A High-Order Compact Limiter Based on Spatially Weighted Projections for the Spectral Volume and the Spectral Differences Method

Raphaël Lamouroux[1] · **Jérémie Gressier**[1] · **Gilles Grondin**[1]

Abstract This paper exposes the theoretical developments needed to design a class of spatially weighted polynomial projections used in the definition of a compact limiter dedicated to high-order methods. The spectral volume framework and its integral representation of the solution is used to introduce the degree reduction of the polynomial interpolation. The degree reduction is conducted through a linear projection onto a smaller polynomial space. A particular care is taken regarding the conservativity property and results in a parametric framework where projections can be monitored with spatial weights. These projections are used to define a simple and compact high-order limiting procedure, the SWeP limiter. Then, numerical evaluations are performed using the spectral differences method for the mono-dimensional Euler equations and demonstrate the high-order behavior of the SWeP limiter.

Keywords Compact limitation · High-order schemes · Spectral differences method · Order reduction · Hyperbolic conservation laws

1 Introduction

The past two decades saw major improvements in the development of high-order computational methods. It has been recognized that high-order schemes are valid candidates to address CFD challenges of the forthcoming years such as aeroacoustic, vortex-dominated flows, separation prediction, or turbulence awareness. One of their key features lies in their ability to deliver a higher level of accuracy for a computational cost lower than that of common

✉ Raphaël Lamouroux
raphael.lamouroux@isae.fr

Jérémie Gressier
jeremie.gressier@isae.fr

Gilles Grondin
gilles.grondin@isae.fr

[1] Institut Supérieur de l'Aéronautique et de l'Espace (ISAE-SUPAERO), Université de Toulouse, 31055 Toulouse Cedex 4, France

"low-order" methods such as finite differences or finite volume methods. Furthermore, most popular high-order methods come with a compact stencil favouring parallel data processing. Incidentally, industrials actors keep a close watch on their elaboration. For a complete and recent review of the status of the high-order methods in CFD, the reader is referred to [1–3].

The work presented in this paper is based on a major high-order method, the spectral volume method (SVM). The SVM is a compact high-order method that belongs to the more general class of Godunov-type finite volume methods [4,5]. Based on the extra degrees of freedom (DoFs) provided by an additional partitioning of the mesh cells, the SV delivers a compact high-order data reconstruction scheme. After Wang et al. [6] introduced the method in 2002, it was further developed through basic 1D [7] and 2D [8,9] numerical evaluation, before Liu et al. [10] extended the method to 3D systems . In 2006, Sun et al. [11] introduced viscous flow treatment in the SVM in order to extend the range of applications to the Navier–Stokes equations. The SVM was compared to the discontinuous Galerkin method by Zhang et al. [12] as they tested the stability, accuracy and convergence of both methods. To the knowledge of the authors, the most demanding computations carried out using SVM are due to Haga et al. [13]. Using the Japanese Earth simulator they completed 3D simulations of the transonic flow around ONERA-M6 wing and captured the λ-shaped shock structure. Different analyses of the SVM properties have also been performed. Van den Abeele et al. [14,15] studied its dissipative and dispersive behaviour and discovered weak instabilities for specific mesh cells partitions. Harris et al. [16] managed to eliminate these weak instabilities using some bi-dimensional optimization studies of the aforementioned cells partitions.

The implementation of the method has also been the subject of different studies. Harris et al. [17] developed an efficient variant of the SVM which avoided Gauss quadrature formula for flux computations in and brought it close to the spectral difference formulation. This methodology also lead to an *hp*-adaptive version of the SVM [18]. More recently, SVM schemes have exhibited some connections with the correction procedure via reconstruction or CPR schemes [19].

Solving a system of conservation laws like the Euler system is not an easy task since solutions may contain discontinuities even if a smooth initial condition is considered. When using high-order schemes, discontinuities tend to create spurious numerical oscillations. Consequences go from the simple noising of the solution to computation failure. To avoid the development of oscillations, two main strategies have emerged: resort to a limiting procedure or add artificial viscosity to the equations. The viscous approach is an active field of research, closely linked to the most recent developments regarding high-order schemes [20], although their origin can be traced back up to the 50's [21]. Nevertheless, the ideas of this paper are related to the first strategy. Most commonly used limiters include total variation bounded (TVB) limiter [22], moment limiter [23,24], MOOD limiter [25,26] and ENO/WENO limiter [27,28]. The main challenge with the design of a limiting procedure is in reaching a middle ground between non-oscillatory and high-order properties. This hindrance has been well overcome with the very new implementations of the WENO framework for the discontinuous Galerkin scheme [29] and the CPR scheme [30]. These studies show that high-order accuracy was preserved for smooth computations and at the same time discontinuities were well handled and oscillations strongly limited. In [29,30], important efforts are carried out to extend the stencil of the WENO scheme to its minimum. A numerical scheme generally falls into the category of the compact schemes if it only uses the information from its neighbouring cells and so are the WENO implementation described in [29,30]. Yet, in these cases, the knowledge of the solution field at *all* the DoFs contained in the immediate neighbouring cells is still needed and the strict compacity of the high-order scheme involved is lost. By *strict compacity* we mean that the unique additional informations usually needed (the informations

that need to be exchanged between processors) are those involved in the Riemann solver used for the fluxes definition, i.e. the interpolated state values from the neighbouring cells at their shared faces. This last property is very advantageous in a HPC context as the computing efficiency is highly dependent on the amount of data exchanged between the different CPUs.

This paper aims at exposing theoretical developments regarding a particular polynomial projection methodology used in the definition of a compact limiting procedure recovering high-order accuracy. The compactness level of the aforementioned limiting procedure is even better than that of most popular high-order schemes (spectral differences, discontinuous Galerkin,…), as the sole information contained in the cell is used to construct the limited solution. The idea of avoiding previously mentioned numerical issues with a reduction of the polynomial degree is present in the MOOD methodology [25] and is also at the basis of our approach. An extension of the SVM is introduced and serves as a framework where spatially weighted projections (SWeP) intuitively arise. Then, a compact limiting procedure based on the previous projections, the SWeP limiter, is proposed. In its setup, the projections are defined to adequately represent discontinuities thanks to the important amount of information per cell that characterizes high-order methods. Finally, so as to demonstrate the effectiveness of our limiter and its easy integration into a high-order scheme, numerical computations of the one-dimensional Euler equations are performed using the spectral differences method.

The present paper is built up as follows. In Sect. 2 we present the effects of the reduction of the polynomial degree of the solution representation for the spectral volume method and the spectral differences method. Section 3 exposes the theoretical development that led to the definition of specific compact projections. Then in Sect. 4, a limiting procedure based on the spatially weighted projections defined in Sect. 3, denoted as the SWeP limiting procedure, is presented. Finally, Sect. 5 gathers numerical simulations involving the proposed SWeP limiter.

2 Order Reduction for the Spectral Volume and the Spectral Differences Method

2.1 General Framework of the Order Reduction

Consider the hyperbolic conservation law over the spatial domain Ω bounded by $\partial\Omega$ and governed by the following partial differential equation in conservative form:

$$\partial_t u = -\nabla \cdot \boldsymbol{F}(u), \tag{1}$$

where $u(x,t)$ represents the scalar or vectorial conservative variable defined on $\Omega \times [0,T]$. Flux $\boldsymbol{F}(u)$ can be a linear or non-linear function of u. Domain Ω is divided into N^{s} cells S_i. That is:

$$\Omega = \bigcup_{i \in I^{\mathrm{s}}} S_i, \tag{2}$$

with $I^{\mathrm{s}} = [\![1, N^{\mathrm{s}}]\!]$. Let p be a positive integer and $\mathbb{P}_p[X]$ be the set of polynomials of degree at most p. For both the spectral volume method (SVM) and the spectral differences method (SDM), a polynomial solution $\hat{u}_h$ is sought in the finite element space of piecewise continuous functions:

$$V_h^p = \left\{ v_h \in L^\infty(\Omega) \mid v_{h\,|S_i} \in \mathbb{P}_p[X],\ i \in I^{\mathrm{s}} \right\}. \tag{3}$$

To be able to represent the N^{s} different polynomials of degree p, each cell hosts N^{c} degrees of freedom (DoFs). The different DoFs are gathered in vector $\boldsymbol{u}^{\mathrm{s}}$, which is for example an element of $\mathbb{R}^{N^{\mathrm{s}} \times N^{\mathrm{c}}}$ in the scalar case. One of the main differences between the SVM and the SDM is in the definition of these DoFs. The SDM uses a pointwise representation whereas the SVM uses an integral representation. The computation of the right hand side of (1) is as well specific to each scheme. As their name indicate, the SVM uses a finite-volume methodology, while the SDM is based on a finite-difference approach. Yet, despite those differences, both schemes can take the following general semi-discrete form:

$$\frac{\mathrm{d}\boldsymbol{u}^{\mathrm{s}}}{\mathrm{d}t} = \mathscr{R}\left(\hat{u}_h\right), \tag{4}$$

where residual $\mathscr{R}\left(\hat{u}_h\right)$ represents the spatial discretization operator. The function mapping the piecewise continuous polynomial $\hat{u}_h$ to the vector of DoFs $\boldsymbol{u}^{\mathrm{s}}$ is denoted by $\mathscr{L}$:

$$\mathscr{L} : \left\{ \begin{array}{ccc} V_h^p & \longrightarrow & \mathbb{R}^{N^{\mathrm{s}} \times N^{\mathrm{c}}} \\ \hat{u}_h & \longmapsto & \boldsymbol{u}^{\mathrm{s}} \end{array} \right. . \tag{5}$$

Let $\mathscr{B}^p = \left\{\phi_k^p \mid k \in [\![1, N^p]\!]\right\}$ be a basis of $\mathbb{P}_p[X]$. In order to uniquely define the polynomial in each cell, it is necessary that the number of DoFs per cell is equal to the cardinal N^p of basis $\mathscr{B}^p$. That is:

$$N^{\mathrm{c}} = N^p = \frac{(p+d)!}{p!d!}, \tag{6}$$

where d is the spatial dimension of the problem. If so, $\mathscr{L}$ is invertible and the semi-discrete scheme (4) becomes:

$$\frac{\mathrm{d}\boldsymbol{u}^{\mathrm{s}}}{\mathrm{d}t} = \mathscr{R} \circ \mathscr{L}^{-1}\left(\boldsymbol{u}^{\mathrm{s}}\right). \tag{7}$$

Here $\mathscr{L}^{-1}$ can be identified as part of the reconstruction step which is often used to interpolate the state of the solution.

The main feature of the SVM and SDM is the compact and high-order behaviour of the spatial discretization operator. However, when dealing with discontinuities (shocks), high-order interpolation often produces numerical oscillations (because of Runge's phenomenon for example) or overshoots. The different TVD [31] and TVB [32] schemes strategies or, as mentioned in the introduction, the use of artificial viscosity or limiters aim at reducing these oscillations. More recently, the extension of the WENO schemes family to the discontinuous Galerkin method (DG) [33,34] or the correction procedure via reconstruction scheme (CPR) [30,35] have contributed to the development of high-order methods for the computation of discontinuous flows. Yet, when using the WENO limiting procedure, the strict compactness of the stencil is lost, reducing the computation parallelization efficiency.

In order to damp the oscillations and the overshoots that can appear with high-order interpolations without altering the compactness of the scheme, we consider a direct reduction of the polynomial degree instead of a concurrent and a posteriori interpolation method. This strategy is based on the idea that the problem addressed, the oscillations, are represented in a polynomial expansion by the polynomials of the highest degree, like they are represented by high order trigonometric functions in a Fourier series. This property is used for example by Persson et al. [36] to activate artificial viscosity in their scheme, or more recently by Sheshadri et al. [37] in the definition of a shock-detection procedure based on the concentration method. Moreover, we will explain later in this section how the degree reduction frees some DoFs and then allow us to choose a more suitable, i.e. less oscillating, representation of the solution. Next sections will provide the technical details of the procedure employed. These details

are for the moment not necessary to consider the consequences of such a reduction. Let $r \in [\![0, p]\!]$ be the *reduction order* and let

$$\mathscr{P} : V_h^p \longrightarrow V_h^{p-r} \tag{8}$$

be a linear projection from V_h^p onto V_h^{p-r} (like for example the L^2 projection). Its composition with $\mathscr{L}^{-1}$ is a lesser order reconstruction step and will be referred to as:

$$\mathscr{L}^+ = \mathscr{P} \circ \mathscr{L}^{-1}. \tag{9}$$

Then, we can define new semi-discrete schemes with:

$$\frac{\mathrm{d}\boldsymbol{u}^{\mathrm{s}}}{\mathrm{d}t} = \mathscr{R} \circ \mathscr{L}^+ \left(\boldsymbol{u}^{\mathrm{s}}\right). \tag{10}$$

Because the polynomial reconstruction used in this scheme is of lower degree, the order of accuracy of the scheme is theoretically lowered too. The scheme will be referred to as a reduced order scheme.

2.2 Consequences of the Order Reduction

In order to evaluate the scheme of reduced order (10), the case of the scalar advection equation with constant velocity is implemented. The following problem is considered:

$$\partial_t u = -a \partial_x u, \quad x \in [0, 1]. \tag{11}$$

with unitary velocity ($a = 1$). A sinusoidal initial condition u^0 is considered:

$$u^0(x) = \frac{1 + \cos(4\pi x)}{2}, \quad x \in [0, 1], \tag{12}$$

We set up a first order reduction with the implementation of the spectral volume method of order 3 reduced to order 2 ($p = 2, r = 1$). The 1D-computational domain has a unit length and is uniformly meshed with 50 cells, each partitioned with the Gauss-Legendre points, generating 150 DoFs. Periodic boundary conditions are considered. The simulation time is $T = 1/4$, so that the signal has moved one fourth of the computational domain downstream. Semi-discrete scheme (10) is advanced in time using the forward Euler scheme. The time step is $\Delta t = 1 \times 10^{-2}$ and the resulting CFL number based on the size of the cells is 5×10^{-1}. Upwind flux is used for the computation of $\mathscr{R}$.

The results are presented in Fig. 1. The solution displays significant spurious oscillations that find an origin in Eq. (9). For an effective reduction, the kernel of $\mathscr{P}$ contains elements different from the null polynomial $\hat{0}$ of V_h^p:

$$\forall r > 0, \quad \left\{\hat{0}\right\} \subsetneq \ker(\mathscr{P}) \tag{13}$$

Let $\boldsymbol{u}^{\mathrm{s}}$ be an element of $\mathbb{R}^{N^{\mathrm{s}} \times N^{\mathrm{c}}}$ and $\boldsymbol{u}^{\mathrm{s,k}}$ be the component of $\boldsymbol{u}^{\mathrm{s}}$ in $\ker\left(\mathscr{L}^+\right) = \mathscr{L}(\ker(\mathscr{P}))$. Then by definition of the kernel:

$$\mathscr{R} \circ \mathscr{L}^+ \left(\boldsymbol{u}^{\mathrm{s}} - \boldsymbol{u}^{\mathrm{s,k}}\right) = \mathscr{R} \circ \mathscr{L}^+ \left(\boldsymbol{u}^{\mathrm{s}}\right) \tag{14}$$

The computation of the residual will provide the exact same value whether $\boldsymbol{u}^{\mathrm{s,k}}$ is taken into account or not. Thus, thanks to Eq. (10) we can write:

$$\frac{\mathrm{d}\boldsymbol{u}^{\mathrm{s,k}}}{\mathrm{d}t} = \frac{\mathrm{d}\boldsymbol{u}^{\mathrm{s}}}{\mathrm{d}t} - \frac{\mathrm{d}\left(\boldsymbol{u}^{\mathrm{s}} - \boldsymbol{u}^{\mathrm{s,k}}\right)}{\mathrm{d}t} = \boldsymbol{0}, \tag{15}$$

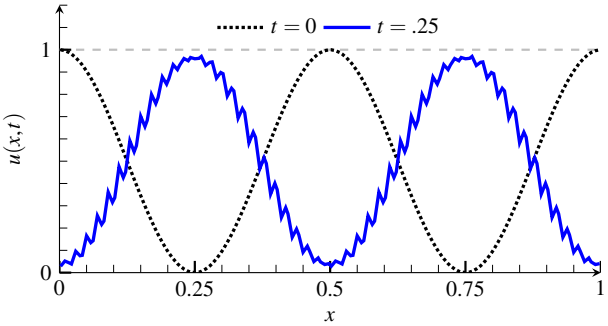

Fig. 1 Computation of the advection equation with the first order reduction of the third order SV scheme. Components from the kernel of $\mathscr{L}^+_{p-r}$ are not advected and generate spurious oscillations

where $\mathbf{0}$ stands for the null vector. Equation (15) indicates us that the elements of the kernel have the distinctive property to be strictly conserved over each cell as long as the time integration scheme cannot affect them. This particularity generates spurious oscillations as those visible in Fig. 1. Indeed, after each time step, the different DoFs have been updated by the time-marching scheme with the residual, which is, except in some very special cases, a piecewise polynomial of degree p. Therefore, the updated solution vector $\boldsymbol{u}^{\mathrm{s}}$ is a representative of an element of V_h^p that can be split into two parts, one in V_h^p and the other in $\ker\left(\mathscr{L}^+\right)$. Thus, new kernel components appear in every cell where they remain stuck and pollute the solution.

Besides, the functions used to construct the basis $\mathscr{B}^2$ have not been presented. However, the element of prime importance here is the only existence of the kernel of $\mathscr{P}$. Different function choices would only affect the condition number of the matrices involved.

This leads to the following conclusion: the sole projection of $\hat{u}_h$ for the computation of the residual is insufficient, $\hat{u}_h$ also needs to be effectively projected before it is used in the time-marching scheme. The components of $\boldsymbol{u}^{\mathrm{s}}$ in the kernel of $\mathscr{L}^+$ have to be removed, otherwise, these elements of the kernel are unnoticed by the numerical scheme and kept stationary over the mesh, hence, polluting the results (see Fig. 1). It also must be considered that to preserve the conservativity of the scheme, the kernel of the projection, whose elements must be removed, should only be composed of zero mass element, i.e. :

$$\forall \hat{u}_h \in \ker\left(\mathscr{P}\right), \quad \int_{\Omega} \hat{u}_h \, \mathrm{d}V = 0. \tag{16}$$

It has to be noted that "mass" generally refers to any conservative quantity of Eq. 1, like the mass, the momentum or the energy if considering the Euler equations.

Eventually, if an order reduction is considered, the projected vector which has to be used in the time-marching scheme is:

$$\boldsymbol{u}^{\mathrm{s,new}} = \boldsymbol{u}^{\mathrm{s}} - \boldsymbol{u}^{\mathrm{s,k}} = \mathscr{L} \circ \mathscr{L}^+ \left(\boldsymbol{u}^{\mathrm{s}}\right), \tag{17}$$

For example with the forward Euler scheme, the reduction order lead to compute solution vector $\boldsymbol{u}^{\mathrm{s}}$ at instant t^{n+1} using:

$$\frac{\boldsymbol{u}^{\mathrm{s}}\left(t^{n+1}\right) - \boldsymbol{u}^{\mathrm{s,new}}}{\Delta t} = \mathscr{R} \circ \mathscr{L}^+ \left(\boldsymbol{u}^{\mathrm{s}}\right), \tag{18}$$

which can also be formulated using (17) as:

$$\frac{\boldsymbol{u}^{\mathrm{s}}\left(t^{n+1}\right)-\boldsymbol{u}^{\mathrm{s,new}}}{\Delta t}=\mathscr{R}\circ\mathscr{L}^{-1}\left(\boldsymbol{u}^{\mathrm{s,new}}\right). \tag{19}$$

The fully discretized scheme takes the exact same form as the WENO scheme presented in [29,30]. The general idea presented in this paper of a limiter using the order reduction is naturally compelled to follow the implementation of the aforementioned WENO limiting procedure. Both consist in:

- the addition of a step of projection of the solution,
- the use of this projected solution in the time-marching scheme.

At this point, the projection applied to perform the reduction of the polynomial interpolation degree has not been defined. We will show in the next section that a particular care has to be taken regarding its definition in order to preserve the conservativity of the scheme. From them on and without lack of generality, the spectral volume framework and its integral representation is going to be employed to define parametric conservative projections of the solution and finally, to introduce a compact limiting procedure which, like the WENO scheme [29,30], is adapted to the spectral volume and the spectral differences methods.

3 Definition of Conservative Projections

The general consideration conducted in Sect. 2 about the reduction of the solution polynomial degree leads to two important facts:

- the reduction of the polynomial degree has to be seen as an independent projection step,
- to ensure conservativity, the elements in the kernel of the projection have to be zero mass elements.

In this section, the integral representation of the SVM is introduced to set up the framework in which parametric projections are defined. Then, algebraic developments lead to a set of constraints that ensure the conservativity of the projections. A numerical example is proposed to illustrate the different remarks. Finally, we expose with some examples of weighted projections how we intend to use these projections in the definition of a compact limiting procedure.

3.1 The Order Reduction Through an Algebraic Projection

According to the conservativity criterion, it is natural to resort to the integral representation of the spectral volume method. Let $\mathscr{B}^p$ be a basis of $\mathbb{P}_p[X]$ and let $\hat{u}_{i,k}$ be the coefficient of the polynomial representation of the solution $\hat{u}_h$ over cell S_i:

$$\hat{u}_h(\boldsymbol{x})=\sum_{k=1}^{N^p}\hat{u}_{i,k}\phi_k^p(\boldsymbol{x}),\quad \boldsymbol{x}\in S_i. \tag{20}$$

The SVM representation uses a partition of each cell of the mesh in N^{c} subcells $C_{i,j}$ called the control volumes (CVs). Using $I^{\mathrm{c}}=[\![1,N^{\mathrm{c}}]\!]$, the spatial mesh can be described in terms of cells and CVs with:

$$\Omega = \bigcup_{i \in I^{\mathrm{s}}} S_i, \tag{21}$$

$$S_i = \bigcup_{j \in I^{\mathrm{c}}} C_{i,j}, \quad i \in I^{\mathrm{s}}. \tag{22}$$

We impose $N^{\mathrm{c}} = N^p$. The integral representation of the polynomial solution over cell S_i is introduced as:

$$\overline{u}_{i,j} = \frac{1}{V_{i,j}} \int_{C_{i,j}} \hat{u}_h \, \mathrm{d}V, \quad i \in I^{\mathrm{s}}, \ j \in I^{\mathrm{c}}, \tag{23}$$

with $V_{i,j}$ being the volume of CV $C_{i,j}$. Then, using the following vectors:

$$\overline{\boldsymbol{u}}_i = \begin{pmatrix} \overline{u}_{i,1} \\ \vdots \\ \overline{u}_{i,N^{\mathrm{c}}} \end{pmatrix}, \quad \hat{\boldsymbol{u}}_i = \begin{pmatrix} \hat{u}_{i,1} \\ \vdots \\ \hat{u}_{i,N^p} \end{pmatrix}, \quad i \in I^{\mathrm{s}}, \tag{24}$$

and introducing the invertible matrix $\boldsymbol{L}_{\mathrm{b}}$ as:

$$(\boldsymbol{L}_{\mathrm{b}})_{jk} = \frac{1}{V_{i,j}} \int_{C_{i,j}} \phi_k^p \, \mathrm{d}V, \quad j \in I^{\mathrm{c}}, \ k \in [\![1, N^p]\!], \tag{25}$$

the coefficients vector $\hat{\boldsymbol{u}}_i$ is mapped to the integral representation $\overline{\boldsymbol{u}}_i$ of the polynomial solution with:

$$\boldsymbol{L}_{\mathrm{b}} \hat{\boldsymbol{u}}_i = \overline{\boldsymbol{u}}_i, \quad i \in I^{\mathrm{s}}. \tag{26}$$

For better readability, although matrix $\boldsymbol{L}_{\mathrm{b}}$ does depend on the shape and size of the cell, subscript i is dropped in its notation. Also, subscript 'b' is here to remind that $\boldsymbol{L}_{\mathrm{b}}$ is invertible and defines a bijection. Note that this is the arbitrary choice of the solution representation. The same bijective transformation could be made to obtain a nodal representation as in the SD methodology.

In the same way as in Sect. 2, r is the reduction order. Let $\mathscr{B}^{p-r}$ be a basis of $\mathbb{P}_{p-r}[X]$. For any reduction order $r > 0$, matrix $\boldsymbol{L}$ is defined similarly to $\boldsymbol{L}_{\mathrm{b}}$ as:

$$\boldsymbol{L} : \begin{cases} \mathbb{R}^{N^{p-r}} & \longrightarrow & \mathbb{R}^{N^{\mathrm{c}}} \\ \hat{\boldsymbol{u}} & \longmapsto & \overline{\boldsymbol{u}} \end{cases}, \tag{27}$$

with:

$$(\boldsymbol{L})_{jk} = \frac{1}{V_{i,j}} \int_{C_{i,j}} \phi_k^{p-r} \, \mathrm{d}V, \quad j \in I^{\mathrm{c}}, \ k \in [\![1, N^{p-r}]\!], \tag{28}$$

We would like to emphasize that here, the coefficients in $\hat{\boldsymbol{u}}$ represents a polynomial of $\mathbb{P}_{p-r}[X]$ and no longer $\mathbb{P}_p[X]$. Since there are too many DoFs to be represented by a polynomial of degree lesser than p, the inverse of $\boldsymbol{L}$ is not defined. Indeed, for a distribution of integral values not in the image space of $\boldsymbol{L}$, there exists no corresponding polynomial representation in $\mathbb{P}_{p-r}[X]$. A standard choice for a projector uses the pseudo-inverse $\boldsymbol{L}^+$ of matrix $\boldsymbol{L}$ defined with:

$$\boldsymbol{L}^+ = \left(\boldsymbol{L}^T \boldsymbol{L}\right)^{-1} \boldsymbol{L}^T, \tag{29}$$

which corresponds to the use of the L^2 projection $\mathscr{P}$ in Sect. 2. The associated projector matrix, which is the algebraic version of $\mathscr{P}$ locally to a cell, is:

$$\boldsymbol{\Pi} = \boldsymbol{L}\boldsymbol{L}^+. \tag{30}$$

We now express the constraint on the mass of the elements of the kernel of $\boldsymbol{\Pi}$. In our finite volume formalism, the unknowns in vector $\overline{\boldsymbol{u}}_i$ defined in Eq. (23) satisfy:

$$\int_{S_i} \hat{u}_h \, \mathrm{d}V = \sum_{j \in I^{\mathrm{c}}} V_{i,j} \overline{u}_{i,j} = \boldsymbol{V} \cdot \overline{\boldsymbol{u}}_i, \quad i \in I^{\mathrm{s}}, \tag{31}$$

where $\boldsymbol{V} = \left(V_{i,1}, \ldots, V_{i,N^{\mathrm{c}}}\right)^T$. Subscript i has also been removed from the notation used for the CV's volume vector $\boldsymbol{V}$. For each element $\overline{\boldsymbol{u}}^*$ in $\mathbb{R}^{N^{\mathrm{c}}}$, a conservative affine space $\pi_{\mathrm{c}}\left(\overline{\boldsymbol{u}}^*\right)$ is defined with:

$$\pi_{\mathrm{c}}\left(\overline{\boldsymbol{u}}^*\right) : \left\{ \overline{\boldsymbol{u}} \in \mathbb{R}^{N^{\mathrm{c}}} \mid \boldsymbol{V} \cdot \left(\overline{\boldsymbol{u}} - \overline{\boldsymbol{u}}^*\right) = 0 \right\}. \tag{32}$$

These conservative affine spaces are the hyperplanes of $\mathbb{R}^{N^{\mathrm{c}}}$ of same mass elements. The nullity constraint on the mass of the elements of $\ker\left(\boldsymbol{\Pi}\right)$ which have to be removed can then be written as:

$$\ker\left(\boldsymbol{\Pi}\right) \subset \pi_{\mathrm{c}}\left(\boldsymbol{0}\right). \tag{33}$$

It should be noted that at this point, nothing ensures identity (33) to hold a priori. One can note that for the standard non-reduced case with $r = 0$ and $\boldsymbol{L}_{\mathrm{b}}$ invertible, condition (33) is satisfied since:

$$\ker\left(\boldsymbol{\Pi}\right) = \ker\left(\boldsymbol{L}_{\mathrm{b}} \boldsymbol{L}_{\mathrm{b}}^{-1}\right) = \{\boldsymbol{0}\} \subset \pi_{\mathrm{c}}\left(\boldsymbol{0}\right). \tag{34}$$

The problems encountered with the reduced scheme are entirely caused by the fact that $\ker\left(\boldsymbol{\Pi}\right) \neq \{\boldsymbol{0}\}$. The same conservative constraint could have been written for the continuous representation or the nodal representation.

Matrix $\boldsymbol{L}$ is completely defined with the choice of a polynomial basis and a cell partition and so is $\boldsymbol{L}^+$. To have control over the mass of the elements of $\ker\left(\boldsymbol{\Pi}\right)$, additional parameters are needed in the inversion process of $\boldsymbol{L}$. There will be obtained with the use of the generalized pseudo-inverse. Let $\boldsymbol{M} \in \mathbb{R}^{N^{\mathrm{c}} \times N^{\mathrm{c}}}$ be a symmetric definite positive matrix. Its associated norm $\|\cdot\|_{\boldsymbol{M}}$ is defined with:

$$\forall \mathbf{x} \in \mathbb{R}^{N^{\mathrm{c}}}, \quad \|\mathbf{x}\|_{\boldsymbol{M}}^2 = \mathbf{x}^T \boldsymbol{M} \mathbf{x}. \tag{35}$$

The generalized pseudo-inverse of $\boldsymbol{L}$ is given by:

$$\boldsymbol{L}_{\boldsymbol{M}}^{+} = \left(\boldsymbol{L}^T \boldsymbol{M} \boldsymbol{L}\right)^{-1} \boldsymbol{L}^T \boldsymbol{M}. \tag{36}$$

Clearly, $\boldsymbol{L}^+$ is obtained with $\boldsymbol{M} = \boldsymbol{I}$, the identity matrix.

The generalized pseudo-inverse $\boldsymbol{L}_{\boldsymbol{M}}^{+}$ gives the solution of the following convex minimization problem. For all $\overline{\boldsymbol{u}}^* \in \mathbb{R}^{N^{\mathrm{c}}}$:

$$\underset{\hat{\boldsymbol{u}} \in \mathbb{R}^{N^p - r}}{\operatorname{argmin}} \left\| \overline{\boldsymbol{u}}^* - \boldsymbol{L} \hat{\boldsymbol{u}} \right\|_{\boldsymbol{M}} = \boldsymbol{L}_{\boldsymbol{M}}^{+} \overline{\boldsymbol{u}}^*. \tag{37}$$

This problem can be reformulated as:

$$\underset{\overline{\boldsymbol{u}} \in \mathrm{im}(\boldsymbol{L})}{\operatorname{argmin}} \left\| \overline{\boldsymbol{u}}^* - \overline{\boldsymbol{u}} \right\|_{\boldsymbol{M}} = \boldsymbol{L} \boldsymbol{L}_{\boldsymbol{M}}^{+} \overline{\boldsymbol{u}}^*, \tag{38}$$

with $\mathrm{im}\left(\cdot\right)$ designating the image space. Equation (38) proves that:

$$\boldsymbol{\Pi}_{\boldsymbol{M}} = \boldsymbol{L} \boldsymbol{L}_{\boldsymbol{M}}^{+}, \tag{39}$$

is the matrix of the projector from $\mathbb{R}^{N^c}$ to im $(\boldsymbol{L})$ according to the norm $\|\cdot\|_{\boldsymbol{M}}$. The projection vector $\boldsymbol{p}_{\boldsymbol{M}}(\overline{\boldsymbol{u}})$ for element $\overline{\boldsymbol{u}}$, normalized for the euclidean norm $\|\cdot\|_{\boldsymbol{I}}$, is given by:

$$\boldsymbol{p}_{\boldsymbol{M}}(\overline{\boldsymbol{u}}) = \frac{\boldsymbol{\Pi}_{\boldsymbol{M}}\overline{\boldsymbol{u}} - \overline{\boldsymbol{u}}}{\left\|\boldsymbol{\Pi}_{\boldsymbol{M}}\overline{\boldsymbol{u}} - \overline{\boldsymbol{u}}\right\|_{\boldsymbol{I}}}. \tag{40}$$

Like that of the standard pseudo-inverse $\boldsymbol{L}^{+}$, the kernel of $\boldsymbol{L}_{\boldsymbol{M}}^{+}$ is different from the zero vector space $\{\boldsymbol{0}\}$ and $\boldsymbol{\Pi}_{\boldsymbol{M}}$ will be used to remove the spurious oscillations due to the components in the kernel of $\boldsymbol{\Pi}_{\boldsymbol{M}}$.

As defined in Eq. (36), the choice for $\boldsymbol{M}$ in the generalized pseudo-inverse definition gives a full range of different matrices $\boldsymbol{L}_{\boldsymbol{M}}^{+}$ to invert Eq. (26) with $\boldsymbol{L}_{\mathrm{b}}$ replaced by $\boldsymbol{L}$, each producing a different representation and having different functions in their kernel. We will now focus on relation (33) and obtain explicit numerical constraints on $\boldsymbol{M}$ that preserve the conservativity of the scheme during the projection step.

3.2 Constraints for Conservative Schemes

As previously stated, in order to conservatively project the vector of unknowns and prevent some of its components to pollute the computations, it is required that:

$$\ker\left(\boldsymbol{\Pi}_{\boldsymbol{M}}\right) \subset \pi_{\mathrm{c}}(\boldsymbol{0}), \tag{41}$$

or in an equivalent manner:

$$\ker\left(\boldsymbol{L}^{T}\boldsymbol{M}\right) \subset \pi_{\mathrm{c}}(\boldsymbol{0}). \tag{42}$$

Because matrix $\boldsymbol{M}$ is invertible, we have:

$$\ker\left(\boldsymbol{L}^{T}\boldsymbol{M}\right) = \left\{\boldsymbol{M}^{-1}\overline{\boldsymbol{u}} \mid \overline{\boldsymbol{u}} \in \ker\left(\boldsymbol{L}^{T}\right)\right\}. \tag{43}$$

It is clear that $\ker\left(\boldsymbol{L}^{T}\right) = \text{im}\,(\boldsymbol{L})^{\perp}$ the orthogonal complement of im $(\boldsymbol{L})$ in $\mathbb{R}^{N^c}$. Then, the condition for conservativity can be reformulated as:

$$\left\{\boldsymbol{M}^{-1}\overline{\boldsymbol{u}} \mid \overline{\boldsymbol{u}} \in \text{im}\,(\boldsymbol{L})^{\perp}\right\} \subset \pi_{\mathrm{c}}(0). \tag{44}$$

The conservativity condition of identity (44) can be expressed, since $\boldsymbol{M}$ is symmetric, as:

$$\forall \overline{\boldsymbol{u}} \in \text{im}\,(\boldsymbol{L})^{\perp}, \quad \boldsymbol{V}\cdot\boldsymbol{M}^{-1}\overline{\boldsymbol{u}} = \boldsymbol{M}^{-1}\boldsymbol{V}\cdot\overline{\boldsymbol{u}} = 0, \tag{45}$$

which exactly express the fact that:

$$\boldsymbol{M}^{-1}\boldsymbol{V} \in \text{im}\,(\boldsymbol{L}). \tag{46}$$

Any projector onto the image of $\boldsymbol{L}$, like $\boldsymbol{\Pi}_{\boldsymbol{I}}$ for example, can then be used to obtain an explicit formulation of the conservativity constraints on matrix $\boldsymbol{M}$:

$$\left(\boldsymbol{\Pi}_{\boldsymbol{I}} - \boldsymbol{I}\right)\boldsymbol{M}^{-1}\boldsymbol{V} = \boldsymbol{0}. \tag{47}$$

This formulation represents a notable progress compared to formulation (42) since the dependence on $\boldsymbol{M}$ has been transferred from the domain of validity of the equation to the equation itself and thus simplifies its resolution.

The simple case of a diagonal matrix $\boldsymbol{M}$ is now investigated:

$$\boldsymbol{M} = \begin{pmatrix} m_1 & & 0 \\ & \ddots & \\ 0 & & m_{N^c} \end{pmatrix}. \tag{48}$$

It is obvious that, for any reduction order $r > 0$ considered, the constant polynomial:

$$\phi : \boldsymbol{x} \in \mathbb{R}^{N^c} \longmapsto 1, \tag{49}$$

is an element of $\mathbb{P}_{p-r}[X]$. As a consequence, vector $\overline{\boldsymbol{u}} = (1, \ldots, 1)^T$, which is the representation of ϕ over the CVs of a cell, satisfies:

$$\forall r \in [\![1, p]\!], \quad \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathrm{im}\,(\boldsymbol{L})\,. \tag{50}$$

The previous relation is the expression of Eq. (46) with matrix $\boldsymbol{M_V}$ as the weighting matrix, where $\boldsymbol{M_V}$ is defined by the sub-volumes of the partition:

$$\boldsymbol{M_V} = \begin{pmatrix} V_1 & & 0 \\ & \ddots & \\ 0 & & V_{N^c} \end{pmatrix}. \tag{51}$$

This demonstrates the existence of at least one matrix $\boldsymbol{M}$ solution of (47), $\boldsymbol{M_V}$, which is a universal parameter matrix for the definition of schemes with a conservative associated projection. This existence can be ensured independently from the dimension of the problem or from the basis reduction order.

3.3 Numerical Validation

We now validate the previous developments with Burgers' equation and the computation of a moving discontinuity since it is very sensitive to flux computation and conservation properties. The following problem is considered:

$$\frac{\partial u}{\partial t} + \frac{\partial u^2}{\partial x} = 0, \quad x \in [0, 1]\,, \tag{52}$$

with a Heaviside function as an initial condition u^0:

$$u^0(x) = \begin{cases} H_0 & \text{if} \quad x < x_0 \\ 0 & \text{if} \quad x \geq x_0 \end{cases}. \tag{53}$$

Like in Sect. 2.2, the spectral volume scheme of order 3 reduced to order 2 is implemented. The mesh is composed of 30 identical cells partitioned with Gauss-Legendre points. Considering two adjacent states u_L and u_R, the solution to the Riemann problem is given by:

$$\begin{aligned} \text{if } u_L \geq u_R : & \begin{cases} \dfrac{u_R^2}{2} & \text{if} \quad \dfrac{u_L + u_R}{2} < 0 \\ \dfrac{u_L^2}{2} & \text{if} \quad \dfrac{u_L + u_R}{2} \geq 0 \end{cases}, \\ \text{if } u_L < u_R : & \begin{cases} \dfrac{u_R^2}{2} & \text{if} \quad u_R < 0 \\ 0 & \text{if} \quad u_L u_R < 0\,. \\ \dfrac{u_L^2}{2} & \text{if} \quad u_L > 0 \end{cases} \end{aligned} \tag{54}$$

For our computation, the parameter were chosen as $H_0 = 1.6$, $x_0 = 0.1$ and $T = 1$. Time discretization is performed using the standard RK3 SSP scheme [38] with a time step of

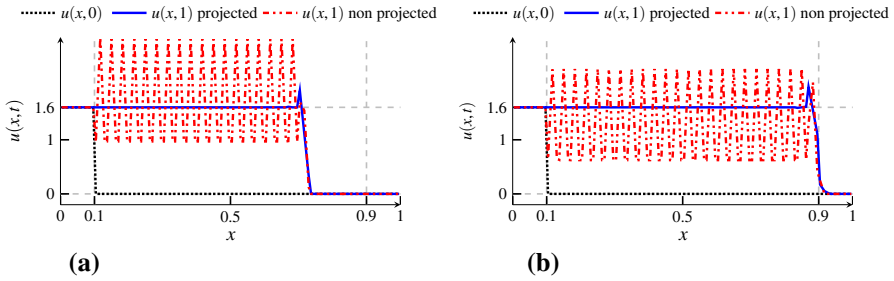

Fig. 2 Discontinuity displacement through the Burgers' equation. *Grey dots* indicates the theoretical final states. In case (**b**) matrix $\boldsymbol{M}$ satisfies the conservativity constraints, unlike in case (**a**). The projection operation is performed using $\boldsymbol{\Pi}_{\boldsymbol{M}}$. **a** $\boldsymbol{M} = \boldsymbol{I}$, non-conservative projection. **b** $\boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{V}}$, conservative projection

$\Delta t = 10^{-2}$ (based on the cells size, the initial CFL is 4.8×10^{-1}). The Rankine-Hugoniot relation on the discontinuity velocity indicates that at instant $t = T$ the discontinuity should be located at $x_s = x_0 + \frac{H_0}{2} T = 0.9$. Boundary conditions $u(0, t) = H_0$ and $u(1, t) = 0$ are imposed and are consistent with the initial condition and total simulation time.

Results are presented in Fig. 2. Both the effect of the use of the projected solution in the time-marching scheme and the effect of different choices for $\boldsymbol{M}$ are observed. Figure 2a presents the computation of the discontinuity evolution using matrix $\boldsymbol{M} = \boldsymbol{I}$. The two different effects are distinguishable:

- when evaluated, the total mass of the solution is correct and matches the expected value of $x_s \times H_0 = 1.44$ (the SVM is conservative) although the discontinuity position is incorrect. This is due to the erroneous evaluation of the flux values: the scheme does not see the different elements with their real mass since the mass of the kernel components, which is not null, are neglected (see Eq. 14). Consequently, the discontinuity is convected with the wrong speed.
- the projection of the solution between each step of the time-marching scheme does suppress the spurious oscillations, providing a smooth solution. But as removed components are non-zero mass elements, the projection is non-conservative and results in this case in a loss of mass.

In Fig. 2b, the computation is made using matrix $\boldsymbol{M} = \boldsymbol{M}_{\boldsymbol{V}}$. There, the position of the discontinuity is accurate, thanks to correct fluxes evaluation. And as the kernel is controlled to have zero-mass elements, the projection is a conservative operation. The use of the projected solution in the time-marching scheme does not change the mass of the solution as confirms numerical evaluation.

We have extended the spectral volume methodology with a projection step that allows us to change the polynomial degree of the interpolation. To ensure the conservativity of the projection, the considered reduction order is linked to a set of constraints for the matrix $\boldsymbol{M}$ that defines the projection. Through the new DoFs contained in $\boldsymbol{M}$, we have access to many acceptable conservative projections of a given element. We will now present a possible use of the matrix $\boldsymbol{M}$ that raises an interest in the definition of a compact limiting procedure.

3.4 Weighted Projections

The reduction of the polynomial degree has two objectives: to reduce the oscillations and to limit the boundary overshoots which appear when trying to interpolate discontinuities.

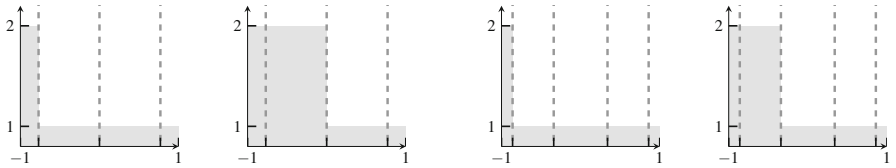

Fig. 3 Discontinuity patterns for fourth and fifth order SVM scheme. **a** Left pattern $\overline{\boldsymbol{u}}^{\mathrm{l}}$ (*left*) and centered pattern $\overline{\boldsymbol{u}}^{\mathrm{c}}$ (*right*) for fourth order scheme. **b** Left pattern $\overline{\boldsymbol{u}}^{\mathrm{l}}$ (*left*) and left-centered pattern $\overline{\boldsymbol{u}}^{\mathrm{lc}}$ (*right*) for fifth order scheme.

The use of the generalized pseudo-inverse and the presence of parameter matrix $\boldsymbol{M}$ in its definition give a full control on the projection and consequently on the projected polynomial.

Considering a reduction of order r of the polynomial degree, the conservativity constraints on $\boldsymbol{M}$ consist in a system of N^{eq} equations with:

$$N^{\mathrm{eq}} = \dim\left(\mathrm{im}\,(\boldsymbol{L})^{\perp}\right) = N^{\mathrm{c}} - N^{p-r}, \tag{55}$$

and leaves $N^{\mathrm{c}} - N^{eq} = N^{p-r}$ components of matrix $\boldsymbol{M}$ as available DoFs.

Only the case of diagonal matrices (48) will now be considered. Matrix $\boldsymbol{M}$ can be seen as a weighting matrix. The coefficients on its diagonal represent the weights of the CVs in the projection. Indeed, problem (38) reformulates as:

$$\forall \overline{\boldsymbol{u}}^{*} \in \mathbb{R}^{N^{\mathrm{c}}}, \quad \underset{\overline{\boldsymbol{u}} \in \mathrm{im}(\boldsymbol{L})}{\mathrm{argmin}} \sum_{j \in N^{\mathrm{c}}} m_j \left(\overline{u}^{*}_{,j} - \overline{u}_{,j}\right)^2 = \boldsymbol{L}\boldsymbol{L}^{+}_{\boldsymbol{M}}\overline{\boldsymbol{u}}^{*}. \tag{56}$$

The heavier the weight of the j-th CV is, the closer $\overline{u}^{*}_{,j}$ and $\overline{u}_{,j}$ will be. It is thus possible to influence the shape of the projection with a well-chosen matrix $\boldsymbol{M}$.

More precisely, it is possible to search for the weighting matrices that limit the overshoots and the oscillations when interpolating a discontinuity. For a given dimension d and a given number of DoFs per cell N^{c}, we can identify a finite number of problematic patterns that represent discontinuities. For example in the 1D case, Fig. 3 shows 2 possible idealized volumic representations of a discontinuity for SVM schemes of different orders. For the fourth order scheme ($p = 3$, $N^{\mathrm{c}} = 4$), we have a left and a centered pattern referred to as $\overline{\boldsymbol{u}}^{\mathrm{l}} = (2, 1, 1, 1)^T$ and $\overline{\boldsymbol{u}}^{\mathrm{c}} = (2, 2, 1, 1)^T$ (see Fig. 3a). For the fifth order scheme ($p = 4$, $N^{\mathrm{c}} = 5$), we have a left and a left-centered pattern, referred to as $\overline{\boldsymbol{u}}^{\mathrm{l}} = (2, 1, 1, 1, 1)^T$ and $\overline{\boldsymbol{u}}^{\mathrm{lc}} = (2, 2, 1, 1, 1)^T$ (see Fig. 3b). The polynomials $\hat{u}^{\mathrm{l}}$, $\hat{u}^{\mathrm{c}}$ and $\hat{u}^{\mathrm{lc}}$ associated to these representations present the different flaws (oscillations and overshoots) already mentioned and are very likely to pollute computations. Obviously, right ($\overline{\boldsymbol{u}}^{\mathrm{r}}$, $\hat{u}^{\mathrm{r}}$) and right-centered ($\overline{\boldsymbol{u}}^{\mathrm{rc}}$, $\hat{u}^{\mathrm{rc}}$) representations and polynomials exist but have not been displayed for symmetry reasons.

We can associate each pattern with a well-chosen weighting matrix. Its role is to provide an acceptable projection of the polynomial that represents the problematic pattern, using the projector associated by relation (39). Different criteria could be taken into account like the total variation, the L^2 norm of the function or its derivative…In our pursuit of a new limiting procedure, we have already mentioned that we were interested in reducing the size of overshoots at both borders of the cell and the amplitude of oscillations. Taking care not to violate the conservativity constraints (46) and considering a first order reduction ($r = 1$), we

have identified the following left and centered weighting matrices for the patterns of Fig. 3a:

$$\boldsymbol{M}^{\mathrm{l}} \approx \begin{pmatrix} 0.35 & 0 & 0 & 0 \\ 0 & 2.30 & 0 & 0 \\ 0 & 0 & 11.8 & 0 \\ 0 & 0 & 0 & 10.0 \end{pmatrix}, \quad \boldsymbol{M}^{\mathrm{c}} \approx \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 0 & 2.00 & 0 & 0 \\ 0 & 0 & 2.00 & 0 \\ 0 & 0 & 0 & 1.00 \end{pmatrix}, \tag{57}$$

and for the patterns of Fig. 3b, we selected the following left and left-centered weighting matrices:

$$\boldsymbol{M}^{\mathrm{l}} \approx \begin{pmatrix} 0.12 & 0 & 0 & 0 & 0 \\ 0 & 0.90 & 0 & 0 & 0 \\ 0 & 0 & 5.00 & 0 & 0 \\ 0 & 0 & 0 & 7.18 & 0 \\ 0 & 0 & 0 & 0 & 5.00 \end{pmatrix}, \quad \boldsymbol{M}^{\mathrm{lc}} \approx \begin{pmatrix} 5.00 & 0 & 0 & 0 & 0 \\ 0 & 1.00 & 0 & 0 & 0 \\ 0 & 0 & 1.50 & 0 & 0 \\ 0 & 0 & 0 & 15.1 & 0 \\ 0 & 0 & 0 & 0 & 4.00 \end{pmatrix}. \tag{58}$$

To strictly satisfy the conservativity constraints, the coefficients in the different weighting matrices should be given with a significant number of digits. To improve readability, approximated values have been given here but exact values are available in Appendix 1. Matrices $\boldsymbol{M}^{\mathrm{r}}$ and $\boldsymbol{M}^{\mathrm{rc}}$ associated to the symmetric patterns are respectively defined by reversing the main diagonal of $\boldsymbol{M}^{\mathrm{l}}$ and $\boldsymbol{M}^{\mathrm{lc}}$. The pseudo-inverse associated to these different matrices are naturally denoted with $\boldsymbol{L}_{\mathrm{c}}^{+}$, $\boldsymbol{L}_{\mathrm{l}}^{+}$,$\boldsymbol{L}_{\mathrm{r}}^{+}$, $\boldsymbol{L}_{\mathrm{lc}}^{+}$ and $\boldsymbol{L}_{\mathrm{rc}}^{+}$.

Figure 4 gathers the projected polynomials obtained with the projectors using the weighting matrices of relations (57) and (58). For each pattern, we can see that the associated projectors provide almost non-oscillating projected polynomial with very limited overshoots (for each pattern, the associated projection is designated by an arrow). It also occurs that the different projectors are not performing well with the projection of a pattern that is different from the one they are associated with.

It should be noted that the projection described here has an even better level of compactness than the most popular high-order methods such as the discontinuous Galerkin, the spectral differences, or the correction procedure by reconstruction: no information from neighbouring cells or from the boundaries are needed to perform the projection. The next section will now describe how these newly-defined projectors will be used in the definition of a compact limiter.

4 The Spatially Weighted Projection (SWeP) limiter

We expose in this section a compact limiter for high-order methods based on the spatially weighted projections described in Sect. 3.4 and thus will be referred to as the *SWeP* limiting procedure. Our approach is inspired by previous studies concerning the development of the WENO procedure for the discontinuous Galerkin [29,33] and the CPR [30] methodologies. As it has already been mentioned in Sect. 2, the common strategy adopted consists in:

- detecting the troubled cells,
- changing the polynomial representation of the solution.

This general limitation procedure only rests upon the available polynomial interpolation of the data, making the involved limiter naturally compatible with any compact high-order framework, as it has been proven with the multiple implementation of the already mentioned WENO framework [29,30]. In our example, the SWeP limiter will be combined with the spectral differences scheme and the minmod detector for which brief reviews are proposed.

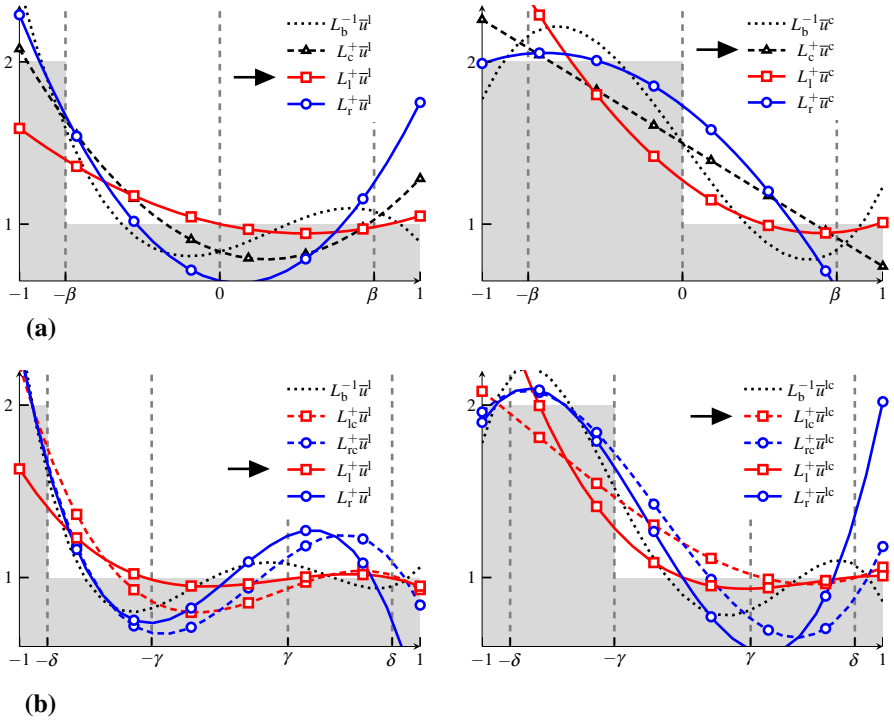

Fig. 4 Results of the different pseudo-inverse defined for a first order reduction of the fourth (**a**) and the fifth (**b**) order SVM scheme. The *arrows* designate the projection associated to the discontinuity pattern. **a** Different projections of the patterns $\overline{\boldsymbol{u}}^{\mathrm{l}}$ and $\overline{\boldsymbol{u}}^{\mathrm{c}}$ (*shaded area*) of the fourth order SVM scheme ($\beta \approx 0.77$). **b** Different projections of the patterns $\overline{\boldsymbol{u}}^{\mathrm{l}}$ and $\overline{\boldsymbol{u}}^{\mathrm{lc}}$ (*shaded area*) of the fifth order SVM scheme ($\gamma \approx 0.34$, $\delta \approx 0.86$)

4.1 Review of the SD Method for the One Dimensional Case

The reader is advised that some notations used in Sect. 3 will be found again in this section. The one-dimensional version of (1) is considered:

$$\partial_t u + \partial_x F(u) = 0, \tag{59}$$

where u represents the scalar or vectorial form of the conservative variables. The computational domain $[a, b]$ is partitioned into N^{s} cells S_i:

$$[a,b] = \bigcup_{i\in I^{\mathrm{s}}} S_i = \bigcup_{i\in I^{\mathrm{s}}} \left[x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}\right], \quad x_{\frac{1}{2}} = a, \quad x_{N^{\mathrm{s}}+\frac{1}{2}} = b. \tag{60}$$

For the sake of simplicity, we will work with a regular meshing of the domain, considering that all cells have the same size Δx.

The spectral differences method of order $p+1$ uses two different sets of points inside each element. First, the set of solution points, defined with the mapping of the Gauss points of order $p+1$ in each cell S_i:

$$X_i^{\mathrm{s}} = \left\{ x_{i,j}^{\mathrm{s}} = x_{i-\frac{1}{2}} + \frac{\Delta x}{2}\left[1 - \cos\left(\frac{2j-1}{2(p+1)}\pi\right)\right], 1 \le j \le p+1 \right\}, \quad i \in I^{\mathrm{s}} \tag{61}$$

and secondly, the set of $p+2$ flux points X_i^{f}, defined with the mapping of the p Gauss-Legendre points of order p to the element S_i with additional end points $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$.

In the spectral differences framework, the unknowns are the conservative variables at the solution points denoted by:

$$u_{i,j}^{\mathrm{s}} = u\left(x_{i,j}^{\mathrm{s}}\right) \tag{62}$$

Then a polynomial interpolation $\hat{u}_i$ of the solution is defined in S_i using the family of $p+1$ Lagrange polynomials $\phi_k^{\mathrm{L,s}}$ associated with the set of solution points X_i^{s}:

$$\forall x \in S_i, \quad \hat{u}_i(x) = \sum_{k=1}^{p+1} u_{i,k}^{\mathrm{s}} \phi_k^{\mathrm{L,s}}(x) \tag{63}$$

The flux is then approximated with:

$$\forall x \in S_i, \quad F(x) \approx \sum_{k=1}^{p+2} F_{i,k}^{\mathrm{f}} \phi_k^{\mathrm{L,f}}(x) \tag{64}$$

where $\phi_k^{\mathrm{L,f}}$ are the Lagrange polynomials associated to the set of flux points X_i^{f} and $F_{i,j}^{\mathrm{f}}$ are the values of the flux constructed at the flux points from the conservative variables given by Eq. (63). As the flux is only element-wise continuous, a Riemann solver is needed to compute the flux values at the boundary of each element. The solution is then updated using:

$$\frac{\mathrm{d}u_{i,j}^{s}}{\mathrm{d}t} = -\left(\partial_x F\right)_{i,j} \tag{65}$$

where the flux spatial derivative at the solution points can be computed thanks to (64). For more information about the SD methodology or its generalization to higher dimensions, the reader is referred to [3].

4.2 The Minmod Troubled Cell Indicator

To evaluate the SWeP limiting procedure, different numerical examples exposed in [29,33] will be computed. In order to allow comparison, the detector used in the mentioned studies is going to be set up: the TVB-modified minmod detector. A brief review of this detector for the one-dimensional scalar case is proposed below. For a more complete investigation, details can be found in [22].

The average $\breve{u}_i$ of the scalar polynomial solution $\hat{u}_i$ over cell S_i is defined with:

$$\breve{u}_i = \frac{1}{\Delta x} \int_{S_i} \hat{u}_i(x)\, \mathrm{d}x, \quad i \in I^{\mathrm{s}}. \tag{66}$$

We further denote:

$$\tilde{u}_i = \hat{u}_i\left(x_{i+\frac{1}{2}}\right) - \breve{u}_i, \quad \tilde{\tilde{u}}_i = \breve{u}_i - \hat{u}_i\left(x_{i-\frac{1}{2}}\right), \tag{67}$$

the difference to the average value at the cell border and:

$$\Delta_+ \breve{u}_i = \breve{u}_{i+1} - \breve{u}_i, \quad \Delta_- \breve{u}_i = \breve{u}_i - \breve{u}_{i-1}, \tag{68}$$

the gap between the average values of two successive cells. The TVB-modified minmod function $\tilde{m}$ is defined as:

$$\tilde{m}(a_1, \ldots, a_l) = \begin{cases} a_1 & \text{if } |a_1| \leq M\Delta x^2, \\ m(a_1, \ldots, a_l) & \text{otherwise,} \end{cases} \tag{69}$$

with:

$$m\left(a_1,\ldots,a_l\right)=\begin{cases} s\,\min\limits_{1\le k\le l}|a_k| & \text{if } s=\operatorname{sign}\left(a_1\right)=\ldots=\operatorname{sign}\left(a_l\right),\\ 0 & \text{otherwise,}\end{cases} \tag{70}$$

where M is the TVB parameter that depends on the solution of the problem. We compute:

$$\tilde{u}_i^{(\mathrm{mod})}=\tilde{m}\left(\tilde{u}_i,\Delta_+\breve{u}_i,\Delta_-\breve{u}_i\right)\quad \tilde{\tilde{u}}_i^{(\mathrm{mod})}=\tilde{m}\left(\tilde{\tilde{u}}_i,\Delta_+\breve{u}_i,\Delta_-\breve{u}_i\right) \tag{71}$$

At this point, a cell is **not** going to be declared as a troubled cell if and only if both following equalities are true:

$$\tilde{u}_i^{(\mathrm{mod})}=\tilde{u}_i,\quad \tilde{\tilde{u}}_i^{(\mathrm{mod})}=\tilde{\tilde{u}}_i \tag{72}$$

The extension to the vectorial case consists in applying this scalar procedure to each vectorial component. The important role played by the TVB constant M can already be noted. Depending on its value, too many or too few troubled cells will be identified, directly affecting the computational cost of the method. Yet, the constant M also gives a convenient way to drive the behaviour of the detector and thus M can be used to provide substantial information about the performance of the limiting procedure in an over- or under-limited simulation context.

4.3 The Compact SWeP Algorithm for One-Dimensional Systems

This subsection gives the details of the SWeP limiting procedure and its integration in the SD framework for the one-dimensional scalar and system case.

4.3.1 Scalar Problem

We assume that S_i is identified as a troubled cell. First, we switch from the point-wise representation of the SD method to the integral representation of the SVM framework. In accordance with the notation used in Sect. 2, vector $\boldsymbol{u}_i^{\mathrm{s}}$ is defined as:

$$\boldsymbol{u}_i^{\mathrm{s}}=\begin{pmatrix} u_{i,1}^{\mathrm{s}}\\ \vdots\\ u_{i,p+1}^{\mathrm{s}}\end{pmatrix} \tag{73}$$

and matrix $\boldsymbol{P}_{jk}^{\mathrm{s}}$ is defined using the basis $\mathscr{B}^p=\left\{\phi_k^p\mid 1\le k\le p+1\right\}$ of $\mathbb{P}_p\left[X\right]$ as:

$$\boldsymbol{P}_{jk}^{\mathrm{s}}=\phi_k^p\left(x_j^{\mathrm{s}}\right),\quad 1\le j,k\le p+1. \tag{74}$$

The integral representation of the polynomial solution is given by:

$$\overline{\boldsymbol{u}}_i=\boldsymbol{L}_{\mathrm{b}}\left[\boldsymbol{P}^{\mathrm{s}}\right]^{-1}\boldsymbol{u}_i^{\mathrm{s}} \tag{75}$$

where $\boldsymbol{L}_{\mathrm{b}}$ is associated to basis $\mathscr{B}^p$ by relation (25).

For a given polynomial degree p and a given reduction order r, we explained in Sect. 3.4 that a finite number of problematic patterns could be identified and associated to a set of weighting matrices. This set of matrices is now referred to as $\mathscr{M}^{p,r}$. The idea of the SWeP limiting procedure is to choose the most acceptable projection of the solution among those

given by the projectors associated to the different matrices in $\mathscr{M}^{p,r}$. For a first order reduction ($r = 1$), the set $\mathscr{M}^{3,1}$ associated to a fourth order scheme is composed of the 3 matrices:

$$\mathscr{M}^{3,1} = \left\{ \boldsymbol{M}^{\mathrm{c}}, \boldsymbol{M}^{\mathrm{l}}, \boldsymbol{M}^{\mathrm{r}} \right\}, \tag{76}$$

and for a fifth order scheme, $\mathscr{M}^{4,1}$ is composed of the 4 matrices:

$$\mathscr{M}^{4,1} = \left\{ \boldsymbol{M}^{\mathrm{lc}}, \boldsymbol{M}^{\mathrm{rc}}, \boldsymbol{M}^{\mathrm{l}}, \boldsymbol{M}^{\mathrm{r}} \right\}. \tag{77}$$

These matrices were described in Sect. 3.4. For better readability, subscript p was omitted in the notation of the weighting matrices. The reader must keep in mind that the size of each matrix in $\mathscr{M}^{3,1}$ or $\mathscr{M}^{4,1}$ depends on the order of the scheme.

Let $\boldsymbol{L}$ be the matrix associated by relation (27) to the reduced basis $\mathscr{B}^{p-r}$. Thanks to the developments presented in Sect. 3 the set of pseudo-inverses $\mathscr{I}^{p,r}$ is defined with:

$$\mathscr{I}^{p,r} = \left\{ \boldsymbol{L}^{+}_{\boldsymbol{M}} = \left(\boldsymbol{L}^T \boldsymbol{M} \boldsymbol{L} \right)^{-1} \boldsymbol{L}^T \boldsymbol{M} \mid \boldsymbol{M} \in \mathscr{M}^{p,r} \right\} \cup \left\{ \boldsymbol{L}_{\mathrm{b}}^{-1} \right\}. \tag{78}$$

To identify the pseudo-inverse that provides the most acceptable projection, we evaluate the quadratic integral of the distance to the average value. This integral works as a selective criteria. The pseudo-inverse in $\mathscr{I}^{p,r}$ that minimize this integral is labelled as optimal and provide the sought polynomial reconstruction:

$$\boldsymbol{L}^{+}_{\mathrm{opt}} = \underset{\boldsymbol{L}^{+} \in \mathscr{I}^{p,r}}{\operatorname{argmin}} \left(\int_{S_i} \left[\widehat{\boldsymbol{L}^{+} \overline{\boldsymbol{u}}_i}\,(x) - \breve{u}_i \right]^2 \mathrm{d}x \right). \tag{79}$$

We added matrix $\boldsymbol{L}_{\mathrm{b}}^{-1}$ to $\mathscr{I}^{p,r}$ because for smooth solutions, the original polynomial representation is a very good candidate. Then, the new point-wise representation of the solution in cell S_i is defined with:

$$\boldsymbol{u}_i^{\mathrm{s,new}} = \boldsymbol{P}^{\mathrm{s}} \boldsymbol{L}^{+}_{\mathrm{opt}} \overline{\boldsymbol{u}}_i. \tag{80}$$

We can now compute the numerical flux F^{new} using the new solution vector $\boldsymbol{u}_i^{\mathrm{s,new}}$. Following remarks of Sect. 2, a time marching scheme is then used to get solution at time level $n + 1$. For example with the forward Euler scheme:

$$\frac{u^s_{i,j}\left(t^{n+1}\right) - u^{\mathrm{s,new}}_{i,j}}{\Delta t} = -\left(\partial_x F^{\mathrm{new}}\right)_{i,j} \tag{81}$$

From a computational point of view, all matrices in the set $\mathscr{I}^{p,r}$ can be pre-computed and the different integrals in (79) are evaluated as a weighted inner-product, reducing the additional cost associated with the use of the SWeP limiter. The quality of this limiting procedure is linked to the diversity of the matrices contained in $\mathscr{M}^{p,r}$ and in the ability of the minimization criteria to select the appropriate projection. Numerical results exposed in Sect. 5 will demonstrate the quality of the sets and criteria just mentioned. The definition of other sets $\mathscr{M}^{p,r}$ for higher-order scheme and different reduction order is currently a topic of research for the authors.

4.3.2 Non-linear System of Conservation Laws

Equation (1) is now a system of conservation laws, u and F are taken as column vectors. To accentuate the limitation of the oscillations by the SWeP limiting procedure, we first employ a local characteristic field decomposition. The Jacobian matrix of the flux function is denoted with $A_i = \frac{\partial F}{\partial u}\big|_{\breve{u}_i}$ and its left and right normalized eigenvectors are $l_i^{(k)}$ and $r_i^{(k)}$, $k = 1, \ldots m$

($m = 3$ for the 1D Euler equations). The $m \times m$ transfer matrix constructed with the right eigenvectors taken as its columns is denoted with:

$$\boldsymbol{R}\left(\breve{u}_i\right) = \left(r_i^{(1)}, r_i^{(2)}, \ldots, r_i^{(m)}\right). \tag{82}$$

Because the eigenvectors are normalized, it is easy to see that:

$$\boldsymbol{R}^{-1}\left(\breve{u}_i\right) = \left(l_i^{(1)}, l_i^{(2)}, \ldots, l_i^{(m)}\right)^T. \tag{83}$$

At time n the values of u at solution points in cell i is represented by the matrix (super-script n has been removed for better readability):

$$\boldsymbol{u}_i^{\mathrm{s}} = \left(u_{i,1}^{\mathrm{s}}, u_{i,2}^{s}, \ldots, u_{i,p+1}^{s}\right). \tag{84}$$

The matrix of the associated local characteristic representation $\boldsymbol{C}_i^s$ used is:

$$\boldsymbol{C}_i^{\mathrm{s}} = \begin{pmatrix} \boldsymbol{c}_i^{(1)^T} \\ \vdots \\ \boldsymbol{c}_i^{(m)^T} \end{pmatrix} = \boldsymbol{R}^{-1}\boldsymbol{u}_i^{\mathrm{s}}. \tag{85}$$

Vectors $\boldsymbol{c}_i^{(l)}$, $1 \leq l \leq m$, are the discretization of the m polynomial characteristics $\hat{c}_i^{(l)}$ over cell S_i. We naturally denote:

$$\hat{\boldsymbol{C}}_i = \begin{pmatrix} \hat{c}_i^{(1)} \\ \vdots \\ \hat{c}_i^{(m)} \end{pmatrix}. \tag{86}$$

In order to use the minmod detector, information is needed from the immediate neighbouring cell:

$$\boldsymbol{C}_{i-1}^{s} = \boldsymbol{R}^{-1}\boldsymbol{u}_{i-1}^{\mathrm{s}}, \quad \boldsymbol{C}_{i+1}^{s} = \boldsymbol{R}^{-1}\boldsymbol{u}_{i+1}^{\mathrm{s}}, \tag{87}$$

where matrix $\boldsymbol{R}$ is the same as the one in (85). Polynomial vectors $\hat{\boldsymbol{C}}_{i-1}$ and $\hat{\boldsymbol{C}}_{i+1}$ are defined in the same manner as $\hat{\boldsymbol{C}}_i$. Then we can compute:

$$\tilde{\boldsymbol{C}}_i^{(\mathrm{mod})} = \tilde{m}\left(\tilde{\boldsymbol{C}}_i, \Delta_+\overline{\boldsymbol{C}}_i, \Delta_-\overline{\boldsymbol{C}}_i\right) \quad \tilde{\tilde{\boldsymbol{C}}}_i^{(\mathrm{mod})} = \tilde{m}\left(\tilde{\tilde{\boldsymbol{C}}}_i, \Delta_+\overline{\boldsymbol{C}}_i, \Delta_-\overline{\boldsymbol{C}}_i\right), \tag{88}$$

using for each components of the vector the previously defined function $\tilde{m}$. At this point, a cell is **not** going to be declared as a troubled cell if and only if both the following vectorial equalities are true:

$$\tilde{\boldsymbol{C}}_i^{(\mathrm{mod})} = \tilde{\boldsymbol{C}}_i, \quad \tilde{\tilde{\boldsymbol{C}}}_i^{(\mathrm{mod})} = \tilde{\tilde{\boldsymbol{C}}}_i \tag{89}$$

Assume that cell S_i is detected as a troubled cell. The scalar SWeP limiting procedure describe in Sect. 4.3.1 is applied to all vectors $\boldsymbol{c}_i^{(l)}$. We obtain m new vectors $\boldsymbol{c}_i^{(l),\mathrm{new}}$, that are combined to obtain the projected conservative variables following:

$$\boldsymbol{u}_i^{\mathrm{s,new}} = \boldsymbol{R}\boldsymbol{C}_i^{\mathrm{s,new}} = \boldsymbol{R}\begin{pmatrix} \boldsymbol{c}_i^{(1),\mathrm{new}^T} \\ \vdots \\ \boldsymbol{c}_i^{(m),\mathrm{new}^T} \end{pmatrix}. \tag{90}$$

As in the scalar case, we can now use the new projected solution in an adequate time-marching scheme.

5 Numerical Results

In this section, numerical computations are presented to demonstrate the efficiency of the compact SWeP limiter. The different examples presented use the system of the one-dimensional Euler equations:

$$\partial_t u + \partial_x F(u) = 0, \tag{91}$$

with:

$$u = \begin{pmatrix} \rho \\ \rho v \\ E \end{pmatrix}, \quad F(u) = \begin{pmatrix} \rho v \\ \rho v^2 + p \\ v(E+p) \end{pmatrix}, \tag{92}$$

where ρ is the density, v the velocity, E the total energy and p the pressure defined with:

$$p = (\gamma - 1)\left(E - \frac{1}{2}\rho v^2\right), \tag{93}$$

and $\gamma = 1.4$ for the air.

The SWeP limiting procedures for the fourth and fifth order SD schemes respectively associated to the sets $\mathscr{I}^{3,1}$ and $\mathscr{I}^{4,1}$ detailed in Sect. 3.4 are evaluated. The TVB constant M of the minmod detector takes the same values as those proposed in [30]. In every case studied, an over-limiting very small value and an adapted value are used to test the permissiveness of the limiter regarding the troubled cell detector. In the figures proposed, solid black line stands for reference solutions and red diamonds for SD solutions computed with the SWeP limiter. The figures present one point per cell, corresponding to the average value of the cell. In every figure a zoom is proposed, then all the DoFs contained in each cell are displayed using blue filled circles, allowing to differentiate specific behaviour like inner-cell oscillations or overshoots.

In some cases, we needed to resort to the positivity-preserving limiter used in [29,30]. It is designed to ensure the positivity of the density and the pressure at any time step. Implementation details can be found in [39,40]. All our meshes are constituted of identical cells. The third-order RK3 SSP time-marching scheme is used. The CFL evaluation is based on the cell size. For all our computations, a CFL number of 0.2 has been taken, except in the first example where the spatial scheme order is evaluated and where the CFL is fixed to 0.05 so that the error due to the time-marching scheme became negligible compared to the error due to the spatial discretization. We used the HLLE Riemann solver [41] for fluxes evaluation. All our computations were performed with uniform meshes.

Apart from the first example, we use a second-order finite-volume spatial scheme with MUSCL field reconstruction [5] coupled with the HLLC Riemann solver and the RK3 SSP time marching scheme to obtain the reference solutions. The mesh sizes used are detailed in the examples description.

Example 1 We first test the different limiters involved with a grid-convergence study. The computational domain is $[0, 2\pi]$. The Euler equations are reduced to an advection problem equations with the initial conditions $\rho(x, 0) = 1 + 0.2\sin(x)$, $v(x, 0) = 1$, $p = 1$. Periodic boundary condition are imposed. The exact solution is the advection of the density profile following $\rho(x, t) = 1 + 0.2\sin(x - t)$ while pressure and velocity profile remain constant. Total simulation time is $t = 2\pi$. At every time step, if the SWeP limiter is enabled, the totality of the mesh cells are declared as troubled so that the minmod detector is not involved in this example.

Table 1 Grid convergence study of example 1 for the advection of a sine density wave with the Euler equations

N^s	SWeP limiter disabled, ($r = 0$)				SWeP limiter enabled, ($r = 1$)			
	L^1 norm	Order	L^∞ norm	Order	L^1 norm	Order	L^∞ norm	Order
$p = 3$								
10	4.91E−5	–	1.83E−5	–	1.48E−3	–	4.84E−4	–
20	3.05E−6	4.00	1.15E−6	3.99	2.33E−4	2.66	7.55E−5	2.68
40	1.90E−7	4.00	7.19E−8	4.00	3.27E−5	2.83	9.72E−6	2.96
80	1.19E−8	4.00	4.50E−9	4.00	4.36E−6	2.91	1.27E−6	2.93
160	7.42E−10	4.00	2.81E−10	4.00	5.63E−7	2.95	1.63E−7	2.96
$p = 4$								
10	1.53E−6	–	6.09E−7	–	2.64E−4	–	1.41E−4	–
20	4.82E−8	4.99	1.93E−8	4.98	2.51E−5	3.39	2.05E−5	2.78
40	1.50E−9	5.00	6.03E−10	5.00	1.98E−6	3.66	2.57E−6	3.00
80	4.69E−11	5.00	1.89E−11	5.00	1.40E−7	3.81	3.03E−7	3.08
160	1.78E−12	4.72	8.55E−13	4.47	8.17E−9	4.10	2.59E−8	3.55

Errors for the density profile. All the mesh cells are declared as troubled cells

Results for different mesh sizes are gathered in Table 1. Firstly, high-order convergence is naturally recovered when the SWeP limiter is turned off. The disparity observed for the finest mesh are attributed to the important smallness of the error recorded. Secondly, the SWeP limiter performs as expected and maintains a high-order of accuracy equal to $p + 1 - r$, although small discrepancies are observed for the L^∞ norm.

Example 2 The Sod shock tube problem is investigated. Computational domain is $[-5, 5]$. Initial condition respects:

$$u(x, 0) = \begin{cases} u_L, & x \leq 0 \\ u_R, & x > 0 \end{cases} \tag{94}$$

with $(\rho_L, v_L, p_L) = (1, 0, 1)$ and $(\rho_R, v_R, p_R) = (0.125, 0, 0.1)$. The Mesh is composed of 100 cells. Results are presented in Fig. 5. The reference solution used a mesh composed of 3000 DoFs.

The different results obtained are very close to those observed in [30]. The relaxation profile is correctly approximated and no important smearing is discernible, even around the contact discontinuity. It is well-known that high-order methods became more diffusive when the polynomial degree of the solution field is lowered. This increase in the diffusive behaviour is confirmed by two different modifications of the solution that appear as the minmod detector, and consequently the SWeP limiter, gets more often activated (with $M = 0.01$): the reduction of the small inner-cell oscillations and the thickening of the contact discontinuity.

Example 3 The other usual problem with a Riemann type initial condition of the form (94) is the Lax problem. It follows $(\rho_L, v_L, p_L) = (0.445, 0.698, 3.528)$ and $(\rho_R, v_R, p_R) = (0.5, 0, 0.571)$. Computational domain is again $[-5, 5]$ with a mesh composed of 100 cells. Results are presented in Fig. 6. The reference solution used a mesh composed of 3000 DoFs.

The SWeP limiter provides accurate solutions. No major differences can be identified when looking at global solutions. Yet, interesting comments concerns the density plateau. As

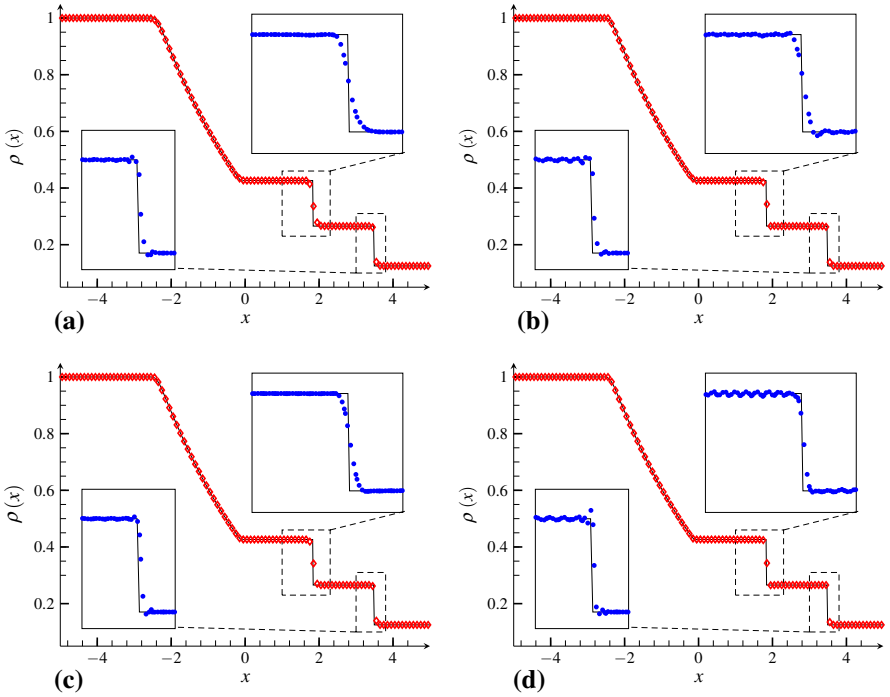

Fig. 5 The Sod problem, $t = 2$, $N^{\mathrm{S}} = 100$. Simulations present the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$. **b** $p = 3$, $M = 5$. **c** $p = 4$, $M = 0.01$. **d** $p = 4$, $M = 10$

in the second example, the diffusive effect of the limiter appears as M diminishes and make the inner-cell oscillations of the plateau shrink.

Example 4 The problem of the interaction of two blast waves is considered here. The initial condition for the computational domain [0, 1] is taken as:

$$u(x,0) = \begin{cases} u_L, & 0 \le x < 0.1 \\ u_M, & 0.1 \le x < 0.9 \\ u_R, & 0.9 \le x < 1 \end{cases} \tag{95}$$

with $(\rho_L, v_L, p_L) = \left(1, 0, 10^3\right)$, $(\rho_M, v_M, p_M) = \left(1, 0, 10^{-2}\right)$ and $(\rho_R, v_R, p_R) = \left(1, 0, 10^2\right)$. Reflective boundary conditions are used at the ends of the domain. More details can be found in [42]. The positivity-preserving limiter is added here to prevent negative density or pressure. Density plot are presented in Fig. 7. The reference solution used a mesh composed of 10^4 DoFs.

In [29,30], the minmod constant takes the two values 0.01 and 200 and provides similar results. We noticed that higher values of M could be used, we reached $M = 5 \times 10^4$ and ran some very under-limited computations. We can notice that there are no important or major changes in the results whether $M = 0.01$ or $M = 5 \times 10^4$. Global results for both scheme order are good, with small imperfections disclosed by the zoom: small inner cell oscillations and an incised maximum around $x = 0.75$. These flaws are here again reduced in the over-limited computations thanks to the diffusive effect of the limiter.

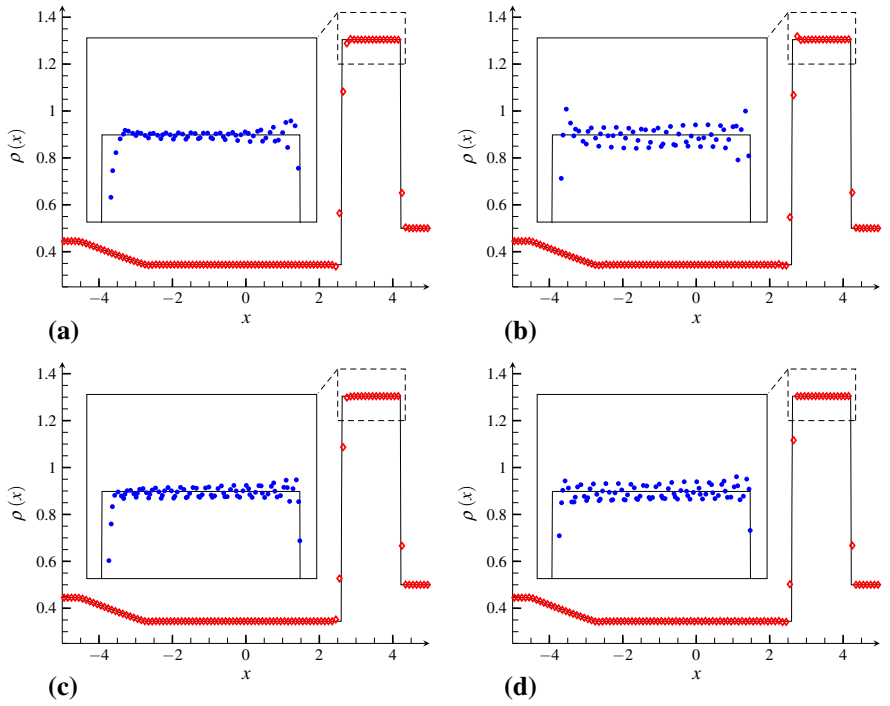

Fig. 6 The Lax problem, $t = 1.7$, $N^s = 100$. Simulations present the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$. **b** $p = 3$, $M = 15$. **c** $p = 4$, $M = 0.01$. **d** $p = 4$, $M = 30$

Example 5 The LeBlanc shock tube problem is considered in this example. The initial condition for the computational domain $[-10, 10]$ is taken as:

$$u(x, 0) = \begin{cases} u_L, & -10 \leq x < 0 \\ u_R, & 0 \leq x < 10 \end{cases} \tag{96}$$

with $(\rho_L, v_L, p_L) = \left(2, 0, 10^9\right)$ and $(\rho_R, v_R, p_R) = \left(10^{-3}, 0, 1\right)$. The positivity preserving limiter has been used in this example for all the computations. The reference solution used a mesh composed of 4×10^4 DoFs.

Density and velocity profile are presented in Figs. 8 and 9. To improve the perceptibility of the scheme influence, we used a coarse mesh composed of 400 cells. Again, the results are globally equivalent. As expected, the fifth-order scheme is better able to capture the discontinuities than the fourth-order scheme, or to reproduce the plateau between the rarefaction wave and the contact discontinuity. The behaviour of the limiter is less marked compared to previous examples. For a scheme of a given order, the different values of the minmod constant M used give noticeably comparable results.

Example 6 The last example is the Shu–Osher problem. It exposes the advantages of the high-order methods as it contains both shocks and complex smooth region structures. The initial condition for the computational domain $[-5, 5]$ is taken as:

$$u(x, 0) = \begin{cases} u_L, & -5 \leq x < -4 \\ u_R, & 4 \leq x < 5 \end{cases} \tag{97}$$

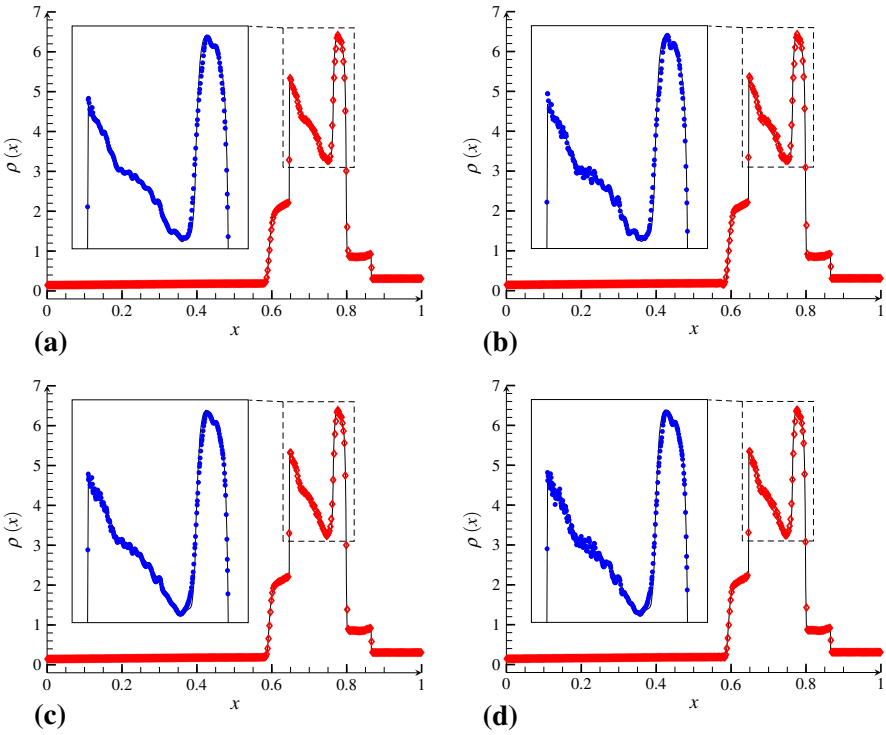

Fig. 7 The blast waves interaction problem, $t = 0.038$, $N^s = 400$. Simulations present the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$. **b** $p = 3$, $M = 5 \times 10^4$. **c** $p = 4$, $M = 0.01$. **d** $p = 4$, $M = 5 \times 10^4$

with $(\rho_L, v_L, p_L) = (3.857143, 2.629369, 10.333333)$ and $(\rho_R, v_R, p_R) = \big(1 + 0.2\sin(5x), 0, 1\big)$. The mesh is composed of 110 cells. Results are presented in Fig. 10. For this test case, the positivity preserving limiter as only be used in the cases where $M = 300$. The reference solution used a mesh composed of 10^4 DoFs.

For $M = 0.01$, we observe that the region presenting the major density variations is captured, but with an underestimation of the solution amplitude. Once again, this has to be linked to the diffusive level of the SWeP limiter. For $M = 300$, fourth order scheme performed very well, but fifth order scheme seems to present small oscillations at the end of the fine structures region. We ran a computation with $M = 150$ (not shown here) and observe the vanishing of these oscillations, suggesting that the value of $M = 300$ is a very under-limited case.

6 Conclusion

In the present study, we developed a general framework based on the spectral volume methodology that provides polynomial reconstruction of variable order using conservative polynomial projections. The exposed projections contain degrees of freedom gathered in a parameter matrix and that were identified as spatial weights. Thanks to these new

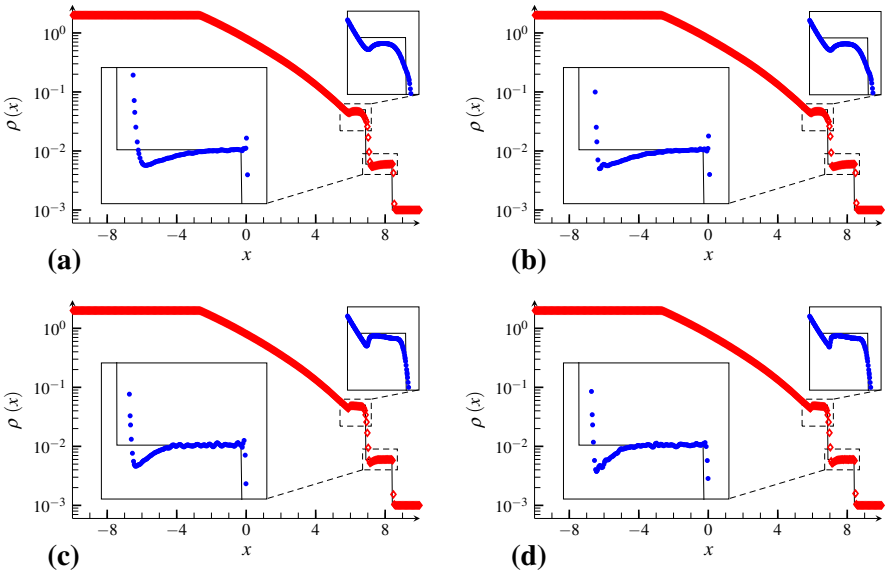

Fig. 8 The LeBlanc problem, $t = 0.0001$, $N^{\mathrm{s}} = 400$. Simulations present density profile for the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$. **b** $p = 3$, $M = 200$. **c** $p = 4$, $M = 0.01$. **d** $p = 4$, $M = 200$

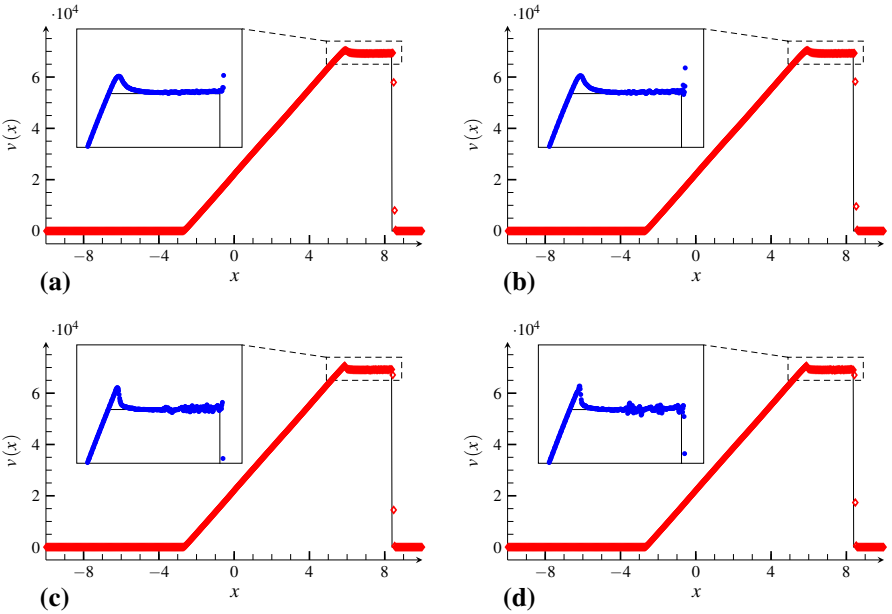

Fig. 9 The LeBlanc problem, $t = 0.0001$, $N^{\mathrm{s}} = 400$. Simulations present velocity profile for the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$. **b** $p = 3$, $M = 200$. **c** $p = 4$, $M = 0.01$. **d** $p = 4$, $M = 200$

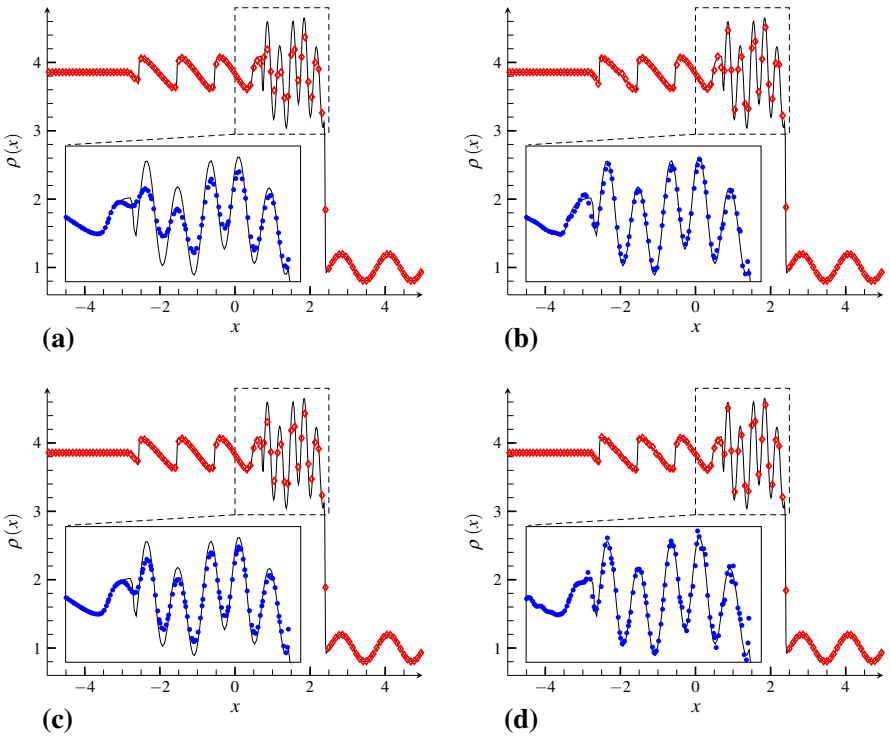

Fig. 10 The Shu–Osher problem, $t = 1.8$, $N^{\mathrm{s}} = 110$. Simulations present the reference solution (*solid black*) and the SWeP scheme (*open red diamond*). Zoomed region display all the DoFs contained in each cell (*filled blue circle*). **a** $p = 3$, $M = 0.01$, **b** $p = 3$, $M = 300$, **c** $p = 4$, $M = 0.01$, **d** $p = 4$, $M = 300$

DoFs, numerical dissipation could implicitly be added to the projections and unwanted side effects coming from the interpolation of a discontinuity have been controlled. Then, a limiting procedure based on these spatially weighted projections have been proposed, the SWeP limiter. Its major attribute is its compactness, competing with that of the standard compact high-order methods. Another asset is its portability. Because the SWeP limiter is based on the polynomial structure of the data in one cell, it can technically be combined with any high-order methodology. A grid-convergence study shows that the SWeP limiter was able to recover optimal high-order convergence rates and different numerical examples have demonstrated the capacity of the SWeP limiting procedure to accurately deliver non-oscillating 1D simulations of flows containing discontinuities. The authors would also like to point out that the extension to higher dimension is theoretically straightforward as long as the methodology is mainly based on algebraic developments. Future works will concern the computations of multi-dimensional cases that involve the SWeP limiter.

Two different components make up the SWeP limiting procedure. Firstly, the set of weighting matrices: it provides the different projectors needed to handle the problematic discrete patterns identified over one cell. As of today, the different matrices are set with an empirical approach. Yet, the development of a deterministic strategy is currently under investigation by the authors. Secondly, the projection selection criteria: used to detect the optimal projection to use, its evaluation should not be too complex as its implementation will directly affect the

computational cost of the SWeP limiter. The criteria used in this paper is for example very easy to compute.

Moreover, even if the limiting procedure described in this paper is compact, the used min-mod detector still need information from its neighbours. A fully compact limiting procedure would use a compact detector, which is still a hot topic of research. The development of a compact detector based on to the weighted projections presented in this paper is contemplated by the authors.

Acknowledgments The authors thank Professor Laurent Joly for many helpful discussions and for carefully reviewing a first draft of the paper. This research was supported in part by the French Ministry of Defence through a financial support of the DGA.

Appendix: Weighting Matrices for the SWeP Limiter

As stated in Sect. 3.4, we provide in this appendix the exact values of the weighting matrices used in our computations. For the patterns of Fig. 3a:

$$\boldsymbol{M}^{\mathrm{l}} = \begin{pmatrix} 0.35 & 0 & 0 & 0 \\ 0 & 2.30 & 0 & 0 \\ 0 & 0 & 11.823189085020825 & 0 \\ 0 & 0 & 0 & 10.00 \end{pmatrix}, \quad \boldsymbol{M}^{\mathrm{c}} = \begin{pmatrix} 1.0 & 0 & 0 & 0 \\ 0 & 2.0 & 0 & 0 \\ 0 & 0 & 2.0 & 0 \\ 0 & 0 & 0 & 1.0 \end{pmatrix}, \tag{98}$$

and for the patterns of Fig. 3b:

$$\begin{aligned} \boldsymbol{M}^{\mathrm{l}} &= \begin{pmatrix} 0.12 & 0 & 0 & 0 & 0 \\ 0 & 0.90 & 0 & 0 & 0 \\ 0 & 0 & 5.00 & 0 & 0 \\ 0 & 0 & 0 & 7.177304710492795 & 0 \\ 0 & 0 & 0 & 0 & 5.00 \end{pmatrix}, \\ \boldsymbol{M}^{\mathrm{lc}} &= \begin{pmatrix} 5.0 & 0 & 0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 & 0 \\ 0 & 0 & 1.5 & 0 & 0 \\ 0 & 0 & 0 & 15.145898361801240 & 0 \\ 0 & 0 & 0 & 0 & 4.0 \end{pmatrix}. \end{aligned} \tag{99}$$

References

1. Ekaterinaris, J.A.: High-order accurate, low numerical diffusion methods for aerodynamics. Prog. Aerosp. Sci. **41**(3), 192–300 (2005)
2. Wang, Z.J.: High-order methods for the Euler and Navier-Stokes equations on unstructured grids. Prog. Aerosp. Sci. **43**(1), 1–41 (2007)
3. Wang, Z.J., Fidkowski, K., Abgrall, R., Bassi, F., Caraeni, D., Cary, A., Deconinck, H., Hartmann, R., Hillewaert, K., Huynh, H.T., et al.: High-order CFD methods: current status and perspective. Int. J. Num. Method Fluids **72**(8), 811–845 (2013)
4. Godunov, S.K.: A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics. Matematicheskii Sbornik **89**(3), 271–306 (1959)
5. Van Leer, B.: Towards the ultimate conservative difference scheme V: a second-order sequel to Godunov's method. J. Comput. Phys. **32**(1), 101–136 (1979)
6. Wang, Z.J.: Spectral (finite) volume method for conservation laws on unstructured grids I: basic formulation. J. Comput. Phys. **178**(1), 210–251 (2002)

7. Wang, Z.J., Liu, Y.: Spectral (finite) volume method for conservation laws on unstructured grids III: one dimensional systems and partition optimization. J. Sci. Comput. **20**(1), 137–157 (2004)
8. Wang, Z.J., Liu, Y.: Spectral (finite) volume method for conservation laws on unstructured grids II: extension to two-dimensional scalar equation. J. Comput. Phys. **179**(2), 665–697 (2002)
9. Wang, Z.J., Zhang, L., Liu, Y.: Spectral (finite) volume method for conservation laws on unstructured grids IV: extension to two-dimensional systems. J. Comput. Phys. **194**(2), 716–741 (2004)
10. Liu, Y., Vinokur, M., Wang, Z.J.: Spectral (finite) volume method for conservation laws on unstructured grids V: extension to three-dimensional systems. J. Comput. Phys. **212**(2), 454–472 (2006)
11. Sun, Y., Wang, Z.J., Liu, Y.: Spectral (finite) volume method for conservation laws on unstructured grids IV: extension to viscous flow. J. Comput. Phys. **215**(1), 41–58 (2006)
12. Zhang, M., Shu, C.-W.: An analysis of and a comparison between the discontinuous Galerkin and the spectral finite volume methods. Comput. Fluids **34**(4), 581–592 (2005)
13. Haga, T., Ohnishi, N., Sawada, K., Masunaga, A.: Spectral volume computation of flowfield in aerospace application using Earth Simulator. In: 24th AIAA Applied Aerodynamics Conference, AIAA Paper 2006-2823, San Fransisco, California (2006)
14. Van den Abeele, K., Broeckhoven, T., Lacor, C.: Dispersion and dissipation properties of the 1D spectral volume method and application to a p-multigrid algorithm. J. Comput. Phys. **224**(2), 616–636 (2007)
15. Abeele, K., Lacor, C., Wang, Z.J.: On the stability and accuracy of the spectral difference method. J. Sci. Comput. **37**(2), 162–188 (2008)
16. Harris, R., Wang, Z. J.: Partition design and optimization for high-order spectral volume schemes. In: 48th AIAA Aerospace Sciences Meeting, AIAA Paper 2009-1333, Orlando, Florida (2009)
17. Harris, R., Wang, Z.J., Liu, Y.: Efficient quadrature-free high-order spectral volume method on unstructured grids: theory and 2D implementation. J. Comput. Phys. **227**(3), 1620–1642 (2008)
18. Harris, R., Wang, Z.J.: High-order adaptive quadrature-free spectral volume method on unstructured grids. Comput. Fluids **38**(10), 2006–2025 (2009)
19. Wang, Z.J., Gao, H.: A unifying lifting collocation penalty formulation including the discontinuous Galerkin, spectral volume/difference methods for conservation laws on mixed grids. J. Comput. Phys. **228**(21), 8161–8186 (2009)
20. Nguyen, N., Peraire, J.: An adaptive shock-capturing hdg method for compressible flows. AIAA Paper 3060 (2011)
21. VonNeumann, J., Richtmyer, R.D.: A method for the numerical calculation of hydrodynamic shocks. J. Appl. Phys. **21**(3), 232–237 (1950)
22. Cockburn, B., Shu, C.-W.: Tvb runge-kutta local projection discontinuous galerkin finite element method for conservation laws. II: General framework. Math. Comput. **52**(186), 411–435 (1989)
23. Biswas, R., Devine, K.D., Flaherty, J.E.: Parallel, adaptive finite element methods for conservation laws. Appl. Numer. Math. **14**(1), 255–283 (1994)
24. Burbeau, A., Sagaut, P., Bruneau, C.-H.: A problem-independent limiter for high-order runge-kutta discontinuous galerkin methods. J. Comput. Phys. **169**(1), 111–150 (2001)
25. 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)
26. 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. Comput. Fluids **64**, 43–63 (2012)
27. Harten, A., Engquist, B., Osher, S., Chakravarthy, S.R.: Uniformly high order accurate essentially non-oscillatory schemes, III. J. Comput. Phys. **71**(2), 231–303 (1987)
28. Liu, X.-D., Osher, S., Chan, T.: Weighted essentially non-oscillatory schemes. J. Comput. Phys. **115**(1), 200–212 (1994)
29. Zhong, X., Shu, C.-W.: A simple weighted essentially nonoscillatory limiter for runge-kutta discontinuous galerkin methods. J. Comput. Phys. **232**(1), 397–415 (2013)
30. Du, J., Shu, C.-W., Zhang, M.: A simple weighted essentially non-oscillatory limiter for the correction procedure via reconstruction (CPR) framework. Appl. Numer. Math. **95**, 173–198 (2015)
31. Harten, A.: High resolution schemes for hyperbolic conservation laws. J. Comput. Phys. **49**(3), 357–393 (1983)
32. Shu, C.-W.: TVB uniformly high-order schemes for conservation laws. Math. Comput. **49**(179), 105–121 (1987)
33. Qiu, J., Shu, C.-W.: Runge-kutta discontinuous galerkin method using weno limiters. J. Sci. Comput. **26**(3), 907–929 (2005)
34. Zhu, J., Qiu, J., Shu, C.-W., Dumbser, M.: Runge-kutta discontinuous galerkin method using weno limiters II: unstructured meshes. J. Comput. Phys. **227**(9), 4330–4353 (2008)

35. Du, J., Shu, C.-W., Zhang, M.: A simple weighted essentially non-oscillatory limiter for the correction procedure via reconstruction (CPR) framework on unstructured meshes. Appl. Numer. Math. **90**, 146–167 (2015)
36. Persson, P.-O., Peraire, J.: Sub-cell shock capturing for discontinuous galerkin methods. In: 44th AIAA Aerospace Sciences Meeting, Vol. 112 (2006)
37. Sheshadri, A., Jameson, A.: Shock detection and capturing methods for high order discontinuous-galerkin finite element methods. In: 32nd AIAA Applied Aerodynamics Conference, Vol. 2688 of AIAA Paper (2014)
38. Gottlieb, S., Shu, C.-W., Tadmor, E.: Strong stability-preserving high-order time discretization methods. SIAM Rev. **43**(1), 89–112 (2001)
39. Zhang, X., Shu, C.-W.: On maximum-principle-satisfying high order schemes for scalar conservation laws. J. Comput. Phys. **229**(9), 3091–3120 (2010)
40. Zhang, X., Shu, C.-W.: On positivity-preserving high order discontinuous galerkin schemes for compressible euler equations on rectangular meshes. J. Comput. Phys. **229**(23), 8918–8934 (2010)
41. Einfeldt, B.: On godunov-type methods for gas dynamics. SIAM J. Numer. Anal. **25**(2), 294–318 (1988)
42. Woodward, P., Colella, P.: The numerical simulation of two-dimensional fluid flow with strong shocks. J. Comput. Phys. **54**(1), 115–173 (1984)