Limitation strategies for high-order discontinuous Galerkin schemes applied to an Eulerian model of polydisperse sprays

Katia Ait-Ameur, Mohamed Essadki, Marc Massot, Teddy Pichard

Limitation strategies for high-order discontinuous Galerkin schemes applied to an Eulerian model of polydisperse sprays

Katia Aït-Ameur* Mohamed Essadki† Marc Massot‡ Teddy Pichard‡

January 5, 2024

Abstract

In this paper, we tackle the modeling and numerical simulation of polydisperse sprays. Starting from a kinetic description for point particles, we focus on an Eulerian high-order geometric method of moment (GeoMOM) in size and consider a system of partial differential equations on a vector of successive fractional size moments of order 0 to $N/2$, $N > 2$, over a compact size interval. These moments correspond to physical quantities, which can be interpreted in terms of the geometry of the interface at small scale. There exists a stumbling block for the usual approaches using high-order moment methods resolved with high-order numerical methods: the transport algorithm does not naturally preserve the moment space. Indeed, reconstruction of moments by polynomials inside computational cells can create N-dimensional vectors which can fail to be moment vectors. We thus propose a new approach, as well as an algorithm, which is arbitrarily high-order in space and time with limited numerical diffusion, including at the boundaries of the state space, where a specific study is proposed. It allows to accurately describe the advection process and naturally preserves the moment space, at a reasonable computational cost. We show that such an approach is competitive compared to second order finite volume schemes, where limiters generate numerical diffusion and clipping at extrema. An accuracy study assesses the order of the method as well as the low level of numerical diffusion on structured meshes. We focus in this paper on cartesian meshes and 2D test cases are presented where the accuracy and efficiency of the approach are assessed.

1 Introduction

The present work aims at proposing and analyzing a high-order numerical scheme for a system of weakly hyperbolic conservation laws modelling sprays of droplets. In practice, this system is constructed by first considering a collisionless kinetic equation on a distribution function of droplet ([43, 18]), and then by extracting the first few moments with respect to the kinetic variables. The resulting system is underdetermined and it is closed using a quadrature-based approach ([40, 36, 9]). This construction has been widely used (see e.g. the previous work [20, 19, 14] and references therein) to model clouds of spherical droplets. More recently, this approach has also been exploited in [34] for the modelling of multi-scale flow with non-spherical liquid inclusion at small scale.

The considered moment system consists of a pressureless gas dynamics (PGD) system augmented with conservation laws on geometric moments. Therefore the study of the moment system exploit the one of the PGD: First, it is necessary to look for solutions in a weak sense ([5, 3, 6, 44, 4]) because, even with reasonable smooth initial and boundary conditions, this problem may involve measures, so-called delta-shocks transported with the flow (see also [52, 32, 30]). Second, the uniqueness of the weak solution is only ensured under additional constraints ([5]). Among those constraints, two properties of the initial and boundary

*INRIA, team LEMON / IMAG, Univ. Montpellier, CNRS, 860 Rue Saint Priest, 34095 Montpellier Cedex 5, France.

†The MathWorks, 2 Rue de Paris, 92196 Meudon.

‡Centre de Mathématiques Appliquées, CNRS, École polytechnique, Institut Polytechnique de Paris, Route de Saclay, 91128 Palaiseau Cedex, France.

values are preserved through space and time: the density remains non-negative and the velocity satisfies a maximum principle, which is closely related to the total variation diminishing (TVD) property. Therefore, these two properties needs also to be preserved by numerical schemes for stability reasons. Enforcing the positivity of the density in the PGD corresponds to enforcing that the solution remains in a convex set, called realizability domain or moment set (as it is the set of moments of the non-negative distributions ; [15, 47, 31, 16, 28]), for the moment model. Two types of solutions are difficult to capture by most numerical approaches, those involving concentration of the solution in a spatial point (delta-shock solution), and those involving void regions where the solution is zero as such a value belongs to the boundary of the realizability domain.

At the numerical level, a first order approach for the PGD preserving positivity of the density and the maximum principle on the velocity has been first proposed in [2] based on kinetic interpretation of the PGD, so-called kinetic finite-volume (KFV) scheme. It has been extended to second order in [7] using linear reconstructions with slope limiters to preserve the TVD property on the velocity, and to the considered moment model in [20]. However its construction is restricted to linear reconstructions, therefore to second order schemes usually involving clipping of extrema, since such a reconstruction can be interpreted as a convex combinaison of the value at each boundary of a cell, while higher order reconstructions do not satisfy such a property. When associated with a strong stability preserving (SSP) Runge Kutta (RK) time discretization ([49, 50, 24]), the discontinuous Galerkin approach (DG ; [11, 12, 17]) has been shown to be a good alternative to construct high-order discretisations for hyperbolic systems with discontinuous solutions.

The DG method produces accurate results if the solution is smooth or contains (relatively) weak discontinuities, otherwise significant oscillations and nonlinear instabilities may occur. To avoid such difficulties with numerical oscillations, the DG method needs to be accompanied by a limitation procedure such as minmod [13], artificial viscosity [35], total variation diminishing [27], weighted essentially nonoscillatory (WENO) [39, 45] techniques or extrema preserving limitations [58, 56]. In addition, there are other techniques for bound-preserving limiters, such as flux corrected transport algorithms [1] and convex limiting approaches [41, 25, 26, 33, 46]. There have been intensive studies on positivity-preserving and maximum-principle-satisfying methods. The genuinely high-order maximum-principle-satisfying DG method has been proposed in [58, 56] for scalar hyperbolic equations. This procedure has been rapidly developed for different problems ever since, for the Euler equations [59, 60], Navier-Stokes equations [57], shallow water equations [54] and fluid flow in porous media [10], among others. Exploiting a quadrature interpretation of the cell reconstructions, a pointwise limitation has been suggested in this framework. This limitation is rather simple to use, both in term of implementation and to obtain theoretical estimates. However, those estimates are only obtained under the constraint that the solution remains away from the void regions, even if this specific limit is interesting and frequently encountered in applications. Such a scheme has also been tested for the PGD in [55] and for another moment model in [48], which does not involve void or concentrated solutions.

The present work aims at constructing, analyzing and testing some limitation strategies for high-order RKDG discretisations applied to the considered weakly hyperbolic geometric moment system, thus combining the difficulties of moment space preservation within the framework PGD solutions, which can be potentially singular or involve void regions. More specifically, we study the impact of such a limitation in the vicinity of void regions. In practice, the considered limitations can be interpreted as projections of the discrete solution onto the set of admissibility, that is the set of vectors satisfying both the bounds on the velocity and the realizability condition on the geometric moments, while preserving the mean value of the solution in a cell. Two choices of projections are focused on: one consists in projecting all the solution toward the cell mean, the other consist in enforcing first the realizability then the overall admissibility. These two projections show different behavior in the void region in terms of accuracy and of numerical diffusion.

The paper is organized as follows. In the next section, the construction of the moment model from a kinetic description is recalled and we present its main features in more details. In Section 3, the DG scheme is presented and the discrete versions of the constraints are specified. Section 4 presents limitation strategies to preserve the realizability and the velocity bounds and their behavior in the vacuum limit is tested on a few test cases. A numerical study is provided in Section 5 to illustrate accuracy and the behavior of the numerical scheme with delta-shocks or vacuum solutions. The last section is devoted to concluding remarks.

2 High-order geometric moment modelling

We present here the construction of the system we aim at solving numerically in the next section, and analyze the properties of its solution we need to preserve at the numerical level.

2.1 Construction of the moment system

The considered system is obtained by evaluating the moments with respect to velocity and size variables of a kinetic model.

2.1.1 Kinetic description

The spray of droplets is described by a NDF $f(t,x,S,v)$, such that $f(t,x,S,v)\mathrm{d}x\,\mathrm{d}S\,\mathrm{d}v$ represents the probable number of droplets located in x, with size S and velocity v. The NDF satisfies a Williams-Boltzmann equation [53], that models the transport of a spray carried by a gaseous flow

$$\partial_t f + div_x(vf) = 0. \tag{1}$$

This model is a simple toy problem, but we aim at modelling a more realistic physics. This can be achieved in two manners from (1). First, we can enrich the description of the droplet, typically having a more precise geometry of the droplets, by considering a more complex phase space than only the size variable $S \in \mathbb{R}^+$ (e.g. modelling mean curvatures, temperature or oscillation of the droplets; see e.g. [34, 20, 19]). Second, considering other physical effects such as the drag and evaporation of the droplet (see e.g. [38, 19, 20] and references therein) can simply be modeled through additional terms in (1), potentially depending on these additional variables. However the present contribution in terms of numerical methods naturally extends to these more complex models as the main difficulties arise at the numerical level from the resolution of the transport operator.

2.1.2 Velocity moments

The kinetic phase space is composed of the size variable S and the velocity v. Concerning the velocity, for simplicity, we make the hypothesis that all droplets at a location x are transported at the same velocity u. This corresponds to approximating the distribution f by

$$f(t,x,S,v) \approx \rho(t,x,S)\delta(v-u(t,x)). \tag{2}$$

With such an approximation, extracting the first two moments of (1) with respect to v, that is integrating (1) against 1 and v yields

$$\partial_t \rho + div_x(\rho u) = 0, \tag{3a}$$
$$\partial_t q + div_x(qu^T) = 0, \tag{3b}$$

such that $q = \rho u$. This system corresponds to the pressureless gas dynamics (PGD) system, that has been widely studied in the literature (see e.g. [2, 5, 8] and references therein), and where the variable S appears as a parameter. Remark that the absence of pressure in our approach can be justified by the absence of collisions in the underlying kinetic model (1), that is the infinite Knudsen limit, which is valid in a lot a realistic configurations (see e.g. [37]).

2.1.3 Size moments

Concerning the size variable S, we use a fractional moment method ([20]). We use the half order moments $\mathbf{m} = (m_0, m_{1/2}, m_1, m_{3/2})^T$

$$m_\alpha(t,x) = \int_0^1 S^\alpha \rho(t,x,S)\mathrm{d}S, \tag{4}$$

where the maximum and minimum admissible sizes are chosen to be 0 and 1 for simplicity. The reason for this choice of $\mathbf{m}$ is that we retreive from those moments the following geometric quantities commonly used in the context of separated phase modeling ([18])

$$\Sigma\hat{G} = 4\pi m_0, \qquad \Sigma\hat{H} = 2\sqrt{\pi}m_{1/2}, \qquad \Sigma = m_1, \qquad \alpha = \frac{1}{6\sqrt{\pi}}m_{3/2}, \tag{5}$$

where the α is the volume fraction of liquid, Σ is the interfacial area density and $\hat{G}$ and $\hat{H}$ are respectively the densities of Gauss and mean curvatures averaged over the surface of a droplet ([34, 20, 19]).

Eventually, we extract the moments with respect to $b(S) := (S^0, S^{1/2}, S^1, S^{3/2})^T$ from (3a) and the moment with respect to S only from (3b) to obtain

$$\partial_t U + div_x(Uu^T) = 0, \qquad m_1 u = q, \tag{6}$$

where $U = (\mathbf{m}^T, q^T)^T$ with the moment vector $\mathbf{m} = (m_0, m_{1/2}, m_1, m_{3/2})^T$. The surface area density $\Sigma = m_1$ acts like a density in this model, and the moment against S of (3b) was used to construct q. Following [20], this choice is more relevant when considering drag or evaporation effects.

2.2 Properties of the moment system

The considered properties are presented for a 1D version of (6) as such a system can be decomposed into two subsystems widely studied in the literature. The analysis of those 1D subsystems provides some constraints on the solution, our numerical scheme has to satisfy. Eventually, we extend those constraints in a multi-D framework.

2.2.1 Bounds on u

The first subsystem rewritten in 1D is the PGD system on m_1 and $q = (m_1 u)$ which simply consists in a 1D version of (6) where the vector $\mathbf{m}$ is replaced by m_1, or of (3) replacing ρ by m_1 independent of S.

This system has been analyzed in [2, 7, 5] and we recall a few results here. One specificity of the PGD (3) is the possible appearance of so-called δ-shocks in the solution. It consists of a Dirac measure of mass m_1 transported at velocity u. For this purpose, one needs to focus on solutions the following weak sense (see [2]).

Definition 2.1. A couple $(m_1, q) \in C\left(]0,T[; \mathcal{M}_{loc}\left(\mathbb{R}\right)\right)^2$ is a duality solution to the 1D equation (3) if

- The mass $m_1 \geq 0$ is non-negative.
- There exists $u \in L^\infty(]0,T[\times\mathbb{R})$ and $\alpha \in L^1_{loc}(]0,T[)$ such that
  - One-sided-Lipschitz condition: $\partial_x u \leq \alpha$.
  - Weak solution: For all $\phi, \psi \in C_c^\infty(]0,T[\times\mathbb{R})$, then

$$\int m_1(\partial_t\phi + u\partial_x\phi) = 0, \quad \int q(\partial_t\psi + u\partial_x\psi) = 0.$$

  - Representation of u: $m_1 u = q$ a.e. with respect to the measure m_1.

Remark 1. In this definition, the velocity u corresponds to the Radon-Nikodym derivative of $q \equiv (m_1 u)$ with respect to m_1. It is therefore $L^\infty(\mathrm{d}m_1)$. It is defined only on $Supp(m_1)$, but this function extends in the complement $\mathbb{R}\backslash Supp(m_1)$ into some $L^\infty(]0,T[\times\mathbb{R})$ function (see the notion of universal representative in [2, 5, 3]).

Remark also that the requirements on u allow to define a unique characteristic curve X in the sense of Filippov ([44, 23]), i.e. absolutely continuous such that $X(t,x,t_0) = \int_{t_0}^t u(X(\tau,x,t_0))\mathrm{d}\tau$ and $X(t_0,x,t_0) = x$. For the numerical application in the next section, we exploit the following property.

Proposition 1 ([3]). *Consider a duality solution to the 1D equation* (3) *with initial data* m_1^0 *and* $(m_1 u)^0$. *Let us denote the essential infimum and supremum of a function* f *with respect to the measure* μ *on the interval* I *by* $\inf_I^\mu f$ *and* $\sup_I^\mu f$, *and define* $u_{\inf} := \inf_{\mathbb{R}}^{m_1^0} u$, $u_{\sup} := \sup_{\mathbb{R}}^{m_1^0} u$. *Then for all* $t > 0$, *globally in* $x \in \mathbb{R}$,

$$u_{\inf} \leq inf_{\mathbb{R}}^{m_1(t,\cdot)} u(t,\cdot), \qquad sup_{\mathbb{R}}^{m_1(t,\cdot)} u(t,\cdot) \leq u_{\sup}, \tag{7}$$

or for all $0 < \tau < t$, *locally around* $x \in (a,b)$ *with* $a < b$,

$$\begin{aligned} inf_{I^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot) &\leq inf_{I^t}^{m_1(t,\cdot)} u(t,\cdot), \\ sup_{I^t}^{m_1(t,\cdot)} u(t,\cdot) &\leq sup_{I^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot), \end{aligned} \tag{8}$$

where the intervals yield $I^\tau = (a - u_{\sup}(t-\tau),\ b - u_{\inf}(t-\tau))$.

These constraints are closely related to the total variation diminishing property (TVD; [27]) on velocity, which is also proved to be satisfied by u in [2]. Numerical schemes violating the discrete equivalent of this property can trigger oscillations or overshoots around discontinuities.

2.2.2 Preservation of initial data set

A second subsystem rewritten in 1D yields

$$\partial_t \mathbf{m} + \partial_x(\mathbf{m}u) = 0, \tag{9}$$

where $\mathbf{m} = (m_0, m_{1/2}, m_1, m_{3/2})^T$ is the vector of moments of n with respect to the basis functions $b(S) = (1, S^{1/2}, S, S^{3/2})^T$ and the velocity u is the one found in the previous paragraph from the PGD. We consider again weak solutions in the sense:

Definition 2.2. Duality solutions to (9) are $\mathbf{m} \in C(]0,T[; \mathcal{M}(\mathbb{R}))^4$ satisfying for all $\phi \in C_c^\infty(]0,T[\times\mathbb{R})$

$$\int m_\alpha(\partial_t\phi + u\partial_x\phi) = 0, \qquad \alpha = 0,\ 1/2,\ 1,\ 3/2.$$

The considered solution preserve the initial states:

Proposition 2. *Suppose that* m_α^0 *are dominated by* m_1^0 *and that* u *is obtained from a duality solution to the 1D equation* (3). *Then the duality solutions to* (9) *satisfy for all borel set* B *and* $t > 0$,

$$\mathbf{m}(t,B) \in Cone\left(\{\mathbf{m}^0(y),\ y \in \mathbb{R}\}\right),$$

where $Cone(\cdot)$ *is the convex cone pointed at the origin generated by all initial* $\mathbf{m}^0$.

Proof. This results from the method of characteristics in the sense of Filippov ([23, 44]). Remark that all the components m_α follow the same characteristic curve which provides the result. The requirement that m_α^0 is dominated by m_1^0 simply provides that $Supp(m_\alpha^0) \subset Supp(m_1^0)$ and therefore following the characteristics provides $Supp(m_\alpha(t,\cdot)) \subset Supp(m_1(t,\cdot))$ and assures the uniqueness of the velocity where m_α is non-zero. □

2.2.3 Moment set and Hankel determinants

This initial set is encompassed into a larger set, that is the set of moments of n with respect to $b(S)$, also called the realizability domain. This set of moments is often studied when constructing moment models because an important part of the physics is put into the nonlinear source terms, which are only defined and numerically evaluated under realizability constraints. This constraint extends the constraint $m_1 \geq 0$ in Definition 2.1.

Definition 2.3. The set of moments or realizability domain yields

$$\mathcal{R} = \left\{ \int_0^1 b(S)\mathrm{d}\mu(S), \qquad \mu \in \mathcal{M}([0,1]) \right\}.$$

This set is characterized by numerical constraints following Hausdorff problem.

Proposition 3. *The vector* $\mathbf{m} = (m_0, m_{1/2}, m_1, m_{3/2})^T \in \mathbb{R}^4$ *is realizable if the following matrices are symmetric non-negative*

$$H^1 = \begin{pmatrix} m_{1/2} & m_1 \\ m_1 & m_{3/2} \end{pmatrix}, \qquad H^2 = \begin{pmatrix} m_0 - m_{1/2} & m_{1/2} - m_1 \\ m_{1/2} - m_1 & m_1 - m_{3/2} \end{pmatrix}, \tag{10a}$$

and all their components are non-negative. This is equivalent to requiring their trace and determinants are non-negative, which reformulates $h_i(U) \geq 0$ *with*

$$\begin{aligned} h_1(U) = m_{1/2} + m_{3/2}, \quad h_2(U) = m_0 - m_{1/2} + m_1 - m_{3/2}, \\ h_3(U) = m_{1/2}m_{3/2} - m_1^2, \quad h_4(U) = (m_0 - m_{1/2})(m_1 - m_{3/2}) - (m_{1/2} - m_1)^2. \end{aligned} \tag{10b}$$

Proof. The fractional realizability condition simply follows from the solution of Hausdorff moment problem ([15, 31, 47, 16, 28, 42]) after using a change of variable $S = r^2$ in the integration. □

Eventually, integrating the solution U to (6) provides a local admissible set defined (by abuse of notations)

$$\left(\int_a^b \mathrm{d}U(t,y) \right) \in \mathcal{A}^\tau_{t,[a,b]} := \left\{ (\tilde{\mathbf{m}}^T, \tilde{q})^T \text{ s.t. } \quad \tilde{\mathbf{m}} \in \mathcal{R}, \quad \tilde{q} \in \left[\tilde{m_1} u^\tau_{\inf}(t),\ \tilde{m_1} u^\tau_{\sup}(t) \right] \right\}, \tag{11a}$$

$$u^\tau_{\inf}(t) = \inf\nolimits_{I^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot), \qquad u^\tau_{\sup}(t) = \sup\nolimits_{I^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot), \tag{11b}$$

as in (8) and where u satisfies $m_1 u = q$.

2.2.4 Extension to multi-D problems

We extend the framework for muldi-D problems, but the analysis is left for future work. When considering a problem of spatial dimension $d > 1$, we consider duality solutions under the following sense.

Definition 2.4. A couple $U = (\mathbf{m}^T, q^T)^T \in C(]0,T[; \mathcal{M}_{loc}(\mathbb{R}^d))^{4+d}$ is a duality solutions of (6) if:

- Every component $m_i \ll m_1$ is absolutely continuous w.r.t. m_1.
- The vector $\mathbf{m} \in \mathcal{R}$ is realizable m_1-a.e.
- There exists $u \in L^\infty(]0,T[\times\mathbb{R}^d)^d$ such that
  - Weak solution: For all component U_i and $\forall \phi \in C_c^\infty(]0,T[\times\mathbb{R}^d)$, then
  $$\int U_i \left(\partial_t \phi + u^T \nabla_x \phi \right) = 0.$$
  - Representation of u: $m_1 u = q$ is satisfied m_1-a.e.

Remark that an entropy condition à la Oleinik has been present in Definition 2.1 in the 1D case. It is missing in the present extension, and we do not perform a proper analysis of the multi-D model. A first result in this direction has been proposed in [6] for the transport equation and extension of this work to (6) is left for future work.

Assuming that there still exists a unique Filippov characteristics passing at every $(t,x) \in]0,T[\times\mathbb{R}^d$, then the admissible set extends in multi-D in the following way:

- The bound (8) on the velocity applies in every direction: for all $n \in \mathbb{S}^d$, at a global level for all $t > 0$,

$$(u^T n)_{\inf} \leq \inf_{\mathbb{R}^d}^{m_1(t,\cdot)} u(t,\cdot)^T n, \qquad \sup_{\mathbb{R}^d}^{m_1(t,\cdot)} u(t,\cdot)^T n \leq (u^T n)_{\sup},$$

or at a local level, for all $0 < \tau < t$ and $a_i < b_i$,

$$\inf_{C^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot)^T n \leq \inf_{C^{(t)}}^{m_1(t,\cdot)} u(t,\cdot)^T n,$$
$$\sup_{C^{(t)}}^{m_1(t,\cdot)} u(t,\cdot)^T n \leq \sup_{C^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot)^T n,$$

where $(u^T n)_{\inf} = \inf_{\mathbb{R}^d}^{m_1^0}(u^T n)$ and $(u^T n)_{\sup} = \sup_{\mathbb{R}^d}^{m_1^0}(u^T n)$, and we define the cell $C^\tau = \prod_i (a_i - (u^T e_i)_{\sup}(t-\tau), b_i - (u^T e_i)_{\inf}(t-\tau))$ that encompasses all possible feet of characteristics at time $t-\tau$ starting in C^0 at time t.

- The realizability $\mathbf{m} \in \mathcal{R}$ naturally extends if this vector is transported along characteristic curves.

This yields an admissible set:

$$\mathcal{A}_{t,C}^\tau := \left\{ (\tilde{\mathbf{m}}^T, \tilde{q}^T)^T \quad \text{s.t.} \quad \tilde{\mathbf{m}} \in \mathcal{R}, \quad \tilde{q}_i \in [\tilde{m}_1 (u^T e_i)_{\inf}^\tau(t),\ \tilde{m}_1 (u^T e_i)_{\sup}^\tau(t)] \right\}, \tag{12}$$
$$(u^T n)_{\inf}^\tau(t) = \inf_{C^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot), \qquad (u^T n)_{\sup}^\tau(t) = \sup_{C^{(t-\tau)}}^{m_1(t-\tau,\cdot)} u(t-\tau,\cdot).$$

2.2.5 Discussion on numerical difficulties

Two types of difficulties are focused on in the numerical section below:

- The appearance of void: at certain location, the moment solution can become zero. Such a value turns one of the inequality in (10) into an equality, and therefore the moment $\mathbf{m} = (m_0, m_{1/2}, m_1, m_{3/2})^t \in \partial\mathcal{R}$ belongs to the boundary of its admissible set. Furthermore, the velocity u is ill-defined in this limit because it only appears multiplied by m_1 in (6). Such an issue has been illustrated in Bouchut [2] through a 1D test case for the PGD, that we extend in the present framework into

$$m_\alpha^0(x) = \int_0^1 S^\alpha \mathrm{d}S = (1+\alpha)^{-1},\ q^0(x) = m_1^0(x) \times \begin{cases} 0.5 & x > 0, \\ -0.5 & x < 0. \end{cases} \tag{13}$$

  This m_α^0 is in the interior of $\mathcal{R}$ and generates a void region in finite time.

- The appearance of δ-shocks: the solution may contain Dirac measures that are propagated with the flow. Again such a solution has been exhibited in Bouchut through a test case rewritten into

$$m_\alpha^0(x) = (1+\alpha)^{-1}, \qquad q^0(x) = m_1^0(x) \times \begin{cases} -0.5 & x > 0, \\ 0.5 & x < 0. \end{cases} \tag{14}$$

The following section presents a high-order scheme preserving the bounds on the velocity, the moment space and capturing void and δ-shocks solutions.

3 RKDG scheme preserving admissibility

The aim is to construct high-order schemes to solve (6). In practice, those numerical schemes need to preserve discrete versions of the conditions (8) and (10). Among the most common approaches to solve hyperbolic PDE, we can list the finite volume schemes and the discontinuous Galerkin schemes as they are able to capture solutions with weak regularity. A version of the first has been proposed in [2] exploiting the underlying kinetic equation. However, the high-order extensions require to impose the conditions (8) and (10) everywhere to the polynomial reconstruction. This turns difficult at second order and prohibitive at third [7]. Here, we focus on the other approach, namely the discontinuous Galerkin schemes, for which admissibility need only to be imposed at finite locations.

3.1 DG space discretization

In order to construct the space discretization of (6), we first remark that the moment method of Section 2 is nothing but a Galerkin approximation of (1) with respect to the kinetic variables (S, v) using $\mathbf{b}$ for the test functions and (2) for approximation function. We extend the Galerkin approximation with space discretization.

For simplicity, we use a Cartesian grid $D = \bigcup_e \Omega_e$ where Ω_e is a product of intervals of the form $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$. Define polynomial test functions $g \in \mathbb{P}_r(\Omega_e)$ of degree r with respect to the space variable x and compute the integral

$$\begin{aligned} 0 &= \int_{\mathbb{R}^d}\int_0^1\int_{\Omega_e} g(x)\mathbf{b}(S,v)(\partial_t f + div_x(vf))(t,x,v,S)\,\mathrm{d}x\,\mathrm{d}S\,\mathrm{d}v \\ &= \frac{d}{dt}\int_{\mathbb{R}^d}\int_0^1\int_{\Omega_e} g(x)\mathbf{b}(S,v)f(t,x,v,S)\,\mathrm{d}x\,\mathrm{d}S\,\mathrm{d}v \\ &\quad - \int_{\mathbb{R}^d}\int_0^1\int_{\Omega_e} (\nabla_x g(x)^T v)\mathbf{b}(S,v)f(t,x,v,S)\,\mathrm{d}x\,\mathrm{d}S\,\mathrm{d}v \\ &\quad + \int_{\mathbb{R}^d}\int_0^1\int_{\partial\Omega_e} g(x)(v^T n(x))\mathbf{b}(S,v)f(t,x,v,S)\,\mathrm{d}x\,\mathrm{d}S\,\mathrm{d}v, \end{aligned} \tag{15}$$

where n denotes the outgoing normal to the boundary $\partial\Omega_e$. In the spirit of [2, 7], following the characteristic curves suggests to decompose f in the last integral into two parts coming from both side of the interface. Considering $x \in \Gamma_{ee'} = \Omega_e \cap \Omega_{e'}$ on the interface between Ω_e and $\Omega_{e'}$ and denoting n the normal directed toward $\Omega_{e'}$, this corresponds to writing

$$(v^T n(x))f(t,x,v,S) = \lim_{\substack{y\to x\\ y\in\Omega_e}} (v^T n(y))_+ f(t,y,v,S) + \lim_{\substack{y\to x\\ y\in\Omega_{e'}}} (v^T n(y))_- f(t,y,v,S), \tag{16}$$

where $a_\pm = (a \pm |a|)/2$ designate the positive or negative part of a. Now, following (2), we approximate f by f_h defined as

$$f_h(t,x,v,S) = \rho_h(t,x,S)\prod_j \delta_{u_{j,h}(t,x)}(v_j) = \sum_e \rho^e(t,x,S)\prod_j \delta_{u_j^e(t,x)}(v_j)\mathbf{1}_{\Omega_e}(x), \tag{17}$$

where the subscript h refers to the functions defined by parts and the superscript e refers to the functions in the cell Ω_e. The functions u^e and ρ^e are chosen such that the moments $x \mapsto U^e(t,x) = ((\mathbf{m}^e)^T, (q^e)^T(t,x)^T \in \mathbb{P}_r(\Omega_e)^{4+d}$

$$U^e(t,x) = \int_0^1\int_{\mathbb{R}^d} \mathbf{b}(S,v)\rho^e(t,x,S)\prod_j \delta_{u_j^e(t,x)}(v_j)\mathrm{d}S = \int_0^1 \mathbf{b}(S,u^e(t,x))\rho^e(t,x,S)\mathrm{d}S \tag{18}$$

are polynomials of degree r over the spatial cell Ω_e. Eventually, only the moment equation is solved, and the fact that U_e is polynomial is sufficient for the construction[1].

Now we choose the g_j such that it forms a basis of polynomials, which is orthogonal with respect to the L^2 scalar product on Ω_e. Denote x_k some quadrature points. In practice, we simply choose in 1D the Gauss-Lobatto points such that they include the boundary of each interval, and they maximize the accuracy in the sense that the space integrals are exact up to degree $2r-1$. Finally, denote l_k the Lagrange polynomials associated to these quadrature points. This structure (quadrature points and Lagrange polynomials) is simply tensorized in multi-D (see [56, 59] and references therein). Reinjecting it in (15) provides

$$0 = M\frac{d}{dt}U - F(U) + E(U), \tag{19a}$$

[1]Such an approximation can clearly be achieved in several ways, as such underlying ρ^e and u^e exists. For instance $\rho^e(t,x,S) = \sum_{i=1}^4 \alpha_i(t,x)\delta_{S_i}(S)$, $u_j^e(t,x) = p_j(t,x)/(\sum_{i=1}^4 \alpha_i(t,x)S_i)$, with $x \mapsto \alpha_i(t,x), p_j(t,x) \in \mathbb{P}_r(\Omega_e)$ and some distinct fixed $S_i \in (0,1)$ provides such U_e. Other choices are possible and this formula does not impact our construction.

where the unknown $(U_{j,k})_{j=1,\dots,4+d} = ((\mathbf{m}_k)^T, (q_k)^T)^T$ in this equation approximates $U(x_k) = ((\mathbf{m}^e)^T, (q^e)^T)(x_k)^T$ at the quadrature points x_k and

$$\left(M\frac{dU}{dt}\right)_{i,j} = \sum_k \left(\int_{\Omega_e} g_i(x) l_k(x) \mathrm{d}x\right) \frac{dU_{j,k}}{dt}, \tag{19b}$$

$$F(U)_{i,j} = \sum_k \left(\int_{\Omega_e} g_i(x) \nabla_x l_k(x)^T \mathrm{d}x\right) u_k U_{j,k}, \tag{19c}$$

where the velocity u_k satisfies $m_{1,k} u_k = q_k \approx q(x_k) = m_1(x_k) u(x_k)$. It is a scalar in 1D, or a vector in multi-D of the same size as $\nabla_x l_k(x)$. The case $m_{1,k} = 0$, which corresponds to the zero mass $m_1 = 0$ case, will be treated in the next section.

For the exchange term E, the boundary $\partial\Omega_e = \cup_{e'} \Gamma_{ee'}$ of the cell is splitted into the interfaces $\Gamma_{ee'} = \Omega_e \cap \Omega_{e'}$ and one remarks that the quadrature points x_k along $\Gamma_{ee'}$ in Ω_e are identical to those on the other side, along $\Gamma_{ee'}$ in $\Omega_{e'}$. Therefore, reinjecting (16) in the last integral of (15) and using the approximation (17) leads to (see also [2, 7])

$$E(U)_{i,j} = \sum_{\substack{e' \text{ s.t.} \\ \Gamma_{ee'} \neq \emptyset}} \sum_{\substack{k \text{ s.t.} \\ x_k \in \Gamma_{ee'}}} \left(\int_{\Gamma_{ee'}} g_i(x) l_k(x) \mathrm{d}x\right) \left[\left(u_k^T n(x_k)\right)_+ U_{j,k} + \left(u_{k'}^T n(x_k)\right)_- U_{j,k'}\right], \tag{19d}$$

where the index k refers to the quadrature points along the edge $\Gamma_{ee'}$ in Ω_e and the index $k' \neq k$ corresponds to the quadrature point in $\Omega_{e'}$ at the same location $x_{k'} = x_k \in \Gamma_{ee'}$. In 1D, this exchange terms reduces to Bouchut fluxes [2].

3.2 Numerical admissibility constraint

The conditions (10) and (8) satisfied by the duality solution at the continuous level need to be transposed at the discrete level.

Realizability For the realizability condition (10), despite providing a discretized version of the underlying kinetic model, the transposition of this condition at the discrete level is essentially driven by the applications we have in mind. Indeed, violating a discrete version of the pointwise realizability criteria (10) would not affect the precision nor the stability of the scheme, as a non-realizable vector would simply be transported at velocity u. The main motivation for imposing this constraint arise from the additional physical effects discussed in Section 2.1.1 that would require a strong imposition of this property.

Velocity bounds The preservation of monotonicity in the solution, and therefore the bounds on velocity, is closely related to the total variation diminishing property (TVD; [27]). In 1D, the velocity u has been shown to be total variation diminishing (TVD) in [2] at the continuous level and preserving the bounds on the total variation at the numerical level is essential for stability. These bounds are applied at the cell level Ω_e to the cell mean values.

Formulation of the requirement at the cell level First, denote $\bar{U}_e$ the integral of the approximation U^e in the cell Ω_e. Using the appropriate Gauss-Lobatto quadrature weigts $\omega_k > 0$

$$\bar{U}_e = \int_{\Omega_e} U^e(x) \mathrm{d}x = \sum_{\substack{k \text{ s.t.} \\ x_k \in \Omega_e}} \omega_k U_k \approx \int_{\Omega_e} \int_0^1 \int_{\mathbb{R}^d} \mathbf{b}(v, S) f(x, v, S) \mathrm{d}v \mathrm{d}S \mathrm{d}x,$$

which approximates the moments of f in the spatial cell Ω_e. Assuming that $U(t^n, \cdot)$ is of the form (18) at time t^n with positive $\rho(t^n, \cdot)$, then we expect the integral of exact solution to satisfy $\bar{U}_e^{n+1} \in \mathcal{A}^{\Delta t}_{t^n, \Omega_e}$ as

defined in (12) for all cell Ω_e. This rewrites:

$$\bar{\mathbf{m}}_e^{n+1} \in \mathcal{R}, \qquad \bar{q}_{j,e}^{n+1} \in \left[\bar{m}_{1,e}^{n+1}(u^T e_j)^n_{\min,e},\ \bar{m}_{1,e}^{n+1}(u^T e_j)^n_{\max,e}\right],$$
$$(u^T n)^n_{\min,e} = \min_{\substack{e' \text{ s.t.} \\ \Omega_e \cap \Omega_{e'} \neq \emptyset}} (\bar{u}^n_{e'})^T n, \qquad (u^T n)^n_{\max,e} = \max_{\substack{e' \text{ s.t.} \\ \Omega_e \cap \Omega_{e'} \neq \emptyset}} (\bar{u}^n_{e'})^T n,$$

where $\bar{u}_e^n$ satisfies $\bar{m}_{1,e}^n \bar{u}_e^n = \bar{q}_e^n$. Assuming that the time step satisfies a condition of the form $\Delta t \leq \max_e \bar{u}_e^n R(\Omega_e)$ where the radius $R(\Omega_e)$ is the maximum distance between two points of Ω_e, this corresponds to imposing (10) and (8) to the approximate solution f^h averaged in a cell Ω_e at time t^{n+1}.

Formulation of the requirement at the node level Exploiting the positivity of the quadrature weigts of Gauss-Lobatto and the convexity of the admissible set, $\bar{U}_e^{n+1} \in \mathcal{A}^{\Delta t}_{t^n,\Omega_e}$ holds if $U_k^{n+1} \in \mathcal{A}^{\Delta t}_{t^n,\Omega_e}$ holds for every quadrature point $x_k \in \Omega_e$, or equivalently

$$\mathbf{m}_k^{n+1} \in \mathcal{R}, \qquad q_{j,k}^{n+1} \in \left[m_{1,k}^{n+1}(u^T e_j)^n_{\min,e},\ m_{1,k}^{n+1}(u^T e_j)^n_{\max,e}\right], \tag{21}$$

implies (20). For the applications below, we impose (21) and we rewrite the discrete admissible set as

$$\mathcal{A}_e^n := \left\{U \ \text{ s.t. } \ \mathbf{m} \in \mathcal{R},\ q_j \in \left[m_1(u^T e_j)^n_{\min,e},\ m_1(u^T e_j)^n_{\max,e}\right]\right\},\ \mathcal{A}^n := \prod_e \mathcal{A}_e^n, \tag{22}$$

using velocities $(u^T e_j)^n_{\min,e}$ and $(u^T e_j)^n_{\max,e}$ given at a time step t^n in the cell Ω_e.

Imposition of the requirement A numerical scheme preserves admissibility if

$$U^n \in \mathcal{A}^{n-1} \quad \Rightarrow \quad U^{n+1} \in \mathcal{A}^n.$$

If admissibility is lost at some time step and at some quadrature point during the simulation, we correct this numerical solution in the following way: for $U \notin \mathcal{A}^n$, we define a corrected value $(\mathcal{P}^n U)$ as a projection onto $\mathcal{A}^n$ and it needs to satisfy

- $(\mathcal{P}^n U) \in \mathcal{A}^n$ is admissible,
- for all Ω_e,

$$\sum_{\substack{k \text{ s.t.} \\ x_k \in \Omega_e}} \omega_k U_k = \bar{U}_e = \sum_{\substack{k \text{ s.t.} \\ x_k \in \Omega_e}} \omega_k (\mathcal{P}^n U)_k. \tag{23}$$

The second criteria aims at imposing conservativity of the scheme. Indeed, the numerical scheme satisfied by the cell-averaged quantities $\bar{U}_e^{n+1}$ with the time discretizations of the next subsection can be rewritten in a conservative (finite volume) manner. Correcting the numerical solution at every time step such that (23) holds, this numerical scheme can still be written in a conservative way, but with modified fluxes.

In practice, we also expect the correction $\|U - \mathcal{P}^n U\|$ to be as small as possible in order not to deteriorate the accuracy of the scheme. Especially, the restriction $\mathcal{P}^n|_{\mathcal{A}^n}$ to the admissible set must be the identity. This additional error is studied in the next two sections.

Following the work of [56, 58, 59, 60], we consider corrections of the form:

$$(\mathcal{P}^n U)_{i,k} = U_{i,k} + \theta^n_{i,e}\left(\bar{U}_{i,e} - U_{i,k}\right), \tag{24}$$

where the convex combination parameters $\theta^n_{i,e}$ are such that $(\mathcal{P}^n U) \in \mathcal{A}^n$. They can be different for the different components i of the vector $U = (\mathbf{m}^T, q^T)^T$, for different cells Ω_e and for different time t^n. But they must be the same for every quadrature points $x_k \in \Omega_e$ among the same cell to satisfy (23).

3.3 SSP Runge-Kutta time discretization

Concerning the time discretization, we exploit the strong stability preserving (SSP) Runge-Kutta (RK) framework ([49, 50, 24]). Such schemes have been originally designed to preserve the TVD property while going higher order.

Rewriting the DG semi-discretization (19a) of the last subsection under the form

$$\frac{dU}{dt} = \mathcal{G}(U)_k = M^{-1}(F - E)(U)_k, \tag{25}$$

then the corrected explicit Euler scheme yields

$$U^{n+1} = \mathcal{P}^n \left(U^n + \Delta t \mathcal{G}(U^n) \right). \tag{26}$$

By construction, $\bar{U}_e^{n+1} = \bar{U}_e^n + \Delta t \overline{\mathcal{G}(U^n)}_e \in \mathcal{A}_e^n$ are admissible without correction and Euler scheme satisfies $(\bar{U}_e^{n+1} - \bar{U}_e^n)/\Delta t = O(\Delta t)$. The considered corrected m-stage SSPRK scheme take the form

$$\left\{ \begin{array}{rl} U^{(0)} & = U^n, \\ U^{(k)} & = \sum\limits_{i=0}^{k-1} \alpha_{i,k} \mathcal{P}^n \left(U^{(k)} + \Delta t \beta_{i,k} \mathcal{G}(U^{(k)}) \right) \quad \text{for } k = 1, \dots, m, \\ U^{n+1} & = U^{(m)}, \end{array} \right. \tag{27}$$

where $\alpha_{i,k}, \beta_{i,k} \geq 0$ and $\sum\limits_{i=0}^{k-1} \alpha_{i,k} = 1$. This scheme preserves admissibility since it is define as a convex combination of admissible values. The coefficients $\alpha_{i,k}, \beta_{i,k}$ are choosen such that if no correction is needed, that is if $\mathcal{P}^n = Id$ for all k in (27), then $(\bar{U}_e^{n+1} - \bar{U}_e^n)/\Delta t = O(\Delta t^p)$ at a certain order p. However, this order of accuracy is a priori not preserved. In this work, we use a SSPRK method of the same order as the one in space ([49, 50, 24, 51]).

4 Projection methods

The projections are defined locally as functions $\mathcal{P}$ depending on $U \in \mathbb{R}^{4+d}$ and $\bar{U} \in \mathcal{A}_e^n$, corresponding respectively to the quadrature value U_k^n and the cell value $\bar{U}_e^n$. Following (24), it takes the form

$$\mathcal{P}(U_i, \overline{U}_i) = \theta_i U_i + (1 - \theta_i)\overline{U}_i, \tag{28}$$

where the coefficients θ_i are defined below.

4.1 Componentwise combinations

We define values of θ_i to enforce the admissibility requirement. This condition rewrites $h_i(U) \geq 0$ for $i = 1, \dots, 4 + 2d$, the first four correspond to realizability, the last to the TVD of u. In practice, we impose $h_i(U) \geq \varepsilon > 0$ with a small value of ε in order to avoid admissibility loss due to round-off error. This parameter is fixed to 10^{-12} such that the admissibility domain is not too reduced, this was found sufficient for our applications.

4.1.1 Realizability projection

For $\overline{\mathbf{m}}$ and $\mathbf{m}$ such that $h_i(\overline{U}) > \varepsilon$ and $h_i(U) < \varepsilon$ for some $i = 1, \dots, 4$, we seek a convex combination parameter $\theta_i \in [0, 1]$ such that the projected values

$$\mathcal{P}^{real} U = \theta_i U + (1 - \theta_i)\overline{U},$$

satisfies $h_i(\mathcal{P}U) = \varepsilon$. It yields respectively

$$\theta_1 = \frac{\varepsilon - h_1(\overline{U})}{h_1(U - \overline{U})}, \qquad \theta_2 = \frac{\varepsilon - h_2(\overline{U})}{h_2(U - \overline{U})}, \tag{29a}$$

$$\theta_3 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \qquad \theta_4 = \frac{-\tilde{b} + \sqrt{\tilde{b}^2 - 4\tilde{a}\tilde{c}}}{2\tilde{a}} \tag{29b}$$

$$\text{with} \quad \left\{ \begin{array}{l} a = h_3(U - \overline{U}), \quad c = h_3(\overline{U}) - \varepsilon, \\ b = \overline{m_{1/2}}(m_{3/2} - \overline{m_{3/2}}) + (m_{1/2} - \overline{m_{1/2}})\overline{m_{3/2}} - 2\overline{m_1}(m_1 - \overline{m_1}), \end{array} \right.$$

$$\left\{ \begin{array}{l} \tilde{a} = h_4(U - \overline{U}), \qquad \tilde{c} = h_4(\overline{U}) - \varepsilon, \\ \tilde{b} = (\overline{m_0} - \overline{m_{1/2}})(m_1 - \overline{m_1} - m_{3/2} + \overline{m_{3/2}}) \\ \qquad +(\overline{m_1} - \overline{m_{3/2}})(m_0 - \overline{m_0} - m_{1/2} + \overline{m_{1/2}}) \\ \qquad -2(\overline{m_{1/2}} - \overline{m_1})(m_{1/2} - \overline{m_{1/2}} - m_1 + \overline{m_1}). \end{array} \right.$$

One verifies that each $\theta_i \in [0, 1]$ is well-defined as long as $h_i(U) < \varepsilon$ and $h_i(\overline{U}) > \varepsilon$.

4.1.2 TVD projection

Similarly, the constraints (22) on q rewrites $h_i(U) \geq 0$ for $i = 5, \ldots, 4 + 2d$ with:

$$h_{4+2j-1}(U) = q_j - m_1(u^T e_j)_{min}, \qquad h_{4+2j}(U) = m_1(u^T e_j)_{max} - q_j. \tag{30}$$

For $\overline{U}$ and $\hat{U}$ such that $h_j(\overline{U}) > \varepsilon$ and $h_j(\hat{U}) < \varepsilon$ for $j = 5, \ldots, 4 + 2d$, we define again a projection of the form

$$\mathcal{P}^{TVD}\hat{U} = \theta\hat{U} + (1 - \theta_j)\overline{U},$$

where $\theta_j \in [0, 1]$ is such that $h_j(\mathcal{P}\hat{U}) = \varepsilon$. This yields

$$\theta_j = \frac{\varepsilon - h_j(\overline{U})}{h_j(\hat{U} - \overline{U})}. \tag{31}$$

4.2 Assembling the projections

We need to have $m_1 \geq 0$ for the intervals $[m_1(u^T e_j)_{\min}, m_1(u^T e_j)_{\max}]$ not to be empty in (22). In practice, the first projection performed always ensures $\mathbf{m} \in \mathcal{R}$, which implies $m_1 \geq 0$.

4.2.1 Minimal projection

It yields the closest admissible vector to a non-admissible one $U \notin \mathcal{A}_e^n$ (red curve on Fig. 1):

$$\mathcal{P}_{\min}(U, \overline{U}) = \underset{V \in \mathcal{A}_e^n}{\operatorname{argmin}} ||U - V|| \tag{32}$$

This projection $\mathcal{P}_{\min}$ does not take the form (28) and does not depend on $\overline{U}$. Hence, this projection does not preserve the conservative character of the DG scheme. Therefore, it is not used afterward for the DG limitation, but only for accuracy comparisons because, assuming that $U \notin \mathcal{A}_e^n$ approximates $U^{ex} \in \mathcal{A}_e^n$ with a certain precision, then

$$\|\mathcal{P}_{\min}(U, \overline{U}) - U^{ex}\| \leq \|\mathcal{P}_{\min}(U, \overline{U}) - U\| + \|U - U^{ex}\| \leq 2\|U - U^{ex}\|. \tag{33}$$

4.2.2 Step-by-step Projection

Inspired of [59, 56], the projection is performed in two step. For this purpose, we exploit the convex cone properties of $\mathcal{A}_e^n$:

$$\begin{aligned} h_i(V) \geq 0 \quad &\Rightarrow \quad h_i(\alpha V) \geq 0 \quad \forall \alpha \geq 0, \\ h_i(V), h_i(W) \geq 0 \quad &\Rightarrow \quad h_i(\theta V + (1-\theta)W) \geq 0 \quad \forall \theta \in [0,1]. \end{aligned}$$

This property directly follows from the fact that $\mathcal{A}_e^n$ is a set of moments of positive distributions, that is a convex cone. This allows to first project $\mathbf{m}$ onto $\hat{\mathbf{m}} \in \mathcal{R}$, then to project $\hat{U} = (\hat{\mathbf{m}}^T, q^T)^T$ onto $\mathcal{A}_e^n$: We first define (blue curve in Fig. 1)

$$\mathcal{P}_1(\mathbf{m}, \overline{\mathbf{m}}) = \theta^1 \mathbf{m} + (1-\theta^1)\overline{\mathbf{m}}, \qquad \text{where} \quad \theta^1 = \min_{i=1,\dots,4} \theta_i, \tag{34}$$

where $\theta_i = 1$ if $h_i(U) \geq \varepsilon$ or equals (29) otherwise, such that $\mathcal{P}^1(\mathbf{m}, \overline{\mathbf{m}})$ satisfies (10). Remark that $\mathcal{P}_1$ matches the components of $\mathcal{P}^{real}$.

Then we project U onto $\mathcal{A}_e^n$ using the functions h_i for $i = 5, \dots, 4+2d$:

$$\mathcal{P}_{SbS}\left(U, \overline{U}\right) = \theta^2 \hat{U} + (1-\theta^2)\overline{U} \quad \text{where} \quad \hat{m} = \mathcal{P}_1(\mathbf{m}, \overline{\mathbf{m}}), \quad \hat{q} = q, \quad \theta^2 = \min_{i=5,\dots,4+2d} \theta_i, \tag{35}$$

where $\theta_i = 1$ if $h_i(\hat{U}) \geq \varepsilon$ or equals (31) otherwise, such that $\mathcal{P}_{SbS}\left(U, \overline{U}\right)$ satisfies (21). This second projection does not alter the realizability of $\mathbf{m}$ due to the convex cone property. Equivalently, this projection corresponds to fixing in (28) $\theta_i = \theta^2$ onto the component q (i.e. $i \geq 5$) and $\theta_i = \theta^1\theta^2$ (i.e. $i \leq 4$) onto the components $\mathbf{m}$. This projection has the form (28) and preserves the conservative character of the DG scheme. The accuracy is debated in Section 4.3.

4.2.3 Straight Projection

In a simpler manner, we can use the same parameter θ for every component (see green curve in Figure 1). This simply consists in

$$\mathcal{P}_{Str}(U, \overline{U}) = \theta^3 \hat{U} + (1-\theta^3)\overline{U}, \qquad \hat{U} = \theta^1 U + (1-\theta^1)\overline{U}, \tag{36}$$

where θ^1 is defined in (34) and θ^3 is such that $\mathcal{P}_{Str}(\hat{U}, \overline{U})$ satisfies (21). It yields $\theta^3 = \min_{i=5,\dots,4+2d} \theta_i$ with either $\theta_i = 1$ if $h_i(\hat{U}) \geq 0$ or it equals (31) otherwise. This projection $\mathcal{P}_{Str}$ shares the same properties as $\mathcal{P}_{SbS}$ according to the conservativity and accuracy. One remarks that $\mathcal{P}_{SbS}(U, \overline{U}) = \mathcal{P}_{Str}(U, \overline{U})$ if $\mathbf{m} \in \mathcal{R}$.

4.3 Behavior near the vacuum regime

This configuration corresponds to having $m_1 > 0$ close to 0. We study the properties of $\mathcal{P}_{SbS}(U, \overline{U})$ and $\mathcal{P}_{Str}(U, \overline{U})$ when $\overline{m_1} \to 0^+$. For simplicity, we conduct this study in 1D in the (m_1, q) plane where the admissibility property simplifies into:

$$\varepsilon \leq m_1, \qquad m_1 u_{\min} + \varepsilon \leq q \leq m_1 u_{\max} - \varepsilon. \tag{37}$$

4.3.1 The minimal projection

The projection $\mathcal{P}_{\min}$ can be derived analytically depending on the position of the non realizable point U in the set of non admissible states:

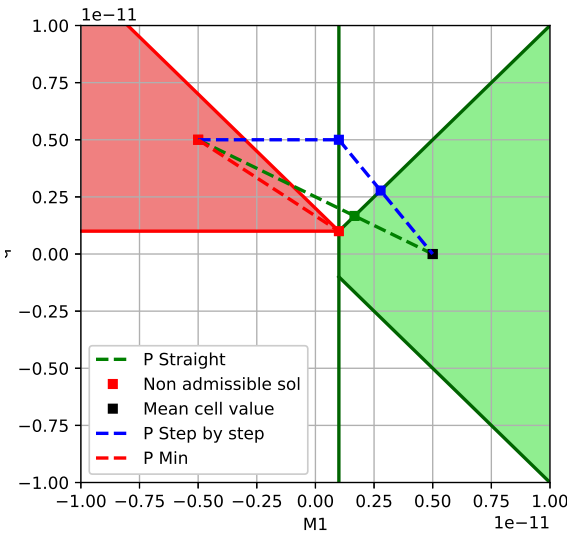

Figure 1: Cut in the (m_1, q) plane of the admissible set and the minimal, step-by-step and straight projections of a non-admissible vector straight with $\varepsilon = 10^{-12}$.

- When U satisfies $m_1 u_{max} - q \leq 0$ and $m_1 + q u_{max} \geq \varepsilon(1 + u_{max}^2)$ in yellow in Fig. 2, then $\mathcal{P}_{\min}(U, \overline{U})$ is the orthogonal projection onto the axis $q = m_1 u_{\max}$

$$\mathcal{P}_{min}(U, \overline{U}) = \left(m_1^{min}, m_1^{min} u_{max}\right)^T, \qquad m_1^{min} = \frac{m_1 + q u_{max}}{1 + u_{max}}.$$

- When U satisfies $m_1 + q u_{max} \leq \varepsilon(1 + u_{max}^2)$ and $q \geq u_{max}\varepsilon$ in red in Fig. 2, then $\mathcal{P}_{\min}(U, \overline{U}) = (\epsilon, \epsilon u_{\max})$.
- When U satisfies $0 \leq q \leq u_{max}\varepsilon$ in blue in Fig. 2, then $\mathcal{P}_{\min}(U, \overline{U}) = (\varepsilon, q)$.

The ones for negative q can be deduced by symmetry.

4.3.2 Qualitative comments for $\mathcal{P}_{SbS}$ and $\mathcal{P}_{Str}$

Concerning the two projections $\mathcal{P}_{SbS}$ and $\mathcal{P}_{Str}$, compared to $\mathcal{P}_{min}$, we can remark:

- When $m_1 \geq \varepsilon$, then $\mathcal{P}_{SbS}(U, \overline{U}) = \mathcal{P}_{Str}(U, \overline{U})$.
- When $0 \leq q \leq \varepsilon u_{max}$ (blue region), then $\mathcal{P}_{SbS}(U, \overline{U}) = \mathcal{P}_{\min}(U, \overline{U})$, while the value of $\mathcal{P}_{Str}(U, \overline{U})$ depend ons the location of $U \notin \mathcal{A}_e^n$ and of $\overline{U} \in \mathcal{A}_e^n$.
- When $q \geq \varepsilon u_{\max}$ and $m_1 \leq \varepsilon$, we have a priori $\mathcal{P}_{SbS}(U, \overline{U}) \neq \mathcal{P}_{Str}(U, \overline{U})$.

4.3.3 Quantitative estimations for $\mathcal{P}_{SbS}$ and $\mathcal{P}_{Str}$

Following the computations in [59, 56], if $U \notin \mathcal{A}_e^n$ is supposed to approximate $U^{ex} \in \mathcal{A}_e^n$ with a certain accuracy (typically $\mathcal{O}(\Delta x^p)$ in the DG framework), then

$$\begin{aligned} \|\mathcal{P}(U, \overline{U}) - U^{ex}\| &\leq \|(\mathcal{P} - \mathcal{P}_{min})(U, \overline{U})\| + \|\mathcal{P}_{min}(U, \overline{U}) - U\| + \|U - U^{ex}\| \\ &\leq \|(\mathcal{P} - \mathcal{P}_{min})(U, \overline{U})\| + 2\|U - U^{ex}\|. \end{aligned}$$

In [59, 56], the authors propose an estimation of this error under the condition that $\overline{U}$ is far from the boundary $\partial\mathcal{A}_e^n$. On the contrary, we study $\|(\mathcal{P} - \mathcal{P}_{min})(U, \overline{U})\|$ for the two projections $\mathcal{P}_{SbS}$ and $\mathcal{P}_{Str}$ when $\overline{U} \to (\varepsilon, \varepsilon u)$ with $u \in [u_{min}, u_{max}]$.

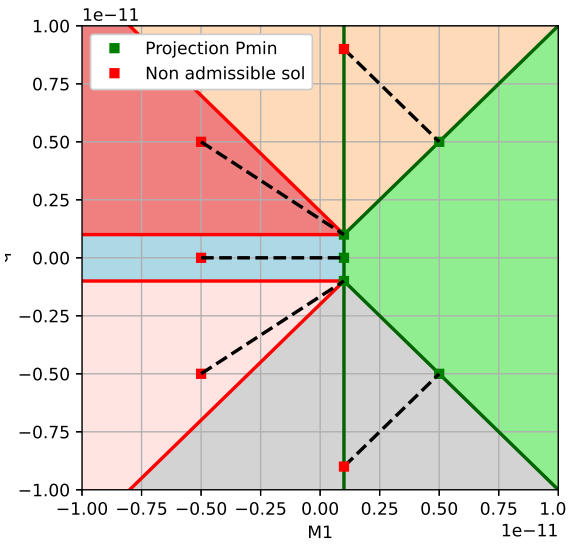

Figure 2: Projection $\mathcal{P}_{\min}$ applied to (m_1, q)

Concerning $\mathcal{P}_{SbS}$, one computes

$$\mathcal{P}_1(m_1, \overline{m_1}) = \epsilon, \quad \mathcal{P}_{SbS}\left(\hat{U}, \overline{U}\right) = \theta_q \hat{U} + (1-\theta_q)\overline{U}, \quad \theta_q = \frac{\overline{m_1} u_{\max} - \overline{q}}{q - \overline{q} - u_{\max}(\epsilon - \overline{m_1})}.$$

- Suppose that $m_1 + q u_{max} \leq \varepsilon(1 + u_{max}^2)$ and $q \geq u_{max}\varepsilon$ (red in Fig. 1) such that $\mathcal{P}_{min}(U, \overline{U}) = (\varepsilon, \varepsilon u_{max})$. This provides the estimation

$$\| \left(\mathcal{P}_{SbS} - \mathcal{P}_{\min}\right)(U, \overline{U})\| \leq \|(1, u_{max})\| \, |1 - \theta_q| \, (\overline{m_1} - \varepsilon),$$

  which tends to zero when $\overline{U} \to (\varepsilon, \varepsilon u)$ with $u \in [u_{min}, u_{max}]$.

- Suppose that $m_1 + q u_{max} \geq \varepsilon(1 + u_{max}^2)$ and $m_1 \leq \varepsilon$ (yellow in Fig. 1). Then

$$\| \left(\mathcal{P}_{SbS} - \mathcal{P}_{\min}\right)(U, \overline{U})\| \leq \|(1, u_{max})\| \left|\theta_q m_1 + \overline{m_1}(1 - \theta_q) - m_1^{min}\right|,$$

  which tends to $\|(1, u_{max})\| |m_1^{min} - \varepsilon|$ when $\overline{U} \to (\varepsilon, \varepsilon u)$ and which is non-zero. In this region, one needs further assumptions on U to control this error. This is typically done in the DG framework by exploiting the value of an exact solution U^{ex} assumed to be close to U.

Concerning $\mathcal{P}_{Str}$, one computes

$$\mathcal{P}_{Str}\left(U, \overline{U}\right) = \theta U + (1-\theta)\overline{U}, \qquad \theta = \min\left(\theta_{m_1}, \theta_q\right), \tag{38}$$
$$\theta_{m_1} = \frac{\epsilon - \overline{m_1}}{m_1 - \overline{m_1}}, \qquad \theta_q = \frac{\overline{m_1} u_{\max} - \overline{q}}{q - \overline{q} - u_{\max}(\epsilon - \overline{m_1})}.$$

We distinguish the two cases $\theta = \theta_{m_1}$ and $\theta = \theta_q$ (see Fig. 3) depending on the location of $\overline{U}$ and U.

- Suppose that $m_1 + q u_{max} \leq \varepsilon(1 + u_{max}^2)$ and $q \geq u_{max}\varepsilon$ (red in Fig. 1) and $\theta = \theta_{m_1}$, then

$$\left\|\left(\mathcal{P}_{Str} - \mathcal{P}_{\min}\right)(U, \overline{U})\right\| = \varepsilon |u - u_{\max}|, \qquad u = \frac{\theta q + (1-\theta)\overline{q}}{\theta m_1 + (1-\theta)\overline{m_1}},$$

  which is always considered negligible.

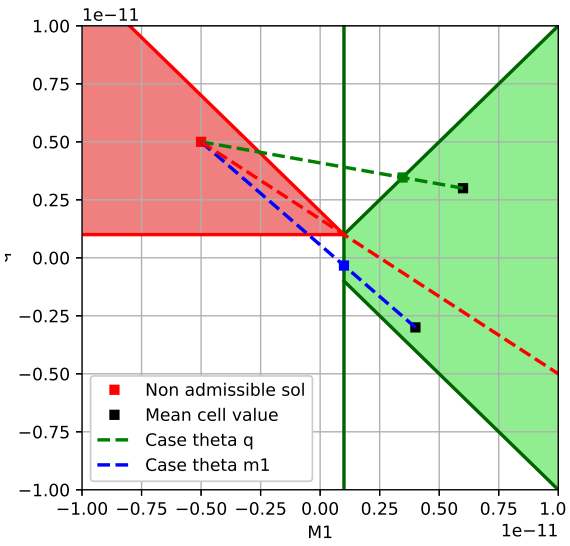

Figure 3: Projection $\mathcal{P}_{Str}$ depending on the location of $\overline{U}$

- Suppose that $m_1 + qu_{max} \leq \varepsilon(1+u_{max}^2)$ and $q \geq u_{max}\varepsilon$ (red in Fig. 1) and $\theta = \theta_q$, then

$$\left\|\left(\mathcal{P}_{Str} - \mathcal{P}_{\min}\right)(U,\overline{U})\right\| \leq \|(1,u_{\max})\| \left|m_1^{Str} - \varepsilon\right|, \quad m_1^{Str} = \theta m_1 + (1-\theta)\overline{m_1}.$$

  Since $\epsilon \leq m_1^{Str} \leq \overline{m_1}$, then this tends to zero when $\overline{m_1} \to \varepsilon$.

- Suppose that $m_1 + qu_{max} \geq \varepsilon(1+u_{max}^2)$ and $m_1 \leq \varepsilon$ (yellow in Fig. 1) and $\theta = \theta_{m_1}$, then

$$\left\|\left(\mathcal{P}_{Str} - \mathcal{P}_{\min}\right)(U,\overline{U})\right\| \leq \varepsilon|u - u_{max}| + u_{max}|\varepsilon - m_1^{min}|.$$

  The first term $\varepsilon|u-u_{max}|$ is negligible, but the second term $|\varepsilon - m_1^{min}|$ is a priori not controled and one needs again further assumptions on $\mathcal{P}_{min}(U,\overline{U})$.

- Suppose that $m_1 + qu_{max} \geq \varepsilon(1+u_{max}^2)$ and $m_1 \leq \varepsilon$ (yellow in Fig. 1) and $\theta = \theta_q$, then

$$\left\|\left(\mathcal{P}_{Str} - \mathcal{P}_{\min}\right)(U,\overline{U})\right\| \leq \|(1,u_{max})\||m_1^{Str} - m_1^{min}|$$

- Suppose that $0 \leq q \leq \varepsilon u_{max}$ and $\theta = \theta_{m_1}$, then $\mathcal{P}_{Str}(U,\overline{U}) = \mathcal{P}_{min}(U,\overline{U})$.

- Suppose that $0 \leq q \leq \varepsilon u_{max}$ and $\theta = \theta_q$, then

$$\left\|(\mathcal{P}_{min} - \mathcal{P}_{Str})(U,\overline{U})\right\| \leq \varepsilon|u - u_{max}| + u_{max}|m_1^{Str} - \varepsilon|,$$

  The first term $\varepsilon|u-u_{max}|$ is negligible and, since $\varepsilon \leq m_1^{Str} \leq \overline{m_1}$, the second term $|\varepsilon - m_1^{Str}| \to 0$ in the limit $\overline{m_1} \to \varepsilon$.

4.3.4 Comparison of m_1 for both projections in the vacuum limit

Suppose now that $\varepsilon \leq \overline{m_1} \leq \varepsilon(1+\delta)$ for some small $\delta > 0$ and that $q \geq \varepsilon(1+\delta)u_{max}$. Then, one easily verifies that

$$m_1^{Str} \leq m_1^{SbS}.$$

This comparison is observed in the numerical examples below and may result in larger numerical diffusion effects on m_1 with $\mathcal{P}_{SbS}$ than with $\mathcal{P}_{Str}$.

5 Numerical validation, accuracy and performance assessment

The strength of the present approach is the availability to reach high-order accuracy (restricted to second order in [29, 22, 21]) while remaining robust with void regime and singularities. This is illustrated in this section through three representative test-cases, with increasing difficulties. The first is a smooth 1D case to study the accuracy of the method. The second and third cases test the robustness of the method respectively when void or δ-shocks appear. The last test cases extend the void and δ-shock studies in a 2D framework.

Our numerical results are compared with a kinetic finite volume scheme (KFV; [29, 22, 21]). It is a finite volume scheme with fluxes (19d) and a MUSCL linear reconstruction. A minmod limiter based on u and the so-called canonical moments formulation is used. It consists in a non-linear transformation of the realizability domain $\mathcal{R}$ onto $[0,1]^4$. However, this strategy is limited to second order.

5.1 Accuracy study for a 1D spray

The first initial condition yields

$$m_\alpha(x,0) = \int_0^1 S^\alpha G(x,1/2)\mathrm{d}S = (\alpha+1)^{-1} G(x,1/2), \qquad q(x,0) = -m_1(x,0), \tag{39}$$

for $\alpha = 0, 1/2, 1, 3/2$, where $G(x,x_c) = \exp\left(-(x-x_c)^2/\sigma^2\right)$ with $\sigma = 0.1$. Periodic conditions are used at the boundary of the $[0,1]$-domain. With such a field, all the moments are transported at velocity $u=-1$. Fig. 4 (left) shows the numerical solution m_0 (the other moments show identical features) with 100 cells at time $t=2$, i.e. after two periods. Fig. 4 (right) shows the relative l^1-distance of the numerical solutions m_0^N obtained with the different schemes to the exact solution at final time $t^N = 2$ as a function of Δx

$$\varepsilon(\Delta x) = \frac{\sum_e \sum_q \omega_q \left| m_{0,q}^N - m_0(x_q, t=2)\right|}{\sum_e \sum_q \omega_q m_0(x_q, t=2)}.$$

Both limitations of Section 4 have no impact for this simulation and give identical result. Therefore, only one is shown and the RKDG schemes yield the desired order. The KFV scheme [29] with the minmod limitation yields a slightly slower convergence than the theoretical second order, and the maximum is clipped. Without limitation, the RKDG and the KFV schemes yield a similar underlying spatial reconstruction. The difference in the behavior is due to the considered limitation and the minmod limitation on the canonical moments affects more the local extrema of the solution.

5.2 Vacuum test case

We extend a test case from [7] in the present moment framework (see (13)). We use the initial condition:

$$m_\alpha(x,0) = (\alpha+1)^{-1},\ q(x,0) = m_1(x,0) \times \begin{cases} -0.4 & \text{if } 0.5 < x \text{ or } x > 1.8, \\ 0.4 & \text{if } 0.5 < x < 1, \\ 1.4 - x & \text{if } 1 < x < 1.8. \end{cases} \tag{40}$$

The first gap in the initial velocity is meant to trigger the low density region and periodic boundary conditions are used.

Fig. 5 displays the numerical solutions m_0 (left) and q (right) at $t^N = 0.5$ with 100 cells. Vacuum is created at the location of the velocity discontinuity in the initial condition. The KFV scheme is more diffusive in this region. Fig. 6 show zooms on this void region with the different schemes and Table 1 presents the minimal value of m_1 obtained with all limitations. The safety parameter used in the limitations (29) and (31) is set to $\varepsilon = 10^{-12}$. At second order with low N, the two projections show identical minimal value of m_1, this corresponds to the case where only one limitation applies. In all the other simulations, the straight limitation $\mathcal{P}^{Str}$ (36) shows lower values of m_1 compared to the step-by-step one $\mathcal{P}^{SbS}$ (35), as observed in

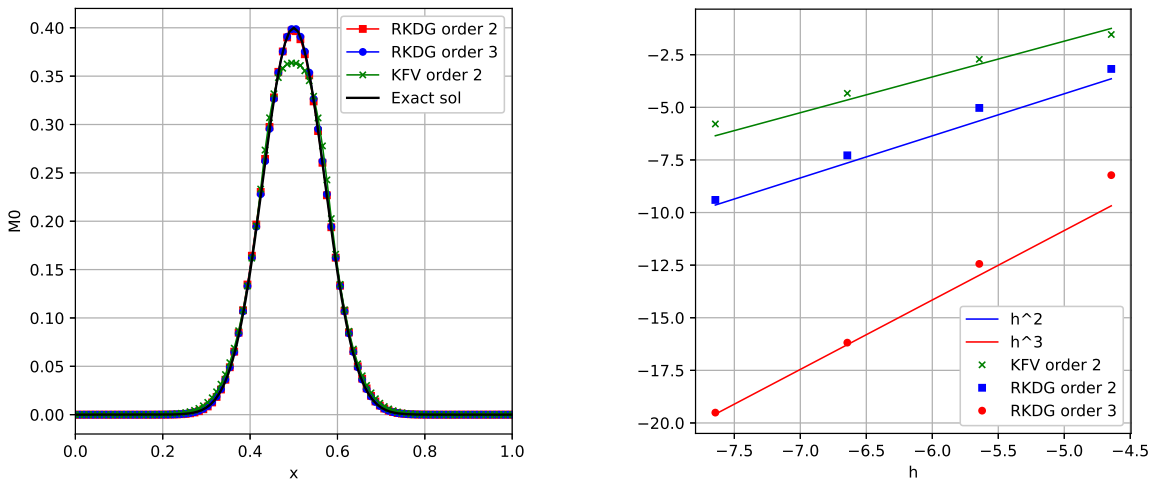

Figure 4: Moment m_0 (left) obtained with the 2nd and 3rd order RKDG schemes and the 2nd order KFV scheme [29] at $t^N = 2$ with the initial condition (39). Relative l^1 distances (right) between the numerical solutions m_0^N at $t^N = 2$ and the exact solution.

Subsection 4.3.4. The even orders, i.e. odd order polynomial reconstructions, show lower values of m_1 in the void region. As in the previous test case, the profile in the higher density region is sharper with the third order scheme and more diffused with the KFV scheme. However, the RKDG schemes present larger overshoots on the sides of this profile and in the middle of this region. This was already observed in [7]. This effect reduces when raising the order of accuracy. As a summary, the present high-order scheme remains

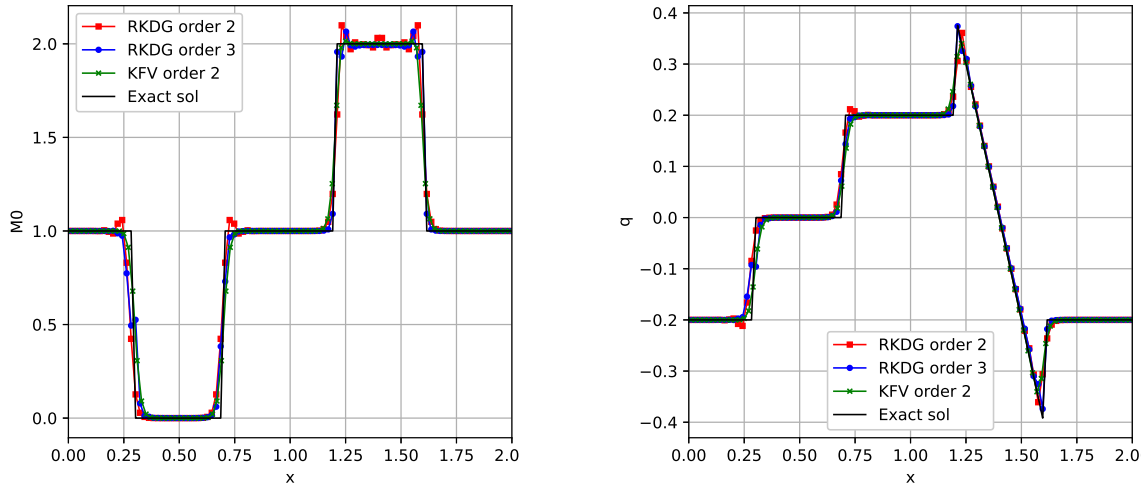

Figure 5: Moments m_0 (left) and q (middle) obtained with the RKDG schemes and the KFV scheme at $t = 0.4$ with the initial conditions (40).

robust when considering solutions with very low m_1.

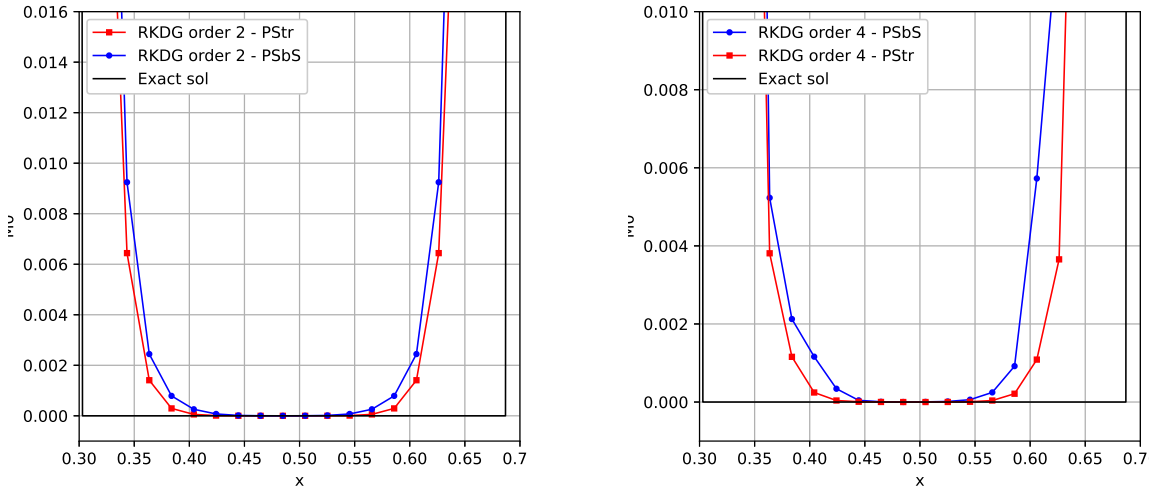

Figure 6: Zoom of the m_0-plot in the vacuum region $x \in [0.3, 0.7]$.

Mesh size	RKDG Order 2		RKDG Order 3	
	$\mathcal{P}_{SbS}$	$\mathcal{P}_{Str}$	$\mathcal{P}_{SbS}$	$\mathcal{P}_{Str}$
$N = 25$	8.015×10^{-3}	8.015×10^{-3}	1.225×10^{-2}	1.225×10^{-2}
$N = 50$	6.497×10^{-5}	6.497×10^{-5}	6.297×10^{-5}	6.510×10^{-5}
$N = 100$	4.221×10^{-9}	4.221×10^{-9}	2.813×10^{-6}	8.459×10^{-6}
$N = 200$	1.775×10^{-12}	1.403×10^{-12}	6.881×10^{-7}	5.688×10^{-8}

Mesh size	RKDG Order 4		KFV Order2
	$\mathcal{P}_{SbS}$	$\mathcal{P}_{Str}$	
$N = 25$	3.188×10^{-3}	5.096×10^{-7}	5.859×10^{-3}
$N = 50$	2.281×10^{-3}	1.448×10^{-5}	6.103×10^{-5}
$N = 100$	2.460×10^{-7}	$10^{-12} = \varepsilon$	5.587×10^{-9}
$N = 200$	2.060×10^{-12}	$10^{-12} = \varepsilon$	$10^{-12} = \varepsilon$

Table 1: Minimal value of m_1 with the different schemes and limitations in the vacuum test case.

5.3 1D δ-shock test case

We extend another test case from [7] using:

$$\begin{aligned} m_\alpha(x,0) &= \int_0^1 S^\alpha \left(G(x_1,x) + \frac{4}{3} G(x_2,x) \mathbf{1}_{[0.5,1]}(S) \right) \mathrm{d}S \\ &= (\alpha+1)^{-1} \left(G(x_1,x) + \frac{4}{3}(1 - 0.5^{\alpha+1}) G(x_2,x) \right), \end{aligned} \tag{41a}$$

$$q(x,0) = m_1(x,0) \left(\mathbf{1}_{\mathbb{R}^-}(x - 0.5) - x \right), \tag{41b}$$

where the Gaussians' parameters are $\sigma = 0.075$, $x_1 = 0.15$ and $x_2 = 0.85$. This initial condition is plotted in Fig. 7. The coefficient $4/3$ is chosen such that m_1 is symmetric in the domain, and since u is antisymmetric, such a configuration triggers a stationnary δ-shock. Remark that the monokinetic assumption (2) is not valid at this location, and the kinetic solution to (1) is composed of the sum of the two distributions defined in (41a) crossing each others at the velocity given (41b). On the contrary for the moment solution to (6), the two masses do not cross each others, but enter into a stationnary δ-shock located in $x = 0.5$. This δ-shock is the sum of the masses coming from both sides of the shock which are symmetric for m_1 but not for m_0, $m_{1/2}$ and $m_{3/2}$. Fig. 8 shows m_0 and m_1 with the different schemes at $t^N = 0.4$. The limitations

mainly activate in the region where the two masses enter in the δ-shock. In this region, the moments $\mathbf{m}$ are far from the boundary $\partial\mathcal{R}$. Therefore, the two limitations gives again identical values and only one (those with $\mathcal{P}_{Str}$) is shown in Fig. 8. Due to the shape of the initial velocity (see Fig. 7, right), it is crucial to impose velocity bounds that are computed locally (as in (8)) rather than globally (as in (7)) in order to filter spurious oscillations. One observes furthermore in Fig. 8 (right) that the velocity profile with the RKDG schemes is less diffused compared to the one with the KFV scheme. Finally, we compare in Table 2 the value of the moment vector inside the δ-shocks with the exact solution. Straightforward computations leads to an exact value

$$m_\alpha^\delta = (\alpha+1)^{-1}\left(\int_{1/6}^{1/2} G(x,x_1)\mathrm{d}x + \frac{4}{3}(1-0.5^{\alpha+1})\int_{1/2}^{5/6} G(x,x_2)\mathrm{d}x\right).$$

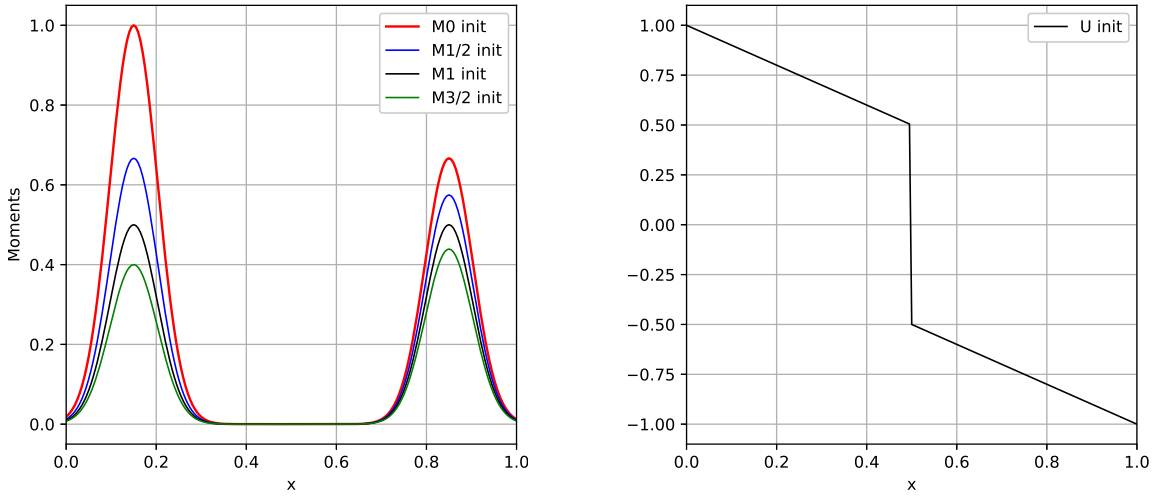

Figure 7: Initial conditions (41) on m_α (left) for $\alpha = 0$ (red), $1/2$ (blue), 1 (black), $3/2$ (green) and on $u \equiv q/m_1$ (right).

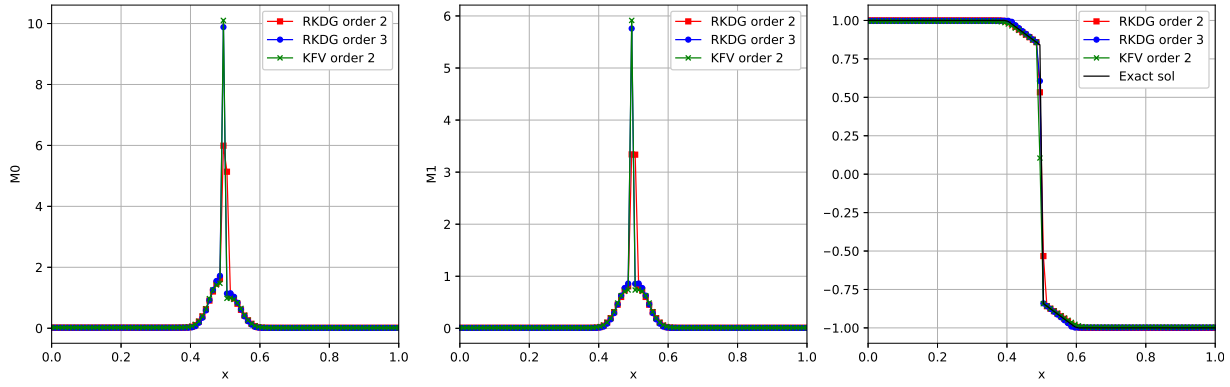

Figure 8: Moments m_0 (left) and m_1 (middle) and velocity profile $u = q/m_1$ obtained with the RKDG schemes and the KFV scheme at $t = 0.4$ with the initial conditions (41).

	m_0	$m_{1/2}$	m_1	$m_{3/2}$	q
Exact	0.083	0.062	0.05	0.042	0
RKDG Order 2	0.1112	0.0828	0.0667	0.056	0.0105
RKDG Order 3	0.0988	0.0722	0.0575	0.0479	0.0071
KFV Order 2	0.101	0.0740	0.0591	0.0493	0.0062

Table 2: Moment vector inside the δ-shock with the different schemes.

5.4 2D δ-shock test case

This case corresponds to (6) with $U = (\mathbf{m}^T, q^T)^T$ where $q = (q_1, q_2)^T$ and $x = (x_1, x_2)^T$ are two-dimensional. The initial condition are:

$$m_\alpha(x,0) = (\alpha+1)^{-1}, \qquad q(x,0) = -0.25 \times m_1(x,0) \times (\text{sign}(x_1),\ \text{sign}(x_2))^T . \tag{42}$$

The computational domain $[-1,1]^2$ is meshed with 100^2 uniform cells with $\Delta x = \Delta y = 1/50$. Dirichlet boundary conditions are imposed. The final time is $t^N = 0.5$. The initial configuration has a uniform mass m_1 over the domain, and the initial velocity profile (42) is such that δ-shocks form along the axis $x = 0$ and $y = 0$. This symmetric setup also leads to a mass concentration at the origin. Fig. 9 shows the numerical solution m_1 obtained with second order RKDG scheme with the straight projection limitation (36). The

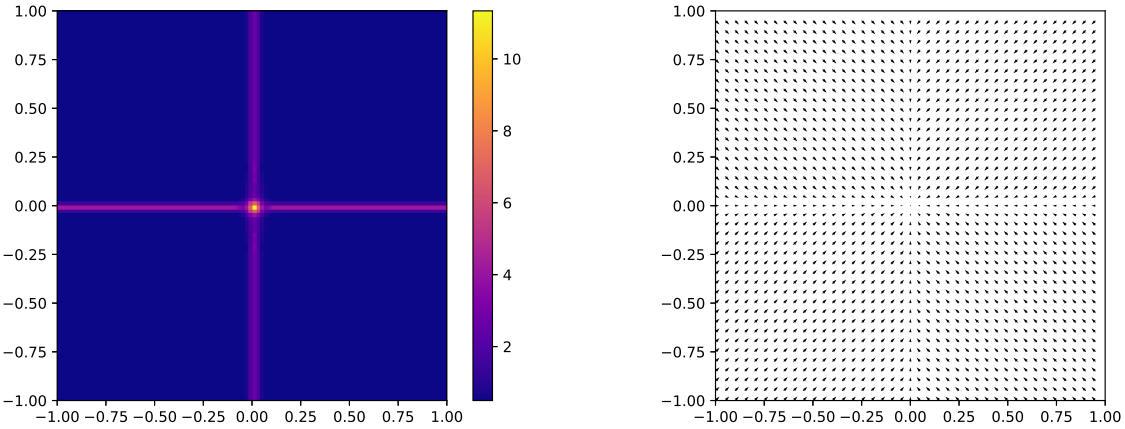

Figure 9: Moment m_1 (left) and velocity field (right) at $t^N = 0.5$ with the initial condition (42).

result is similar to the 1D case in Subsection 5.3. As expected, the mass accumulates along the axes and at the origin. The robustness and stability of the RKDG scheme with the limitation procedure are preserved in 2D even in presence of δ-shock singularities.

This case is modified in Section SM1.1 and SM1.2 of the supplementary material with asymmetric δ-shocks or δ-shocks not aligned with the mesh.

5.5 2D δ-vacuum test case

Eventually, we consider the initial condition:

$$m_\alpha(x,0) = (\alpha+1)^{-1}, \qquad q(x,0) = 0.4 \times m_1(x,0) \times (\text{sign}(x_1),\ \text{sign}(x_2))^T . \tag{43}$$

The numerical parameters are identical to the previous case. The initial mass m_1 is uniform. The initial velocity is directed outwards the axes $x = 0$, $y = 0$ such that it generates vacuum region $m_\alpha \approx 0$ around these two axes. Fig. 10 m_0 with the second order RKDG scheme with the straight limitation (36). The numerical results agree qualitatively with the expected solution. This case is modified to generate a vaccuum not aligned with mesh in Section SM1.3 of the supplementary material.

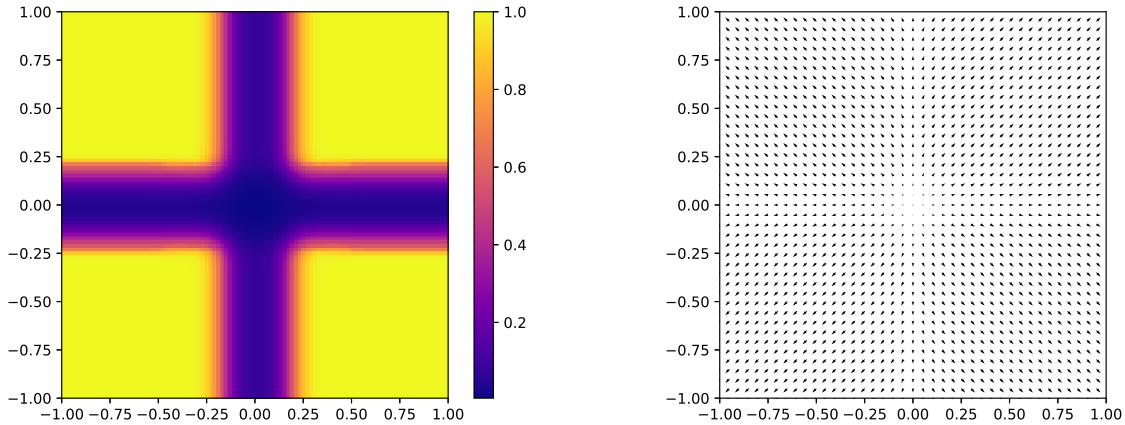

Figure 10: Moment m_0 (left) and velocity field (right) at $t^N = 0.5$ with the initial condition (43).

6 Conclusion

The purpose of this contribution has been to construct high-order RKDG schemes able to cope efficiently with the peculiarities of a weakly hyperbolic system of moments equations, which is frequently encountered in fluid mechanics and combustion applications, where a spray of polydisperse droplets is to be found either coupled to a gaseous flow field or as a small scale modeling in two-scale gas-liquid flows. The key feature of the proposed method has been to maintain high-order for smooth solutions but also to cope efficiently with convex admissibility set preservation, be it the moment space or the maximum principle on velocity, in the presence of singularities and void. A specific limitation procedure has been introduced in order to limit numerical diffusion when vacuum is created and to handle properly singularities. The proposed strategy has been shown to be competitive in terms of accuracy and of robustness compared to a reference kinetic finite volume scheme [7, 29, 21] at second order but also has the ability to reach third and fourth order within the same paradigm. Such a feature is essential for applications where spray dynamics experience preferential concentration and where vacuum is always present, thus influencing the local mixture fraction and the evaporation rate for combustion applications in particular but it has a much broader range of applications. The present piece of work is also the building block for the construction of a numerical strategy for more complex multi-variate cases such as in [34] for oscillating droplets flows. The aim is to simulate a cloud of non spherical oscillating droplets by taking into account the geometrical dynamics described by the phase variables. This is work in progress.

A Additional numerical results

Here we include some additional 2D numerical results in order to illustrate the behavior of the second order RKDG scheme associated to the straight projection (36). In all the numerical experiments below, the computational domain $[-1, 1]^2$ is divided into 100×100 uniform cells with $\Delta x = \Delta y = 1/50$. Dirichlet boundary conditions are imposed and the computations are performed until the final time t^N. In all the cases, we solve a Riemann problem where the initial data in each quadrant are constants. The quadrants are indexed conventionally: from the first to the fourth in a counterclockwise direction, starting from the right upper one.

A.1 Asymmetric δ-shock case

The initial condition is:

$$\begin{aligned} m_\alpha(x,0) &= (\alpha+1)^{-1}\left(\mathbf{1}_{(\mathbb{R}^+)^2}+\mathbf{1}_{(\mathbb{R}^-)^2}+\frac{4}{3}(1-0.5^{\alpha+1})(\mathbf{1}_{\mathbb{R}^+\times\mathbb{R}^-}+\mathbf{1}_{\mathbb{R}^-\times\mathbb{R}^+})\right)(x), \\ q(x,0) &= -0.1\times m_1(x,0)\times(\mathrm{sign}(x_1),\ \mathrm{sign}(x_2))\,. \end{aligned} \tag{44}$$

We start from the configuration where the mass m_1 is equally distributed over the domain, but not the other moments. The initial velocity profile (44) is chosen to generate stationnary δ-shocks along the axes $x=0$ and $y=0$. Fig. 11 shows the moments m_0 and m_1 at final time $t^N=0.25$. The moments m_0, $m_{1/2}$ and

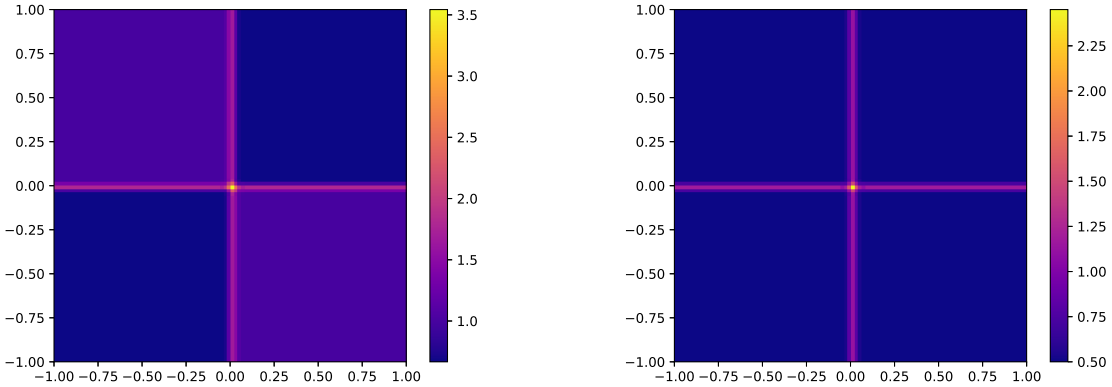

Figure 11: Moments m_0 (left) and m_1 (right) at $t^N=0.25$ with the initial condition (44).

$m_{3/2}$ are different in the quadrants next to each others. The velocity profile at the end of the simulation is the same as the one observed in the 2D δ-shock test case with equally distributed mass (42).

A.2 Diagonal δ-shock case

We use the initial condition:

$$\begin{aligned} m_\alpha(Rx,0) &= (\alpha+1)^{-1}, \\ q(Rx,0) &= -0.1\times m_1(x,0)\times(\mathrm{sign}(x_1),\ \mathrm{sign}(x_2))\,. \end{aligned} \tag{45}$$

with $Rx=(\cos(\pi/3)x_1+\sin(\pi/3)x_2,\ -\sin(\pi/3)x_1+\cos(\pi/3)x_2)$. The initial mass m_1 is equally distributed over the domain and the initial velocity (45) is chosen to generate δ-shocks along the diagonals $y=x\sqrt{3}$ and $y=x/\sqrt{3}$. The velocity is heading towards the two diagonals and the solution of the moment system (6) yields δ-singularities located at the origin and the two axes. This is illustrated on Fig. 12 which shows m_1 and $u\equiv q/m$ at final time $t^N=0.25$. The numerical solution exhibits the expected behavior, i.e. stationary δ-shocks along the two diagonals $y=x\sqrt{3}$ and $y=x/\sqrt{3}$ (see Fig. 12 (left) for moment m_1). The result is similar to the 2D case (42) after applying a rotation of angle $\pi/3$. The robustness and stability of the RKDG scheme and its limitation procedure are preserved even in the case of δ-shock singularities that are not aligned with the mesh.

A.3 Diagonal vacuum case

Eventually, we consider the initial condition:

$$m_\alpha(Rx,0)=(\alpha+1)^{-1}, \qquad q(Rx,0)=0.4\times m_1(x,0)\times(\mathrm{sign}(x_1),\ \mathrm{sign}(x_2))\,. \tag{46}$$

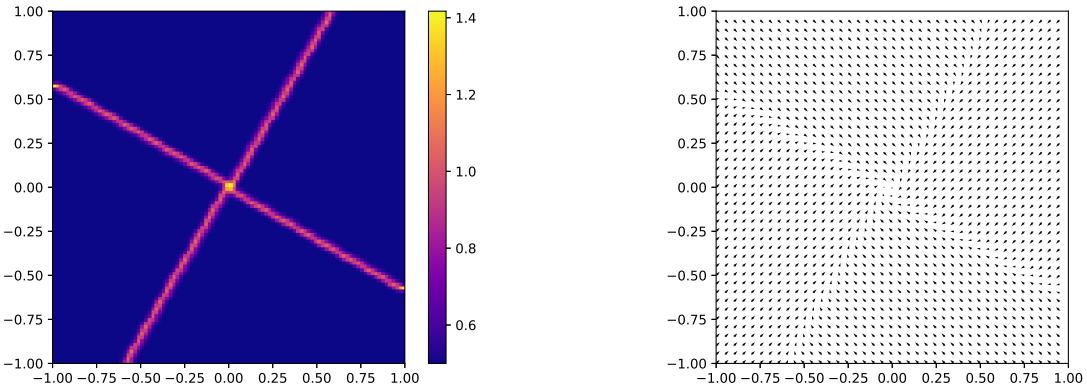

Figure 12: Moment m_1 (left) and velocity (right) at $t^N = 0.25$ with the initial condition (45).

The initial mass m_1 is equally distributed over the domain and the initial velocity (46) generates vacuum $m_\alpha \approx 0$ along the diagonals $y = x\sqrt{3}$ and $y = x/\sqrt{3}$. Indeed, the velocity is heading outwards these axes. Fig. 13 shows the moment m_1 and the velocity $u \equiv q/m_1$ at final time $t^N = 0.5$. Again, the behavior of the

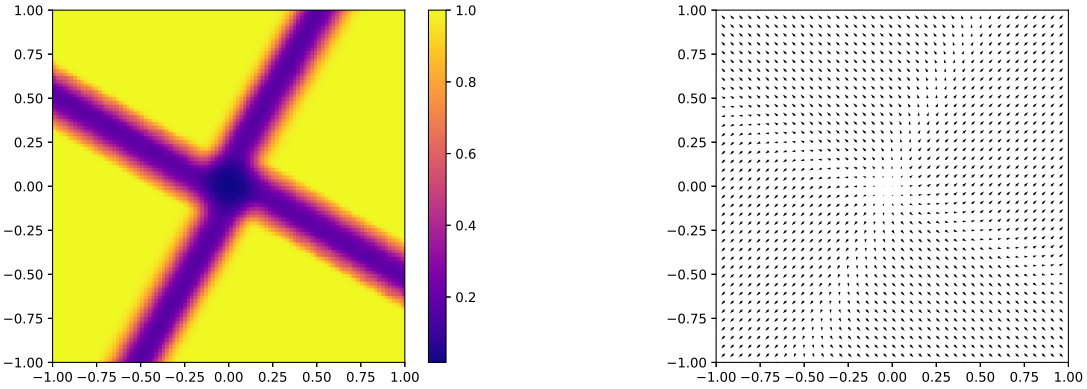

Figure 13: Moment m_0 (left) and velocity (right) at $t = 1$ with the initial condition (46).

numerical method agree qualitatively with the structure of the expected solution. Indeed, we can observe that the numerical solution approximates well the vacuum that are not aligned with the mesh. Because of the presence of vacuum, the limitation procedure is activated in order to avoid non-realizable vector of moments as in the 2D case of (43) after applying a rotation of angle $\pi/3$.

References

[1] R. Anderson, V. Dobrev, T. Kolev, D. Kuzmin, M. Quezada de Luna, R. Rieben, and V. Tomov. High-order local maximum principle preserving MPP discontinuous Galerkin finite element method for the transport equation. *J. Comput. Phys.*, 334:102–124, 2017.

[2] F. Bouchut. On zero pressure gas dynamics. In *Advances in kinetic theory and computing*, volume 22 of *Ser. Adv. Math. Appl. Sci.*, pages 171–190. World Sci. Publ., 1994.

[3] F. Bouchut and F. James. One-dimensional transport equations with discontinuous coefficients. *Nonlinear Anal. TMA*, 32(7):891–933, 1998.

[4] F. Bouchut and F. James. Solutions en dualité pour les gaz sans pression. *C. R. Acad. Sci. Paris, Série 1*, pages 1073–1078, 1998.

[5] F. Bouchut and F. James. Duality solutions for pressureless gases, monotone scalar conservation laws, and uniqueness. *Commun. Partial Diff. Eq.*, 24:2173–2189, 1999.

[6] F. Bouchut, F. James, and S. Mancini. Uniqueness and weak stability for multi-dimensional transport equations with one-sided Lipschitz coefficient. *Ann. Scuola Normale Superiore di Pisa*, 4(1):1–25, 2005.

[7] F. Bouchut, S. Jin, and X. Li. Numerical approximations of pressureless and isothermal gas dynamics. *SIAM J. Num. Anal.*, 41:135–158, 2003.

[8] Y. Brenier and E. Grenier. Sticky particles and scalar conservation laws. *SIAM J. Numer. Anal.*, 35(6):2317–2328, 1998.

[9] C. Chalons, D. Kah, and M. Massot. Beyond pressureless gas dynamics: quadrature-based velocity moment models. *Commun. Math. Sci.*, 10(4):1241–1272, 2012.

[10] Nattaporn Chuenjarern, Ziyao Xu, and Yang Yang. High-order bound-preserving discontinuous Galerkin methods for compressible miscible displacements in porous media on triangular meshes. *J. Comput. Phys.*, 378:110–128, 2019.

[11] B. Cockburn and C.-W. Shu. The Runge-Kutta Discontinuous Galerkin Method for Conservation Laws V. *J. Comput. Phys.*, 141(2):199–224, 1998.

[12] B. Cockburn and C.-W. Shu. *Discontinuous Galerkin Methods: Theory, computation and applications.* Springer, 2000.

[13] Bernardo Cockburn and Chi-Wang Shu. TVB Runge-Kutta Local Projection Discontinuous Galerkin Finite Element Method for Conservation Laws II: General Framework. *Math. Comp.*, 52(186):411–435, 1989.

[14] P. Cordesse. *Contribution to the study of combustion instabilities in cryotechnic rocket engines : coupling diffuse interface models with kinetic-based moment methods for primary atomization simulations.* PhD thesis, CentraleSupélec, 2020.

[15] R. Curto and L. A. Fialkow. Recusiveness, positivity, and truncated moment problems. *Houston J. Math.*, 17(4):603–634, 1991.

[16] H. Dette and W. J. Studden. *The theory of canonical moments with applications in statistics, probability, and analysis.* John Wiley & Sons Inc., 1997.

[17] D. Di Pietro and A. Ern. *Mathematical Aspects of Discontinuous Galerkin Methods*, volume 69 of *SMAI Mathématiques et Applications.* Springer, 2012.

[18] D. A. Drew. Evolution of geometric statistics. *SIAM J. Appl. Math.*, 50(3):649–666, 1990.

[19] F. Drui. *Modélisation et simulation Eulériennes des écoulements diphasiques à phases séparées et dispersées : développement d'une modélisation unifiée et de méthodes numériques adaptées au calcul massivement parallèle.* PhD thesis, Université Paris-Saclay, 2017.

[20] M. Essadki. *Contribution to a unified Eulerian modeling of fuel injection: from dense liquid to polydisperse evaporating spray.* PhD thesis, CentraleSupélec, 2018.

[21] Mohamed Essadki, Stephane de Chaisemartin, Frédérique Laurent, and Marc Massot. High-order moment model for polydisperse evaporating sprays towards interfacial geometry. *SIAM Journal on Applied Mathematics*, 78(4):2003–2027, 2018.

[22] Mohamed Essadki, Stéphane de Chaisemartin, Marc Massot, Frédérique Laurent, Adam Larat, and Stéphane Jay. Adaptive Mesh Refinement and High-Order Geometrical Moment Method for the Simulation of Polydisperse Evaporating Sprays. *Oil & Gas Science and Technology - Revue d'IFP Energies nouvelles*, 71(5), 2016.

[23] A. F. Filippov. *Differential Equations with Discontinuous Righthand Sides*. Springer, 1988.

[24] S. Gottlieb and C.-W. Shu. Total variation diminishing runge-kutta methods. *Math. Comp.*, 67:73–85, 1998.

[25] J.-L. Guermond, B. Popov, and I. Tomas. Invariant domain preserving discretization-independent schemes and convex limiting for hyperbolic systems. *Comput. Meth. Appl. Mech. Eng.*, 347:143–175, 2019.

[26] H. Hajduk. Monolithic convex limiting in discontinuous Galerkin discretizations of hyperbolic conservation laws. *Computers Math. Appl.*, 87:120–138, 2021.

[27] A. Harten. High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.*, 49(3):357–393, 1983.

[28] F. Hausdorff. Summationmethoden und momentfolgen. *Math. Z.*, (9):74–109, 1921.

[29] D. Kah, F. Laurent, M. Massot, and S. Jay. A high-order moment method simulating evaporation and advection of a polydisperse spray. *J. Comput. Phys.*, 231(2):394–422, 2012.

[30] H. C. Kranzer and B. L. Keyfitz. A strictly hyperbolic system of conservation laws admitting singular shocks. In *Nonlinear Evolution Equations that Change Type*, volume 27, pages 107–125. IMA Volumes in Mathematics and its Applications, Springer-Verlag, 1990.

[31] J.-B. Lasserre. *Moment, positive polynomials, and their applications*, volume 1. Imperial college press, 2009.

[32] Ph. Le Floch. An existence and uniqueness result for two nonstrictly hyperbolic systems. In *Nonlinear Evolution Equations that Change Type*, volume 27, pages 126–138. IMA Volumes in Mathematics and its Applications, Springer-Verlag, 1990.

[33] Yimin Lin, Jesse Chan, and Ignacio Tomas. A positivity preserving strategy for entropy stable discontinuous Galerkin discretizations of the compressible Euler and Navier-Stokes equations. *J. Comput. Phys.*, 475:111850, 2023.

[34] A. Loison, S. Kokh, M. Massot, and T. Pichard. Two-phase flow reduced-order model: a two-scale model of polydisperse oscillating droplets. *Submitted, arXiv:2308.15641*, 2023.

[35] Yu Lv, Yee Chee See, and Matthias Ihme. An entropy-residual shock detector for solving conservation laws using high-order discontinuous Galerkin methods. *J. Comput. Phys.*, 322:448–472, 2016.

[36] D. Marchisio and R. Fox. Solution of population balance equations using the direct quadrature method of moments. *J. Aerosol Sci.*, 36:43–73, 2005.

[37] M. Massot, F. Laurent, S. de Chaisemartin, L. Fréret, and D. Kah. Eulerian multi-fluid models: modeling and numerical methods. In *Modelling and Computation of Nanoparticles in Fluid Flows*, Lectures Notes of the von Karman Institute. NATO RTO-EN-AVT-169, 2009. Available at http://hal.archives-ouvertes.fr/hal-00423031/en/.

[38] M. Massot, F. Laurent, D. Kah, and S. de Chaisemartin. A robust moment method for evaluation of the disappearance rate of evaporating sprays. *SIAM J. Appl. Math.*, 70:3203–3234, 2010.

[39] A. Mazaheri, C.-W. Shu, and V. Perrier. Bounded and compact weighted essentially nonoscillatory limiters for discontinuous Galerkin schemes: Triangular elements. *J. Comput. Phys.*, 395:461–488, 2019.

[40] R. McGraw. Description of aerosol dynamics by the quadrature method of moments. *Aerosol Sci. Tech.*, (27):255–265, 1987.

[41] W. Pazner. Sparse invariant domain preserving discontinuous Galerkin methods with subcell convex limiting. *Comput. Meth. Appl. Mech. Eng.*, 382:113876, 2021.

[42] T. Pichard. A moment closure based on a projection on boundary of the realizability domain: 1d case. *Kin. rel. mod.*, 13(6):1243–1280, 2020.

[43] S. B. Pope. The evolution of surfaces in turbulence. *Int. J. Engng Sci.*, 26(5):445–269, 1988.

[44] F. Poupaud and M. Rascle. Measure solutions to the linear multi-dimensional transport equation with non-smooth coefficients. *Commun. Partial Diff. Eq.*, 22(1-2):225–267, 1997.

[45] Jianxian Qiu and Chi-Wang Shu. Runge-Kutta discontinuous Galerkin method using WENO limiters. *SIAM J. Sci. Comput.*, 26(3):907 – 929, 2005.

[46] Andrés M Rueda-Ramírez, Benjamin Bolm, Dmitri Kuzmin, and Gregor J Gassner. Monolithic Convex Limiting for Legendre-Gauss-Lobatto Discontinuous Galerkin Spectral Element Methods, 2023.

[47] K. Schmüdgen. *The moment problem.* Springer, 2018.

[48] F. Schneider. Kershaw closures for linear transport equations in slab geometry ii: High-order realizability-preserving discontinuous-galerkin schemes. *J. Comput. Phys.*, 322, 02 2016.

[49] C.-W. Shu. Total-variation diminishing time discretizations. *SIAM J. Sci. Stat. Comp.*, 9:1073–1084, 1988.

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

[51] R. J. Spiteri and S. J. Ruuth. A new class of optimal high-order strong-stability-preserving time discretization methods. *SIAM J. Numer. Anal.*, 40:469–491, 2002.

[52] D. Tan, T. Zhang, and Y. Zheng. Delta-shock waves as limits of vanishing viscosity for hyperbolic systems of conservation laws. *J. Diff. Eq.*, 112:1–32, 1994.

[53] F. A. Williams. Spray combustion and atomization. *Physics of Fluids*, 1:541–545, 1958.

[54] Y. Xing, X. Zhang, and C.-W. Shu. Positivity-preserving high-order well-balanced discontinuous Galerkin methods for the shallow water equations. *Adv. Water Resources*, 33, 12 2010.

[55] Y. Yang, D. Wei, and C.-W. Shu. Discontinuous Galerkin method for Krause's consensus models and pressureless Euler equations. *J. Comput. Phys.*, pages 109–127, 2013.

[56] X. Zhang. *Maximum-Principle-Satisfying and Positivity-Preserving High-Order Schemes for Conservation Laws.* PhD thesis, Brown, 2011.

[57] X. Zhang. On positivity-preserving high-order discontinuous Galerkin schemes for compressible Navier–Stokes equations. *J. Comput. Phys.*, 328:301–343, 2017.

[58] X. Zhang and C.-W. Shu. On maximum-principle-satisfying high-order schemes for scalar conservation laws. *J. Comput. Phys.*, 229(9):3091–3120, 2010.

[59] X. Zhang and C.-W. Shu. On positivity-preserving high-order discontinuous Galerkin schemes for compressible Euler equations on rectangular meshes. *J. Comput. Phys.*, 229:8918–8934, 2010.

[60] X. Zhang and C.-W. Shu. Positivity-preserving high-order discontinuous Galerkin schemes for compressible Euler equations with source. *J. Comput. Phys.*, 230(4):1238–1248, 2011.