A Posteriori Correction of High-Order Discontinuous Galerkin Scheme through Subcell Finite Volume Formulation and Flux Reconstruction

François Vilar[a]

[a] *IMAG, Univ Montpellier, CNRS, Montpellier, France*

Abstract

In this paper, we present a new limiter for discontinuous Galerkin (DG) schemes, based on subcell resolution through reconstructed flux correction, for hyperbolic conservation laws. This limitation technique is constructed by means of a subcell Finite Volume (FV) formulation. The algorithm is thus simple, and is able to retain the very high accurate subcell resolution of DG schemes.

The main idea of this correction procedure is to preserve as much as possible the high accuracy and the very precise subcell resolution of DG schemes. Consequently, an *a posteriori* correction will only be applied locally at the subcell scale where it is needed, but still ensuring the scheme conservation. To do so, we first reformulate DG schemes as a subcell FV scheme provided the use of the correct numerical flux. This very simple development provides us with the so-called DG reconstructed flux. This theoretical part will serve as a basis for our limiter framework.

Practically, at each time step, we compute a DG candidate solution and check if this solution is admissible (for instance positive, non-oscillating, ...). If it is the case, we go further in time. Otherwise, we return to the previous time step and correct locally, at the subcell scale, the numerical solution. This is why it is refereed to as *a posteriori* limitation. To this end, each cell is subdivided into subcells. Then, if the solution is locally detected as bad, we substitute the DG reconstructed flux on the subcell boundaries by a robust first-order or second-order TVD numerical flux. And for subcell detected as admissible, we keep the high-order reconstructed flux which allows us to retain the very high accurate resolution and conservation of the DG scheme. Furthermore, only the solution inside troubled subcells and its first neighbors will have to be recomputed, elsewhere the solution remains unchanged. Numerical results on various type problems and test cases will be presented, both in 1D and 2D on Cartesian grids, to assess the very good performance of the design limiting algorithm.

Keywords: Discontinuous Galerkin scheme, a posteriori limitation, subcell limitation, arbitrary high-order, DG subcell FV formulation, Flux reconstruction (CPR), hyperbolic conservation laws, subcell conservative scheme

Email address: `francois.vilar@umontpellier.fr` (François Vilar)

1. Introduction

The Discontinuous Galerkin (DG) method, initially introduced by Reed and Hill in the context of neutron transport [52], has become these last decades one of the most widely used numerical scheme, especially in the context of computational fluid dynamics. A major development of DG schemes was carried out by Cockburn, Shu *et al* in a series of seminal papers [12, 11, 9, 10]. Theoretically, DG methods allow to reach any arbitrary order of accuracy, while keeping the stencil compact, along with other good properties as L_2 stability and *hp*-adaptivity. Discontinuous Galerkin scheme is extremely accurate with a very precise subcell resolution. It is even superconvergent in some cases, see for instance [45, 63]. Nonetheless, accuracy is not the only issue to be addressed. Robustness is of fundamental importance. It is well known that high-order DG schemes can produce spurious oscillations in the presence of discontinuities. These non-physical oscillations may generate non-admissible solution (negative density or pressure in the case of gas dynamics for instance), which may lead to nonlinear instability or crash of the code. Those schemes thus require some stabilization techniques. This fundamental issue has been extensively tackled in the past, and there is thus a vast literature on that topic. In the following, a brief review of the different existing limitation techniques is given. For a lot more detailed description of the state of the art limiter, we refer to [18] and the references within.

To overcome this issue of spurious oscillations, a wide range of techniques already exists. These techniques mainly rely on two different paradigms that we referred to as *a priori* and *a posteriori*. In the so-called *a priori* limitation, the correction procedure is applied before advancing the numerical piecewise polynomial solution further in time. So first, a troubled zone indicator is used to find where a limitation is required. Then, sufficient efforts are made on the numerical solution or on the numerical scheme to be sure that one will be able to carry the computation out to the next time step. Among others *a priori* limitation techniques, we could mention the artificial viscosity technique [46, 57, 19] where some dissipative mechanism is added in shock regions. Some other very popular limiting techniques can be gathered and referred to as slope and moment limiters [10, 3, 4, 38, 62, 34, 41, 44]. In the former ones, as in [12, 10], the polynomial approximated solution is flattened around its mean value to control the solution jumps at cell interfaces. A smooth extrema detector is then generally used to prevent the limitation technique to spoil the accuracy in regions where no limiting is required. Moments limiters, mainly based on [3] and further developed in [4], can be seen as the extension of the aforementioned slope limiters to the case of very high orders of accuracy. In those limiting strategies, the different moments of the polynomial solution are successively scaled in a decreasing sequence, from the higher degree to the lower one, allowing the preservation of the solution accuracy, as well as ensuring the solution boundedness near discontinuities. The high-order DG limiter [38], generalized moment limiter [62], hierarchical Multidimensional Limiting Process (MLP) [34, 33] and vertex-based hierarchical slope limiters [41, 44] all derive from [12, 3, 4], and thus fall into this category. Now, another limiting strategy that deserves to be mentioned is the (H)WENO limiting procedure [47, 2, 64, 65], where the DG polynomial is substituted in troubled regions by a reconstructed (H)WENO polynomial. Last but not least, some original subcell Finite Volume (FV) shock capturing techniques in the frame of DG schemes [28, 6, 56, 13] have recently gained in popularity. In [28], the authors use a convex combination between high-order DG schemes and first-order finite volumes on a subgrid, allowing them to retain the very high accurate resolution of DG in smooth areas and ensuring the scheme robustness in the presence of shocks. Similarly, in [56, 13], after having detected the troubled zones, cells are then subdivided into subcells, and a robust first-order finite volume scheme is performed on the subgrid

in troubled cells. Alternatively, some robust high-order scheme as MUSCL or WENO could either be used to avoid too much accuracy discrepancy.

The *a priori* paradigm has already and extensively proved in the past its high capability and feasibility, as in the aforementioned articles. Those techniques are *a priori* in the sense that only the data at time t^n are needed to perform the limitation procedure. Then, the limited solution is used to advance the numerical scheme in time to t^{n+1}. The "worst case scenario" has to be generally considered as a precautionary principle. The paradigm of *a posteriori* limitation is different in the way that first an unlimited candidate solution is computed at the new time step. The unlimited solution is then checked according to some criteria (for instance positivity, discrete maximum principle, ...). If the solution is considered admissible, we go further in time. Otherwise, we return to the previous time step and correct locally the numerical solution by making use of a more robust scheme. Because the troubled zone detection is performed *a posteriori*, the correction can be done only where it is absolutely necessary. Furthermore, let us emphasize that in *a posteriori* correction procedures, the maximum principle preservation or positivity preservation is included without any additional effort, while it is generally not the case of *a priori* limitations. Any other property can be added as long as the admissibility set is convex, as entropy stability for instance. Their scalability to any order of accuracy is also perfectly natural, as it does not imply to modify different moments of a polynomial which may be of different orders of magnitude. Also, because *a posteriori* corrections rely on a robust scheme, generally first-order finite volume scheme, the corrected DG scheme is less sensitive to stability issue, as if the high-order method starts developing instabilities it will trigger the *a posteriori* correction and the robust scheme is then alternately used.

Recently, some new *a posteriori* limitations have arisen. Let us mention the so-called MOOD technique, [8, 14, 15]. Through this procedure, the order of approximation of the numerical scheme is locally reduced in an *a posteriori* sequence until the solution becomes admissible. In [18, 17], a subcell FV technique similar to the one presented in [56] has been applied to the *a posteriori* paradigm. Practically, if the numerical solution in a cell is detected as bad, the cell is then subdivided into subcells and a first-order finite volume, or alternatively other robust scheme (second-order TVD FV scheme, WENO scheme, ...), is applied on each subcell. Then, through these new subcell mean values, a high-order polynomial is reconstructed on the primal cell. In contrary to [56], in [18, 17] the authors make use of a lot more subcells than degrees of freedom of the DG solution, namely $2k+1$ subcells for $k+1$ degrees of freedom for a $(k+1)^{\text{th}}$-order DG scheme in the one-dimensional case. Different arguments have motivated such choice. First, the subcell finite volume scheme optimal CFL condition will match the one of the Runge-Kutta discontinuous Galerkin method on the primal cell. The second reason is that the more subcells are used the more accurate the subcell correction will theoretically be. However, doing so, to be able to reconstruct a polynomial of degree k with $2k+1$ subcell mean values, the authors are constrained to make use of a least square procedure, loosing at the same time the subcell mean values computed through the robust corrected scheme. It is then impossible to prove that the numerical solution is for instance positive at the subcell level. Nevertheless, this correction procedure has the benefit to be very simple and robust, and is able to preserve the high accuracy of DG schemes in smooth areas. We want to emphasize that the correction procedure presented in this paper belongs to same family as this aforementioned *a posteriori* technique, as it relies on a subgrid decomposition and finite volume correction.

In all the aforementioned limitation techniques, *a priori* and *a posteriori*, in the troubled cells the

high-order DG polynomial is either globally modified in the cell, or even discard as it is in the (H)WENO limiter or any *a posteriori* correction technique. One of the main advantage of high-order scheme is to be able to use coarse grids while still being very precise. But even in the case where the troubled zone, as the vicinity a shock for instance, is very small regarding the characteristic length of a cell, the DG polynomial will be globally modified. In the present paper, we then introduce a conservative technique to overcome this issue, by modifying the DG numerical solution only locally at the subcell scale. Let us now list the different objectives of the designed correction procedure. First, to avoid the occurrence of non-admissible solution, we want the corrected scheme to be maximum principle preserving, or in the context of systems positivity-preserving. And we want to prevent the code from crashing (for instance avoiding NaN in the code). It is also essential for the corrected scheme to be conservative. Secondly, we would like to essentially avoid the appearance of spurious oscillations. To do so, as it is generally done, we will enforce a discrete maximum principle. Thirdly, we want to retain as much as possible the high accuracy and subcell resolution of DG schemes, by minimizing the number of subcells in which the solution has to be recomputed. Practically, we want the correction procedure to only modify the DG solution in troubled subcell regions without impacting the solution elsewhere in the cell. Finally, we want the whole procedure to be totally parameter free, and to behave properly from 2nd order to any order of accuracy.

To design such correction procedure, we first need to reformulate DG schemes as a subcell FV method provided the use of the correct numerical fluxes. This very simple development provides us with the so-called DG reconstructed flux. We will demonstrate that this theoretical part is consistent with the work presented in [29]. Let us also emphasize that the question of reformulating DG schemes into finite volume schemes have also been recently addressed by means of Residual Distribution (RD) schemes. Indeed, in a series of papers [50, 48, 51, 49], R. Abgrall and his co-authors managed to prove that almost any numerical scheme, as DG and flux reconstruction schemes for instance, can be recast as a residual distribution scheme. Reversely, in the very recent paper [49], it has been proved that RD schemes can be rewritten as a finite volume method, which permits consequently to formulate DG schemes as a finite volume method.

The DG reconstructed flux obtained in the first section will help us in correcting discontinuous Galerkin schemes. Practically, at each time step, we compute a DG candidate solution and check if this solution is admissible. If it is the case, we go further in time. Otherwise, we return to the previous time step and correct locally, at the subcell scale, the numerical solution. In the subcells where the solution was detected as bad, we substitute the DG reconstructed flux on the subcell boundaries by a robust first-order or second-order TVD numerical flux. And for subcells detected as admissible, we keep the high-order reconstructed flux which allows us to retain the very high accurate resolution and conservation of DG schemes. Consequently, only the solution inside troubled subcells and their first neighbors will have to be recomputed. Elsewhere, the solution remains unchanged. This correction procedure is then extremely local.

To present this *a posteriori* correction, the remainder of this paper is organized as follows: In Section 2, in the very simple case of one-dimensional scalar conservation laws, we briefly recall the derivation of discontinuous Galerkin scheme, to finally express it as a specific FV scheme on a subgrid. This theoretical section will allow us to introduce our correction technique in Section 3. Last, numerical results both in 1D and 2D on Cartesian grids, provided in Section 4, will demonstrate the effectiveness of the presented technique.

2. DG as a subcell Finite Volume scheme

This section is devoted to the recall of discontinuous Galerkin schemes and their equivalency with a finite volume method on a subgrid. To remain as simple as possible, one-dimensional Scalar Conservation Laws (SCL) will be considered. Let $u = u(x,t)$, for $x \in \omega \subset \mathbb{R}$, and $t \in [0,T]$, be the solution of the following system

$$\begin{cases} \dfrac{\partial u}{\partial t} + \dfrac{\partial F(u)}{\partial x} = 0, & (x,t) \in \omega \times [0,T], \quad (1a) \\ u(x,0) = u^0(x), & x \in \omega, \quad (1b) \end{cases}$$

where u^0 is the initial data and $F(u)$ is the flux function. For the subsequent discretization, let us introduce the following notation. Let $\{\omega_i\}_i$ be a partition of the computational domain ω. Here, $\omega_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]$ denotes a generic computational cell of size Δx_i. We also introduce a partition of the time domain $0 = t^0 < t^1 < ... < t^n < ... < t^N = T$ and the time step $\Delta t^n = t^{n+1} - t^n$. In order to obtain a $(k+1)^{\text{th}}$ order discretization, let us consider a piecewise polynomial approximated solution $u_h(x,t)$, where its restriction to cell ω_i, namely $u_{h|\omega_i} = u_h^i$, belongs to $\mathbb{P}^k(\omega_i)$ the set of polynomial of degree up to k. The numerical solution then writes

$$u_h^i(x,t) = \sum_{m=1}^{k+1} u_m^i(t)\, \sigma_m(x), \tag{2}$$

where $\{\sigma_m\}_m$ is a basis of $\mathbb{P}^k(\omega_i)$. The coefficients u_m^i present in (2) are the solution moments to be computed through a local variational formulation on ω_i. To this end, one has to multiply equation (1a) by $\psi \in \mathbb{P}^k(\omega_i)$, a polynomial test function, and integrate it on ω_i. By means of an integration by parts and substituting the solution u by its approximated polynomial counterpart u_h^i, one gets

$$\int_{\omega_i} \frac{\partial u_h^i}{\partial t}\, \psi \, \mathrm{d}x = \int_{\omega_i} F(u_h^i) \frac{\partial \psi}{\partial x} \, \mathrm{d}x - \Big[\mathcal{F}\, \psi\Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \forall\, \psi \in \mathbb{P}^k(\omega_i). \tag{3}$$

The terms $\int_{\omega_i} F(u_h^i) \frac{\partial \sigma_p}{\partial x} \, \mathrm{d}x$ and $[\mathcal{F}\, \sigma_p]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}$ are respectively referred to as volume and surface integrals. In (3), the numerical flux function $\mathcal{F}$, in addition to ensure the scheme conservation, is the cornerstone of any finite volume or DG scheme regarding fundamental considerations as stability, positivity and entropy among others. In the context of DG schemes, this numerical flux is defined as a function of the two states on the left and right of each interface, *i.e.* $\mathcal{F}_{i+\frac{1}{2}} = \mathcal{F}(u_h^i(x_{i+\frac{1}{2}}, t), u_h^{i+1}(x_{i+\frac{1}{2}}, t))$. This function is generally obtained through the resolution of an exact or approximated Riemann problem. In the remainder of this paper, for sake of simplicity, we make use of the very well-known local Lax-Friedrichs numerical flux which reads $\mathcal{F}(u,v) = \frac{1}{2}\left(F(u) + F(v) - \gamma(u,v)\,(v-u)\right)$, where $\gamma(u,v) = \max(|F'(u)|, |F'(v)|)$.

Lately, some new schemes which can be closely related to discontinuous Galerkin Spectral Element Method (DGSEM) have been introduced and have gained in popularity. These schemes, firstly introduced in the context of finite difference by means of Summation By Parts (SBP) operators and Simultaneous Approximation Term (SAT), are now generally referred to as entropy stable schemes, see for instance [20, 21, 5, 23, 7]. In [20], the authors found a remarkable equivalence of general diagonal norm high-order summation-by-parts operators to a subcell based finite volume formulation. This equivalence allows them to construct provably entropy stable schemes by a specific choice of

the subcell finite volume fluxes. It also demonstrates the scheme conservation at the subcell scale. This subcell finite volume formulation is remarkable in the sense that one can directly impact on the scheme properties, as entropy stability for instance, by choosing the proper subcell finite volume fluxes. Let us now present a similar subcell finite volume formulation for general DG schemes.

For the following proof, in equation (3) we need to substitute in the volume integral the exact interior flux function $F(u_h^i)$ with some polynomial approximation F_h^i. To this end, we define $F_h^i \in \mathbb{P}^{\,\alpha}(\omega_i)$, where $\alpha \in \mathbb{N}^*$, as the L_2 projection of function $F(u_h^i)$ onto $\mathbb{P}^{\,\alpha}(\omega_i)$ as follows

$$\int_{\omega_i} F_h^i \, \psi \, \mathrm{d}x = \int_{\omega_i} F(u_h^i) \, \psi \, \mathrm{d}x, \qquad \forall \, \psi \in \mathbb{P}^{\,\alpha}(\omega_i). \tag{4}$$

And as long as $\alpha \geq k-1$, and that the volume integral in the right-hand side of the L_2 projection (4) is computed similarly to the volume integral in DG schemes (3), namely by an exact integration or by the same quadrature rule for instance, the scheme (3) rewrites

$$\int_{\omega_i} \frac{\partial \, u_h^i}{\partial t} \, \psi \, \mathrm{d}x = \int_{\omega_i} F_h^i \, \frac{\partial \, \psi}{\partial x} \, \mathrm{d}x - \Big[\mathcal{F} \, \psi \Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \forall \, \psi \in \mathbb{P}^k(\omega_i). \tag{5}$$

In DG schemes, volume integrals are generally computed by means of a quadrature rule. And for the purpose of accuracy, it is possible to demonstrate that to design a $(k+1)^{\text{th}}$ order numerical scheme, a quadrature rule exact at least for polynomial up to degree $2k$ is required, see [9].

Let us note that DG schemes formulation (5) also holds for the so-called collocated and nodal DG. In those methods, the numerical solution $u_h^i(x,t)$, defined by means of $k+1$ point values $\{u_m^i\}_m$, writes $u_h^i(x,t) = \sum_{m=1}^{k+1} u_m^i \, L_m^i(x)$, where $L_m^i(x)$ are the Lagrangian basis functions associated to the solution points. And similarly, the approximated flux function $F_h^i \in \mathbb{P}^k(\omega_i)$ is developed onto the same basis as $F_h^i(x,t) = \sum_{m=1}^{k+1} F(u_m^i) \, L_m^i(x)$, where $F(u_m^i)$ is simply the analytical flux function apply to the solution point value u_m^i.

Now, through analytical integration or provided with the chosen quadrature rule, the volume integral in (5) is computed exactly as $F_h^i \, \frac{\partial \, \psi}{\partial x} \in \mathbb{P}^{\,\alpha+k-1}$. We can then perform an integration by part and get what is generally referred to as the strong form of DG schemes

$$\int_{\omega_i} \frac{\partial \, u_h^i}{\partial t} \, \psi \, \mathrm{d}x = - \int_{\omega_i} \frac{\partial \, F_h^i}{\partial x} \, \psi \, \mathrm{d}x + \Big[(F_h^i - \mathcal{F}) \, \psi \Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \forall \, \psi \in \mathbb{P}^k(\omega_i). \tag{6}$$

Remark 2.1. *We will see in the remainder that the following development holds for a polynomial flux F_h^i of degree $\alpha \leq k+1$. We have then the condition $\alpha \in [\![k-1, k+1]\!]$. For instance, by means of analytical integration or if one wants to use a quadrature rule exact for polynomials up to $2k+2$ in order to reduce aliasing effect, equation* (6) *and the remaining demonstration would still be true with F_h^i the L_2 projection onto $\mathbb{P}^{k+1}(\omega_i)$ of function $F(u_h^i)$. Alternatively, for a collocated flux approximation, one could use $k+2$ collocation points, as for instance the flux points introduced below. Because the more general case covered by the following development is $F_h^i \in \mathbb{P}^{k+1}(\omega_i)$, the polynomial flux F_h^i will then be assumed to be of degree $k+1$ in the following. Obviously, all these considerations vanish in the linear case, as $F_h^i = F(u_h^i) \in \mathbb{P}^k(\omega_i)$.*

That being said, let us introduce the subcell decomposition of cell ω_i. Let $\{\widetilde{x}_{m+\frac{1}{2}}\}_{m=0,\dots,k+1}$ be the k+2 flux points. These points allow us to defined $\{S_m^i\}_m$, the $k+1$ subcells as $S_m^i = [\widetilde{x}_{m-\frac{1}{2}}, \widetilde{x}_{m+\frac{1}{2}}]$ for $m = 1, \dots, k+1$, see Figure 1.

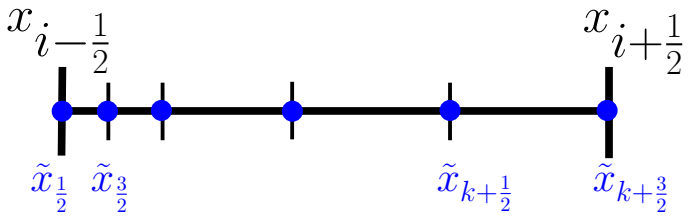


Figure 1: Subcell decomposition of ω_i through $k+2$ flux points.

Remark 2.2. *Please note that these flux points can be chosen totally arbitrarily, and do not have to be related to any quadrature rule nor diagonal norm matrix as it is generally the case in entropy stable schemes [20, 21]. Furthermore, these flux points do not depend on the choice of the DG basis functions. Let us insist on the fact that any cell subdivision would lead to exactly the same theoretical result, namely the same DG scheme. It would only affect the definition of the subcell finite volume numerical fluxes defined in the remainder of the present development. However, it does have an impact of the correction procedure, see Section 4.*

Now, the critical step of this demonstration is the introduction of some very specific basis functions that we refer from now on to as subresolution basis functions. These particular functions can be seen as the L_2 projection onto $\mathbb{P}^k(\omega_i)$ of the subcell indicator functions $\mathbb{1}_m(x)$, which are equal to one if $x \in S_m^i$ and zero otherwise. Let then $\{\phi_m\}_{m=1,\dots,k+1}$ be the $\mathbb{P}^k(\omega_i)$ basis functions defined such that

$$\int_{\omega_i} \phi_m \, \psi \, \mathrm{d}x = \int_{S_m^i} \psi \, \mathrm{d}x, \qquad \forall \, \psi \in \mathbb{P}^k(\omega_i). \tag{7}$$

These $k+1$ conditions provide us with an easy to solve linear system, see Appendix A for explicit formula. Furthermore, it is straightforward to prove condition (7) enforces that

$$\sum_{m=1}^{k+1} \phi_m(x) = 1, \qquad \forall \, x \in \omega_i. \tag{8}$$

Now, because equation (6) holds for any polynomial function ψ of degree k, let us substitute ϕ_m for ψ in DG schemes

$$\int_{\omega_i} \frac{\partial \, u_h^i}{\partial t} \, \phi_m \, \mathrm{d}x = - \int_{\omega_i} \frac{\partial \, F_h^i}{\partial x} \, \phi_m \, \mathrm{d}x + \Big[(F_h^i - \mathcal{F}) \, \phi_m \Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \text{for} \;\; m = 1, \dots, k+1.$$

Recalling that both $\frac{\partial \, u_h^i}{\partial t}$ and $\frac{\partial \, F_h^i}{\partial x}$ belong to $\mathbb{P}^k(\omega_i)$, by means of condition (7) it follows that

$$\frac{\partial \, \overline{u}_m^i}{\partial t} = -\frac{1}{|S_m^i|} \left(\Big[F_h^i \Big]_{\widetilde{x}_{m-\frac{1}{2}}}^{\widetilde{x}_{m+\frac{1}{2}}} - \Big[\phi_m \, \left(F_h^i - \mathcal{F} \right) \Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \right), \tag{9}$$

where $|S_m^i| = |\widetilde{x}_{m+\frac{1}{2}} - \widetilde{x}_{m-\frac{1}{2}}|$ is the subcell size, and $\overline{u}_m^i$ stands for the subcell mean value

$$\overline{u}_m^i = \frac{1}{|S_m^i|} \int_{S_m^i} u_h^i \, \mathrm{d}x. \tag{10}$$

The final step is the introduction of the $k+2$ subcell finite volume fluxes $\{\widehat{F}^i_{m+\frac{1}{2}}\}_{m=0,\dots,k+1}$, from now on referred to as reconstructed fluxes, located at the $k+2$ flux points. These reconstructed fluxes are defined through the following linear system

$$\widehat{F}^i_{m+\frac{1}{2}} - \widehat{F}^i_{m-\frac{1}{2}} = \Big[F_h^i\Big]_{\widetilde{x}_{m-\frac{1}{2}}}^{\widetilde{x}_{m+\frac{1}{2}}} - \Big[\phi_m \, \big(F_h^i - \mathcal{F}\big)\Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \text{for } \; m = 1, \dots, k+1, \tag{11a}$$

$$\widehat{F}^i_{\frac{1}{2}} = \mathcal{F}_{i-\frac{1}{2}} \qquad \text{and} \qquad \widehat{F}^i_{k+\frac{3}{2}} = \mathcal{F}_{i+\frac{1}{2}}. \tag{11b}$$

This linear system is straightforward to solve. Indeed, substituting subscript m by p in (11a) and summing for p from 1 to m leads to

$$\widehat{F}^i_{m+\frac{1}{2}} = F_h^i(\widetilde{x}_{m+\frac{1}{2}}) - \Big(1 - \sum_{p=1}^{m} \phi_p(x_{i-\frac{1}{2}})\Big)\Big(F_h^i(x_{i-\frac{1}{2}}) - \mathcal{F}_{i-\frac{1}{2}}\Big) - \Big(\sum_{p=1}^{m} \phi_p(x_{i+\frac{1}{2}})\Big)\Big(F_h^i(x_{i+\frac{1}{2}}) - \mathcal{F}_{i+\frac{1}{2}}\Big).$$

One finally gets the following values for the reconstructed fluxes

$$\widehat{F}^i_{m+\frac{1}{2}} = F_h^i(\widetilde{x}_{m+\frac{1}{2}}) - C^{i-\frac{1}{2}}_{m+\frac{1}{2}}\Big(F_h^i(x_{i-\frac{1}{2}}) - \mathcal{F}_{i-\frac{1}{2}}\Big) - C^{i+\frac{1}{2}}_{m+\frac{1}{2}}\Big(F_h^i(x_{i+\frac{1}{2}}) - \mathcal{F}_{i+\frac{1}{2}}\Big), \tag{12}$$

where $C^{i\pm\frac{1}{2}}_{m+\frac{1}{2}}$, the correction coefficients, are defined by means of relation (8) as

$$C^{i-\frac{1}{2}}_{m+\frac{1}{2}} = \sum_{p=m+1}^{k+1} \phi_p(x_{i-\frac{1}{2}}) \qquad \text{and} \qquad C^{i+\frac{1}{2}}_{m+\frac{1}{2}} = \sum_{p=1}^{m} \phi_p(x_{i+\frac{1}{2}}). \tag{13}$$

These reconstructed fluxes (12) are nothing but the interior polynomial flux $F_h^i(\widetilde{x}_{m+\frac{1}{2}})$ with some correction terms taking into account the difference between the boundary values of this interior flux and the numerical fluxes. Furthermore, making use of condition (8), it is obvious that the boundary correction coefficients ensure relations (11b) since

$$C^{i-\frac{1}{2}}_{\frac{1}{2}} = 1 \qquad \text{and} \qquad C^{i-\frac{1}{2}}_{k+\frac{3}{2}} = 0,$$
$$C^{i+\frac{1}{2}}_{\frac{1}{2}} = 0 \qquad \text{and} \qquad C^{i+\frac{1}{2}}_{k+\frac{3}{2}} = 1.$$

For sake of conciseness, we restrict ourselves to a symmetric distribution of the flux points $\{\widetilde{x}_{m+\frac{1}{2}}\}_m$ around the cell center $x_i = \frac{1}{2}(x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}})$, namely $\widetilde{x}_{m+\frac{1}{2}} = (x_{i-\frac{1}{2}} + x_{i+\frac{1}{2}}) - \widetilde{x}_{k+\frac{3}{2}-m}$. It immediately follows that $C^{i+\frac{1}{2}}_{m+\frac{1}{2}} = C^{i-\frac{1}{2}}_{k+\frac{3}{2}-m}$, for $m = 0, \dots, k+1$. Let us then focus on the expression of $C^{i-\frac{1}{2}}_{m+\frac{1}{2}}$. By means of the subresolution basis functions definition, see Appendix A, we are able to explicitly express the $k+2$ correction coefficients. We set $\boldsymbol{B} \in \mathbb{R}^{k+1}$ to be the vector defined as

$$B_j = (-1)^{j+1} \binom{k+j}{j} \binom{k+1}{j},$$

where $\binom{p}{j}$ stands for the binomial coefficient $\binom{p}{j} = \dfrac{p!}{j!\,(p-j)!}$. Let us note that vector $\boldsymbol{B}$ only depends on the degree of approximation k, and not on the flux points position. By introducing $\{\widetilde{\xi}_{m+\frac{1}{2}}\}_m$ the flux points counterpart in the referential element $[0,1]$, as $\widetilde{\xi}_{m+\frac{1}{2}} = \frac{\widetilde{x}_{m+\frac{1}{2}} - x_{i-\frac{1}{2}}}{x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}}$, the correction coefficients finally write

$$C_{m+\frac{1}{2}}^{i-\frac{1}{2}} = 1 - \begin{pmatrix} \widetilde{\xi}_{m+\frac{1}{2}} \\ (\widetilde{\xi}_{m+\frac{1}{2}})^2 \\ \vdots \\ (\widetilde{\xi}_{m+\frac{1}{2}})^{k+1} \end{pmatrix} \cdot \boldsymbol{B}. \tag{14}$$

For further details of calculation, we refer to Appendix A. Let us gather all these results into the following theorem.

Theorem 2.1. *Provided the analytical calculation of volume integrals, or alternatively by means of quadrature rule, discontinuous Galerkin schemes expressed in cell ω_i as follows*

$$\int_{\omega_i} \frac{\partial\, u_h^i}{\partial t}\, \psi \,\mathrm{d}x = \int_{\omega_i} F(u_h^i) \frac{\partial\, \psi}{\partial x} \,\mathrm{d}x - \Big[\mathcal{F}\,\psi\Big]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}}, \qquad \forall\, \psi \in \mathbb{P}^k(\omega_i), \tag{15}$$

can be recast into $k+1$ subcell finite volume schemes as, for $m = 1, \ldots, k+1$

$$\frac{\partial\, \overline{u}_m^i}{\partial t} = -\frac{1}{|S_m^i|} \left(\widehat{F}^i_{m+\frac{1}{2}} - \widehat{F}^i_{m-\frac{1}{2}} \right), \tag{16}$$

where the $k+2$ finite volume fluxes $\widehat{F}^i_{m+\frac{1}{2}}$, referred to as reconstructed fluxes, are defined by

$$\widehat{F}^i_{m+\frac{1}{2}} = F_h^i(\widetilde{x}_{m+\frac{1}{2}}) - C_{m+\frac{1}{2}}^{i-\frac{1}{2}} \left(F_h^i(x_{i-\frac{1}{2}}) - \mathcal{F}_{i-\frac{1}{2}} \right) - C_{m+\frac{1}{2}}^{i+\frac{1}{2}} \left(F_h^i(x_{i+\frac{1}{2}}) - \mathcal{F}_{i+\frac{1}{2}} \right). \tag{17}$$

In this last expression, for $\alpha \in [\![k-1, k+1]\!]$, the polynomial flux F_h^i is either a L_2 projection of $F(u_h^i)$ onto $\mathbb{P}^\alpha(\omega_i)$, or collocated at $(\alpha+1)$ given points, as it is the case for collocated and nodal DG schemes. Simple explicit expression of the correction coefficients can be found in equation (14) *or in Appendix A.*

This theorem allows us to rewrite DG schemes into subcell finite volume schemes provided with the corrected subcell fluxes. Furthermore, it proves the subcell conservation property of DG schemes.

Let us note that in this work, the $k+2$ subcell finite volume fluxes $\widehat{F}^i_{m+\frac{1}{2}}$ are named the reconstructed fluxes. We have borrowed this denomination from H. Huynh seminal paper [29] wherein he introduces a wide range of new numerical schemes referred to as Flux Reconstruction (FR) schemes which covered DG schemes among others. Furthermore, Theorem 2.1 is consistent with the result of the aforementioned paper of H. Huynh. Indeed, through the $k+2$ reconstructed fluxes $\widehat{F}^i_{m+\frac{1}{2}}$, one can define a $\mathbb{P}^{k+1}(\omega_i)$ polynomial reconstructed flux $\widehat{F}_h^i = \sum_{m=0}^{k+1} \widehat{F}^i_{m+\frac{1}{2}}\, \widetilde{L}_{m+\frac{1}{2}}(x)$, where $\widetilde{L}_{m+\frac{1}{2}}(x)$ are the Lagrangian basis functions associated to the $k+2$ flux points $\widetilde{x}_{m+\frac{1}{2}}$. By means of the subresolution functions defined in (7), equations (16) can be recast into $\partial_t u_h^i(x,t) + \partial_x \widehat{F}_h^i(x,t) = 0,\ \forall x \in \omega_i$.

In the end, by introducing $k+1$ solution points $\{x_m^i\}_{m=1,\dots,k+1}$ in cell ω_i, the numerical scheme reduces to the following very simple expression

$$\frac{\partial\, u_h^i(x_m,t)}{\partial t} + \frac{\partial\, \widehat{F}_h^i(x_m,t)}{\partial x} = 0, \qquad \text{for } \; m = 1,\dots,k+1, \tag{18}$$

where the numerical solution is evaluating pointwisely at some solution points through the computation of the spatial derivative of the reconstructed flux. To use consistent notation with [29], let us rewrite this reconstructed flux as

$$\widehat{F}_h^i(x,t) = F_h^i(x,t) + \left(\mathcal{F}_{i-\frac{1}{2}} - F_h^i(x_{i-\frac{1}{2}})\right) g_{LB}(x) + \left(\mathcal{F}_{i+\frac{1}{2}} - F_h^i(x_{i+\frac{1}{2}})\right) g_{RB}(x), \tag{19}$$

where $g_{LB}(x)$ and $g_{RB}(x)$ are respectively the left and right correction polynomial functions taking into account the flux discontinuities, and are defined through the correction coefficients introduced previously as $g_{LB}(x) = \sum_{m=0}^{k+1} C_{m+\frac{1}{2}}^{i-\frac{1}{2}}\, \widetilde{L}_{m+\frac{1}{2}}(x)$ and $g_{RB}(x) = \sum_{m=0}^{k+1} C_{m+\frac{1}{2}}^{i+\frac{1}{2}}\, \widetilde{L}_{m+\frac{1}{2}}(x)$. Through a simple analysis, it is possible to prove that g_{RB} is nothing but the left $\mathbb{P}^{k+1}$ Radau polynomial, while g_{LB} is the right $\mathbb{P}^{k+1}$ Radau polynomial, which is perfectly consistent with the results presented in [29]. Let us briefly recall what flux reconstruction schemes are. Recently, H. Huynh [29, 30] has introduced a new approach referred to as flux reconstruction which unifies several existing schemes. The collocated $\mathbb{P}^k$ interior flux F_h^i is corrected, similarly to equation (19), to take into account the flux discontinuities. With appropriate choices of correction terms, one can recover collocated nodal DG, spectral difference scheme, as well as spectral volume method. Furthermore, the FR versions are generally simpler and more economical than the original versions, due to the fact that the numerical solution is simply advance in time pointwisely at some solution points through the computation of the reconstructed flux spatial derivative. It has also paved the way to a wide range of new numerical schemes that are stable and super convergent. This framework has recently grown in popularity, see [60, 1, 35, 25], and is now sometimes referred to as Correction Procedure via Reconstruction (CPR), as in [61, 22, 31, 16, 32] for instance. Further stability analysis have been carried out through the reformulation of CPR methods using SBP operators with SAT boundary treatment, as in [26, 27]. In [27], for Burgers equation, SBP CPR methods are further extended to the non-diagonal norm matrix case, hence covering the case of modal basis. In [51], entropy stability of flux reconstruction schemes has also been addressed by means of residual distribution schemes.

The main differences between the aforementioned flux reconstruction approach and the results stated in Theorem 2.1 are that the reconstructed fluxes are used here as numerical fluxes for the subcell finite volumes schemes, and not pointwisely to advance in time the numerical solution point values. Discontinuous Galerkin scheme then can be reinterpreted as a finite volume scheme on subcells with a particular definition of the numerical fluxes to be used. This demonstration is not restricted to the frame of flux collocation, and covers the case of general DG schemes wherein integrals are either analytically calculated or approached through the use of quadrature rules. Furthermore, the interior polynomial counterpart of $F(u_h^i)$, namely F_h^i, can be assumed more generally to belong to $\mathbb{P}^{k+1}(\omega_i)$, which can potentially reduce aliasing effect. Finally, the correction coefficients are simple and explicitly defined, without any need of Radau polynomials.

As said in the introduction, the question of reformulating DG schemes as a subcell finite volume method has also been recently addressed by R. Abgrall through residual distribution schemes, [49]. The present theoretical result can then be seen as an alternative simple development, with

explicit formula for any order of accuracy, in the one-dimensional case. As we plan to extend this demonstration, along with the corresponding correction procedure, to the 2D unstructured grid case, we can expect to get consistent results with the aforementioned paper.

3. *A posteriori* subcell limitation

By means of Theorem 2.1, we have now all the tools we need to design a subcell *a posteriori* limitation for DG schemes. In few words, the reconstructed fluxes $\widehat{F}^i_{m+\frac{1}{2}}$ will be modified in a robust way in subcells where the unlimited DG scheme has failed. Let us mention that until now, only the semi-discrete version of schemes and their corresponding analysis were presented. To achieve high-accuracy in time, we make use of SSP Runge-Kutta time integration method [54]. But, in the light of the fact that these multistage time integration methods write as convex combinations of first-order forward Euler scheme, the correction DG procedure will be presented for the simple case of this latter time numerical scheme, for sake of simplicity. DG schemes (3) provided with first-order forward Euler time integration writes, $\forall\, \psi \in \mathbb{P}^k(\omega_i)$

$$\int_{\omega_i} u_h^{i,n+1}\, \psi \,\mathrm{d}x = \int_{\omega_i} u_h^{i,n}\, \psi \,\mathrm{d}x + \Delta t \left(\int_{\omega_i} F(u_h^{i,n}) \frac{\partial\, \psi}{\partial x} \,\mathrm{d}x - \left[\mathcal{F}^n\, \psi \right]_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \right). \tag{20}$$

Now, the numerical solution u_h^i on cell ω_i being assumed to be a $\mathbb{P}^k$ polynomial as follows

$$u_h^i(x,t) = \sum_{m=1}^{k+1} u_m^i(t)\, \sigma_m(x), \tag{21}$$

with $\{\sigma_m\}_m$ any basis of $\mathbb{P}^k(\omega_i)$, we can show that this polynomial solution is uniquely defined through its mean values on $k+1$ subcells, and reversely. It is actually very simple to go from one representation to another, namely from $\{u_m^i\}_m$, the solution moments, to $\{\overline{u}_m^i\}_m$, the subcell mean values. Indeed, defining the non-singular matrix $\mathbf{\Pi}$ such that

$$\pi_{mp} = \frac{1}{|S_m^i|} \int_{S_m^i} \sigma_p \,\mathrm{d}x, \tag{22}$$

it immediately follows

$$\mathbf{\Pi} \begin{pmatrix} u_1^{i,n} \\ \vdots \\ u_{k+1}^{i,n} \end{pmatrix} = \begin{pmatrix} \overline{u}_1^{i,n} \\ \vdots \\ \overline{u}_{k+1}^{i,n} \end{pmatrix}. \tag{23}$$

Remark 3.1. *This can be related to the concept of histopolant, and histopolation basis functions, introduced in [53, 24]. Obviously, if one makes use of these particular histopolation basis functions in the DG scheme, this projection step can be skipped.*

From now on, we consider that, through relation (23), we have access to the solution submean values, and that by means of the submean values we can also reconstruct the unique associated polynomial, as displayed in Figure 2. We have now all the tools we need to introduce the correction procedure. First, we assume that at time t^n the numerical solution u_h^n is satisfactory in the sens that, on any cell ω_i, the subcell mean values are admissible regarding some criteria yet to be defined. Then, we compute u_h^{n+1} a candidate solution through the unlimited DG scheme. The third step is

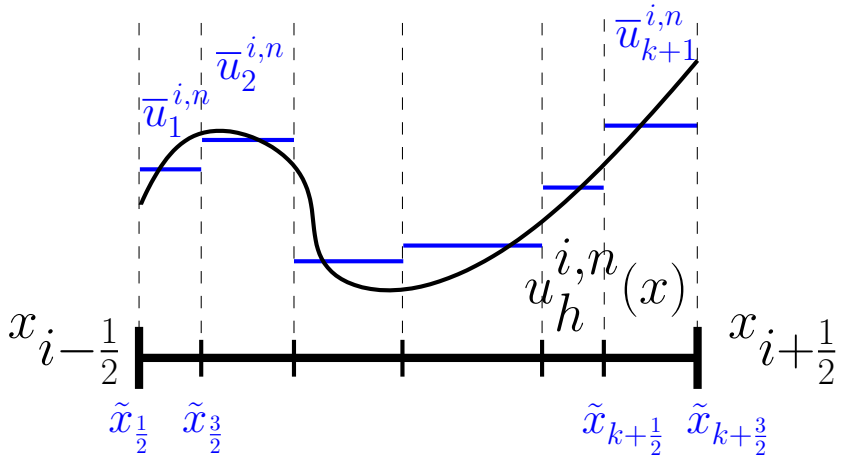


Figure 2: Polynomial solution and its associated submean values.

crucial. Indeed, we then have to check if the new unlimited solution is admissible. If it is the case, we can go further in time without any special treatment, otherwise we have to return to time t^n and recompute the solution locally by means of a more robust scheme. This step is crucial in the sens that it will tell us if and where a new computation would be required.

3.1. Troubled zone detector

As said in the introduction, there is a very wide panel of already existing limitation techniques available in the literature. This also applies to troubled zone detectors. The present work focuses on the treatment of problematic cells and not on the detection of such cells. We have consequently compared many detection techniques and finally kept the ones that seem to suit the most this *a posteriori* subcell context.

Similarly to other *a posteriori* techniques [14, 18], we mainly make use of two detection criteria, namely one ensuring the physical admissibility of the numerical solution (PAD) and another addressing the apparition of spurious oscillations (NAD). We use the same denomination that in the abovementioned papers. Let us then recall these two criteria.

Physical admissibility detection (PAD).

- Check if the different mean values $\overline{u}_m^{i,n+1}$ lie in a chosen convex physical admissible set (maximum principle for SCL, positivity of the pressure and density for Euler, ...). Entropy stability can even be added to this admissible set.
- Check if there is any NaN values

Those are the minimum requirements if one wants to enforce code robustness. Now, in order to tackle the issue of spurious oscillations, we make use of a local maximum principle. Indeed, through the respect of the CFL, the solution in cell ω_i at time t^{n+1} has to remain in the bounds of the solution at the previous time step t^n wherein $\bigcup_{j=i-1}^{i+1}\omega_j$. This condition is reformulated in the following detection criterion.

Numerical admissibility detection (NAD).

- Check if the following Discrete Maximum Principle (DMP) is ensured:

$$\min_{p\in[\![1,k+1]\!]}(\overline{u}_p^{i-1,n},\overline{u}_p^{i,n},\overline{u}_p^{i+1,n}) \leq \overline{u}_m^{i,n+1} \leq \max_{p\in[\![1,k+1]\!]}(\overline{u}_p^{i-1,n},\overline{u}_p^{i,n},\overline{u}_p^{i+1,n}),$$

  for $m = 1,\dots,k+1$.

This NAD criterion will enable us to address the issue of spurious oscillations at the cell level. However, as we will see in the numerical results section, this local maximum principle will not be triggered by spurious oscillation within the cell itself, namely at the subcell scale. These subcell oscillations, while present, are controlled by the NAD. They thus cannot lead to instabilities. However, if one is concerned with these subcell oscillations, one can make use of the following subcell discrete maximum principle, which it is worth to be said has no physical meaning in the context of DG schemes on cell ω_i, but only for subcell numerical scheme on S_m^i.

Subcell numerical admissibility detection (SubNAD).

- Check if for $m = 1,\dots,k+1$:

$$\min(\overline{u}_{m-1}^{i,n},\overline{u}_m^{i,n},\overline{u}_{m+1}^{i,n}) \leq \overline{u}_m^{i,n+1} \leq \max(\overline{u}_{m-1}^{i,n},\overline{u}_m^{i,n},\overline{u}_{m+1}^{i,n}),$$

  where $\overline{u}_0^{i,n} = \overline{u}_{k+1}^{i-1,n}$ and $\overline{u}_{k+2}^{i,n} = \overline{u}_1^{i+1,n}$.

Let us enlighten that both NAD and SubNAD criteria rely on a maximum principle based on subcell mean values. And because these maximum principles are not concerned with the whole polynomial set of values, it is very well-known that one has to relax them to preserve scheme accuracy in the presence of smooth extrema. After numerous comparison, here is the chosen detection procedure responsible of detecting smooth extrema.

Detection of smooth extrema. This smooth extrema detection criterion is based on an idea of Van Leer [42], which has then been generalized in [4]. This is also the criterion used in the generalized moment limiter [62], as well as in the hierarchical slope limiter [40]. In all these aforementioned limitation techniques, the numerical solution is supposed to exhibit a smooth extrema if at least the linearized version of the numerical solution spatial derivative, *i.e.*

$$v_h(x) = \overline{\partial_x u_i}^{n+1} + (x - x_i)\,\overline{\partial_{xx} u_i}^{n+1}, \tag{24}$$

presents a monotonous profile. In (24), $\overline{\partial_x u_i}^{n+1}$ and $\overline{\partial_{xx} u_i}^{n+1}$ are nothing but the mean values on cell ω_i respectively of $\partial_x u_h^i$ and $\partial_{xx} u_h^i$. In practice, the DMP relaxation used here works as a vertex-based limiter on $v_h(x)$. Hence, we set $v_L = \overline{\partial_x u_i}^{n+1} - \frac{\Delta x_i}{2}\,\overline{\partial_{xx} u_i}^{n+1}$ to be the left boundary value of $v_h(x)$ on cell ω_i, as well as $v^L_{\min\setminus\max} = \min\setminus\max\,(\overline{\partial_x u_{i-1}}^{n+1}, \overline{\partial_x u_i}^{n+1})$ the minimum and maximum values of the mean derivative around $x_{i-\frac{1}{2}}$. We then define the left detection factor α_L as following

$$\alpha_L = \begin{cases} \min(1, \dfrac{v^L_{\max} - \overline{\partial_x u_i}^{n+1}}{v_L - \overline{\partial_x u_i}^{n+1}}), & \text{if } v_L > \overline{\partial_x u_i}^{n+1}, \\ 1, & \text{if } v_L = \overline{\partial_x u_i}^{n+1}, \\ \min(1, \dfrac{v^L_{\min} - \overline{\partial_x u_i}^{n+1}}{v_L - \overline{\partial_x u_i}^{n+1}}), & \text{if } v_L < \overline{\partial_x u_i}^{n+1}. \end{cases} \tag{25}$$

Defining $v_R = \overline{\partial_x u_i}^{n+1} + \frac{\Delta x_i}{2}\overline{\partial_{xx} u_i}^{n+1}$ and $v^R_{\min\setminus\max} = \min\setminus\max(\overline{\partial_x u_i}^{n+1}, \overline{\partial_x u_{i+1}}^{n+1})$, the right detection factor α_R is obtained in a similar manner than in (25). Finally, taking the minimum of the two, *i.e.* $\alpha = \min(\alpha_L, \alpha_R)$, we consider that the numerical solution presents a smooth profile on cell ω_i if $\alpha = 1$. In this particular case, the NAD or SubNAD criterion is relaxed allowing the preservation of smooth extrema along with the order of accuracy for smooth problems, see Section 4. We have presented here the detection based on the linearized first derivative of the solution. This would work for any higher order derivative. Actually, in practice, if any of them presents a monotonous profile, the DMP is relaxed. Also, for numerical schemes of accuracy higher than third order, such relaxation procedure can also be applied at the subcell level to also preserve smooth extrema even within a cell. For schemes from first to third order, the subcell version of $v_h(x)$, namely $v_h^m(x) = \overline{\partial_x u}_m^{i,n+1} + (x - x_m)\,\overline{\partial_{xx} u}_m^{i,n+1}$ where $x_m = \frac{1}{2}(\widetilde{x}_{m-\frac{1}{2}} + \widetilde{x}_{m+\frac{1}{2}})$, is a continuous linear function on cell ω_i. We thus lack degrees of freedom to detect a smooth extrema contained in a cell. This technique will thus be used only for $k > 3$. Now that we have described the troubled subcell detector, the correction procedure will be presented in the next subsection.

3.2. Correction

The very simple idea that forms the basis of this correction procedure is the following: if the unlimited DG scheme has produced a numerical solution $u_h^{i,n+1}$ on cell ω_i, which is not admissible in subcell S_m^i in regards to the detection criteria presented previously, the subcell mean value $u_m^{i,n+1}$ will be recomputed by means of a more robust scheme. To do so, and because unlimited DG scheme is equivalent to subcell finite volume scheme with the appropriate high-order reconstructed fluxes, see Theorem 2.1, we substitute on the boundaries of subcell S_m^i the high-order reconstructed fluxes with some first-order finite volume numerical fluxes. The submean value $u_m^{i,n+1}$ will then be recomputed by means of a simple and robust first-order finite volume scheme. This concept is depicted in Figure 3, where the troubled subcell is colored red.

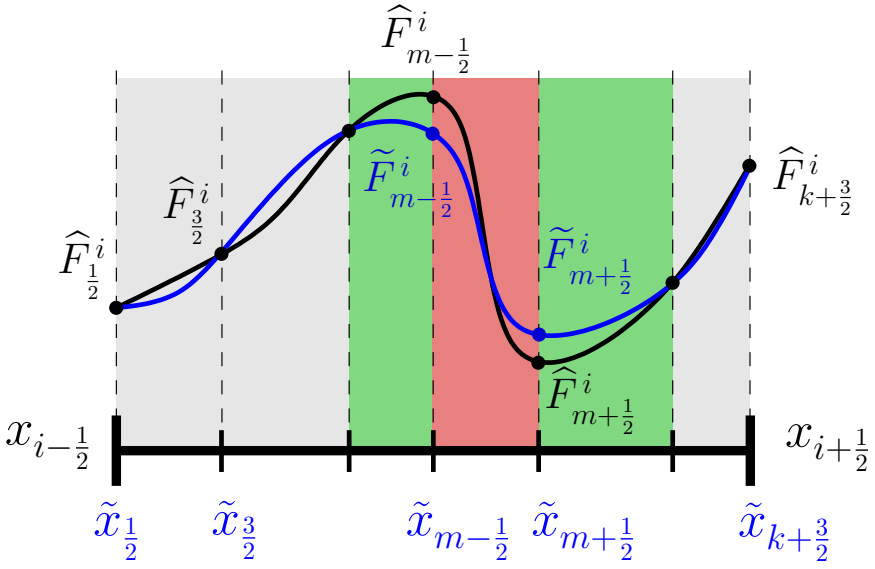


Figure 3: Correction of the reconstructed flux.

Because it is of fundamental importance to preserve scheme conservation, the first left and right neighboring subcells, colored green in Figure 3, have to be also recomputed since we have modify

the reconstructed fluxes $\widehat{F}^i_{m-\frac{1}{2}}$ and $\widehat{F}^i_{m+\frac{1}{2}}$. The submean values $\overline{u}^{i,n+1}_{m-1}$ and $\overline{u}^{i,n+1}_{m+1}$ are then computed through a finite-volume-like scheme with one first-order numerical flux and one high-order reconstructed flux. For the remaining subcells, colored gray in Figure 3, because the corresponding reconstructed fluxes have not been modified, there is no need to recompute them. The corresponding submean values are hence the values obtained through the unlimited DG scheme. This *a posteriori* limitation of DG schemes is summarized in the following flowchart:

1. Compute the candidate solution $u_h^{i,n+1}$ by means of unlimited DG schemes (20)
2. Project $u_h^{i,n+1}$ through (23) to get all the submean values $\overline{u}_m^{i,n+1}$
3. Check $\overline{u}_m^{i,n+1}$ through the troubled zone detection criteria plus relaxation
4. If $\overline{u}_m^{i,n+1}$ is admissible go further in time, otherwise modify the corresponding reconstructed flux values through a first-order numerical flux as following

$$\begin{cases} \widetilde{F}^i_{m+\frac{1}{2}} = \mathcal{F}(\overline{u}_m^{i,n}, \overline{u}_{m+1}^{i,n}) & \text{if } S_m^i \text{ or } S_{m+1}^i \text{ are marked,} \\ \widetilde{F}^i_{m+\frac{1}{2}} = \widehat{F}^i_{m+\frac{1}{2}} & \text{otherwise,} \end{cases}$$

   where $S_0^i = S_{k+1}^{i-1}$ and $S_{k+2}^i = S_1^{i+1}$, as well as $\overline{u}_0^{i,n} = \overline{u}_{k+1}^{i-1,n}$ and $\overline{u}_{k+2}^{i,n} = \overline{u}_1^{i+1,n}$.
5. Through the corrected reconstructed flux, recompute the submean values for tagged subcells and their first neighboring subcells as

$$\overline{u}_m^{i,n+1} = \overline{u}_m^{i,n} - \frac{\Delta t}{|S_m^i|}(\widetilde{F}^i_{m+\frac{1}{2}} - \widetilde{F}^i_{m-\frac{1}{2}}).$$

6. By means of $\mathbf{\Pi}^{-1}$ and equation (23), get the new corrected polynomial solution $u_h^{i,n+1}$
7. Return to point 3.

In light of this correction procedure flowchart, it is clear that the DG solution will only be affected at the subcell scale. Furthermore, the limited scheme is conservative at the subcell level by construction.

Remark 3.2. *In the present correction, when needed we substitute the reconstructed flux value* $\widehat{F}^i_{m+\frac{1}{2}}$ *with a first-order numerical flux* $\mathcal{F}(\overline{u}_m^{i,n}, \overline{u}_{m+1}^{i,n})$. *Obviously, other choices are possible and may even be more appropriate, as for instance a second-order TVD numerical flux or even a WENO numerical flux. We have presented the case of first-order correction for sake of simplicity. Some results with second-order corrections will also be presented in the numerical results section. Practically, for the 2nd-order correction, if a subcell is marked as bad, the subcell mean value would be recomputed using the former submean value along with the left and right subcell mean values (which can potentially belong to the left or right DG cell) through a 2nd-order FV scheme with a classical TVD minmod limiter. This means that the reconstructed flux values on the subcell boundaries are substituted by 2nd-order TVD numerical fluxes, instead of first-order FV numerical fluxes for the 1st-order correction. For higher-order FV or (W)ENO correction, a wider stencil including more neighboring subcells would be needed to compute the corrected reconstructed flux.*

The multi-dimensional extension of the theoretical reformulation of DG schemes, as well as the corresponding limitation technique, is carried out in a 1D tensor product manner. Let us emphasize that, even if everything follows quite naturally from the 1D case, projection or collocation of the numerical fluxes on the cell boundaries have to be also carried out to suit the present theory. Details are given in Appendix B.

4. Numerical results

In this numerical results section, we make use of several widely addressed and challenging test cases to demonstrate the performance and robustness of the DG *a posteriori* correction presented. In all following test cases, if not stated differently, the simple case of local Lax-Friedrichs numerical flux will be used for both the DG scheme and the reconstructed flux correction. Also, for sake of security and to avoid as much as possible recursive steps in the limitation procedure, see correction flowchart in Section 3, when a subcell is detected as bad we also mark as bad the first neighboring subcells. It actually also enhances the quality of the numerical solutions, see Figure 7. The troubled subcell detection is definitely the part to be improved, and should be the topic of a paper on its own.

Regarding the cell decomposition into subcells, this has no impact on the reformulation of DG schemes into subcell finite volume method, see Figure 9(a). However, for the correction procedure, the subdivision does has an impact, see Figure 9(b). Indeed, the use of a non-uniform subdivision, for instance by means of the Gauss-Lobatto points, leads to better results compared to a uniform subdivision. This is more likely the manifestation of the Runge phenomenon in the context of histopolation, as the histopolation basis functions underlying the submean value representation, are more oscillatory for a uniform cell subdivision. Another possible explanation could be that the amount of correction present in the reconstructed flux definition, equation (17), is higher close to the cell boundaries. And thus, the subcell should be more refined near the cell interfaces to take that into account. In all following test cases, if not stated differently, we subdivide the cells by means of Gauss-Lobatto points.

Regarding the time integration, we make use of the classical third-order SSP Runge-Kutta scheme, see for instance [54], with a small enough CFL number. In cases where we compute rates of convergence, a time step $\Delta t \leq \Delta x^{\frac{k+1}{3}}$ is used in order to make the time error negligible in comparison to the spatial discretization error. Otherwise, we define $cfl = \frac{C_e}{2\,k+1}$, where $C_e = 0.2$ for sake of safety. The time step is then chosen as $\Delta t = cfl\,\frac{\Delta x}{\max_u |f'(u)|}$. As said in the introduction, the choice of the CFL is less critical with the designed limitation as if the high-order DG scheme starts developing instabilities, it will trigger the *a posteriori* correction and first-order FV scheme would then be used. Even if the $(k+1)^{\text{th}}$-order DG with 3rd-order RK time integration might be unstable for this CFL, the corrected scheme has in practice proved to be stable.

Let us emphasize that in all figures to come, if not stated otherwise, the solution subcell mean values are displayed at the centroid of each subcell. There is thus one dot per subcell, as in Figure 4 for instance where nine values per cell are displayed. Furthermore, if not stated differently, the simple case of first-order correction is used.

4.1. 1D scalar conservation laws

Let us first assess the performance and accuracy of DG schemes plus correction in the simple case of 1D scalar conservation laws.

4.1.1. Linear advection of a smooth signal

Let us consider the linear advection $\partial_t u + a\,\partial_x u = 0$, where the velocity is set to $a = 1$. We start from a smooth initial condition $u^0(x) = sin(2\pi\,x)$, and consider periodic boundary conditions. We assess the scheme accuracy after one period, namely at time $t = 1$. In Figure 4, the numerical solution of the ninth order scheme on only five cells is plotted. One can see that with only 5 cells,

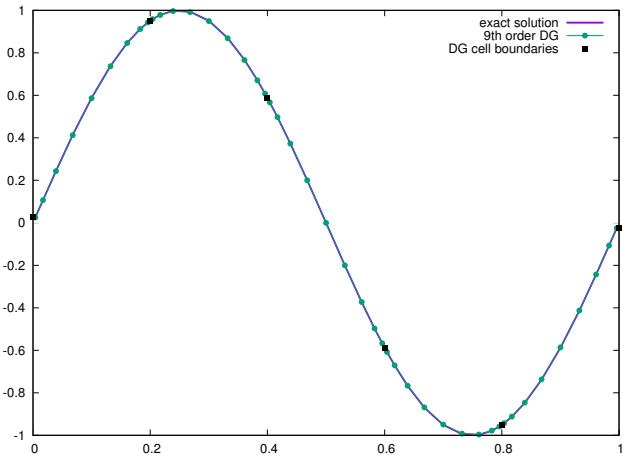


Figure 4: 9th order DG solution for the linear advection case on a 5 cells mesh after one period.

the corrected DG scheme is extremely accurate. Actually, the correction procedure is not activated in this case, which proves that the relaxation criterion on smooth extrema works properly. The rates of convergence are gathered in Table 1 and do exhibit a convergence to nine.

	L_1		L_2	
h	$E^h_{L_1}$	$q^h_{L_1}$	$E^h_{L_2}$	$q^h_{L_2}$
$\frac{1}{20}$	8.07E-11	9.00	8.97E-11	9.00
$\frac{1}{40}$	1.58E-13	9.00	1.75E-13	9.00
$\frac{1}{80}$	3.08E-16	-	3.42E-16	-

Table 1: Convergence rates for the linear advection case for a 9th order DG scheme

4.1.2. Linear advection of a square signal

To assess the efficiency of the correction presented, let us start with the very simple case of the advection of a square signal, namely $u^0(x) = 1$ if $x \in [0.4, 0.6]$ and $u^0(x) = 0$ elsewhere. In Figure 5, we compare the unlimited and corrected versions of the 9th order DG scheme on 10 cells after one period. One can see that only few subcells require a correction, and there are located near the discontinuities. In Figure 5, only the subcells corrected during the last iteration are marked in red. Except for the early stage of the calculation, the same zones are concerned with the correction, namely after the two discontinuities. During the first iterations, a lot more correction is however required to ensure the numerical solution admissibility, see Figure 6. In Figure 6(a), the four central cells are totally marked. This is yet due to the fact that for sake of safety we also mark the neighboring subcells of a bad subcell. Otherwise, only one subcell over two is marked, as in Figure 6(b). This however leads to a more oscillating solution in the end, see Figure 7(b). In all following test problems, this safety feature is thus used.

Let us note that even if the numerical solution is perfectly monotonous at the cell level, one can notice very small oscillations at the subcell scale, see Figure 7(a). These oscillation can not however lead to instability as they will trigger the NAD criterion if they grow too much. Nevertheless, if

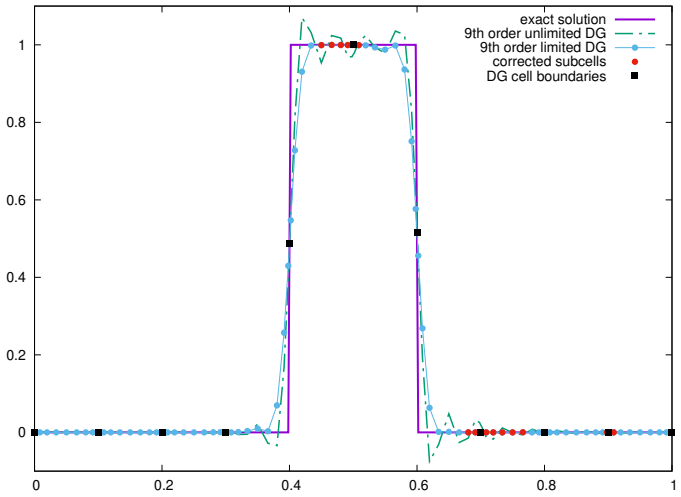
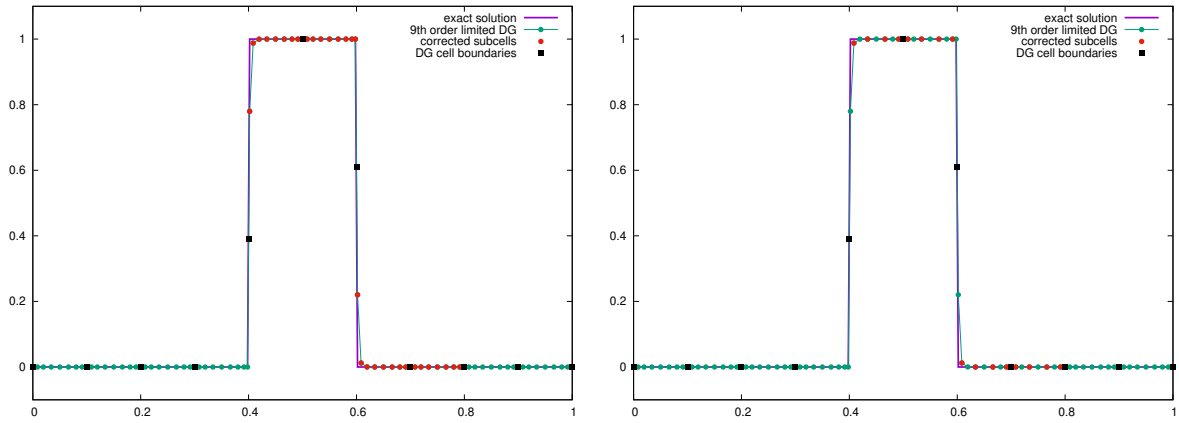


Figure 5: 9th order DG solution for the linear advection case on a 10 cells mesh after one period.



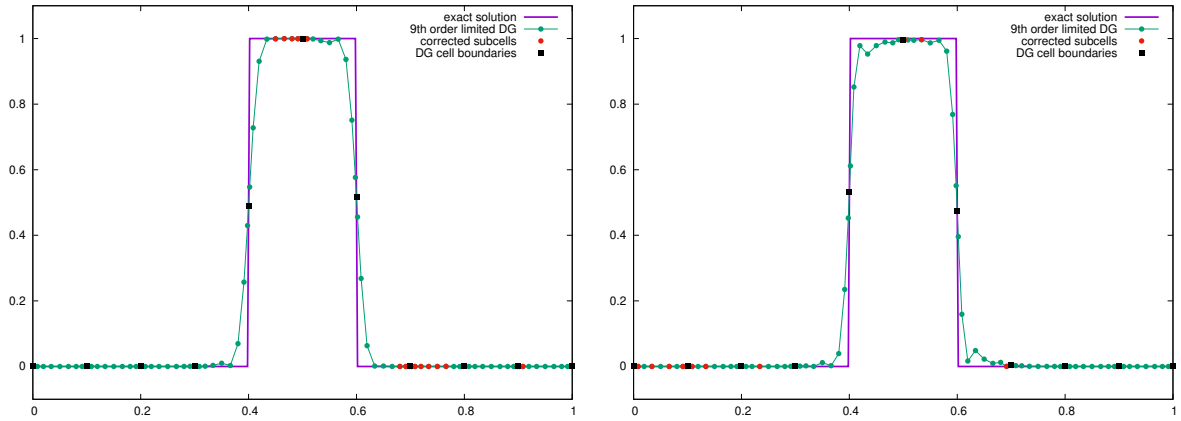
(a) Neighboring subcells also marked.

(b) Only the troubled subcells are marked.

Figure 6: 9th order DG solution for the linear advection case on 10 cells after one iteration.

one wants to address these subcell oscillations, an subcell discrete maximum principle (SubNAD) has to be used instead of a cell discrete maximum principle (NAD). In Figure 8, the subcell mean values now exhibit a monotonous profile. However, it has been done at the expense of accuracy. To reduce this loss of precision, we may want to use a second-order correction, namely on marked subcell boundaries the reconstructed fluxes are substituted with second-order TVD finite volume numerical fluxes instead of first-order ones, see Figure 8.

Now, as said in the introduction of this section, the cell subdivision does only have an impact on the corrected version of the scheme. In Figure 9, we compare the numerical results provided three types of cell subdivision: a uniform one where $\widetilde{x}_{m+\frac{1}{2}} = x_{i-\frac{1}{2}} + \frac{m}{k+1}\,|\omega_i|$, a cosinus distribution similarly to Tchebychev quadrature points where $\widetilde{x}_{m+\frac{1}{2}} = x_{i-\frac{1}{2}} + (1-\cos(\frac{m\,\pi}{k+1}))\,\frac{|\omega_i|}{2}$, and a Gauss-Lobatto distribution where the flux points $\widetilde{x}_{m+\frac{1}{2}}$ are nothing but the Gauss-Lobatto quadrature points. In



(a) Neighboring subcells also marked.

(b) Only the troubled subcells are marked.

Figure 7: 9th order DG solution for the linear advection case on 10 cells after one period.

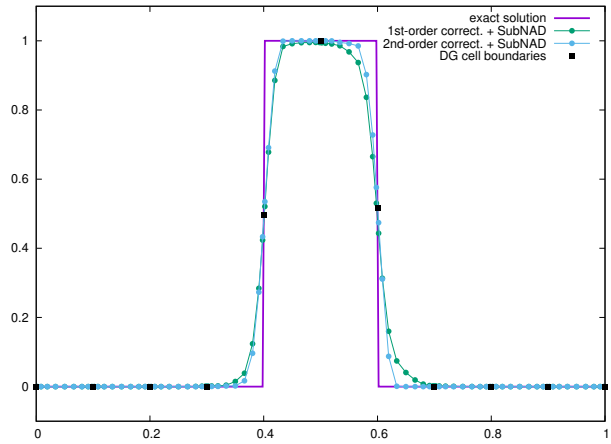
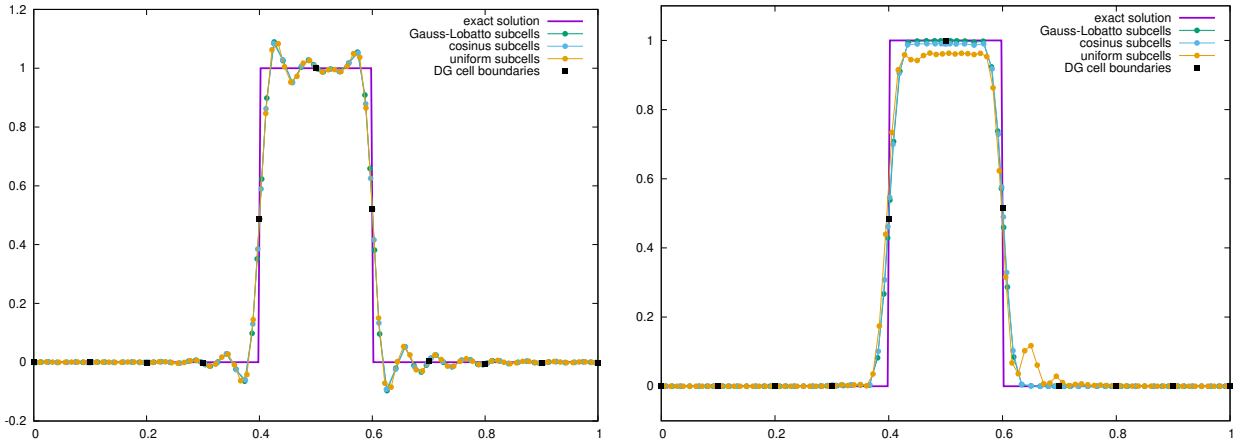


Figure 8: 9th order DG solution on a 10 cells mesh after one period: correction plus subcell DMP detection (SubNAD).

Figure 9(a), the results displayed were obtained by means of the subcell finite volume formulation with non-corrected reconstructed fluxes. The piecewise polynomial solutions at the cell subdivision points are plotted. The results depicted in Figure 9(a) numerically corroborate the fact that the cell subdivision has absolutely no influence on the numerical scheme. Now, in Figure 9(b), the results obtained through the corrected reconstructed fluxes are shown. The difference is now noteworthy. In the light of Figure 9, it is clear that a uniform subcell distribution leads to the worse results. Although the two other cell subdivision produce similar results, the Gauss-Lobatto subcell distribution does yield the best results. This is thus the subdivision used in all remaining test cases.

We recall that in the subcell finite volume limitation presented in [18], if a subcell is detected as bad, the numerical solution in all subcells inside this DG cell are recomputed through a finite volume scheme. The DG polynomial has then been totally discard. In the present correction procedure, only few subcells need to be recomputed. Elsewhere, we keep the solution computed through the unlimited DG scheme. In Figure 10, we compare these two limitation techniques. As expected, this

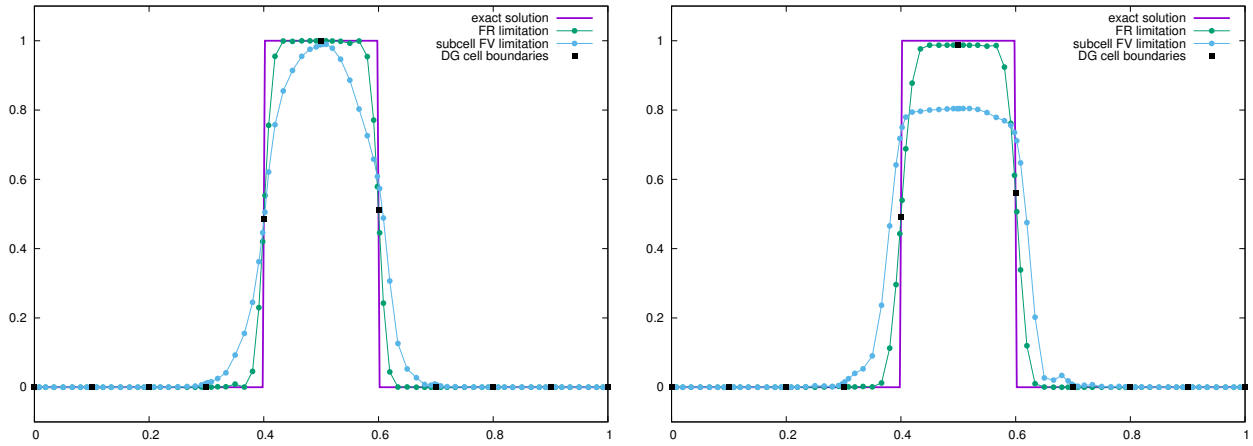


(a) Without correction: subcell boundary values.

(b) With correction: subcell mean values.

Figure 9: 9th order DG solution for the linear advection case on a 10 cells mesh after ten periods: comparison between different subcell distribution.

new limitation outperforms the one presented in [18].



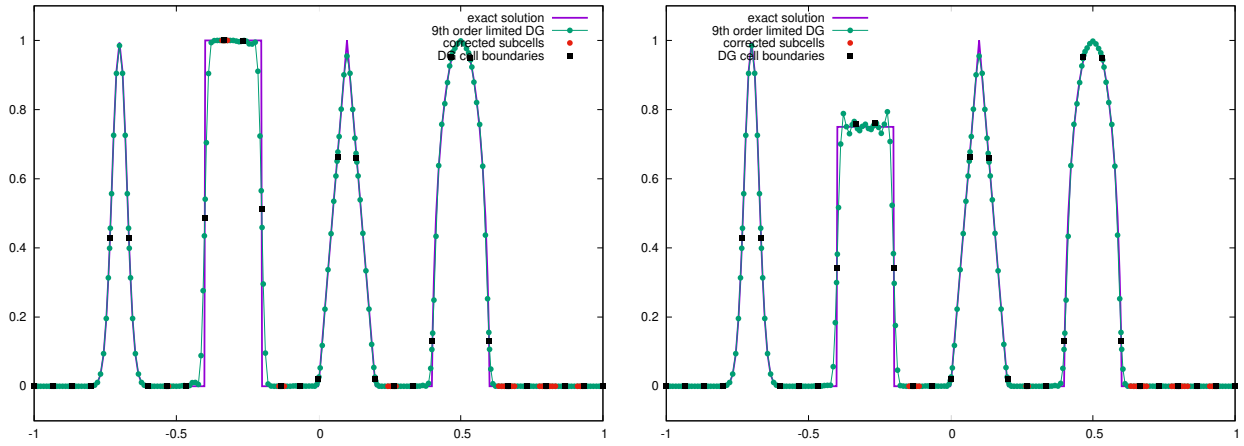
(a) After 1 period.

(b) After 50 periods.

Figure 10: 9th order DG solution on 10 cells: comparison between the subcell finite volume limitation and the subcell flux reconstruction correction.

4.1.3. Linear advection of a composite signal

Let us now address the classical case of the linear advection of a composite signal, introduced in [37]. This signal is composed by the succession of a Gaussian, rectangular, triangular and parabolic signals. However, because the different signal extrema are always zero and one, a limitation procedure only ensuring the preservation of the maximum principle (PAD criterion) would be enough, as depicted in Figure 11(a). In order to assess the relevancy of the spurious oscillations detection, namely the NAD criterion, we modify this test case as in Figure 11(b) to make it more challenging. In Figure 12(a), we plot the numerical solution obtained with a ninth order corrected DG scheme

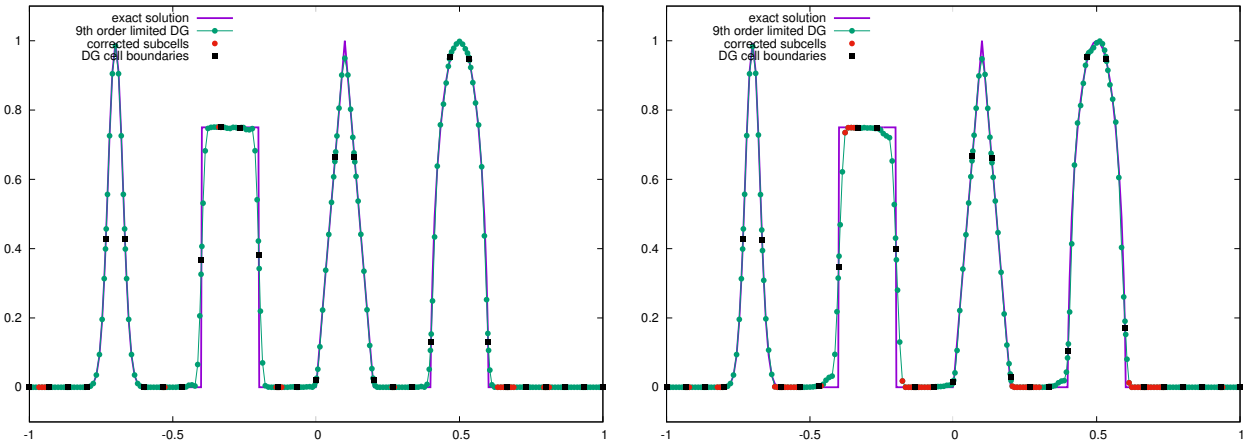


(a) Classical composite signal.

(b) Modified composite signal.

Figure 11: 9th order DG solution on 30 cells after 4 periods: correction only making use of the PAD criterion.

using only 30 cells after 4 periods, for this modified test case. One can see that even by means of this very coarse grid, the numerical solution is extremely precise and robust. In Figure 12(a), on the rectangular signal for instance, we can observe some slight subcell oscillations. It is however possible to use the SubNAD criterion to ensure the monotonicity of the solution even within the cell. A second-order correction is preferred to avoid to much accuracy loss, see Figure 12(b).



(a) PAD + NAD and 1st order correction.

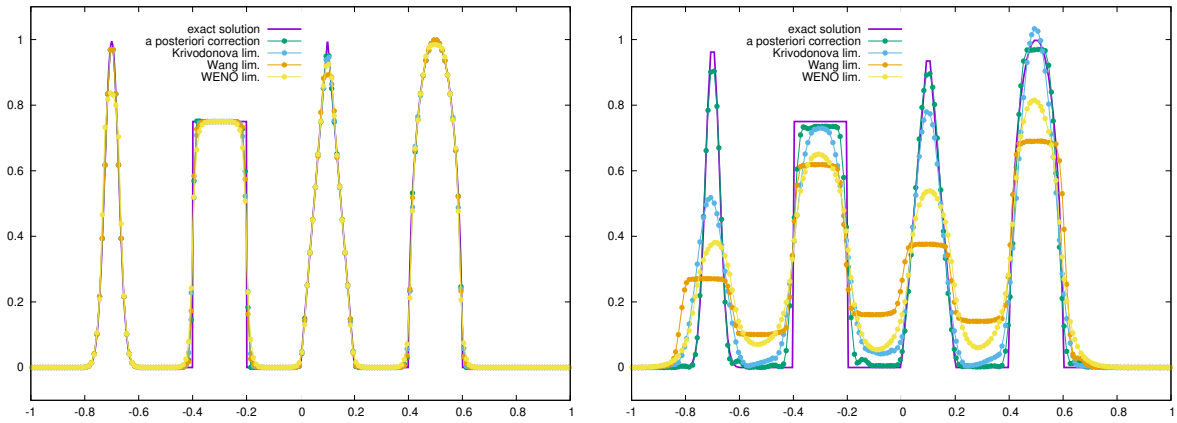
(b) PAD + SubNAD and 2nd order correction.

Figure 12: 9th order DG solution on 10 cells after 4 periods: comparison between first-order correction and second-order correction with SubNAD criterion.

Now, even if this paper is not concerned with doing an exhaustive review of limiting techniques or comparison between them, we show some comparative tests to assess the performance of this *a posteriori* limitation. To this end, we make use of some well-known *a priori* limiting strategies, as the L. Krivodonova limiter [38], the Z.J. Wang moment limiter [62] and the X. Zhong and C.-W. Shu WENO limiter [64]. Let us emphasize that for the latter one, we have experienced that its scalability to very high-order schemes (above 4th order) highly depends on the coefficients in the nonlinear weights definition. It thus makes this technique not very practical for very high-order

DG schemes. Consequently, we will compare with our *a posteriori* correction technique all these three limiters on 4th order DG, and then only the Krivodonova and Wang limiters for 9th-order DG. Let us emphasize that in the following comparisons, the *a posteriori* limiting strategy is done using 1st-order correction of the reconstructed fluxes, to be as fair as possible.

We can see on Figure 13(a) that using 200 cells, as it is generally done, all these different limiting strategies work quite nicely for a 4th order DG scheme. However, the strong advantage of very high-order scheme being to retain a good resolution making use of coarse grids, to assess the difference between those limiters we now use only 30 cells. In Figure 13(b), one can see that Krivodonova limiter as well as our correction technique are still behaving properly. Our limiting strategy seems to be the one producing the best result.



(a) 200 cells: cell mean values.

(b) 50 cells: subcell mean values.

Figure 13: 4th order DG solutions provided different limitations for the linear advection case after 4 periods.

Now, to see the scalability of those limiters to very high-order schemes, we make use of the same test case but with a 9th order scheme on only 30 cells, similarly to Figure 12. In Figure 14, we can clearly observe that the present correction procedure outperforms the other limiting strategies.

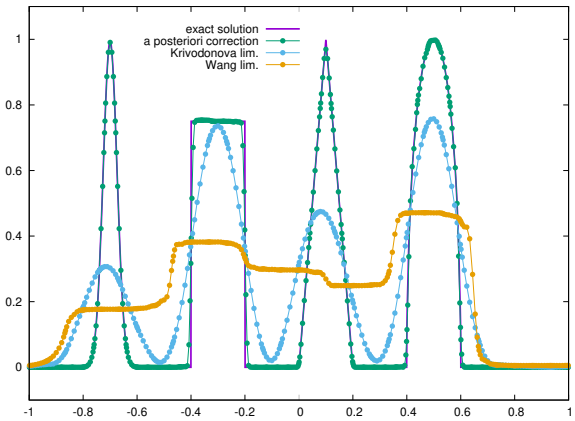
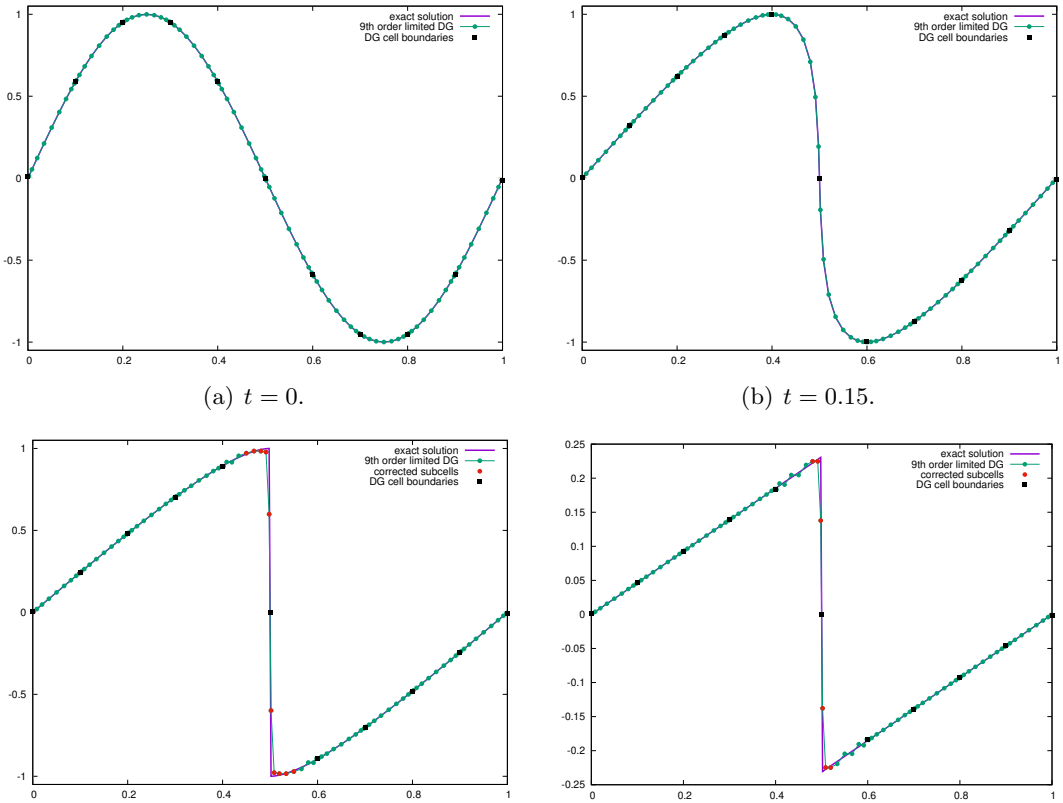


Figure 14: 9th order DG solutions provided different limitations for the linear advection case on 30 cells after 4 periods.

4.1.4. Burgers equation with a smooth initial solution

Let us consider the non-linear Burgers case $\partial_t u + \partial_x(\frac{u^2}{2}) = 0$. Starting from the smooth initial condition $u^0(x) = sin(2\pi\, x)$ on $[0, 1]$, a stationary discontinuity located at $x = \frac{1}{2}$ forms at time $t_c = \frac{1}{2\,\pi}$. In Figure 15, the numerical solution obtained with the corrected 9th order DG scheme on 10 cells is plotted at different times. Before the critical time t_c, one can see that the correction procedure is not active, as in Figure 15(b). It only activates for $t \geq t_c$ and remains active since then in the two DG cells surrounding the discontinuity, see Figures 15(c) and 15(d). This test case proves that even in this extremely coarse mesh case, the discontinuous Galerkin scheme plus correction is very precise as well as robust.



(a) $t = 0$. (b) $t = 0.15$.

(c) $t = 0.25$. (d) $t = 2$.

Figure 15: 9th order DG solution for Burgers equation on a 10 cells mesh.

4.1.5. Burgers equation with shock and expansion waves collision

To demonstrate the performance and robustness of the DG *a posteriori* correction presented, we introduce a new test case in the context of 1D Burgers equation which consists in a shock wave and an expansion wave that will finally meet and form a new shock propagating at non-linear speed. The initial data is characterized by three states as following

$$u^0(x) = \begin{cases} 0 & \text{if } x < x_L, \\ -1 & \text{if } x \in]x_L, x_R[, \\ \frac{1}{2} & \text{if } x > x_R, \end{cases} \tag{26}$$

where $x_R > x_L$. Following the characteristic lines, the entropic weak solution exhibits a shock wave initially located at $x = x_L$ and propagating at speed $S = -\frac{1}{2}$, and an expansion fan located initially at x_R. Before these two waves meet, the solution writes

$$u^0(x) = \begin{cases} 0 & \text{if } x < -\frac{t}{2} + x_L, \\ -1 & \text{if } x \in]-\frac{t}{2} + x_L, -t + x_R], \\ \dfrac{x - x_R}{t}, & \text{if } x \in [-t + x_R, +\frac{t}{2} + x_R], \\ \frac{1}{2} & \text{if } x \geq \frac{t}{2} + x_R, \end{cases} \tag{27}$$

This solution holds until the discontinuity meets the expansion fan, namely at time $t_c = 2(x_R - x_L)$. For $t \geq t_c$, these two waves form a new shock moving to the left and located at $x_s(t) = x_R - \sqrt{2(x_R - x_L)\, t}$. Finally, for $t \geq t_c$ the entropic solution is defined as following

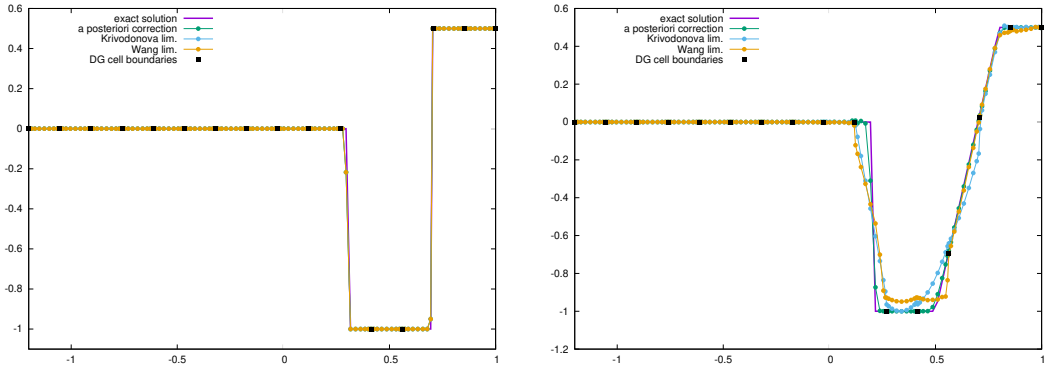
$$u^0(x) = \begin{cases} 0 & \text{if } x < x_R - \sqrt{2(x_R - x_L)\, t}, \\ \dfrac{x - x_R}{t}, & \text{if } x \in]x_R - \sqrt{2(x_R - x_L)\, t}, +\frac{t}{2} + x_R], \\ \frac{1}{2} & \text{if } x \geq \frac{t}{2} + x_R. \end{cases} \tag{28}$$

Here, we take $x_L = 0.3$ and $x_R = 0.7$, while the computational domain is set to be $[-1.2, 1]$. In this set up, the expansion and shock waves meet at time $t = 0.8$. At the final time $t = 3.2$, the shock is located at $x = -0.9$ and the shock left and right values are respectively 0 and $\frac{1}{2}$.

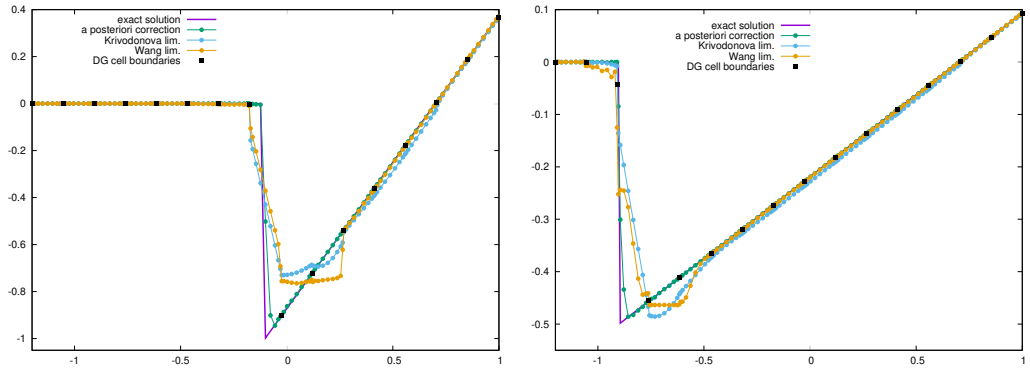
In Figure 16, we compare the results obtained through a 9th order DG scheme provided Krivodonova and Wang *a priori* limiters, along with this new *a posteriori* correction, on a 15 cells mesh at different times. Anew, the *a posteriori* correction seems the only one to capture the correct solution, in this very high-order very coarse grid case. It also illustrates once more the very good behavior of DG schemes provided with the presented *a posteriori* limiting strategy.

4.1.6. Buckley non-convex flux problem

We make use of the challenging Buckley problem to illustrate some well-known problems of discontinuous Galerkin schemes, as entropy and aliasing issues, see Figure 17. The Buckley equation is defined as $\partial_t u + \partial_x F(u) = 0$, where the non-convex flux function writes $F(u) = \frac{4\,u^2}{4\,u^2 + (1-u)^2}$. Since the flux function is now a complex rational function, it is not practical to analytically integrate the volume integrals in DG schemes. Consequently, we make use a quadrature rule exact for polynomial up to $2\,k$, as it is generally done. In [36], G.-S. Jiang and C.-W. Shu proved a cell entropy inequality for DG schemes for scalar conservation laws. However, this demonstration relies on the exact calculation of the integrals, and thus does not applied if one uses quadrature rules. This remark is depicted in Figure 17(a). We start from the initial solution $u^0(x) = 1$ if $x \in [-\frac{1}{2}, 0]$ and $u^0(x) = 0$ elsewhere. In Figure 17(a), we plot 3rd-order numerical solutions obtained by means of an unlimited DG scheme using quadrature rule, for different mesh resolution. It is clear the numerical solution does not converge to the entropic solution of the problem. Let us note that for this test case, global Lax-Friedrichs numerical flux is used to ensure first-order scheme to be entropic. In Figure 17(a), one can also observe strong spurious oscillations for $x \in [-0.7, -0.2]$. These oscillations are due to the so-called aliasing phenomenon. For DG schemes, collocation of the interior flux or approximated calculation of the volume integral through quadrature rule both generate errors if any of the modes arising from the nonlinear terms lies outside the polynomial functions space. Increasing the order of approximation will also increase this phenomenon, see Figure 17(b). Over integration could reduce this, but for very high-order schemes the number of quadrature points required to damp these errors

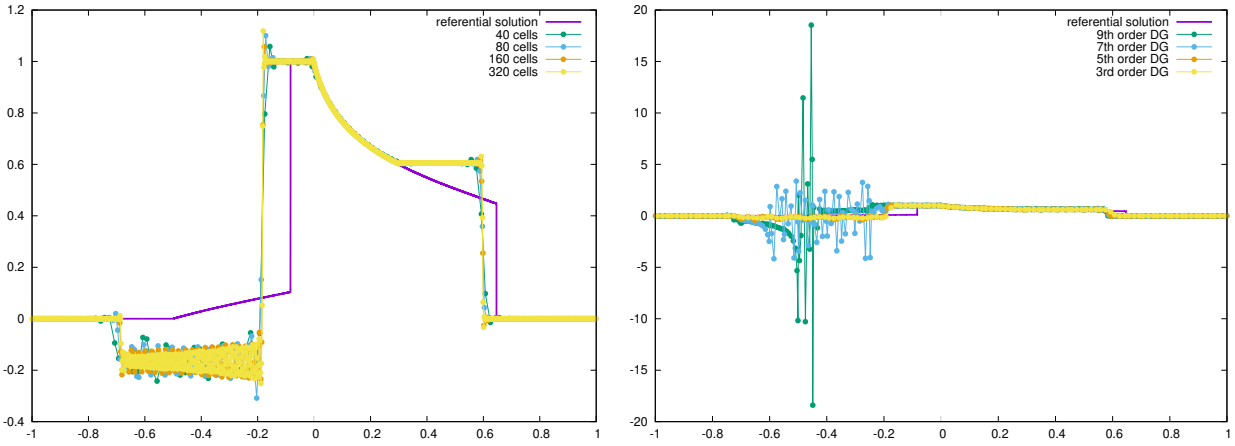


(a) $t = 0$. (b) $t = 0.2$.



(c) $t = 0.8$. (d) $t = 3.2$.

Figure 16: 9th order DG solutions provided different limitations for Burgers equation on a 15 cells mesh.



(a) Non-entropic behavior. (b) Aliasing phenomenon.

Figure 17: Unlimited DG solutions for the Buckley non-convex flux case.

make this solution not very practical. Now, we make use of this challenging test case to assess the efficiency of the presented correction. In Figure 18, we compare, for a 9th order DG scheme on 40

cells, the use of NAD and SubNAD criteria. In Figure 18, one can see that both treatments cure aliasing phenomenon. However, it seems that the use of NAD criterion is not enough for the numerical solution to converge to the entropic one. Making use of the SubNAD criterion allows us to recover the entropic solution. Both first and second order corrections would converge to the correct solution, second-order treatment only produces a slightly more accurate result. It is worth saying that actually both NAD and SubNAD criteria would make the scheme to converge to the correct entropic solution. With the NAD criterion, the convergence is simply slower than with SubNAD. Another option would be to add a discrete entropy condition to the detection part, to check if some submean value break any entropy condition.

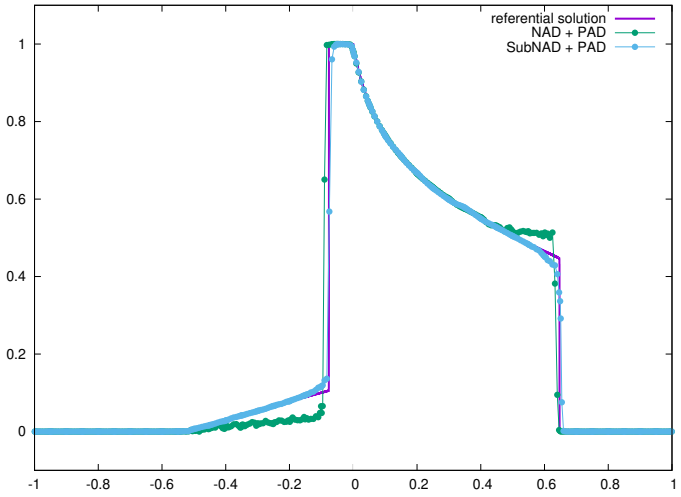
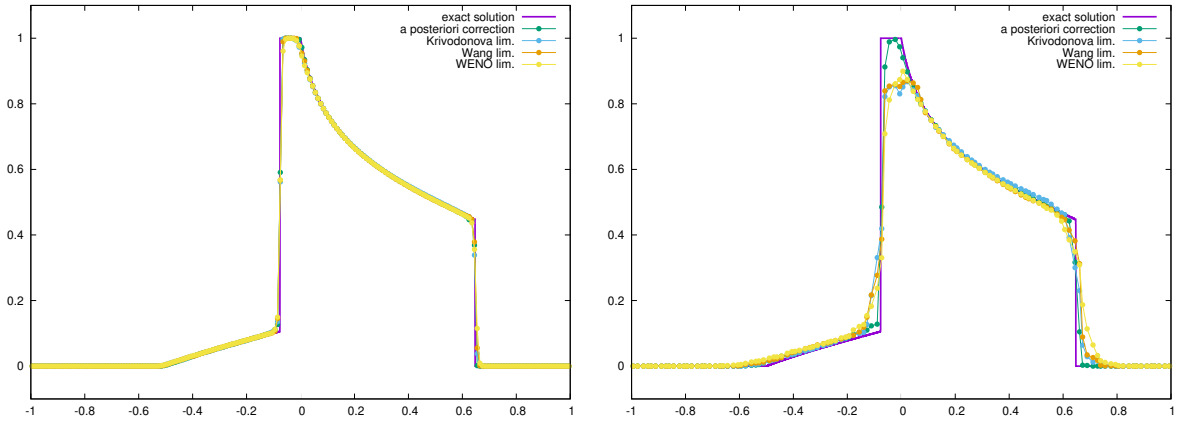


Figure 18: 9th order DG with 40 cells for the Buckley non-convex flux case with different detection criteria.

Let us now perform some comparisons with existing *a priori* limitations. In Figure 19, 4th-order results are presented for different mesh resolutions.



(a) 200 cells: cell mean values.

(b) 30 cells: subcell mean values.

Figure 19: 4th order DG solutions provided different limitations for Buckley problem at time $t = 0.4$.

Once again, one can see on Figure 19(a) that all the limitations tested perform well on a 200 cells

mesh, and seem to be able capture the entropic solution. Reducing the number of cells used, we observe that only the subcell *a posteriori* correction procedure can capture the solution peak, see Figure 19(b).

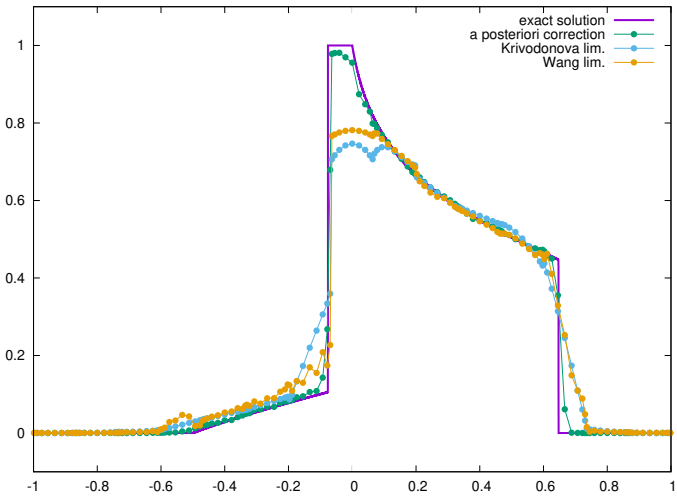


Figure 20: 9th order DG solutions provided different limitations for Buckley problem on 15 cells at time $t = 0.4$.

Similar conclusion for 9th order DG on 15 cells, see Figure 20.

4.2. 1D Euler system

Although the whole theory presented here has been introduced in the simple case of scalar conservation laws, the extension to the system case is perfectly straightforward. The only part which may differ is the troubled detection part. For the physical admissibility detection (PAD), we consider that a solution is physically admissible if the density and the internal energy are strictly positive. For the numerical admissibility detection (NAD), the natural system counterpart would be to apply the previously introduced detection criteria to the Riemann invariants. However, in the non-linear system case, those quantities are not easy to get nor to manipulate. We could have use a linearized version of the Riemann invariants, as in [59] for instance, but for sake of simplicity we naively apply the NAD or SubNAD criteria only to the density. Also, the simple local Lax-Friedrichs (Rusanov) numerical flux will be used in the computations.

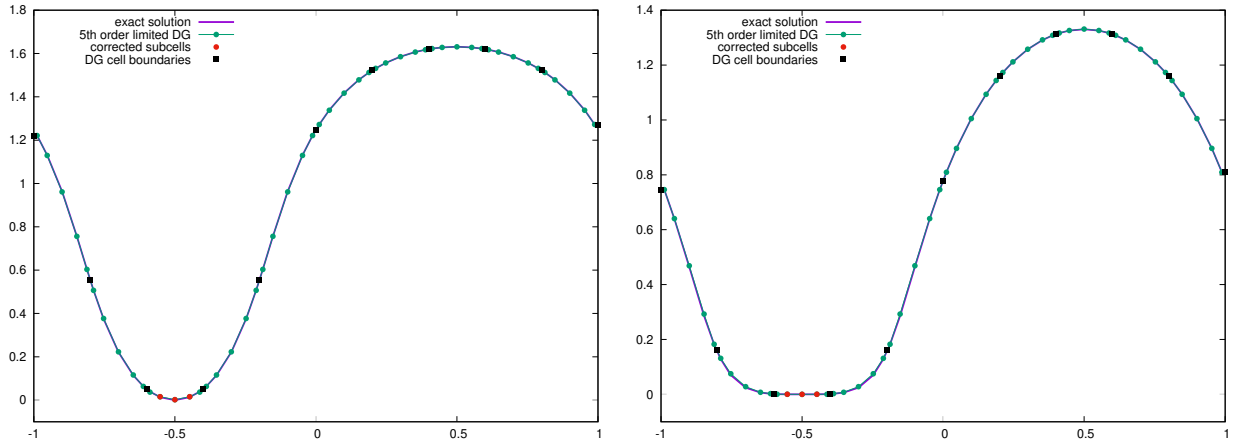
4.2.1. Smooth isentropic flow problem

To test the accuracy of DG scheme plus correction, we make use of a smooth test case initially introduced in [58]. This example has been derived in the isentropic case, for the perfect gas equation of state with the polytropic index $\gamma = 3$. In this special situation, the characteristic curves of the Euler equations become straight lines, and the governing equations reduce to two Burgers equations. It is then simple to solve analytically this problem. Here, we modify the initial data to yield a more challenging example, as

$$\rho^0(x) = 1 + 0.9999999\sin(\pi x), \quad u^0(x) = 0, \quad p^0(x) = \rho^0(x)^\gamma, \quad x \in [-1, 1],$$

provided with periodic conditions. This means that initially $\rho^0(-\frac{1}{2}) = 1.E-7$ and $p^0(-\frac{1}{2}) = 1.E-21$. The density and pressure being so close to zero, any numerical scheme not ensuring

a positivity preservation would fail. This is the case of unlimited DG schemes. In Figure 21, numerical solutions obtained by means of the 5th order corrected DG scheme are depicted at time $t = 0.1$, using 10 cells. One can remark that the *a posteriori* correction is indeed active in the low density/pressure zone.



(a) Density. (b) Internal energy.

Figure 21: 5th order corrected DG solution for the smooth flow problem on 10 cells.

In Table 2, we gather the global errors and rates of convergence related to the 5th order scheme, along with the time average percentage of corrected subcells. The results confirm the expected fifth-order rate of convergence, even though the solution has been locally corrected.

	L_1		L_2		Average % of
h	$E^h_{L_1}$	$q^h_{L_1}$	$E^h_{L_2}$	$q^h_{L_2}$	corrected subcells
$\frac{1}{10}$	8.59E-4	5.80	1.17E-3	5.82	4.21 %
$\frac{1}{20}$	1.48E-5	4.35	2.02E-5	4.18	6.87 %
$\frac{1}{40}$	9.09E-7	4.88	1.38E-6	4.87	3.31%
$\frac{1}{80}$	3.09E-8	4.95	4.73E-8	4.86	2.50 %
$\frac{1}{160}$	1.00E-9	-	1.63E-9	-	1.12%

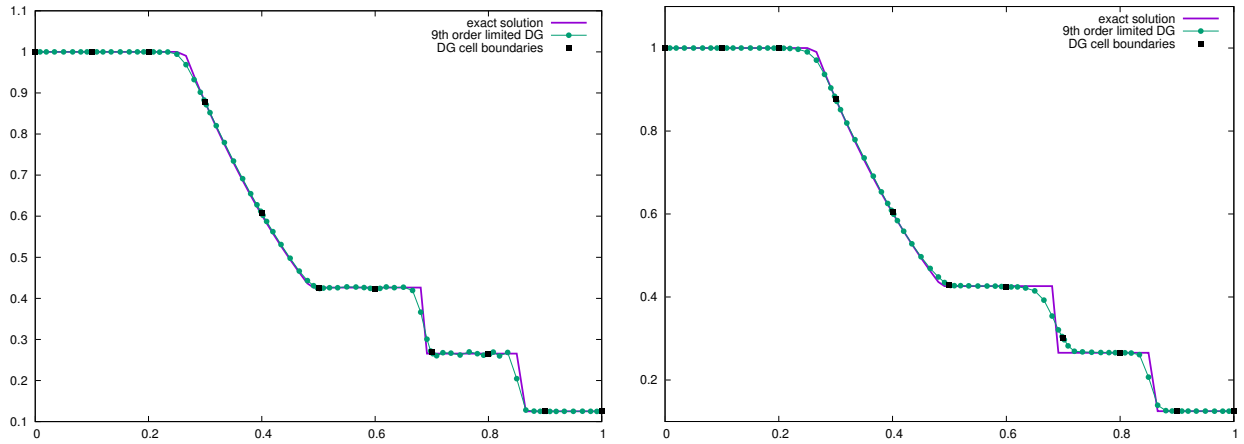
Table 2: Rate of convergence computed on the pressure in the case of the smooth isentropic problem at time $t = 0.1$, for the 5th corrected DG scheme.

We can also notice in the light of Table 2 that refining the mesh, the percentage number of subcells to be recomputed decreases. One could have expected such conclusion since the number of subcells located in the very low density/pressure region remains essentially constant, while the global number of subcells increases.

4.2.2. Sod shock tube problem.

We consider now the classical Sod shock tube problem, [55]. At the initial time, two states are separated by an interface located at $x = 0.5$. The left state is a high pressure fluid characterized by $(\rho^0_L, p^0_L, u^0_L) = (1, 1, 0)$, the right state is a low pressure fluid defined by $(\rho^0_R, p^0_R, u^0_R) = (0.125, 0.1, 0)$. The gamma gas law is defined with $\gamma = \frac{7}{5}$. In Figure 22, results obtained by means of 9th order DG

scheme using only 10 cells are displayed. In Figure 22(a), the standard correction is used, namely making use of PAD and NAD criteria along with first-order flux correction. With only 10 cells, we can already remark how accurate the scheme is. The shock is captured in only one cell, and actually in only one subcell. This again demonstrates the very high potential of high-accurate scheme, as well as the presented subcell correction procedure. One can yet remark small subcell oscillations. This phenomenon can be tackled by use of a subcell discrete maximum principle (SubNAD), as in Figure 22(b). The solution is now monotonous even inside the cell, at the cost of a slight accuracy loss especially in the contact discontinuity region, even with 2nd-order correction.



(a) PAD + NAD and 1st order correction.

(b) PAD + SubNAD and 2nd order correction.

Figure 22: 9th order DG solution for the Sod shock tube problem on 10 cells: comparison between first-order correction and second-order with SubNAD criterion.

Let us emphasize that this *a posteriori* limitation procedure is not limited to the case of very high-order of accuracy. It also preforms very well at second or third order. See for instance Figure 23 where the third-order numerical solution on 100 cells is depicted.

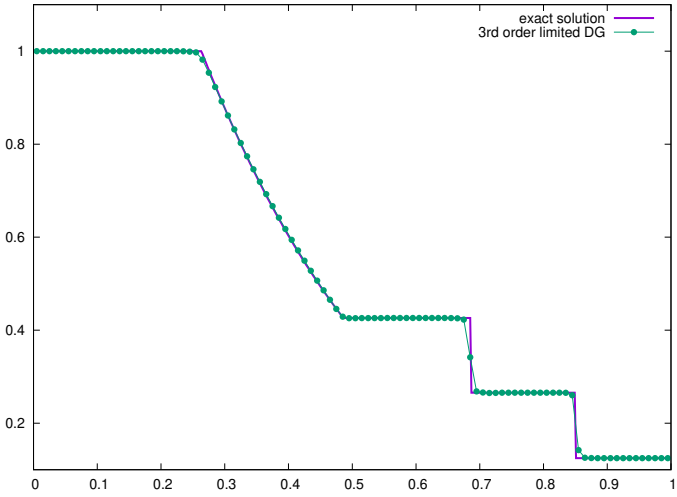


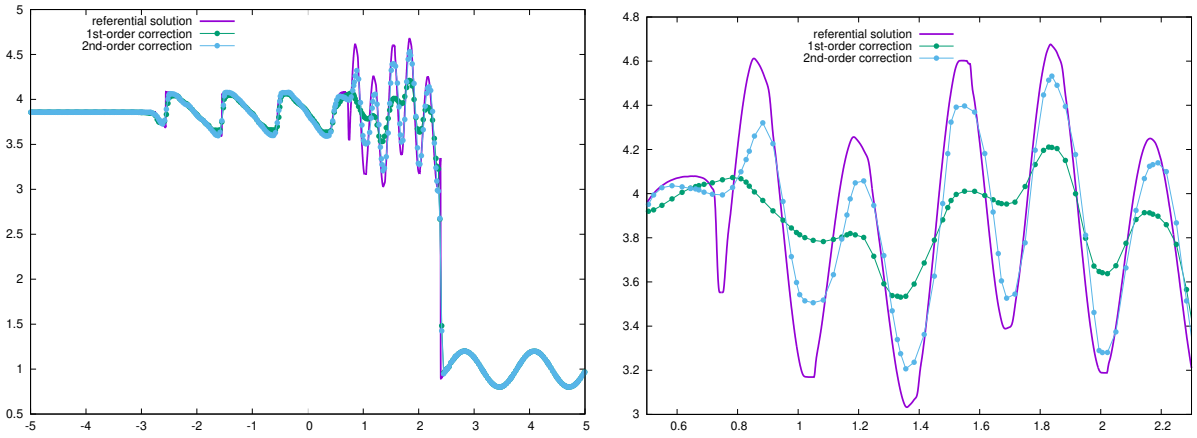
Figure 23: 3rd-order DG solutions for the Sod shock tube problem on 100 cells: cell mean values.

4.2.3. Shock acoustic-wave interaction problem.

The next test case, introduced initially by C.-W. Shu and S. Osher in [54], consists in the interaction of a shock wave and an acoustic wave. The initial data read

$$(\rho^0,\, u^0,\, p^0) = \begin{cases} (3.857143,\, 2.629369,\, 10.333333), & x < -4, \\ (1 + 0.2\sin(5\,x),\, 0,\, 1)\,, & x \geq -4. \end{cases}, \qquad x \in [-5, 5].$$

In this test case, it is critical to achieve high-order accuracy if one hopes to capture this very oscillatory solution. In Figure 24, 7th-order results on 50 cells are presented for both first and second order corrections, with SubNAD detection criterion. In Figure 24(b), the difference between the two corrections is tremendous. Indeed, in this test case, especially for coarse grids, the wave interaction takes place in only one or two cells. Consequently, high accuracy is indeed required to capture correctly the solution. One could expect even better results with higher order corrections, by means for instance of high-order WENO numerical fluxes in the subcell limitation.



(a) Global view. (b) Zoom on $[0.5, 2.3]$.

Figure 24: 7th-order corrected DG solutions on 50 cells for the oscillating shock tube problem: comparison between first and second order corrections.

To demonstrate one last time that the presented correction does also exhibit excellent results for lower order schemes, we display in Figure 25 the third-order numerical solution for a 200 cells mesh.

4.2.4. Blast waves interaction problem.

The blast waves interaction problem is a standard low energy problem involving shocks, generally used to assess the robustness of gas dynamics schemes. The initial data read

$$\rho^0(x) = 1, \quad u^0(x) = 1, \quad p^0(x) = \begin{cases} 10^3, & 0 < x < 0.1, \\ 10^{-2}, & 0.1 < x < 0.9, \\ 10^2, & 0.9 < x < 1.0. \end{cases}, \quad x \in [0, 1].$$

The fluid under consideration is described by the ideal gas equation of state with $\gamma = 1.4$, and reflective conditions are applied to the left and right boundaries of the domain. The resulting solution at time $t = 0.038$ is quite complex, but yet there are no smooth regions where high-order schemes may express their potential and outclass low order schemes. However, to still assess the advantage of high-order on low-order, as well as depict the good behavior of the presented

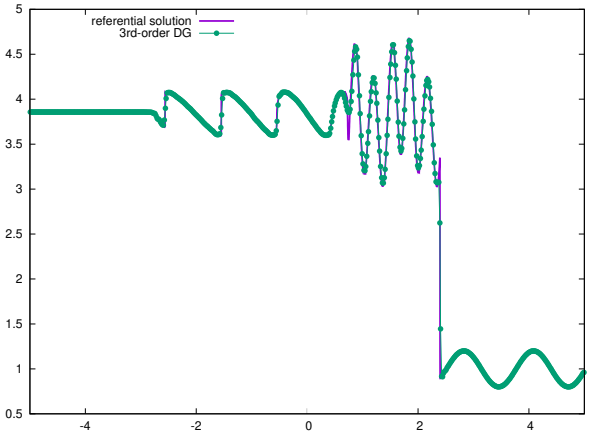


Figure 25: 3rd-order corrected DG solutions on 200 cells for the oscillating shock tube problem: cell mean values.

limitation technique, let us first consider a quite coarse grid. In Figure 26, the density computed with 60 uniform cells, with DG schemes from third-order to ninth-order schemes, are compared with referential solution. This reference solution has been obtained using a second-order scheme with 10000 grid points.

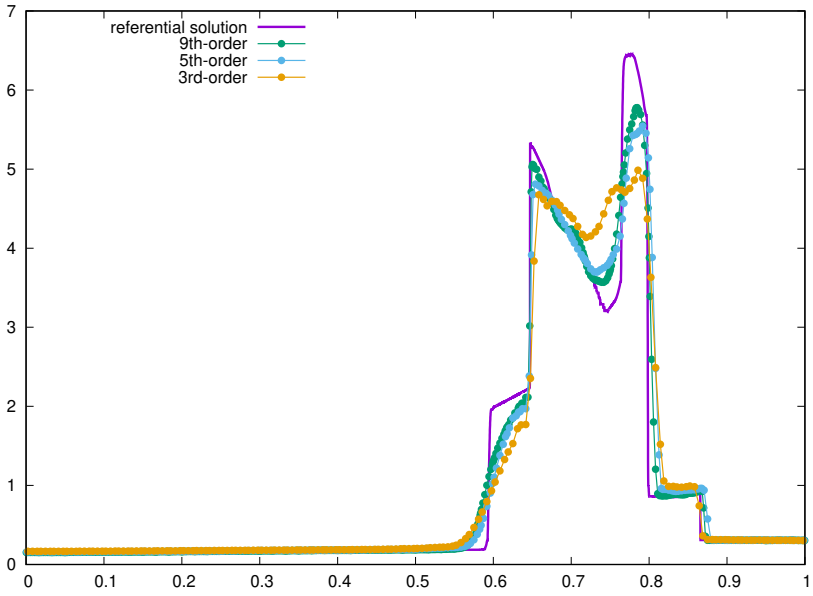


Figure 26: Corrected DG solutions, from 3rd to 9th order, on 60 cells for the blast waves problem.

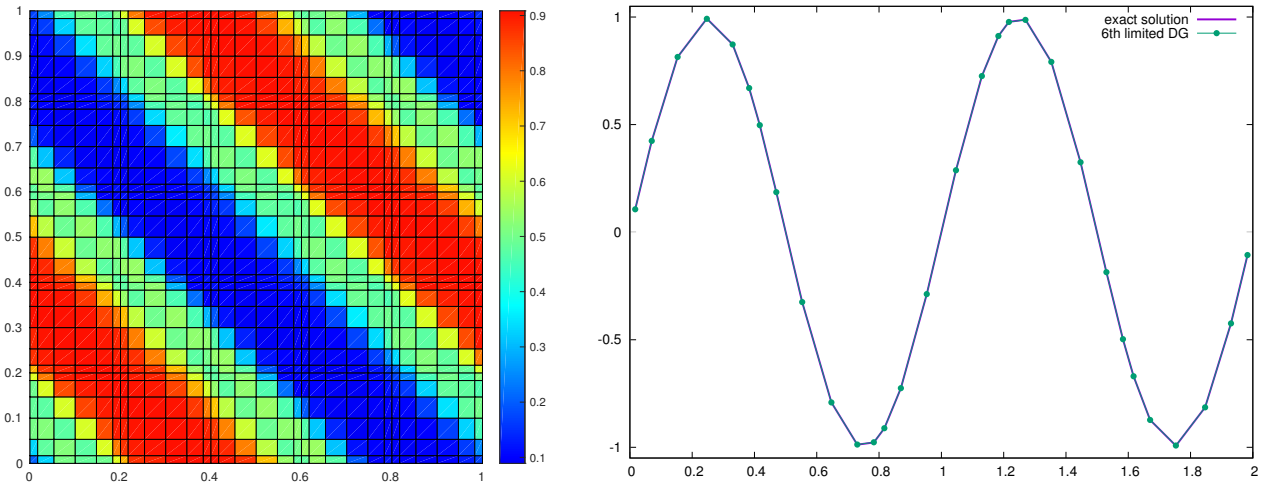
One can note, even in this totally non-smooth solution case, a substantial gain in accuracy increasing the order of approximation. And even though the difference between fifth and ninth order results are slight for this test case, it has the benefit to demonstrates the strong robustness of the presented correction, for any order of approximation.

4.3. 2D scalar conservation laws

To conclude this numerical results section, we present some two-dimensional computational tests for the simple case of 2D scalar conservation laws.

4.3.1. Linear advection of a smooth signal

Let us consider the linear advection case $\partial_t u + \boldsymbol{A} \cdot \nabla u = 0$, where the velocity is set to $\boldsymbol{A} = (1,1)^{\mathrm{t}}$. We start from a smooth initial condition $u^0(x) = sin(2\pi\,(x+y))$. We consider periodic boundary conditions. We assess the scheme accuracy after one period, namely at time $t = 1$. In Figure 27, the numerical solution obtained by means of sixth-order scheme on a 5×5 Cartesian grid is displayed. In all figures to come, we plot subcell mean values. This is why, in Figure 27(a), one can notice 6×6 subcells per cell.



(a) Solution map.

(b) Solution profile along $x = y$.

Figure 27: Linear advection with 6th DG scheme and 5×5 grid after one period.

We see that with only a 5×5 grid, the numerical scheme is extremely accurate. The rate of convergence are gathered in Table 3 and do exhibit a convergence to six.

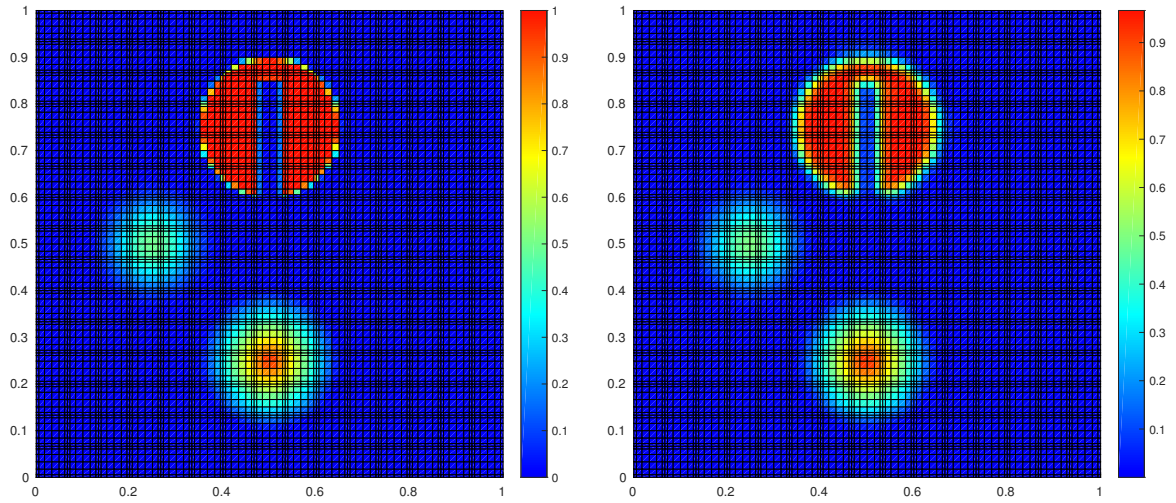
	L_1		L_2	
h	$E^h_{L_1}$	$q^h_{L_1}$	$E^h_{L_2}$	$q^h_{L_2}$
$\frac{1}{5}$	2.10E-6	6.23	2.86E-6	6.24
$\frac{1}{10}$	2.79E-8	6.00	3.77E-8	6.00
$\frac{1}{20}$	3.36E-10	-	5.91E-10	-

Table 3: Convergence rates for the linear advection case for a 6th order DG scheme

4.3.2. Solid body rotation of a composite signal

We make use of the classical test case taken from [43], which corresponds to the two-dimensional extension of the composite signal presented in Section 4.1.3. Let us then consider a divergence-free velocity field corresponding to a rigid rotation, defined by $\boldsymbol{A} = (\frac{1}{2} - y, x - \frac{1}{2})^{\mathrm{t}}$. We apply this solid

body rotation to the initial data displayed in Figure 28(a), which includes both a plotted disk, a cone and a smooth hump.



(a) Exact solution. (b) Numerical solution.

Figure 28: 6th order corrected DG solution for the rigid rotation case on a 15×15 grid after one full rotation.

In the light of the results depicted in Figures 28 and 29, we can note the tremendous precision of the numerical solution using only 15×15 cells, which is furthermore totally oscillation free. This anew demonstrates the very high capability of the correction procedure presented.

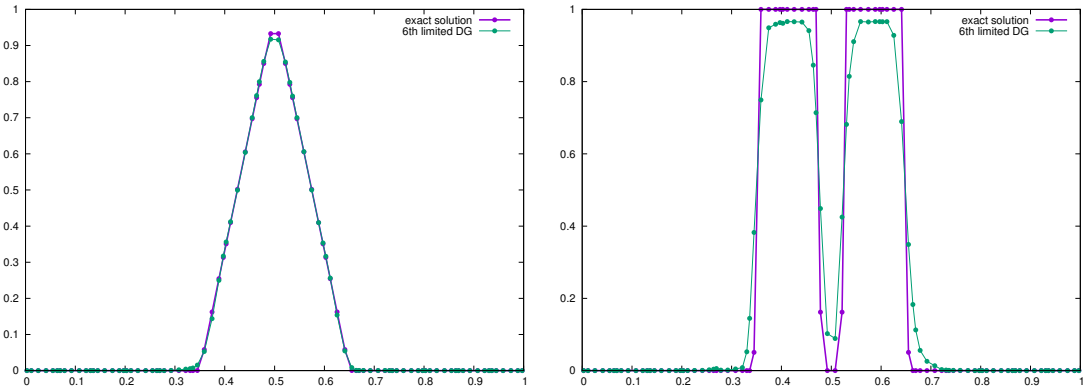
4.3.3. Burgers equation with a smooth initial solution

Let us consider the non-linear Burgers case $\partial_t u + \partial_x F(u) + \partial_y G(u) = 0$, where the fluxes write $F(u) = G(u) = \frac{1}{2}\,u^2$. Starting from the smooth initial condition $u^0(x) = sin(2\pi\,(x,y))$ on $[0,1]^2$, two stationary discontinuities form along the lines $\{(x,y) \in [0,1]^2,\, x+y = \frac{1}{2}\}$ and $\{(x,y) \in [0,1]^2,\, x+y = \frac{3}{2}\}$. It is worth mentioning that unlimited DG scheme crashes in this case. Indeed, as depicted in Figure 30, before the formation of the shocks the analytical solution is very well reproduced. But not long after the apparition of the two discontinuities, spurious oscillations amplify and make the code crash. In Figure 31, the numerical solution obtained with the 6th order limited DG scheme is plotted at time $t = 0.5$. Firstly, the computational code is now able to simulate the test problem until the final time, and the depicted results again prove that even in this extremely coarse mesh case, the corrected DG scheme is very precise as well as robust.

4.3.4. KPP problem

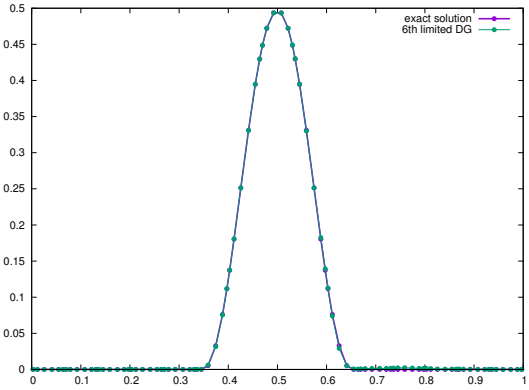
To close this numerical results section, we now turn our attention to non-linear conservation laws with non-convex fluxes. To this end, we consider the KPP problem proposed by Kurganov, Petrova, Popov (KPP) in [39] to test the convergence properties of some WENO schemes in the context of non-convex fluxes. For this particular problem, the flux functions are given by $F(u) = \sin(u)$ and $G(u) = \cos(u)$. Considering the computational domain $[-2,2] \times [-2.5, 1.5]$, the initial condition reads as follows

$$u^0(x) = \begin{cases} \frac{7\pi}{2} & \text{if } x < \frac{1}{2}, \\ \frac{\pi}{4} & \text{if } x > \frac{1}{2}. \end{cases}$$



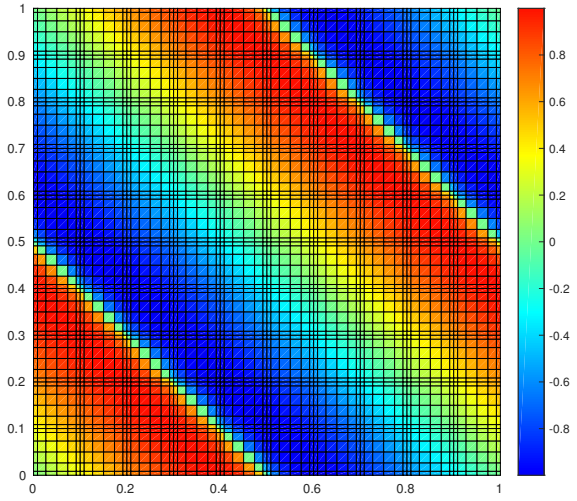
(a) Solution profile for $y = 0.25$.

(b) Solution profile for $y = 0.75$.

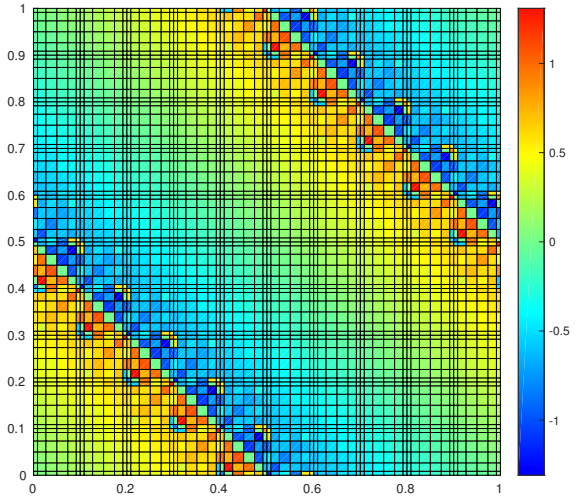


(c) Solution profile for $x = 0.25$.

Figure 29: 6th order corrected DG solution for the rigid rotation case on a 15×15 grid after one full rotation.

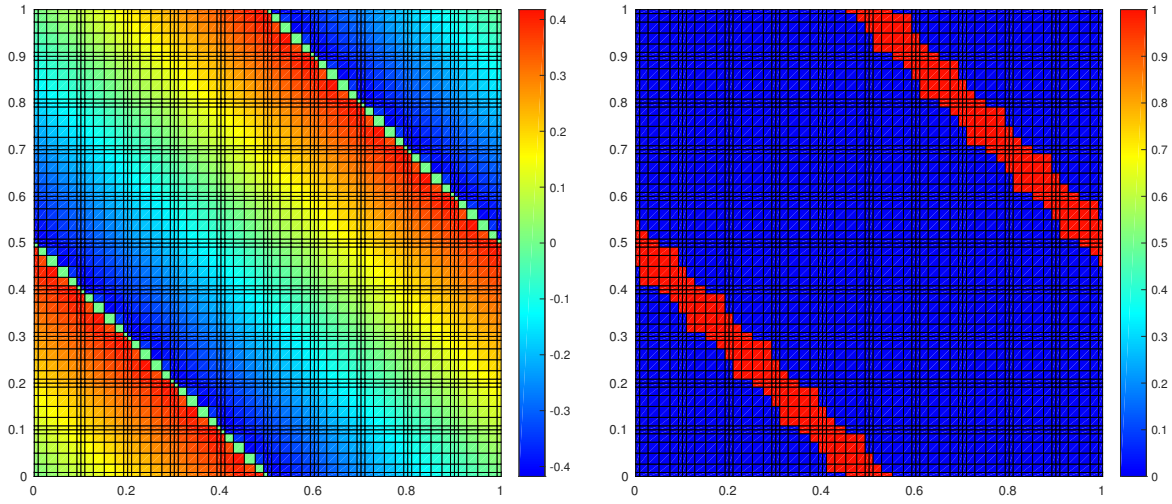


(a) $t = 0.007$.



(b) $t = 0.25$.

Figure 30: 6th order unlimited DG solution for 2D Burgers equation on a 10×10 mesh.



(a) Solution map. (b) Triggered subcells map.

Figure 31: 6th order corrected DG solution for 2D Burgers equation on a 10×10 mesh at time $t = 0.5$.

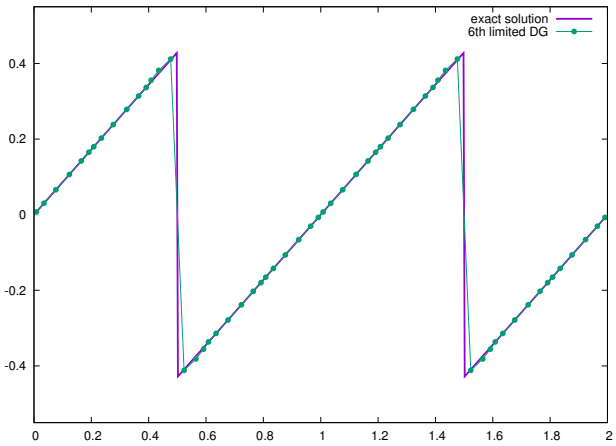


Figure 32: 6th order corrected DG solution for 2D Burgers equation on a 10×10 mesh at $t = 0.5$: profile along $x = y$.

This test is very challenging to many high-order schemes as the solution has a two-dimensional composite wave structure. And generally, to be able to capture such rotation composite structure, very fine grids are used. Here, by means of 6th order limited DG scheme, we make use of a 30×30 Cartesian mesh, which is very coarse in this quite complex situation. Let us note that similarly to the 1D Buckley test case, we use here global Lax-Friedrichs numerical flux as well as SubNAD criterion for the purpose of entropy. Results, displayed in Figure 33, are once more very satisfactory.

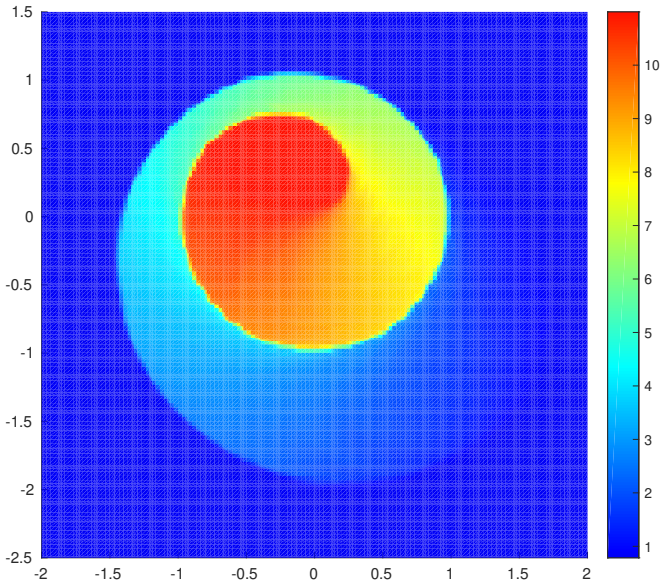


Figure 33: 6th order corrected DG solution for the KPP problem on a 30×30 at time $t = 1$: subcell mean values.

5. Conclusion

The aim of this paper is to present a new correction technique for discontinuous Galerkin schemes. This *a posteriori* correction procedure relies on the expression of DG schemes as a finite volume scheme on a subgrid. By means of this theoretical part, we modify at the subcell level the so-called reconstructed fluxes only where the unlimited DG scheme has failed. Consequently, only very few subcells require this particular treatment. For the remaining subcells, the submean values obtained through the unlimited DG method are kept, as they have been detected as admissible by the troubled zone criteria. This correction procedure allows us to retain as much as possible the very precise subcell resolution of DG schemes, along as addressing the issues of spurious oscillations, non-entropic behavior and aliasing phenomenon. A wide number of test cases have been used to depict the good performance and robustness of the presented correction technique.

Regarding the potential advantages of an *a posteriori* limiting strategy compared to *a priori* limiters, because the troubled zone detection is performed *a posteriori*, the correction can be done only where it is absolutely necessary. Furthermore, maximum principle preservation or positivity preservation is included without any additional effort, while it is generally not the case of *a priori* limitations. Any other property can be added as long as the admissibility set is convex (entropy can thus be added). Their scalability to any order of accuracy is also perfectly natural. Let us finally emphasize that the new correction procedure presented along this paper is totally parameter free.

In the future, we intend to extend this *a posteriori* correction technique to the unstructured grid case. To achieve such extension, we have to be able to rewrite DG schemes as a subcell finite volume scheme also on unstructured cells. Higher order subcell corrections, as WENO for instance, could also be investigated. We also plan to use this reconstructed flux correction framework in a Flux-Corrected Transport (FCT) frame, in order to obtain an automatic very high-order and monotonicity preserving scheme. Finally, the troubled zone detection should also be the topic of a paper on its own, with potentially incorporation of an entropy criterion.

Acknowledgment

The author acknowledges the financial support of the ANR-17-CE23-0019 Fast4HHO.
The author also warmly thanks Michael Dumbser (Univ. Di Trento), Gregor Gassner (Univ. of Cologne) and Raphaël Loubère (Univ. of Bordeaux) for the very fruitful discussions we had.

Appendices

A. Subresolution basis functions and correction coefficients

This appendix aims at giving further details related to the subresolution basis functions and the correction coefficients. First, let us introduce the following change of variable

$$\begin{aligned} [0,1] &\longrightarrow \omega_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}], \\ \xi &\longrightarrow x = x_{i-\frac{1}{2}} + \xi\,(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}). \end{aligned}$$

Let then set $\phi_m(x(\xi)) = \varphi_m(\xi) = \sum_{p=0}^{k} a_p^{(m)}\,\xi^p$, such that

$$\int_0^1 \varphi_m\,\xi^j\,\mathrm{d}\xi = \int_{\widetilde{\xi}_{m-1}}^{\widetilde{\xi}_m} \xi^j\,\mathrm{d}\xi, \qquad \text{for } \; j = 0,\dots,k.$$

These $k+1$ equations lead to the linear system $\mathsf{H}\,\boldsymbol{C}^{(m)} = \mathsf{\Lambda}\,\boldsymbol{D}^{(m)}$, where H is an Hilbert matrix, $\boldsymbol{C}^{(m)}$ is the vector containing the $a_p^{(m)}$ coefficients, $\mathsf{\Lambda}$ is the diagonal matrix such that $\mathsf{\Lambda}_{jj} = \frac{1}{j}$ and $\boldsymbol{D}^{(m)}$ is defined as follows

$$\boldsymbol{D}^{(m)} = \begin{pmatrix} \widetilde{\xi}_m - \widetilde{\xi}_{m-1} \\ (\widetilde{\xi}_m)^2 - (\widetilde{\xi}_{m-1})^2 \\ \vdots \\ (\widetilde{\xi}_m)^{k+1} - (\widetilde{\xi}_{m-1})^{k+1} \end{pmatrix}.$$

Recalling that Hilbert matrices are symmetric, hence also H^{-1} is. The subresolution basis functions finally read as follows

$$\varphi_m(\xi) = \boldsymbol{D}^{(m)} \cdot \mathsf{\Lambda}\mathsf{H}^{-1} \begin{pmatrix} 1 \\ \xi \\ \vdots \\ \xi^k \end{pmatrix}. \tag{A.1}$$

Let us note that the expression of the inverse of an Hilbert matrix is quite complex. However, recalling the correction coefficients definition

$$C_{i-\frac{1}{2}}^{(m)} = \sum_{p=m+1}^{k+1} \varphi_p(0) \qquad \text{and} \qquad C_{i+\frac{1}{2}}^{(m)} = \sum_{p=1}^{m} \varphi_p(1),$$

one can see that only the subresolution functions values in zero and one are needed. Moreover, if we consider a symmetric distribution of the flux points around the cell center, *i.e.* $\xi_p = 1 - \xi_{k+1-p}$,

for $p = 0, \ldots, k+1$, it is then easy to understand only $\varphi_p(0)$ is required, as $\varphi_p(1) = \varphi_{k+2-p}(0)$, for $p = 1, \ldots, k+1$. It also implies that $C^{(m)}_{i+\frac{1}{2}} = C^{(k+1-m)}_{i-\frac{1}{2}}$, for $m = 0, \ldots, k+1$. In the end, the correction coefficients can be put into the following expression

$$C^{(m)}_{i-\frac{1}{2}} = \left(\sum_{p=m+1}^{k+1} \boldsymbol{D}^{(m)} \right) \boldsymbol{.} \, \Lambda \mathsf{H}^{-1} \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{A.2}$$

Consequently, in regards to equation (A.2), only the first column of H^{-1} is needed. By introducing the vector $\boldsymbol{B} = \Lambda \mathsf{H}^{-1} (1, 0, \ldots, 0)^{\mathrm{t}}$, and noting that $\boldsymbol{B} \boldsymbol{.} (1, \ldots, 1)^{\mathrm{t}} = 1$, one gets the following very simple expression of $C^{(m)}_{i-\frac{1}{2}}$

$$C^{(m)}_{i-\frac{1}{2}} = 1 - \begin{pmatrix} \widetilde{\xi}_m \\ (\widetilde{\xi}_m)^2 \\ \vdots \\ (\widetilde{\xi}_m)^{k+1} \end{pmatrix} \boldsymbol{.} \, \boldsymbol{B}.$$

B. Two-dimensional extension

The multi-dimensional extension is carried out in a 1D tensor product manner. Even though only the 2D version of the scheme and limiter is presented here, the 3D case follows in a similar fashion. Let us first recall discontinuous Galerkin scheme in the case of a two-dimensional scalar conservation laws. Let $u = u(\boldsymbol{x}, t)$, for $\boldsymbol{x} \in \omega \subset \mathbb{R}^2$, and $t \in [0, T]$, be solution of the following two-dimensional scalar conservation laws

$$\begin{cases} \dfrac{\partial u}{\partial t} + \nabla \boldsymbol{.} \, \boldsymbol{F}(u) = 0, & (\boldsymbol{x}, t) \in \omega \times [0, T], \quad \text{(B.1a)} \\ u(\boldsymbol{x}, 0) = u^0(\boldsymbol{x}), & \boldsymbol{x} \in \omega, \quad \text{(B.1b)} \end{cases}$$

where u^0 is the initial data and $\boldsymbol{F}(u) = (F(u), G(u))^{\mathrm{t}}$ stands for the 2D flux function. Similarly to the 1D case, let $\{\omega_{i,j}\}_{i,j}$ be a partition of the computational domain ω. Here, $\omega_{i,j} = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}] \times [y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}]$ denotes a generic computational cell. We also introduce a partition of the time domain $0 = t^0 < t^1 < \ldots < t^n < \ldots < t^N = T$ and the time step $\Delta t^n = t^{n+1} - t^n$. In order to obtain a $(k+1)^{\text{th}}$ order discretization, let us consider a piecewise polynomial approximated solution $u_h(x, y, t)$, where its restriction to cell $\omega_{i,j}$, namely $u_{h|\omega_{i,j}} = u_h^{i,j}$, belongs to $\mathbb{P}^k_x([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}]) \times \mathbb{P}^k_y([y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}])$. Here, $\mathbb{P}^k_x$ and $\mathbb{P}^k_y$ are the set of polynomials of degree up to k respectively in variable x and y. The numerical solution $u_h^{i,j}$ then writes

$$u_h^i(x, y, t) = \sum_{m,n=1}^{k+1} u^{i,j}_{m,p}(t) \, \sigma^x_m(x) \, \sigma^y_p(y), \tag{B.2}$$

where $\{\sigma^x_m\}_m$ and $\{\sigma^y_p\}_p$ are basis respectively of $\mathbb{P}^k_x([x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}])$ and $\mathbb{P}^k_y([y_{j-\frac{1}{2}}, y_{j+\frac{1}{2}}])$. Performing a local variation formulation of equation (B.1) with any test function $\psi \in \mathbb{P}^k_x \times \mathbb{P}^k_y$, and introducing the numerical fluxes $\mathcal{F}$ and $\mathcal{G}$, one get the general form of DG schemes

$$\int_{\omega_{i,j}} \frac{\partial\, u_h^{i,j}}{\partial t}\,\psi\,\mathrm{d}V = \int_{\omega_{i,j}} \left(F(u_h^{i,j})\frac{\partial\,\psi}{\partial x} + G(u_h^{i,j})\frac{\partial\,\psi}{\partial y}\right)\,\mathrm{d}V - \int_{\partial\omega_{i,j}} \psi\ (\mathcal{F}\,n_x + \mathcal{G}\,n_y)\ \mathrm{d}S, \tag{B.3}$$

where $\boldsymbol{n} = (n_x, n_y)^{\mathrm{t}}$ is the unit outward normal of $\partial\omega_{i,j}$. On a Cartesian grid, this last expression can be rewritten as

$$\begin{aligned}\int_{\omega_{i,j}} \frac{\partial\, u_h^{i,j}}{\partial t}\,\psi\,\mathrm{d}V \ &= \int_{\omega_{i,j}} \left(F(u_h^{i,j})\frac{\partial\,\psi}{\partial x} + G(u_h^{i,j})\frac{\partial\,\psi}{\partial y}\right)\,\mathrm{d}V \\ &\quad - \int_{y_{j-\frac12}}^{y_{j+\frac12}} \Big[\mathcal{F}\,\psi\Big]_{x_{i-\frac12}}^{x_{i+\frac12}}\mathrm{d}y - \int_{x_{i-\frac12}}^{x_{i+\frac12}} \Big[\mathcal{G}\,\psi\Big]_{y_{j-\frac12}}^{y_{j+\frac12}}\mathrm{d}x.\end{aligned} \tag{B.4}$$

Following the 1D case steps, we will now rewrite DG schemes (B.4) as a finite volume scheme on a subgrid. To this end, we need to substitute the interior fluxes $F(u_h^{i,j})$ and $G(u_h^{i,j})$ with some polynomial counterparts $F_h^{i,j}$ and $G_h^{i,j}$, which can either be a collocated version of the interior fluxes as in nodal DG for instance, or their L_2 projection to fit the standard DG framework. Similarly to the 1D, see Remark 2.1, we consider interior polynomial fluxes $F_h^{i,j} \in \mathbb{P}_x^{\,\alpha}\times\mathbb{P}_y^k$ and $G_h^{i,j} \in \mathbb{P}_x^k\times\mathbb{P}_y^{\,\alpha}$, with $\alpha \in [\![k-1, k+1]\!]$.

Remark B.1. *Likewise the interior fluxes, we also need to replace in* (B.4) *the numerical fluxes* $\mathcal{F}_{i\pm\frac12}$ *and* $\mathcal{G}_{j\pm\frac12}$ *by some polynomial counterparts* $\mathcal{F}_h^{\,i\pm\frac12}$ *and* $\mathcal{G}_h^{\,j\pm\frac12}$ *respectively in* $\mathbb{P}_y^k$ *and* $\mathbb{P}_x^k$*, again through collocation or* L_2 *projection.*

Providing these polynomial fluxes, and by means of an integration by parts, DG schemes (B.4) rewrite

$$\begin{aligned}\int_{\omega_{i,j}} \frac{\partial\, u_h^{i,j}}{\partial t}\,\psi\,\mathrm{d}V \ &= -\int_{\omega_{i,j}} \psi\,\left(\frac{\partial\, F_h^{i,j}}{\partial x} + \frac{\partial\, G_h^{i,j}}{\partial y}\right)\mathrm{d}V \\ &\quad + \int_{y_{j-\frac12}}^{y_{j+\frac12}} \Big[\left(F_h^{i,j} - \mathcal{F}_h\right)\,\psi\Big]_{x_{i-\frac12}}^{x_{i+\frac12}}\mathrm{d}y - \int_{x_{i-\frac12}}^{x_{i+\frac12}} \Big[\left(G_h^{i,j} - \mathcal{G}_h\right)\,\psi\Big]_{y_{j-\frac12}}^{y_{j+\frac12}}\mathrm{d}x.\end{aligned} \tag{B.5}$$

The next step relies on the subdivision of cell $\omega_{i,j}$ into $(k+1)^2$ subcells, as in Figure B.34, where a generic subcell $S_{m,p}^{i,j}$ is defined as $S_{m,p}^{i,j} = [\widetilde{x}_{m-\frac12}, \widetilde{x}_{m+\frac12}]\times[\widetilde{y}_{p-\frac12}, \widetilde{y}_{p+\frac12}]$. Likewise the 1D case, we introduce similar subresolution basis functions $\{\phi_m^x(x)\}_{m=1,\dots,k+1}$ and $\{\phi_p^y(y)\}_{m=1,\dots,k+1}$. Setting in DG schemes (B.5) the test function to be $\psi(x,y) = \phi_m^x(x)\,\phi_p^y(y)$, it immediately follows that

$$\begin{aligned}|S_{m,p}^{i,j}|\,\frac{\partial\,\overline{u}_{m,p}^{i,j}}{\partial t} = \ &-\int_{\widetilde{y}_{p-\frac12}}^{\widetilde{y}_{p+\frac12}} \left(\Big[F_h^{i,j}(.,y)\Big]_{\widetilde{x}_{m-\frac12}}^{\widetilde{x}_{m+\frac12}} - \Big[\left(F_h^{i,j}(.,y) - \mathcal{F}_h(.,y)\right)\,\phi_m^x(.)\Big]_{x_{i-\frac12}}^{x_{i+\frac12}}\right)\mathrm{d}y \\ &-\int_{\widetilde{x}_{m-\frac12}}^{\widetilde{x}_{m+\frac12}} \left(\Big[G_h^{i,j}(x,.)\Big]_{\widetilde{y}_{p-\frac12}}^{\widetilde{y}_{p+\frac12}} - \Big[\left(G_h^{i,j}(x,.) - \mathcal{G}_h(x,.)\right)\,\phi_p^y(.)\Big]_{y_{j-\frac12}}^{y_{j+\frac12}}\right)\mathrm{d}x,\end{aligned} \tag{B.6}$$

where $|S_{m,p}^{i,j}| = |\widetilde{x}_{m+\frac12} - \widetilde{x}_{m-\frac12}|\times|\widetilde{y}_{p+\frac12} - \widetilde{y}_{p-\frac12}|$ is the subcell volume, and $\overline{u}_{m,p}^{i,j}$ the mean value of the polynomial solution $u_h^{i,j}$ on subcell $S_{m,p}^{i,j}$. Finally, we introduce the polynomial reconstructed fluxes $\widehat{F}_h^{i,j} \in \mathbb{P}_x^{k+1}\times\mathbb{P}_y^k$ and $\widehat{G}_h^{i,j} \in \mathbb{P}_x^k\times\mathbb{P}_y^{k+1}$ such that, for $m = 0,\dots,k+1$ and $p = 0,\dots,k+1$

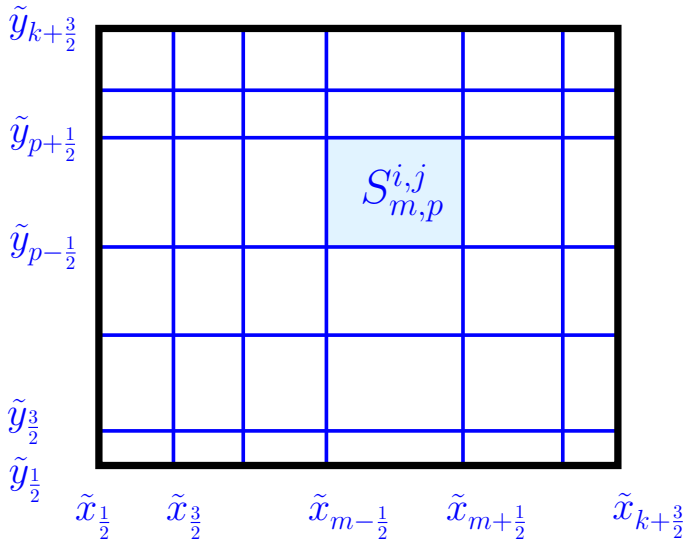


Figure B.34: Subdivision of cell $\omega_{i,j}$ into $(k+1)^2$ subcells.

$$\begin{cases} \widehat{F}_h^{i,j}(\widetilde{x}_{m+\frac{1}{2}},y) = F_h^{i,j}(\widetilde{x}_{m+\frac{1}{2}},y) - C_{m+\frac{1}{2}}^{i-\frac{1}{2}}\left(F_h^{i,j}(x_{i-\frac{1}{2}},y) - \mathcal{F}_h^{i-\frac{1}{2}}(y)\right) - C_{m+\frac{1}{2}}^{i+\frac{1}{2}}\left(F_h^{i,j}(x_{i+\frac{1}{2}},y) - \mathcal{F}_h^{i+\frac{1}{2}}(y)\right), \\ \widehat{G}_h^{i,j}(x,\widetilde{y}_{p+\frac{1}{2}}) = G_h^{i,j}(x,\widetilde{y}_{p+\frac{1}{2}}) - D_{p+\frac{1}{2}}^{j-\frac{1}{2}}\left(G_h^{i,j}(x,y_{j-\frac{1}{2}}) - \mathcal{G}_h^{j-\frac{1}{2}}(x)\right) - D_{p+\frac{1}{2}}^{j+\frac{1}{2}}\left(G_h^{i,j}(x,y_{j+\frac{1}{2}}) - \mathcal{G}_h^{j+\frac{1}{2}}(x)\right), \end{cases} \tag{B.7}$$

where the correction coefficients $C_{m+\frac{1}{2}}^{i\pm\frac{1}{2}}$ and $D_{p+\frac{1}{2}}^{j\pm\frac{1}{2}}$ write

$$\begin{cases} C_{m+\frac{1}{2}}^{i-\frac{1}{2}} = \displaystyle\sum_{q=m+1}^{k+1} \phi_q^x(x_{i-\frac{1}{2}}) \qquad \text{and} \qquad C_{m+\frac{1}{2}}^{i+\frac{1}{2}} = \sum_{q=1}^{m} \phi_q^x(x_{i+\frac{1}{2}}), \\ D_{p+\frac{1}{2}}^{j-\frac{1}{2}} = \displaystyle\sum_{q=p+1}^{k+1} \phi_q^y(y_{j-\frac{1}{2}}) \qquad \text{and} \qquad D_{p+\frac{1}{2}}^{j+\frac{1}{2}} = \sum_{q=1}^{p} \phi_q^y(y_{j+\frac{1}{2}}). \end{cases} \tag{B.8}$$

In the end, DG schemes (B.4) can be reformulated as a finite volume scheme on a subgrid, provided the fluxes defined in (B.7), as

$$\frac{\partial\, \overline{u}_{m,p}^{i,j}}{\partial t} = -\frac{1}{|S_{m,p}^{i,j}|}\left(\int_{\widetilde{y}_{p-\frac{1}{2}}}^{\widetilde{y}_{p+\frac{1}{2}}} \left[\widehat{F}_h^{i,j}(.,y)\right]_{\widetilde{x}_{m-\frac{1}{2}}}^{\widetilde{x}_{m+\frac{1}{2}}} \mathrm{d}y + \int_{\widetilde{x}_{m-\frac{1}{2}}}^{\widetilde{x}_{m+\frac{1}{2}}} \left[\widehat{G}_h^{i,j}(x,.)\right]_{\widetilde{y}_{p-\frac{1}{2}}}^{\widetilde{y}_{p+\frac{1}{2}}} \mathrm{d}x\right). \tag{B.9}$$

To end up with a more concise and elegant formulation of equation (B.9), let us introduce the following quantities

$$\begin{cases} \widehat{F}^{i,j}_{m+\frac{1}{2},p} = \dfrac{1}{|\widetilde{y}_{p+\frac{1}{2}} - \widetilde{y}_{p-\frac{1}{2}}|} \displaystyle\int_{\widetilde{y}_{p-\frac{1}{2}}}^{\widetilde{y}_{p+\frac{1}{2}}} \widehat{F}^{i,j}_h(\widetilde{x}_{m+\frac{1}{2}}, y)\,\mathrm{d}y, & \forall\, m \in [\![0, k+1]\!],\ p \in [\![1, k+1]\!], \\ \widehat{G}^{i,j}_{m,p+\frac{1}{2}} = \dfrac{1}{|\widetilde{x}_{m+\frac{1}{2}} - \widetilde{x}_{m-\frac{1}{2}}|} \displaystyle\int_{\widetilde{x}_{m-\frac{1}{2}}}^{\widetilde{x}_{m+\frac{1}{2}}} \widehat{G}^{i,j}_h(x, \widetilde{y}_{p+\frac{1}{2}})\,\mathrm{d}x, & \forall\, m \in [\![1, k+1]\!],\ p \in [\![0, k+1]\!]. \end{cases} \tag{B.10}$$

These definitions allow us to rewrite (B.9) in a more compact form as

$$\frac{\partial\, \overline{u}^{i,j}_{m,p}}{\partial t} = -\frac{\left(\widehat{F}^{i,j}_{m+\frac{1}{2},p} - \widehat{F}^{i,j}_{m-\frac{1}{2},p}\right)}{|\widetilde{x}_{m+\frac{1}{2}} - \widetilde{x}_{m-\frac{1}{2}}|} - \frac{\left(\widehat{G}^{i,j}_{m,p+\frac{1}{2}} - \widehat{G}^{i,j}_{m,p-\frac{1}{2}}\right)}{|\widetilde{y}_{p+\frac{1}{2}} - \widetilde{y}_{p-\frac{1}{2}}|}. \tag{B.11}$$

Now, by means of this new formulation of DG schemes (B.11), we can finally introduce in few words the 2D version of the *a posteriori* correction procedure. Making use of the same troubled zone detectors presented in the 1D configuration, if a subcell mean value $u^{n+1,\,i,j}_{m,p}$ computed through the unlimited DG scheme is marked as non-admissible, this value will be reevaluated through a first-order finite volume scheme. In other words, the reconstructed fluxes $\widehat{F}^{i,j}_{m-\frac{1}{2},p}$ and $\widehat{F}^{i,j}_{m+\frac{1}{2},p}$ will be respectively replaced by $\mathcal{F}(u^{n,i,j}_{m-1,p}, u^{n,i,j}_{m,p})$ and $\mathcal{F}(u^{n,i,j}_{m,p}, u^{n,i,j}_{m+1,p})$, and $\widehat{G}^{i,j}_{m,p-\frac{1}{2}}$ and $\widehat{G}^{i,j}_{m,p+\frac{1}{2}}$ replaced by $\mathcal{G}(u^{n,i,j}_{m,p-1}, u^{n,i,j}_{m,p})$ and $\mathcal{G}(u^{n,i,j}_{m,p}, u^{n,i,j}_{m,p+1})$. The face neighboring subcells also require to be recomputed for this correction procedure to ensure a conservation property at the subcell scale, see Figure B.35.

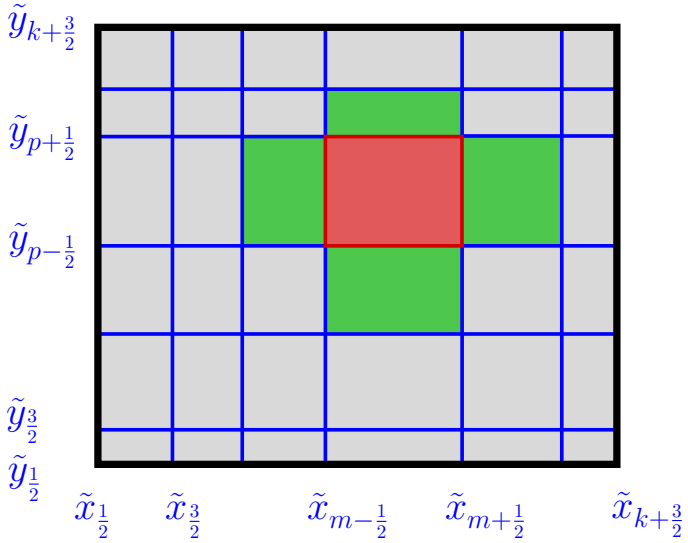


Figure B.35: Troubled subcell correction: in red the troubled subcell and in green the first face neighbors to also recompute.

Remark B.2. *It is worth mentioning that since* $\forall\, x,\ \widehat{F}_h^{i,j}(x,.) \in \mathbb{P}_y^k$, *the* $y-$*mean value* $\widehat{F}_{m,p}^{i,j}$ *defined in* (B.10) *can be modified without impacting the other* $y-$*mean values* $\widehat{F}_{m,q}^{i,j},\ \forall\, q \in [\![1, k+1]\!] \setminus \{p\}$. *Similar consideration holds for* $\widehat{G}_{m,p}^{i,j}$. *This is critical for the scheme conservation.*

References

[1] Y. Allaneau and A. Jameson. Connections between the filtered discontinuous Galerkin method and the flux reconstruction approach to high order discretizations. *Comput. Meth. Appl. Mech. Engrg.*, 200:3628–3636, 2011.

[2] D. Balsara, C. Altmann, C.D. Munz, and M. Dumbser. A sub-cell based indicator for troubled zones in RKDG schemes and a novel class of hybrid RKDG+HWENO schemes. *J. Comp. Phys.*, 226:586–620, 2007.

[3] R. Biswas, K. Devine, and J.E. Flaherty. Parallel adaptive finite element methods for conservation laws. *Applied Numerical Mathematics*, 14:255–284, 1994.

[4] A. Burbeau, P. Sagaut, and C.-H. Bruneau. A problem-independent limiter for high-order Runge Kutta discontinuous Galerkin methods. *J. Comp. Phys.*, 169:111–150, 2001.

[5] M. H. Carpenter, T. Fisher, E. Nielsen, and S. Frankel. Entropy Stable Spectral Collocation Schemes for the Navier–Stokes Equations: Discontinuous Interfaces. *SIAM J. Sci. Comput.*, 36:B835–B867, 2014.

[6] E. Casoni, J. Peraire, and A. Huerta. One-dimensional shock-capturing for high-order discontinuous Galerkin methods. *Int. J. Numer. Meth. Fluids*, 71:737–755, 2013.

[7] T. Chen and C.-W. Shu. Entropy stable high order discontinuous Galerkin methods with suitable quadrature rules for hyperbolic conservation laws. *J. Comp. Phys.*, 345:427–461, 2017.

[8] S. Clain, S. Diot, and R. Loubère. A high-order finite volume method for hyperbolic systems: Multi-dimensional Optimal Order Detection (MOOD). *J. Comp. Phys.*, 230:4028–4050, 2011.

[9] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws IV: The multidimensional case. *Math. Comp.*, 54:545–581, 1990.

[10] B. Cockburn, S. Hou, and C.-W. Shu. The Runge-Kutta Discontinuous Galerkin Method for Conservation Laws V: Multidimensional Systems. *J. Comp. Phys.*, 141:199–224, 1998.

[11] B. Cockburn, S.-Y. Lin, and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws III: One-dimensional systems. *J. Comp. Phys.*, 84:90–113, 1989.

[12] B. Cockburn and C.-W. Shu. TVB Runge-Kutta local projection discontinuous Galerkin finite element method for conservation laws II: General framework. *Math. Comp.*, 52:411–435, 1989.

[13] J. N. de la Rosa and C. D. Munz. Hybrid DG/FV schemes for magnetohydrodynamics and relativistichydrodynamics. *Comp. Phys. Commun.*, 222:113–135, 2018.

[14] S. Diot, S. Clain, and R. Loubère. 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.

[15] S. Diot, R. Loubère, and S. Clain. The MOOD method in the three-dimensional case: very-high-order finite volume method for hyperbolic systems. *Int. J. Numer. Meth. Fluids*, 73:362–392, 2013.

[16] J. Du, C.-W. Shu, and M. Zhang. A simple weighted essentially non-oscillatory limiter for the correction procedure via reconstruction (CPR) framework. *Appl. Num. Math.*, 95:173–198, 2015.

[17] M. Dumbser and R. Loubère. A simple robust and accurate a posteriorisub-cell finite volume limiter for the discontinuousGalerkin method on unstructured meshes. *J. Comp. Phys.*, 319:163–199, 2016.

[18] M. Dumbser, O. Zanotti, R. Loubère, and S. Diot. A posteriori subcell limiting of the discontinuous Galerkin finite element method for hyperbolic conservation laws. *J. Comp. Phys.*, 278:47–75, 2014.

[19] M. Feistauer, V. Dolejsi, and V. Kucera. On the discontinuous Galerkin method for the simulation of compressible flow with wide range of Mach numbers. *Comput. Vis. Sci.*, 10:17–27, 2007.

[20] T. C. Fisher and M. H. Carpenter. High-order entropy stable finite difference schemes for nonlinear conservation laws: Finite domains. *J. Comp. Phys.*, 252:518–557, 2013.

[21] G. Gassner. A Skew-Symmetric Discontinuous Galerkin Spectral Element Discretization and Its Relation to SBP-SAT Finite Difference Methods. *SIAM J. Sci. Comput.*, 35:1233–1253, 2013.

[22] H. Gao and Z. J. Wang. A conservative correction procedure via reconstruction formulation with the Chain-Rule divergence evaluation. *J. Comp. Phys.*, 232:7–13, 2013.

[23] G. J. Gassner, . R.Winters, and D. A. Kopriva. Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations. *J. Comp. Phys.*, 327:39–66, 2016.

[24] M. Gerritsma. Edge Functions for Spectral Element Methods. In *Spectral and High Order Methods for Partial Differential Equations*, pages 199–207. Springer, 2010.

[25] D. De Grazia, G. Mengaldo, D. Moxey, P. E. Vincent, and S. J. Sherwin. Connections between the discontinuous Galerkin method and high-order flux reconstruction schemes. *Int. J. Numer. Meth. Fluids*, 75:860–877, 2014.

[26] H. Ranocha and P. Öffner and T. Sonar. Summation-by-parts operators for correction procedure via reconstruction. *J. Comp. Phys.*, 311:299–328, 2016.

[27] H. Ranocha and P. Öffner and T. Sonar. Extended skew-symmetric form for summation-by-parts operators and varying Jacobians. *J. Comp. Phys.*, 342:13–28, 2017.

[28] A. Huerta, E. Casoni, and J. Peraire. A simple shock-capturing technique for high-order discontinuous Galerkin methods. *Int. J. Numer. Meth. Fluids*, 69:1614–1632, 2012.

[29] H. T. Huynh. A Flux Reconstruction Approach to High-Order Schemes Including Discontinuous Galerkin Methods. In *18th AIAA Computational Fluid Dynamics Conference, Miami*, 2007.

[30] H. T. Huynh. High-Order Methods Including Discontinuous Galerkin Methods by Reconstructions on Triangular Meshes. In *49th AIAA Aerospace Sciences Meeting, Orlando*, 2011.

[31] H. T. Huynh, Z. J. Wang, and P. E. Vincent. High-order methods for computational fluid dynamics: A brief review of compact differential formulations on unstructured grids. *J. Comp. Phys.*, 98:209–220, 2014.

[32] H. T. Huynh, Z. J. Wang, and P. E. Vincent. High-order methods for computational fluid dynamics: A brief review of compact differential formulations on unstructured grids. *J. Comp. Phys.*, 98:209–220, 2014.

[33] J. S. Park and C. Kim. Hierarchical multi-dimensional limiting strategy for correction procedure via reconstruction. *J. Comp. Phys.*, 308:57–80, 2016.

[34] J. S. Park and S.-H. Yoon and C. Kim. Multi-dimensional limiting process for hyperbolic conservation laws on unstructured grids. *J. Comp. Phys.*, 229:788–812, 2010.

[35] A. Jameson, P. E. Vincent, and P. Castonguay. On the Non-linear Stability of Flux Reconstruction Schemes. *J. Sci. Comput.*, 50:434–445, 2012.

[36] G.-S. Jiang and C.-W. Shu. On cell entropy inequality for discontinuous galerkin method for a scalar hyperbolic equation. *Mathematics of Computation*, 62:531–538, 1994.

[37] G.-S. Jiang and C.-W. Shu. Efficient implementation of weighted eno schemes. *J. Comp. Phys.*, 126:202–228, 1996.

[38] L. Krivodonova. Limiters for high-order discontinuous Galerkin methods. *J. Comp. Phys.*, 226:879–896, 2007.

[39] A. Kurganov, G. Petrova, and B. Popov. Adaptive semi-discrete central-upwind schemes for non convex hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 29:2381–2401, 2007.

[40] D. Kuzmin. A vertex-based hierarchical slope limiter for p-adaptative discontinuous Galerkin methods. *J. Comp. Appl. Math.*, 233:3077–3085, 2009.

[41] D. Kuzmin. Slope limiting for discontinuous Galerkin approximations with a possibly non-orthogonal Taylor basis. *Int. J. Numer. Meth. Fluids*, 71:1178–1190, 2013.

[42] B. Van Leer. Towards the ultimate conservative difference scheme. V-A second-order sequel to Godunov's method. *J. Comput. Phys.*, 32:101–136, 1979.

[43] R. J. LeVeque. High-resolution conservative algorithms for advection in compressible flow. *SIAM J. Numer. Anal.*, 33:627–665, 1996.

[44] L. Li and Q. Zhang. A new vertex-based limiting approach for nodal discontinuous Galerkin methods on arbitrary unstructured meshes. *Computers and Fluids*, 159:316–326, 2017.

[45] X. Meng, C.-W. Shu, Q. Zhang, and B. Wu. Superconvergence of discontinuous Galerkin method for scalar nonlinear conservation laws in one space dimension. *SIAM J. Numer. Anal.*, 50:2336–2356, 2012.

[46] P.-O. Persson and J. Peraire. Sub-cell shock capturing for discontinuous Galerkin methods. *AIAA paper*, 2006.

[47] J. Qiu and C.-W. Shu. Runge Kutta discontinuous Galerkin method using WENO limiters. *SIAM J. Sci. Comput.*, 26:907–929, 2005.

[48] R. Abgrall. A Review of Residual Distribution Schemes for Hyperbolic and Parabolic Problems: The July 2010 State of the Art. *Commun. Comput. Phys.*, 11:1043–1080, 2012.

[49] R. Abgrall. Some Remarks About Conservation for Residual Distribution Schemes. *Comput. Methods Appl. Math.*, 18:327–351, 2018.

[50] R. Abgrall and C.-W. Shu. Development of residual distribution schemes for the discontinuous Galerkin method: The scalar case with linear elements. *Commun. Comput. Phys.*, 5:376–690, 2019.

[51] R. Abgrall and E. le Mélédo and P. Öffner. On the connection between residual distribution schemes and flux reconstruction. hal-01820176, arXiv:1807.01261, 2018.

[52] W. H. Reed and T. R. Hill. Triangular Mesh Methods for the Neutron Transport Equation. Technical Report LA-UR-73-479, Los Alamos National Laboratory, 1973.

[53] N. Robidoux. Polynomial histopolation, superconvergent degrees of freedom and pseudospectral discrete hodge operators. Unpublished, 2008. Available at `http://people.math.sfu.ca/~nrobidou/public_html/prints/histogram/histogram.pdf`.

[54] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *J. Comp. Phys.*, 77:439–471, 1988.

[55] G. A. Sod. A survey of several finite difference methods for systems of non-linear hyperbolic conservation laws. *J. Comp. Phys.*, 27:1–31, 1978.

[56] M. Sonntag and C. D. Munz. Shock capturing for discontinuous Galerkin methods using finite volume subcells. In *Finite Volumes for Complex Applications VII*, pages 945–953. Springer, 2014.

[57] H. van der Ven and J.J.W. van der Vegt. Space time discontinuous Galerkin finite element method with dynamic grid motion for inviscid compressible flows, II. Efficient flux quadrature. *Comput. Meth. Appl. Mech. Eng.*, 191:4747–4780, 2002.

[58] F. Vilar, P.-H. Maire, and R. Abgrall. Cell-centered discontinuous Galerkin discretizations for two-dimensional scalar conservation laws on unstructured grids and for one-dimensional Lagrangian hydrodynamics. *Computers and Fluids*, 46(1):498–604, 2011.

[59] F. Vilar, P.-H. Maire, and R. Abgrall. A discontinuous Galerkin discretization for solving the two-dimensional gas dynamics equations written under total Lagrangian formulation on general unstructured grids. *J. Comp. Phys.*, 276:188–234, 2014.

[60] P. E. Vincent, P. Castonguay, and A. Jameson. A New Class of High-Order Energy Stable Flux Reconstruction Schemes. *J. Sci. Comput.*, 47:50–72, 2011.

[61] Z. J. Wang and H. Gao. A unifying lifting collocation penalty formulation including the discontinuous Galerkin, spectral volume/difference methods for conservation laws on mixed grids. *J. Comp. Phys.*, 228:8161–8186, 2009.

[62] M. Yang and Z.J. Wang. A parameter-free generalized moment limiter for high-order methods on unstrucured grids. *Adv. Appl. Math. Mech.*, 4:451–480, 2009.

[63] Y. Yang and C.-W. Shu. Analysis of optimal superconvergence of discontinuous Galerkin method for linear hyperbolic equations. *SIAM J. Numer. Anal.*, 50:3110–3133, 2012.

[64] X. Zhong and C.-W. Shu. A simple weighted essentially nonoscillatory limiter for Runge Kutta discontinuous Galerkin methods. *J. Comp. Phys.*, 232:397–415, 2013.

[65] J. Zhu, X. Zhong, C.-W. Shu, and J. Qiu. Runge Kutta discontinuous Galerkin method using a new type of WENO type limiters on unstructured meshes. *J. Comp. Phys.*, 248:200–220, 2013.