16 and regular meshes made of $I = 25, 50, 75,$ and 100 cells and report in Table 3 the errors when the first-, second-, and sixth- order accurate MOOD NL and explicit TM schemes are used. Note that no bad cells are detected for this test case which implies that the unlimited or *a posteriori* MOOD limited schemes behave strictly equivalently. In the same way, the simulations for the regular solution on the subsonic/subsonic branch lead to the same conclusions and we omit those tables.

Table 3 Errors and convergence rates for the numerical density — Regular solution of Euler system for the Nonlinear (NL) and Time-Marching (TM) schemes. \mathcal{B} represents the number of bad cells detected by MOOD loop.

	I	Nonlinear scheme					Time Marching scheme				
		E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}	E_1	\mathcal{O}_1	E_∞	\mathcal{O}_∞	\mathcal{B}
\mathbb{P}_0	25	7.8E-03	—	1.3E-02	—	0	7.8E-02	—	1.9E-01	—	0
	50	3.9E-03	1.0	6.6E-03	1.0	0	6.8E-02	1.1	1.6E-01	1.0	0
	75	2.6E-03	1.0	4.4E-03	1.0	0	6.0E-02	1.0	1.5E-01	1.0	0
	100	2.0E-03	1.0	3.3E-03	1.0	0	5.4E-02	1.0	1.3E-01	1.0	0
\mathbb{P}_1	25	9.0E-05	—	9.3E-04	—	0	1.8E-03	—	7.4E-03	—	0
	50	1.9E-05	2.3	2.6E-04	1.8	0	1.4E-03	2.0	5.4E-03	2.4	0
	75	7.7E-06	2.2	1.2E-04	1.9	0	1.1E-03	2.0	4.1E-03	2.3	0
	100	4.1E-06	2.2	6.9E-05	1.9	0	8.6E-04	2.0	3.2E-03	2.3	0
\mathbb{P}_5	25	1.5E-07	—	3.4E-06	—	0	1.8E-07	—	6.2E-07	—	0
	50	2.9E-09	5.7	1.3E-07	4.7	0	1.0E-07	4.3	3.6E-07	4.0	0
	75	2.4E-10	6.1	1.6E-08	5.2	0	5.7E-08	4.8	2.1E-07	4.7	0
	100	4.0E-11	6.3	3.4E-09	5.4	0	3.3E-08	5.1	1.2E-07	5.0	0

6.3.2 Discontinuous case

A solution with a steady shock wave is simulated to prove that the oscillatory character of the unlimited scheme is cured by the MOOD approach leading to a stable and more accurate solution than the first order scheme. To fix the position at $x_c = 0.6$, we set

$$A = \int_{x_L}^{x_R} \rho^0(x) dx = \int_{x_L}^{x_R} ((\rho_L - \rho_R)(1-x)^{0.4550} + \rho_R) dx = 2.8975.$$

The exact solution for the density variable is depicted in Fig. 16. We enforce the position by using the additional restriction (31). Notice that the restriction is not necessary for the time marching method if we use ρ^0 and $u^0 = D/\rho^0$ as initial conditions. p^0 derives from relation (16) and E^0 is deduced from relation (18).

Parachute and first-order accurate schemes In Fig. 17 we present the numerical solutions for density variable and $I = 25, 50$, and 100 cells for the NL scheme with polynomial degree 0. From this plot we observe a convergence towards the exact solution when the mesh size increases without any spurious oscillation. This validates the robust and least accurate scheme that is used in the following as the parachute scheme of the cascade. On the right panel of Fig. 17 are displayed the logscale of the cell error for meshes of size $I = 50, 75, 100$, and 200. Because the scheme is nominally of first order of accuracy we observe on the smooth regions that the error decreases by a factor 2 from the 50 to 100 sizes. Obviously the maximal error is located in the vicinity of the shock and do not decrease when the mesh is refined, only the “numerical region of influence” reduces in size, nonetheless the number of cells in this region remains constant, about 4. The same study is carried out for the first-order accurate scheme with cascade $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ and the results are reported in Fig. 18. The same conclusions applies and the error in the smooth parts of the flow are about two orders of magnitude smaller.

a posteriori MOOD NL schemes. Next we test the nominally sixth-, second-, and first-order accurate NL scheme, that is, schemes using a maximal polynomial degree of 5, 1, and 0 and a MOOD loop to drop the degree to 0 if needed. The cascade of degrees are $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ for the sixth-order scheme and $\mathbb{P}_1 \rightarrow \mathbb{P}_0$ for the second-order one. The numerical densities obtained for 75 cells are shown in Fig. 19 and the cells are colored according to their polynomial degree. The

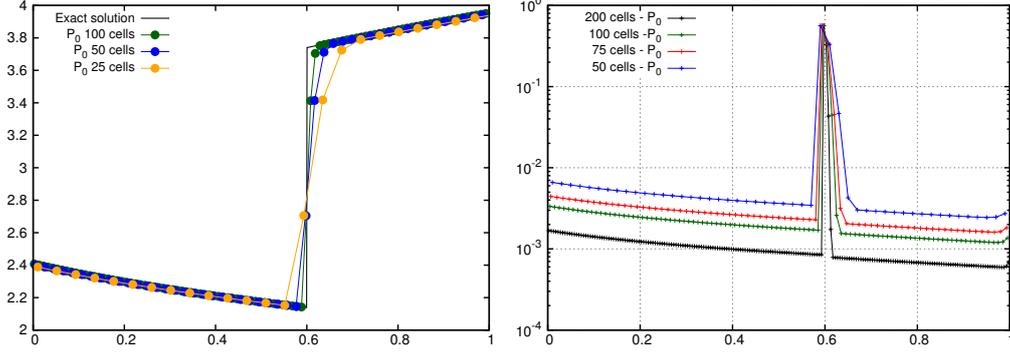


Fig. 17 Numerical solutions for density variable for the NL scheme *a posteriori* limited with maximal degree 0 — Left panel: density for $I = 25, 50,$ and 100 cells — Right panel: error for each cell for $I = 50, 75, 100,$ and 200 cells.

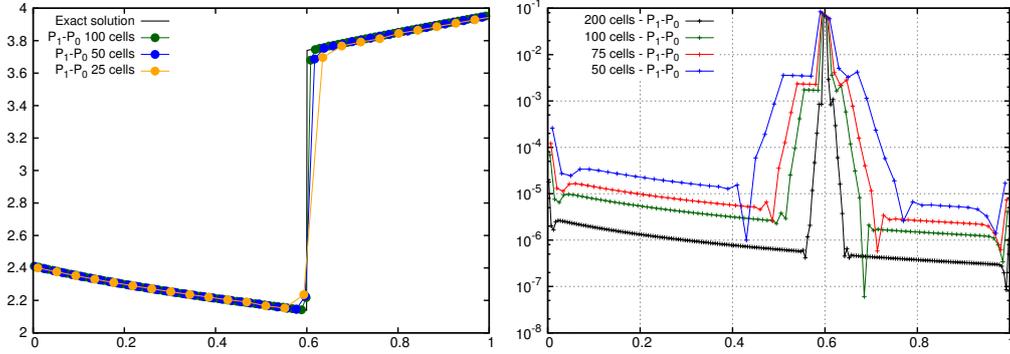


Fig. 18 Numerical solutions for density variable for the NL scheme *a posteriori* limited with maximal degree 1 — Left panel: density for $I = 25, 50$ and 100 cells — Right panel: error for each cell for $I = 50, 75, 100$ and 200 cells.

first-order accurate scheme needs one iteration while the second- and sixth-order accurate schemes demands two and nine respectively. We conclude that all the schemes produce a non-oscillatory solution, capture the correct shock location, employ the most accurate reconstructions far from the shock region, and decrement the polynomial degree of only few cells in the vicinity of the shock wave. We observe that the first iteration of any high-accurate scheme presents spurious oscillations. Then, the MOOD loop further decreases the polynomial degree up to capturing a monotone solution. The exact same behavior is observed when the mesh size is increased or decreased, so we skip these figures. The sixth-order accurate scheme can reproduce the sharp shock wave obtained by the second-order accurate scheme, by restricting its spread over one cell, while the first-order accurate scheme demands about three cells.

An interesting behavior can be observed when the cascade skips the \mathbb{P}_1 reconstruction, that is when the choices are restricted to $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$. Formally the scheme is still nominally sixth order accurate. The results are displayed in Fig. 20 and we observe that the spread of the shock is resembling the first-order scheme from top-panel of Fig. 19. We conclude that it is rather important to let the numerical method to choose \mathbb{P}_1 reconstructions when appropriate.

Next, we present in Fig. 21 the logscale of the cell error produced by the three schemes (first-, second-, and sixth-order accurate) for 75, 100, and 200 cells at convergence, that is for the final MOOD iteration. The left panel results are associated to the right-most curves of Fig. 20. We observe that the highest the maximal polynomial degree, the lower the error in smooth regions,

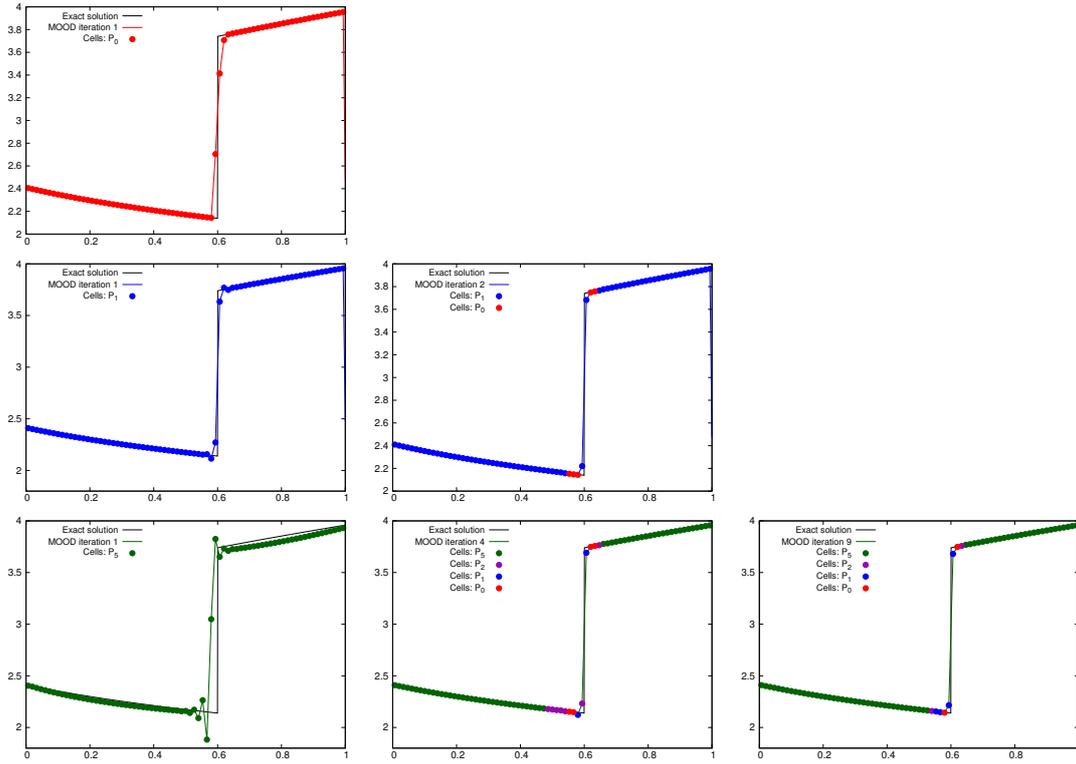


Fig. 19 Numerical solutions for density variable and $I = 75$ cells for the NL scheme *a posteriori* limited with maximal degree 0 (top line), 1 (middle line), and 5 with cascade $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ (bottom line) — Left to right panels: iteration 1, intermediate, and final. The right-most panel of each line is the last iteration of the MOOD loop.

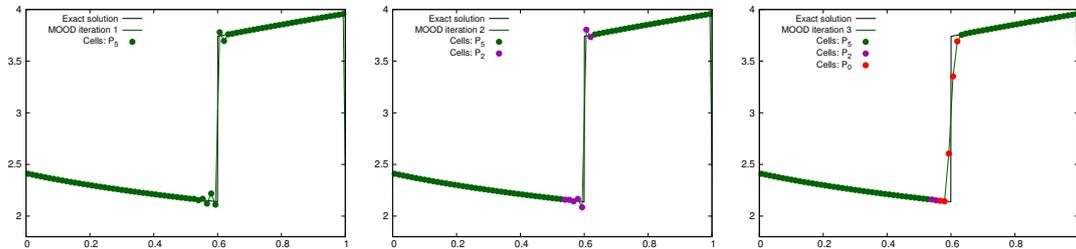


Fig. 20 Numerical solutions for density variable and $I = 75$ cells for the NL scheme *a posteriori* limited with maximal degree 5 with cascade $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_0$ — Left to right panels: iteration 1, intermediate, and final. The right-most panel of each line is the last iteration of the MOOD loop.

i.e. away from the shock wave. We notice a “characteristics error value” of Δx^k , where $k = 1, 2,$ and 6 for the three schemes, for the error produced in the smooth regions when enough cells are employed. Obviously, the maximal error is systematically reached in the vicinity of the shock wave, and, there, all schemes behave alike. However, moving away of 5 cells from this region, the high-accurate schemes are performing much better, and, the influence of the presence of a shock follows some sort of a “light cone”. The width of this cone does not seem to be related to the number of iterations needed to reach the steady-state solution.

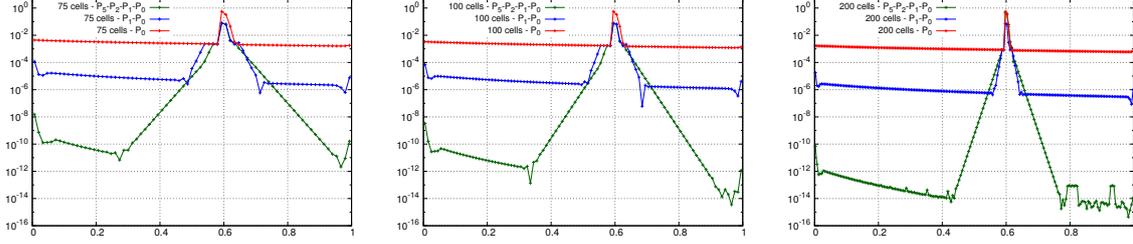


Fig. 21 Numerical errors for density variable and $I = 75, 100,$ and 200 cells for the NL scheme *a posteriori* limited with maximal degree 0 (red), 1 (blue), and 5 (green). Only the errors at final iteration are shown. The 75 cell results correspond to the right-most curves of Fig. 19.

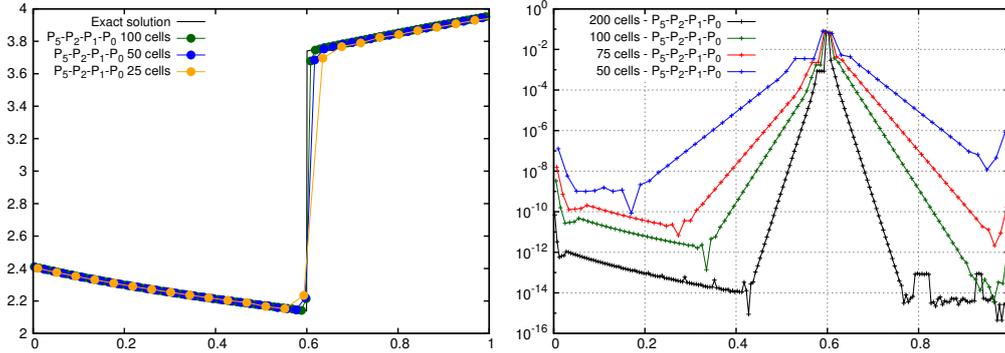


Fig. 22 Numerical solutions for density variable for the NL scheme *a posteriori* limited with maximal degree 5 with cascade $\mathbb{P}_5 \rightarrow \mathbb{P}_2 \rightarrow \mathbb{P}_1 \rightarrow \mathbb{P}_0$ — Left panel: density for $I = 25, 50,$ and 100 cells — Right panel: error for each cell for $I = 50, 75, 100,$ and 200 cells.

Mesh convergence for the sixth-order NL scheme. In Fig. 22 we present the numerical solutions for density variable and $I = 25, 50,$ and 100 cells for the NL scheme *a posteriori* limited with maximal degree 5 and an *a posteriori* MOOD correction loop. From this plot we observe a convergence towards the exact solution when the mesh size increases and no spurious oscillations are observed. This shows that the nominally sixth-order accurate numerical method is able to converge towards the exact solution. Obviously the convergence rate is still limited to the first-order one but, as previously seen, away from the shock wave, better accuracy can be expected. On the right panel of Fig. 22 are displayed the logscale of the cell error for meshes of size $I = 50, 75, 100,$ and 200 . Because the scheme is nominally of sixth order of accuracy we observe on the smooth region to the left away from the cone of influence, a drop of error of a factor ~ 100 when the mesh size is doubled (from 50 to 100 and 100 to 200). This roughly corresponds to 2^6 , the factor expected for a sixth-order accurate scheme.

7 Conclusion and perspectives

A new strategy to construct high accurate finite volume numerical schemes has been presented to solve the one-dimensional steady-state linear, Burgers, and Euler equations. High accuracy is gained via polynomial reconstructions whereas, in the presence of steep gradients or shock waves, the numerical schemes are stabilized with an *a posteriori* MOOD strategy. The technique consists in detecting bad cells on a candidate solution, and further recomputing the solution using lower

degree polynomial reconstructions in the vicinity of the bad cells. Doing so we define: (1) a chain of detectors which determines if a cell is valid or not, (2) a cascade of polynomial degree, that is a list of ordered polynomial degrees that one wishes to try, and (3) a parachute scheme that always produces a valid solution. The chain detector is based on classical physical admissibility, local smoothness, plateau, and extrema detection. The parachute scheme is always the first-order finite volume scheme. The cascades are of two kinds, a nominally second-order accurate one $\mathbb{P}_1 \rightarrow \mathbb{P}_0$, and a sixth-order accurate one $\mathbb{P}_5 \rightarrow \mathbb{P}_2(\rightarrow \mathbb{P}_1) \rightarrow \mathbb{P}_0$.

To perform the simulations, two solvers are considered: a non-linear implicit solver and a so-called explicit pseudo-time marching solver. The one-dimensional scalar linear equation, the Burgers scalar non-linear equation, and the one-dimensional Euler system with source terms were simulated. For each system of equation, several exact regular and irregular solutions are derived. The unknowns of these problems are not only the cell-centered finite volume values of the solution, but also the cell polynomial degree (CPD) map.

In the vicinity of a discontinuous solution, a low-order scheme must be employed (*i.e.* for instance using a polynomial of degree 0) to avoid Gibbs phenomenon, whereas the maximal degree is expected to produce a valid and accurate solution on smooth parts of the flow. An iterative MOOD loop is used to compute a candidate solution associated to its CPD map, then the chain detection checks for bad cells and modifies the CPD map accordingly. A new candidate solution is then recomputed and again checked for possible detected troubled cells.

Numerical experiments show that the optimal order of accuracy is systematically reached for smooth solutions. For non-smooth solutions, the *a posteriori* MOOD stabilization leads to non-oscillatory solutions while maintaining reasonable accuracy on the smooth parts of the flow. As expected the troubled cells are mostly located in the vicinity of discontinuities. We have numerically observed that the stabilization technique performs for a nominally second-order scheme but also for the sixth-order one without visible spurious effects.

The goal of the study is to present a proof of concept that the *a posteriori* MOOD stabilization strategy can be applied to the steady-state case. In the near future we plan to extend this approach to multi-dimensional situations taking into account classical boundary conditions (inflow, outflow, wall, symmetry) and to efficient solvers. The results for the one-dimensional geometry indicate that the method performs well and motivate to tackle the more complex situations involving two- and also three-dimensional geometries.

Acknowledgments

The material of this research has been partly built during the SHARK workshops taking place in Ofir, Portugal, <http://www.math.univ-toulouse.fr/SHARK-FV/>. The authors would like to acknowledge the financial support of Campus France through the PHC Pessoa labeled 26922YH and entitled “Investigation on very high-order finite volume numerical schemes for fluid hydrodynamics simulation”. The research was also supported by the Research Center CMAT of the University of Minho, Portugal, with the Portuguese Funds from the Fundação para a Ciência e a Tecnologia, through the Project PEstOE/MAT/UI0013/2014.

References

1. Abgrall, R.: On essentially non-oscillatory schemes on unstructured meshes: analysis and implementation. *J. Comput. Phys.* **144**, 45–58 (1994)
2. Abgrall, R., Shu, C.: Development of residual distribution schemes for discontinuous galerkin method: the scalar case with linear elements. *Commun. Comput. Phys.* **5**, 376–390 (2009)
3. Barth, T.: Recent developments in high order k-exact reconstruction on unstructured meshes. *AIAA Paper* 93-0668 (1993)

4. Barth, T., Frederickson, P.: High order solution of the Euler equations on unstructured grids quadratic reconstruction. AIAA paper 90-0013 pp. 1–10 (1990)
5. Barth, T., Jespersen, D.: The design and application of upwind schemes on unstructured meshes. AIAA paper 89-0366 pp. 1–12 (1989)
6. Bassi, F., Rebay, S.: A high-order accurate discontinuous finite element method for the numerical solution of the compressible Navier-Stokes equations. *J. Comput. Phys.* **131**, 267–279 (1997)
7. Bassi, F., Rebay, S.: High-order accurate discontinuous finite element solution of the 2D Euler equations. *J. Comput. Phys.* **138**, 251–285 (1997)
8. Beam, R., Warming, R.: An implicit finite-difference algorithm for hyperbolic systems in conservation law form. *J. Comput. Phys.* **22**(1), 87–110 (1976)
9. Beam, R., Warming, R.: Upwind second-order difference schemes and applications in aerodynamic flows. *AIAA J.* **14**(9), 1241–1249 (1976)
10. Breviglieri, C., Azevedo, J., Basso, E., Souza, M.: Implicit high-order spectral finite volume method for inviscid compressible flows. *AIAA Journal* **48**(10), 2365–2376 (2010)
11. Chen, W., Chou, C., Kao, C.: Fast sweeping methods for steady state problems for hyperbolic conservation laws. *J. Comput. Phys.* **234**, 452–471 (2012)
12. Chou, C., Shu, C.: High order residual distribution conservative finite difference weno schemes for steady state problems on non-smooth meshes. *J. Comput. Phys.* **214**, 698–724 (2006)
13. Clain, S., Diot, S., Loubère, R.: A high-order finite volume method for systems of conservation laws multi-dimensional optimal order detection (MOOD). *J. Comput. Phys.* **230**(10), 4028 – 4050 (2011). DOI 10.1016/j.jcp.2011.02.026. URL <http://www.sciencedirect.com/science/article/pii/S002199911100115X>
14. Clain, S., Figueiredo, J.: The mood method for the non-conservative shallow-water system. *Computers & Fluids* <http://dx.doi.org/10.1016/j.compfluid.2016.11.013> (Available online 24 November 2016) (2017)
15. Dervieux, A.: Steady Euler simulation using unstructured meshes. von Karman Institut for Fluid Dynamic, Lecture Series 1985-04 **1**(4) (1985)
16. Diot, S., Clain, S., Loubère, R.: Improved detection criteria for the multi-dimensional optimal order detection (MOOD) on unstructured meshes with very high-order polynomials. *Computers and Fluids* **64**, 43 – 63 (2012). DOI 10.1016/j.compfluid.2012.05.004. URL <http://www.sciencedirect.com/science/article/pii/S0045793012001909>
17. Diot, S., Loubère, R., Clain, S.: The MOOD method in the three-dimensional case: Very-high-order finite volume method for hyperbolic systems. *International Journal of Numerical Methods in Fluids* **73**, 362–392 (2013)
18. Haga, T., Sawada, K., Wang, Z.: An implicit lu-sgs scheme for the spectral volume method on unstructured tetrahedral grids. *Commun. Comput. Phys.* **6**(5), 978–996 (2009)
19. Harten, A., Lax, P., van Leer, B.: On upstream differencing and godunov-type schemes for hyperbolic conservation laws. *SIAM Review* **25**(1), 35–61 (1983)
20. Hu, C., Shu, C.: Weighted essentially non-oscillatory schemes on triangular meshes. *J. Comput. Phys.* **150**, 97–127 (1999)
21. Hu, G.: An adaptive finite volume method for 2d steady Euler equations with weno reconstruction. *J. Comput. Phys.* **252**, 591–605 (2013)
22. Hu, G., Li, R., Tang, T.: A robust high-order residual distribution type scheme for steady Euler equations on unstructured grids. *J. Comput. Phys.* **229**, 1681–1697 (2010)
23. Hu, G., Li, R., Tang, T.: A robust weno type finite volume solver for steady Euler equations on unstructured grids. *Commun. Comput. Phys.* **9**(3), 627–648 (2011)
24. Jameson, A., Mavriplis, D.: Finite volume solution of the two-dimensional Euler equations on a regular triangular mesh. *AIAA Journal* **24**(4), 611–618 (1986)
25. Jameson, A., Schmidt, W., Turkel, E.: Numerical solution of Euler Equations by finite volume methods using Runge-Kutta time-stepping schemes. AIAA paper 81-1259 pp. 1–17 (1981)
26. Jiang, G., Shu, C.: Efficient implementation of weighted eno. *J. Comput. Phys.* **126**, 202–228 (1996)
27. Knoll, D., Keyes, D.: Jacobian-free newton-krylov methods: a survey of approaches and applications. *J. Comput. Phys.* **193**, 357–397 (2004)
28. Li, R., Wang, X., Zhao, W.: A multigrid block lower-upper symmetric gauss-seidel algorithm for steady Euler equation on unstructured grids. *Numer. Math. Theor., Meth. and Appl.* **1**, 92–112 (2008)
29. Loubère, R., Dumbser, M., Diot, S.: A new family of high order unstructured mood and ader finite volume schemes for multidimensional systems of hyperbolic conservation laws. *Communication in Computational Physics* **16**, 718–763 (2014)
30. Maciel, E.: Explicit and implicit tvd high resolution schemes in 2d. *Journal of WSEAS Transactions on Applied and Theoretical Mechanics* **7**(3), 182–209 (2012)
31. Mavriplis, D.: Multigrid solution of the two-dimensional Euler equations on unstructured triangular meshes. *AIAA Journal* **26**(7), 824–831 (1988)
32. Michalak, K., Ollivier-Gooch, C.: Matrix-explicit gmres for a higher-order accurate inviscid compressible flow solver. Proc. 18th AIAA CFD Conf. AIAA paper 2007–3943 (2007)
33. Michalak, K., Ollivier-Gooch, C.: Limiters for unstructured higher-order accurate solutions of the Euler equations pp. AIAA paper 2008–776 (2008)

34. Mitchell, C., Walters, R.: K-exact reconstruction for the navier-stokes equation on arbitrary grids. AIAA Paper 93-0536 (1993)
35. Nejat, A., Ollivier-Gooch, C.: Effect of discretization order on preconditioning and convergence of a high-order unstructured newton-gmres solver for the Euler equations. *J. Comp. Phys.* **227**(4), 2366–2386 (2008)
36. Nejat, A., Ollivier-Gooch, C.: A high-order accurate unstructured finite volume newton-krylov algorithm for inviscid compressible flows. *J. Comp. Phys.* **227**(4), 2592–2609 (2008)
37. Nielsen, E., Anderson, W., Walters, R., Keyes, D.: Application of newton-krylov methodology to a three-dimensional unstructured Euler code. AIAA paper 95-1733 pp. 1–10 (1995)
38. Pesch, L., Vegt, J.: A discontinuous galerkin finite element discretization of the Euler equations for compressible and incompressible fluids. *J. Comput. Phys.* **227**, 5426–5446 (2008)
39. Pulliam, T.: Early development of implicit methods for computational fluid dynamics at nasa ames. *Computers & Fluids* **38**, 491–495 (2009)
40. Pulliam, T., Lomax, H.: Simulation of three- dimensional compressible viscous flow on the illiac iv computer. AIAA paper (79-0206) (1979)
41. Pulliam, T., Steger, J.: Implicit finite-difference simulations of three dimensional compressible flow. *AIAA J.* **18**(2), 159 (1980)
42. Steger, J.: Implicit finite difference simulation of flow about arbitrary geometries with application to airfoils. *AIAA J.* **16**(7), 679–686 (1978)
43. van Leer, B.: Towards the ultimate conservative difference scheme II: Monotonicity and conservation combined in a second order scheme. *J. Comput. Phys.* **14**, 361–370 (1974)
44. van Leer, B.: Towards the ultimate conservative difference scheme V: A second order sequel to Godunov’s method. *J. Comput. Phys.* **32**, 101–136 (1979)
45. Venkatakrishnan, V.: On the accuracy of limiters and convergence to steady state solutions. AIAA paper 93-0880 pp. 1–10 (1993)
46. Venkatakrishnan, V.: Convergence to steady state solutions of the Euler equations on unstructured grids with limiters. *J. Comput. Phys.* **118**, 120–130 (1995)
47. Venkatakrishnan, V., Mavriplis, D.: Implicit solvers for unstructured meshes. *J. Comput. Phys.* **105**(1), 83–91 (1993)
48. X.D. Liu, S.O., Chan, T.: Weighted essentially non-oscillatory schemes. *J. Comput. Phys.* **115**, 200–212 (1994)
49. Yee, H., Kutler, P.: Application of second-order-accurate total variation diminishing (tvd) schemes to the Euler equations in general geometries. AIAA paper **NASA-TM-85845** (1985)
50. Yee, H., Warming, R., Harten, A.: Implicit total variation diminishing (tvd) schemes for steady-state calculations. *J. Comput. Phys.* **57**, 327–360 (1985)
51. Zanotti, O., Dumbser, M., Loubère, R., S.Diot: A posteriori subcell limiting for discontinuous galerkin finite element method for hyperbolic system of conservation laws. *J. Comput. Phys.* **278**, 47–75 (2014)
52. Zhang, S., Shu, C.: A new smoothness indicator for the weno schemes and its effect on the convergence to steady state solution. *J. Sci. Comput.* **31**, 273–305 (2007)
53. Zhu, J., Qiu, J., Shu, C., Dumbser, M.: Runge-Kutta discontinuous galerkin method using WENO limiter II: Unstructured meshes. *J. Comput. Phys.* **227**, 4330–4353 (2008)