On the entropy dissipation of systems of quadratures

Teddy Pichard, Frédérique Laurent

On the entropy dissipation of systems of quadratures

Teddy Pichard[1] and Frédérique Laurent[2]

[1]CMAP, CNRS, École polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France

[2]Laboratoire EM2C, CNRS, CentraleSupélec, Université Paris-Saclay, 91192 Gif-sur-Yvette, France

Abstract

The method of moments in kinetic theory is a popular discretization technique with respect to the kinetic velocity variable. It can be seen as a non-linear Galerkin semi-discretization in velocity. Among these methods, the quadrature-based methods exploiting the theory of orthogonal polynomials are well suited for numerical purposes. However, two important properties of the original kinetic equation are lost during such approximations: the strong hyperbolicity, corresponding to transport phenomena, and the dissipation of entropy, corresponding to the trend of the solution towards an equilibrium. These two properties are closely related through symmetrization techniques.

In this work, we aim to clarify the treatment of these two properties in the derivation of quadrature-based moment systems, and to prove H-theorems, namely dissipation of entropy and equilibrium representation, after such derivations. From this study, we develop of quadrature-based closures adapted to specific entropies. These closures are of two types, either with fixed quadrature points (or velocities), namely the discrete velocity methods (DVM), and with varying ones namely the quadrature-based method of moments (QMOM). To adapt the closures to specific entropies, the number of quadrature points is increased compared to the number of moments, resulting in augmented systems (ADVM or AQMOM) with more unknowns than equations, and the additional parameters are constrained to match the entropy requirements. Similarly, the quadrature-based entropies are of two types, either those based on a symmetrization criterion on the flux vector, which are only intended to have an entropy, or those corresponding to a quadrature formula of the kinetic entropy, which are intended to reproduce the kinetic trend towards the equilibrium. At each step of these developments, based on the considered entropy, we provide definitions for the flux vectors, adapted relaxation operators (with common conservation properties), we compute the associated entropic variables, and highlight the corresponding symmetrization property of the flux and equilibria. Finally, we provide an entropy-dissipative discretization for such moment systems.

1 Introduction

Kinetic equations arise from the mesoscopic modelling of particles moving in a medium and interacting with each others. We focus on a generic 1D-1V equation consisting of the transport of a distribution function f in space $x \in \Omega \subset \mathbb{R}$ at velocity $v \in \mathbb{R}$ evolving over the full real line with a collision operator. Such a model is generally expected to satisfy certain physical and mathematical properties that must to be preserved from the derivation of the model to the numerical simulation. In particular, in this work, we focus on three properties of the kinetic equation of its solution: the conservation of mass, momentum, and energy, i.e. the first three moments of f in velocity, the transport phenomena (characterized afterward by the hyperbolicity property) and the H-theorem, which describes the trend of a system towards an equilibrium. Regarding the H-theorem, we study in particular which entropy is dissipated by the system and its associated local equilibrium, represented by the entropy minimizer (see e.g. [11, 27, 28] for general descriptions of these

models, properties and extensions). When constructing numerical solvers for the kinetic equation, a first semi-discretization with respect to the kinetic velocity variable v is often performed in order to account for these two properties. We focus in particular on the quadrature techniques (see e.g. [45, 34] for general techniques and [67, 63, 65, 64] for their applications to kinetic equations) with respect to v, and we propose modifications of such methods adapted to better match the properties of the original kinetic equation.

The semi-discretization of such an equation with respect to v can be complicated due to the non-compactness of the velocity domain. Naive methods often lead to some form of violation of one or more of the above kinetic properties. The brute-force techniques are based either on probabilistic techniques, such as direct simulation Monte Carlo (DSMC ; see e.g. [11, 44, 20, 79] and references therein) solvers, or on deterministic ones, such as discrete velocity methods (DVM ; see e.g. [22, 21, 85, 69] and references therein). They commonly lead to non-physical equilibria, and the physical ones are only found at (inaccessible) convergence. This is illustrated in the next section, which describes the DVM as the starting point of the present approach. These naive approaches rely on the reduction of the velocity domain to a bounded or discrete one, while among the most efficient alternatives, many are based on velocity integrals. On the one hand, spectral methods using the Fourier transform provide both good error estimates and a framework adapted to the kinetic properties. These have emerged from the mathematical community, and it has been the source of a large literature (see e.g. [77, 43, 38, 78, 26] among others) during the last decades. On the other hand, the method of moments emerged from the physical community as a tool to obtain out-of-equilibrium thermodynamics in fluid models (see e.g. [94, 75, 35, 64, 49]). The work presented here falls into the latter framework.

The method of moment is interpreted below as a non-linear, weak form (in a measure sense instead of L^2) of Galerkin discretizations. Rewriting the kinetic equation in a weak form (in velocity only), this method consists in reducing the test function space to a finite dimensional polynomial space, while reducing the solution set to a manifold of the same dimension. Such a solution manifold must be carefully chosen, first for the resulting finite dimensional system to be well-posed and well-conditioned, and second for this approximation to preserve the considered kinetic properties (see e.g. the reviews [24, 83]). This choice corresponds to the so-called closure problem, the solution of which is generally obtained by solving a moment problem (see e.g. [3, 4, 90, 86, 61]). One of the first moment approximations was proposed by Harold Grad in [48]. The resulting system of moments is hyperbolic and dissipates an entropy only close to the equilibrium. It is therefore ill-adapted for transport phenomena away from this regime. A large recent literature aims at modifying this closure to obtain alternative stability property by various techniques (see e.g. [92, 25, 37]). The entropy-based closure ([62, 75]) was developed based on entropic studies using symmetrization techniques ([42, 47, 70, 58, 89, 59]). By design, it provides entropy dissipation. However, it presents two difficulties: First, this approximation is ill-defined in the vicinity of the Maxwellian regimes ([57]). This issue was also circumvented by various modifications, either of the definition of the closure itself ([88, 52]) or of the entropy it is defined from ([2, 1, 5]). Second, in order to compute this approximation, one needs to solve numerical optimization problems, which become ill-conditioned near the boundary of its domain of definition ([51, 7, 6]). Several other attempts, typically based on the study of the so-called realizability, i.e. the convex cone of moment vectors, have also been made in the literature (see e.g. [66, 81, 82]), but no general theory encompasses all of them and none preserve all the kinetic properties.

In the present work, we focus on the quadrature-based closures, that were originally proposed for aerosols ([67, 63, 65]). This technique can be seen as an approximation of the distribution function by a sum of Dirac measures, with moment-dependent quadrature points and weights. This construction is based on a well-established mathematical theory of numerical integration and associated algorithms ([93, 45, 50]). Several extensions aimed at improving the algorithms ([63]), extending it to multi-D ([102, 41]) or replacing the Dirac measures by Gaussians ([96, 103, 29]) to extend the range of physical phenomena captured. The original QMOM closure ([67]) leads to a weakly hyperbolic system of moments (see e.g. the computations in [30, 54, 82]). Such systems still exhibit transport phenomena and entropy dissipation (see e.g. [17]), but their solutions have a weak (measure) regularity and the considered entropy is not strictly convex in the moments. This problem has become particularly popular in recent years, and various works aimed at imposing the strong hyperbolicity property to the resulting system of moments by modifying this construction. This objective

was achieved in [41, 39] by using algorithms from the theory of orthogonal polynomials. A generalization of this technique was proposed in [40] and an extension to multi-D problems using techniques closely related to Grad's method is in progress. Strong hyperbolicity was also obtained in [73, 74] using similar techniques based on the theory of orthogonal polynomials. Other formal attempts in this direction were tested numerically in [12] using techniques closely related to the entropy-based closure (covered in the present construction), exploiting the framework of [60]. And it was enforced in [99] by combining constructions from [71, 72, 87] and numerics from [45, 93, 81]. Finally, mathematical studies were given in [54, 105] for the closure [103]and in [104] for the closure [39] with proofs of the strict hyperbolicity and entropy decay.

The present work aims at clarifying which notion of entropy should be sought in the construction of the quadrature-based closure. Detailed computations of the entropy of the quadrature-based moment closures are given, relating the notions of strong hyperbolicity, symmetrization and entropy dissipation to explicit formulae for the entropy, entropic variables, symmetrizer and appropriate entropy-dissipating relaxation operators for each closure. Eventually, alternative closures are proposed, which aim to better approximate the underlying kinetic entropy decay. They consist in using more quadrature points than moments, and these additional constraints on these parameters are proposed to obtain a chosen symmetrization property, which corresponds to choosing an approximated (strictly convex) entropy to dissipate.

The next section summarizes the state of the art on the considered kinetic equation, its properties, the moment approximations and the quadrature-based closures studied, namely DVM and QMOM. Section 3 recalls how symmetrization relates to entropy dissipation in the present framework. Section 4 presents the construction of the augmented DVM (ADVM), i.e. the augmentation of the number of quadrature points in the DVM to better approximate the kinetic entropy. Similarly, Section 5 presents the construction of augmented QMOM (AQMOM) technique to further improve the entropy approximation. An entropy-dissipating numerical scheme adapted to these entropic and quadrature-based moment closure is provided in Section 6. The last section gathers conclusive remarks and comments.

2 State of the art

In this section, we recall the standard construction of kinetic, moment and quadrature-based moment equations.

2.1 Kinetic equation

Consider the 1D kinetic equation

$$\partial_t f + v\partial_x f = c(f). \tag{2.1}$$

The left-hand side is a linear transport opertor applied to the distribution function f at the variable velocity $v \in \mathbb{R}$ in the time-space domain $(t, x) \in [0, T) \times \mathbb{R}$. In particular, transport phenomena are observed at all velocities.

The operator $c(f)$ on the right-hand side operator models collisions of the particles with each others or interactions with the background. It is typically a Boltzmann or a BGK operator.

In the following, we are interested in the behaviour of the solution with respect to the variable v. The solution f is assumed to be integrable in with respect to $v \in \mathbb{R}$, i.e. we seek $f(t, x) \in L^1(\mathbb{R})$ satisfying (2.1) in either a strong or a weak sense. Further classical assumptions about $f(t, x)$ are made below.

2.1.1 H-theorem at the kinetic level

This model is assumed to follow an H-theorem:

Property 2.1 (H-theorem at the kinetic level)**.**

- ***Entropy dissipation***: *There exists a strictly convex function $\eta : \mathbb{R}^+ \to \mathbb{R}$ such that*

$$\partial_t h(f) + \partial_x g(f) = s(f) \leq 0, \qquad h(f) = \int_{\mathbb{R}} \eta(f(v))dv, \tag{2.2a}$$

$$g(f) = \int_{\mathbb{R}} v\eta(f(v))dv, \qquad s(f) = \int_{\mathbb{R}} \eta'(f(v))c(f)(v)dv. \tag{2.2b}$$

- ***Equilibrium***: *The entropy production $s(f) = 0$ cancels if and only if $\eta'(f) \in \mathcal{I}$ belongs to the space of collision invariants defined by*

$$\phi \in \mathcal{I} \qquad \Leftrightarrow \qquad \forall g \geq 0, \quad \int_{\mathbb{R}} \phi c(g)dv = 0.$$

In particular, $\eta'(f) \in \mathcal{I}$ implies that $f = f^{eq}(f)$ is of the form

$$f(v) = f^{eq}(f)(v) = (\eta^*)'(\phi), \tag{2.3}$$

for some $\phi \in \mathcal{I}$. Here, η^* is the Legendre-Fenchel transform of η. It satisfies $(\eta')^{-1} = (\eta^*)'$.

Therefore, two additional hypotheses about $f \equiv f(t,x)$ are made at all (t,x):

$$\eta(f) \in L^1(\mathbb{R}) \quad \text{and} \quad f\phi \in L^1(\mathbb{R}) \text{ for all } \phi \in \mathcal{I}.$$

In the following, we use:

- The collision invariants are $\mathcal{I} = Span(1, v, v^2) = \mathbb{P}_2(\mathbb{R})$, corresponding to the conservation of mass, momentum and energy.
- The model dissipates the Boltzmann entropy $\eta(f) = f\log f - f$, which gives $\eta'(f) = \log f$ and $(\eta^*)' = \exp$. When $\mathcal{I} = \mathbb{P}_2(\mathbb{R})$, the equilibrium is represented by a Maxwellian $f = f^{eq}(f)$ given by

$$f^{eq}(f) = \frac{\rho(f)}{\sqrt{2\pi T(f)}} \exp\left(-\frac{(v-u(f))^2}{2T(f)}\right), \qquad \rho(f) = \int_{\mathbb{R}} f(v)dv, \tag{2.4a}$$

$$u(f) = \frac{1}{\rho(f)} \int_{\mathbb{R}} vf(v)dv, \qquad T(f) = \frac{1}{\rho(f)} \int_{\mathbb{R}} (v-u(f))^2 f(v)dv. \tag{2.4b}$$

The aim of the present work is to preserve these properties through the semi-discretization with respect to the variable v.

2.1.2 Collision operator at the kinetic level

For simplicity, the collision operator c in (2.1) is chosen to be a BGK ([76, 53]) relaxation operator at the kinetic level, and we extend its construction at the moment level in the next sections.

First, denote

$$\mathcal{I}' := \left\{ \left| \begin{array}{ll} L^1(\mathbb{R}) & \to \mathbb{R} \\ g & \mapsto \int_{\mathbb{R}} g(v)\phi(v)dv, \end{array} \qquad \forall \phi \in \mathcal{I} \right. \right\},$$

$$\mathcal{I}^\perp := \left\{ g \in L^1(\mathbb{R}) \quad \text{such that} \quad \int_{\mathbb{R}} g(v)\phi(v)dv = 0 \quad \forall \phi \in \mathcal{I} \right\} = \bigcap_{\psi \in \mathcal{I}'} Ker\ \psi,$$

where the orthogonality notation is only formal.

The relaxation operator is constructed proportional (with a positive rate $\tau > 0$) to the difference between the distribution function f and the minimizer of the entropy h under the collision invariant constraints $c(f) \in \mathcal{I}^\perp$, i.e.

$$c(f) = \frac{f^{min}(f) - f}{\tau}, \qquad \text{such that} \qquad f^{min}(f) = \underset{g-f\in\mathcal{I}^\perp}{\operatorname{argmin}}\ h(g). \tag{2.5}$$

This optimization problem has been the source of a large literature ([13, 14, 68, 57, 88, 52]). In the present case, this minimum takes the form $f^{min} = f^{eq}$ defined in (2.3) where $\phi \in \mathcal{I}$ is the only polynomial of degree 2 in v such that the moments of $f^{min}(f)$ up to order 2 are the same as those of f. For this reason, in the following, the notations are unified to

$$f^{eq} = f^{min} = M.$$

One verifies that this operator satisfies the conservation and the dissipation properties. The proof is recalled to clarify the framework and notations used in the next sections.

Proposition 2.1. *Consider $\boldsymbol{\mathcal{I}'} \subset \mathcal{L}(E;\mathbb{R})$ a finite dimensional space of linear forms, and $\boldsymbol{\mathcal{I}}^\perp = \bigcap_{\boldsymbol{\psi}\in\boldsymbol{\mathcal{I}'}} Ker\boldsymbol{\psi}$. Consider a convex function $H : E \to \mathbb{R}$ and a function $\mathbf{C} : E \to E$ defined by*

$$\mathbf{C}(\mathbf{U}) = \frac{1}{\tau}\left(\mathbf{M}(\mathbf{U}) - \mathbf{U}\right), \qquad \mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^\perp}{argmin}\ H(\mathbf{V}). \tag{2.6a}$$

Then:

- ***Entropy dissipation**: For all $\mathbf{U}$ such that this minimum exists:*

$$dH(\mathbf{U})(\mathbf{C}(\mathbf{U})) \leq 0. \tag{2.6b}$$

- ***Equilibrium**:*
  - ***Conservativity**: For all $\boldsymbol{\psi} \in \boldsymbol{\mathcal{I}'}$, for all $\mathbf{U}$ such that this minimum exists:*

$$\boldsymbol{\psi}(\mathbf{M}(\mathbf{U})) = \boldsymbol{\psi}(\mathbf{U}). \tag{2.6c}$$

  - *If H is strictly convex, then for all $\boldsymbol{\psi} \in \mathcal{L}(E;\mathbb{R})\backslash\boldsymbol{\mathcal{I}'}$, there exists $\mathbf{U} \in E$ such that*

$$\boldsymbol{\psi}(\mathbf{M}(\mathbf{U})) \neq \boldsymbol{\psi}(\mathbf{U}). \tag{2.6d}$$

Proof. • Since $\mathbf{M}(\mathbf{U})$ minimizes the convex entropy H under the constraints $\mathbf{M}(\mathbf{U}) - \mathbf{U} \in \boldsymbol{\mathcal{I}}^\perp$ and that $\mathbf{U} - \mathbf{U} = 0 \in \boldsymbol{\mathcal{I}}^\perp$, i.e. $\mathbf{U}$ satisfies these constraints, then

$$H(\mathbf{M}(\mathbf{U})) \leq H(\mathbf{U}).$$

By definition,

$$dH(\mathbf{U})(\mathbf{C}(\mathbf{U})) = \lim_{\theta\to 0} \frac{H\left(\mathbf{U} + \theta(\mathbf{M}(\mathbf{U}) - \mathbf{U})\right) - H\left(\mathbf{U}\right)}{\theta}$$

and by convexity

$$\frac{H\left(\mathbf{U} + \theta(\mathbf{M}(\mathbf{U}) - \mathbf{U})\right) - H\left(\mathbf{U}\right)}{\theta} \leq H(\mathbf{M}(\mathbf{U})) - H(\mathbf{U}) \leq 0,$$

which provides (2.6b).

• The conservation property (2.6c) holds by construction $\mathbf{C}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}^\perp$. Finally, by contradiction, assuming that there exists another $\boldsymbol{\psi} \in \mathcal{L}(E;\mathbb{R})\backslash\boldsymbol{\mathcal{I}'}$ satisfying (2.6c), then this implies that

$$M(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^\perp}{\operatorname{argmin}}\ H(\mathbf{V}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^\perp\oplus Ker(\boldsymbol{\psi})}{\operatorname{argmin}}\ H(\mathbf{V}).$$

By construction, the space $\boldsymbol{\mathcal{I}}^\perp \oplus Ker(\boldsymbol{\psi}) \supsetneq \boldsymbol{\mathcal{I}}^\perp$ is strictly larger than $\boldsymbol{\mathcal{I}}^\perp$. Then this equality violates the strict convexity assumption on H.

□

This proposition is presented in a generic space in order to be applicable to both the present kinetic framework and the moment framework in the next sections. For the kinetic case, one has

$$\mathbf{U} = f, \qquad H(\mathbf{U}) = h(f), \qquad dH(\mathbf{U})(\mathbf{C}(\mathbf{U})) = s(f),$$

as in Property 2.1 and for all $\boldsymbol{\psi} \in \boldsymbol{\mathcal{I}'}$ (finite dimensional), there exists $p \in \mathbb{P}_2(\mathbb{R})$ such that

$$\boldsymbol{\psi}(\mathbf{U}) = \int_{\mathbb{R}} p(v) f(v) dv.$$

Remark 2.1. The last two properties (2.6c) and (2.6d) imply the equilibrium condition. Indeed, they provide $dH(\mathbf{U})(\mathbf{M}(\mathbf{U}) - \mathbf{U}) = 0$ if and only if $dH(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ or, in the kinetic framework, $s(f) = 0$ if and only if $\eta'(f) \in \mathcal{I}$ by a representation theorem.

We refer e.g. to [23], and references therein, for further study of the BGK operator.

2.2 Moment approximations

Here, we construct the moment framework, in which our approximations lie. For this purpose, the solution $f(t,x) \in L^1(\mathbb{R})$ is rewritten $df(t,x) = f(t,x,\cdot)dv \in \mathcal{M}(\mathbb{R})$ as a Borel measure over $\mathbb{R}$ with a density, and we study approximations of $df(t,x)$.

2.2.1 System of moments

The approximations considered are based on a weak formulation of (2.1) with respect to the variable v. Formally, f satisfies the following equation for all test functions $\phi \in L^\infty_{loc}(\mathbb{R})$ in $v \in \mathbb{R}$ in either a weak or strong sense in (t,x)

$$\partial_t \left(\int_{\mathbb{R}} \phi(v) df(t,x)(v) \right) + \partial_x \left(\int_{\mathbb{R}} v\phi(v) df(t,x)(v) \right) = \frac{1}{\tau} \left(\int_{\mathbb{R}} \phi(v) dc(f)(t,x) \right). \tag{2.7}$$

Here, this formal construction requires the boundedness of all the integrals and the well-defined collision operator c (in this weak measure sense) as a function of the measure f. In view of the kinetic L^1 framework in the last section, these are appropriate assumptions.

In a Galerkin framework, we restrict the space of test functions to a finite dimensional polynomial space $\mathbb{P}_N(\mathbb{R}) \subset L^\infty_{loc}(\mathbb{R})$. Denoting $\mathbf{b}(v) = \left(1, v, \dots, v^N\right)^T$, a basis of test functions, then the weak formulation (2.7) can be rewriten

$$\partial_t \mathbf{U} + \partial_x \mathbf{F} = \mathbf{C}, \tag{2.8a}$$

$$\mathbf{U} = \int_{\mathbb{R}} \mathbf{b}(v) df(v), \quad \mathbf{F} = \int_{\mathbb{R}} v\mathbf{b}(v) df(v), \quad \mathbf{C} = \frac{1}{\tau} \left(\int_{\mathbb{R}} \mathbf{b}(v) dc(f)(v) \right). \tag{2.8b}$$

At this level, the vector of moments $\mathbf{U}$ of the exact solution f satisfies (2.8a) and locally belongs to the set of moments

$$\mathbf{U}(t,x) \in \mathcal{R} := \left\{ \int_{\mathbb{R}} \mathbf{b}(v) df(v), \quad f \in \mathcal{M}(\mathbb{R})_+ \right\} \subset \mathbb{R}^{N+1}. \tag{2.9}$$

We refer e.g. to [4, 61, 86, 32, 81] for studies of such sets.

Now, in order to have a unique solution, we need to reduce the solution space into one of the same dimension as $\mathbb{P}_N(\mathbb{R})$. This is the so-called moment closure, which is generally interpreted as rewriting $\mathbf{F}$ and $\mathbf{C}$ as functions of $\mathbf{U}$ in order to obtain a closed system of balance laws

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) = \mathbf{C}(\mathbf{U}). \tag{2.10}$$

Eventually, we seek a solution $\mathbf{U}$ to (2.10) either in a weak or strong sense, but this approximation can be defined over a smaller set

$$\mathcal{R}_{app} \subsetneq \mathcal{R}$$

than the entire set of moments. This is the case for DVM closure below.

2.2.2 Hyperbolicity and H-theorem at the moment level

Consider a closure, i.e. that $\mathbf{F} : \mathcal{R}_{app} \to \mathbb{R}^{N+1}$ and $\mathbf{C} : \mathcal{R}_{app} \to \mathbb{R}^{N+1}$ are given functions of $\mathbf{U} \in \mathcal{R}_{app}$. We recall here how the transport phenomena and the entropy dissipation presented at the kinetic level in Section 2.1 are characterized at the moment level, i.e. for the closed system (2.10).

Remark that $\mathbf{U} \in \mathcal{R}_{app} \subset \mathbb{R}^{N+1}$ now belongs to a finite dimensional space. Therefore, the differential $dH(\mathbf{U})$ can be assimilated to the Jacobian $H_{\mathbf{U}}(\mathbf{U})$ and this formalism is preferred in the rest of the paper.

For the transport phenomena, they are described by the hyperbolicity property:

Definition 2.2 (Hyperbolicity properties)**.**

- The system (2.10) is strongly, resp. weakly, hyperbolic if the Jacobian $\mathbf{F}_{\mathbf{U}}$ is diagonalizable, resp. trigonalizable, with real eigenvalues.
- The system (2.10) is symmetrizable if there exists a diffeomorphism $\mathbf{V}$ over $\mathcal{R}_{app}$ such that $\mathbf{V}_{\mathbf{U}}$ is symmetric positive definite and $\mathbf{F}_{\mathbf{U}}\mathbf{V}_{\mathbf{U}}^{-1}$ is symmetric. Symmetrizability implies strong hyperbolicity.

The eigenvalues of $\mathbf{F}_{\mathbf{U}}$ are the local wave speeds. In particular, these speeds of propagation were all velocities $v \in \mathbb{R}$ at the kinetic level, while they are now restricted to a finite number (at most $N+1$).

For the entropy dissipation, it is described by:

Property 2.2 (H-theorem at the moment level)**.**

- ***Entropy dissipation****: A (potentially strictly) convex function $H : \mathcal{R}_{app} \to \mathbb{R}$ is an entropy for the system* (2.10) *if there exists a function $G : \mathcal{R}_{app} \to \mathbb{R}$ satisfying $G_{\mathbf{U}} = H_{\mathbf{U}}\mathbf{F}_{\mathbf{U}}$ such that*

$$\partial_t H(\mathbf{U}) + \partial_x G(\mathbf{U}) = S(\mathbf{U}) := H_{\mathbf{U}}(\mathbf{U})\mathbf{C}(\mathbf{U}) \leq 0. \tag{2.11}$$

- ***Equilibrium****: The entropy production $S(\mathbf{U}) = 0$ cancels if and only if $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$, where*

$$\boldsymbol{\mathcal{I}}' := \left\{ \mathbf{V} \in \mathbb{R}^{N+1}, \quad \textit{such that} \quad \forall \mathbf{U} \in \mathcal{R}_{app}, \quad \mathbf{V}^T\mathbf{C}(\mathbf{U}) = 0 \right\}. \tag{2.12}$$

In the following, we use again

$$\begin{aligned} \boldsymbol{\mathcal{I}}' &= \left\{ \mathbf{V} \in \mathbb{R}^{N+1}, \quad \text{such that} \quad \mathbf{V}^T\mathbf{b} \in \mathbb{P}_2(\mathbb{R}) \right\}, \\ \boldsymbol{\mathcal{I}}^{\perp} &= \left\{ \mathbf{U} \in \mathbb{R}^{N+1}, \quad \text{such that} \quad \mathbf{V}^T\mathbf{U} = 0 \ \forall \mathbf{V} \in \boldsymbol{\mathcal{I}}' \right\}. \end{aligned}$$

Here, we identify $\boldsymbol{\mathcal{I}}'$ to $\boldsymbol{\mathcal{I}}$ in this finite-dimensional setting and the orthogonality notation $\boldsymbol{\mathcal{I}}^{\perp}$ can be understood with respect to the common scalar product.

In the next sections, the existence of an entropy is studied using a symmetrization technique, so we recall the following result.

Proposition 2.2. *There exists a strictly convex entropy for* (2.10) *if and only if it is symmetrized by a diffeomorphism $\mathbf{V}$ such that $\mathbf{V}(\mathbf{U})^T\mathbf{C}(\mathbf{U}) \leq 0$. In such a case, the equilibrium is characterized by the equivalence $\mathbf{V}(\mathbf{U})^T\mathbf{C}(\mathbf{U}) = 0$ if and only if $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$.*

We refer e.g. to [46] for such an equivalence result, remarking that the entropic variables are $\mathbf{V} = H_{\mathbf{U}}^T$.

Remark 2.3.

- As illustrated below for the QMOM system, there may exist non-strictly convex entropy for a weakly hyperbolic system. Such entropies do not provide a diffeomorphism relating the conserved variables $\mathbf{U}$ and the entropic variables $\mathbf{V}$.
- The entropy dissipation is often defined in a perturbative framework, which presents small discrepancies compared to the condition (2.11). We refer e.g. to [58, 89, 59, 101, 100] for well-posedness studies about systems of the form (2.10). This was also recently applied to systems of quadratures in [105, 104]. The present definition was found to be sufficient for our purposes, but the computations can be adapted.

The main objective of this paper is motivated by the lack of connection between the equilibria represented at the kinetic level and the ones at the moment levels with quadrature-based closures. In particular, vectors satisfying $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ at the moment level generally have no reason to be related to vectors of the form $\int \mathbf{b}(v)M(v)dv$, i.e. to equilibria $\eta'(M) \in \mathcal{I}$ at the kinetic level. The entropy-based closures (see e.g. [62, 75]) were constructed by enforcing the equality of these two sets, but this relation fails for other closures. We focus on the DVM, which are interpreted as a quadrature-based closures, and which allow simplifying computations of the Jacobians of the fluxes.

2.2.3 Collision operator at the moment level

A naive approach to close the collision operator in (2.10) consists in injecting the considered approximation f^{app} of the exact solution f of (2.1) inside the definition (2.4). This yields

$$\mathbf{C}^{naive}(\mathbf{U}) = \frac{1}{\tau}\left(\int_{\mathbb{R}} \mathbf{b}(v)M(v)dv - \mathbf{U}\right), \tag{2.13a}$$

where M is defined in (2.4) with (ρ, u, T) defined from the moments of order 0 to 2 in $\mathbf{U}$. Such a choice is generally unsatisfactory, since it often leads to a violation of the entropy dissipation property of the model (see next sections).

When the moment system has a strictly convex entropy H, a better alternative is to construct the relaxation operator after the moment approximation from this resulting entropy property, as in (2.5):

$$\mathbf{C}^{relax}(\mathbf{U}) = \frac{\mathbf{M}(\mathbf{U}) - \mathbf{U}}{\tau}, \qquad \text{such that} \qquad \mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} H(\mathbf{V}). \tag{2.13b}$$

In such a case, Proposition 2.1 applies and provides all the desired properties. This construction is only valid if the minimization problem (2.13b) has a unique solution. However, it can be extended when considering a non-strictly convex entropy (therefore not symmetrizing the system), e.g. in a weakly hyperbolic framework, or when such a minimum is non-unique by simply selecting one of them.

2.3 Systems of quadratures

We recall the construction of the quadrature-based closures and reformulate some of their properties.

2.3.1 The quadrature-based approximations

The quadrature-based closures consist, at all (t, x), in using the same approximation of f in the definitions (2.8b). Such approximations $f \approx f^{app}$ take the form of a sum of Dirac measures

$$f \approx f^{app} = \sum_{i=1}^{J} m_i \delta_{v_i}, \tag{2.14a}$$

where the masses $\mathbf{m} = (m_i)_{i=1,\dots,J}$ and the positions $\mathbf{v} = (v_i)_{i=1,\dots,J}$ satisfy the moment constraints

$$\mathbf{U} = \sum_{i=1}^{J} m_i \mathbf{b}(v_i). \tag{2.14b}$$

The flux vector is given from these parameters by

$$\mathbf{F} = \sum_{i=1}^{J} m_i v_i \mathbf{b}(v_i). \tag{2.14c}$$

Concerning the operator $\mathbf{C}$, the relaxation operator $\mathbf{C}^{relax}$ is defined in (2.13b) based on the entropy property and therefore depends on the choice of the quadrature approximation. It is given below for every quadrature approximation separately. The naive collision operator (2.13a) is general to all such approximations

$$\mathbf{C}^{naive} = \frac{1}{\tau}\left(\int_{\mathbb{R}} \mathbf{b}(v)\tilde{M}(\mathbf{m},\mathbf{v})(v)dv - \mathbf{U}\right), \tag{2.14d}$$

where the Maxwellian is rewritten $\tilde{M}(\mathbf{m},\mathbf{v})(v)$ as a function of the masses m_i and positions v_i instead of f. It is of the form (2.4), where (ρ, u, T) are replaced by

$$\rho = \sum_{i=1}^{J} m_i, \qquad u = \frac{1}{\rho}\sum_{i=1}^{J} m_i v_i, \qquad T = \frac{1}{\rho}\sum_{i=1}^{J} m_i (v_i - u)^2. \tag{2.14e}$$

In order to compute the closure $(\mathbf{F}, \mathbf{C})$ as a function of $\mathbf{U}$, one needs to invert the relation (2.14b) to compute the masses $\mathbf{m}$ and positions $\mathbf{v}$ as functions of $\mathbf{U}$. In particular, $(\mathbf{m}, \mathbf{v})$ needs to be uniquely defined from $\mathbf{U}$. These are $2J$ parameters for $N+1$ equations (2.14b). Depending on the considered approximation, some of these parameters are left free and some are fixed.

We denote generically $\mathbf{p}$ a vector of independent parameters, composed of masses m_i and positions v_i. The choices made for the construction of the DVM and QMOM approximations are described in the next two subsections, and some modifications are given in the next two sections to modify the entropy dissipation property.

2.3.2 Discrete velocity methods (DVM)

The discrete velocity methods are interpreted here as a quadrature-based moment closure with fixed quadrature points. It is constructed by fixing:

Hypothesis: The number of Dirac measures $J = N+1$ in (2.14) is the number of moments, and the positions $v_0 < v_1 < \cdots < v_N$ are all fixed a priori and different. Only the masses are left free.

With this hypothesis, the parameters are the masses $\mathbf{p} = \mathbf{m} = (m_i)_{i=0,\dots,N}$, and (2.14) rewrites as follows

$$\mathbf{U} = L\mathbf{m}, \qquad L_{i,j} = v_i^j. \tag{2.15}$$

Remark 2.4. The matrix L is a Vandermonde matrix with different quadrature points v_i, then it is invertible. The matrix L^{-1} sends the monomials $\mathbf{b}$ on $\mathbf{l} = L^{-1}\mathbf{b}$ the vector of Lagrange interpolating polynomials in v_i.

The parameters $\mathbf{p} = \mathbf{m} = L^{-1}\mathbf{U}$ are in bijection with the conserved moments $\mathbf{U}$, such that the flux $\mathbf{F}$ and the collisions $\mathbf{C}$ are uniquely defined. Multiplying (2.10) by L^{-1} corresponds to replacing $\mathbf{U}$, $\mathbf{F}$ and $\mathbf{C}$ by

$$\begin{aligned} \tilde{\mathbf{U}} = L^{-1}\mathbf{U}, \quad \tilde{\mathbf{F}}(\tilde{\mathbf{U}}) &= L^{-1}\mathbf{F}(L\tilde{\mathbf{U}}) \qquad \text{and} \quad \tilde{\mathbf{C}}(\tilde{\mathbf{U}}) = L^{-1}\mathbf{C}(L\tilde{\mathbf{U}}), \\ &= Diag(v_0, \dots, v_N)\tilde{\mathbf{U}} \end{aligned} \tag{2.16a}$$

and the balance laws are

$$\partial_t \tilde{\mathbf{U}} + \partial_x \tilde{\mathbf{F}}(\tilde{\mathbf{U}}) = \tilde{\mathbf{C}}(\tilde{\mathbf{U}}). \tag{2.16b}$$

Let us define the set of admissible DVM solutions

$$\tilde{\mathbf{U}} \in \tilde{\mathcal{R}}_{app} = (\mathbb{R}^+)^{N+1}, \qquad \mathbf{U} \in \mathcal{R}_{app} = L\tilde{\mathcal{R}}_{app} = L(\mathbb{R}^+)^{N+1},$$

as the set of vectors (2.15) with non-negative masses $m_i \geq 0$.

Remark 2.5. This set $\mathcal{R}_{app}$ of admissible DVM solutions is smaller than the set of moments

$$L(\mathbb{R}^+)^{N+1} = \mathcal{R}_{app} \subsetneq \mathcal{R}.$$

Since these solutions are represented by sums of Dirac measures with non-negative masses, they belong to $\mathcal{R}$. Conversely, a vector of moments of a measure f belongs to $L(\mathbb{R}^+)^{N+1}$ only if

$$\int_{\mathbb{R}} \mathbf{l}(v) df(v) \in (\mathbb{R}^+)^{N+1}.$$

This criterion is not satisfied by all positive Borel measures $f \in \mathcal{M}(\mathbb{R})_+$, and could be relaxed to work also with with negative masses.

We study the properties of the modified system (2.16):

Proposition 2.3 (Symmetric hyperbolicity of DVM)**.** *For all sets of strictly convex functions η_i, define*

$$H(\tilde{\mathbf{U}}) = \sum_{i=0}^{N} \eta_i(\tilde{U}_i), \qquad G(\tilde{\mathbf{U}}) = \sum_{i=0}^{N} v_i \eta_i(\tilde{U}_i).$$

Then H is strictly convex and satisfies $G_{\tilde{\mathbf{U}}} = H_{\tilde{\mathbf{U}}} \tilde{\mathbf{F}}_{\tilde{\mathbf{U}}}$. Therefore, (2.16) is symmetrizable, and strongly hyperbolic. Its waves speeds are $(v_0, \dots, v_N) = Sp(\tilde{\mathbf{F}}_{\tilde{\mathbf{U}}})$, i.e. the fixed velocities.

Proof. Compute the Hessians

$$H_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = Diag(\eta_0''(\tilde{U}_0), \dots, \eta_N''(\tilde{U}_N)),$$
$$G_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = Diag(v_0\eta_0''(\tilde{U}_0), \dots, v_N\eta_N''(\tilde{U}_N)).$$

They are symmetric and $H_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}}$ is positive definite. Compute the Jacobians

$$H_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = (\eta_0'(\tilde{U}_0), \dots, \eta_N'(\tilde{U}_N)), \qquad G_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = (v_0\eta_0'(\tilde{U}_0), \dots, v_N\eta_N'(\tilde{U}_N)),$$
$$\tilde{\mathbf{F}}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = Diag(v_0, \dots, v_N).$$

One verifies that $G_{\tilde{\mathbf{U}}} = H_{\tilde{\mathbf{U}}} \tilde{\mathbf{F}}_{\tilde{\mathbf{U}}}$. Differentiating this equality leads to

$$H_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}} \tilde{\mathbf{F}}_{\tilde{\mathbf{U}}} = G_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}} - H_{\tilde{\mathbf{U}}} \tilde{\mathbf{F}}_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}},$$

which is a symmetric matrix. One identifies the entropic variables, its Jacobian, and the symmetrization matrix

$$\mathbf{V}(\tilde{\mathbf{U}}) = H_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}), \qquad \mathbf{V}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) = H_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}),$$
$$\tilde{\mathbf{F}}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}}) \mathbf{V}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}})^{-1} = \mathbf{V}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}})^{-1} \left(G_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}} - H_{\tilde{\mathbf{U}}} \tilde{\mathbf{F}}_{\tilde{\mathbf{U}},\tilde{\mathbf{U}}} \right) \mathbf{V}_{\tilde{\mathbf{U}}}(\tilde{\mathbf{U}})^{-1}.$$

□

After the identification of the candidate entropy for the DVM system, we can construct the operator (2.13b):

$$\tilde{\mathbf{C}}^{relax}(\tilde{\mathbf{U}}) = \frac{\tilde{\mathbf{M}}(\tilde{\mathbf{U}}) - \tilde{\mathbf{U}}}{\tau}, \qquad \tilde{\mathbf{M}}(\tilde{\mathbf{U}}) = \underset{\mathbf{V} - \tilde{\mathbf{U}} \in L^{-1} \boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} H(\mathbf{V}).$$

The unique solution of this optimization problem is of the form

$$\tilde{M}(\tilde{\mathbf{U}})_i = (\eta_i^*)'(\alpha_0 + \alpha_1 v_i + \alpha_2 v_i^2), \tag{2.17}$$

where the Lagrange multipliers $(\alpha_i)_{i=1,\dots,3} \in \mathbb{R}^3$ are the only coefficients such that

$$\sum_{i=0}^{N} (\eta_i^*)'(\alpha_0 + \alpha_1 v_i + \alpha_2 v_i^2) \begin{pmatrix} 1 \\ v_i \\ v_i^2 \end{pmatrix} = \sum_{i=0}^{N} \tilde{U}_i \begin{pmatrix} 1 \\ v_i \\ v_i^2 \end{pmatrix}.$$

Property 2.3 (H-theorem for DVM)**.**

- ***Entropy dissipation:***
  - *For* $\mathbf{C}^{naive}$*, compute*

$$S^{naive}(\tilde{\mathbf{U}}) = \mathbf{V}(\tilde{\mathbf{U}})^T \left(\int_{\mathbb{R}} \mathbf{l}(v) \tilde{M}(\mathbf{p}(\tilde{\mathbf{U}}))(v) dv - \tilde{\mathbf{U}} \right)$$
$$= \sum_{i=0}^{N} \int_{\mathbb{R}} l_i(v) \eta_i'(\tilde{U}_i) \left(M(\tilde{\mathbf{U}})(v) dv - \sum_{j=0}^{N} \tilde{U}_j \delta_{v_j} \right).$$

    The criterion $S^{naive}(\tilde{\mathbf{U}}) \leq 0$ *is not satisfied for all admissible* $\tilde{\mathbf{U}}$*.*
  - *For* $\mathbf{C}^{relax}$*, Proposition 2.1 applies.*
- ***Equilibrium:***
  - *For* $\mathbf{C}^{naive}$*, only* $\mathbf{V}(\tilde{\mathbf{U}}) \in L^{-1}\boldsymbol{\mathcal{I}'}$ *implies* $S^{naive}(\tilde{\mathbf{U}}) = 0$*. But having* $S^{naive}(\tilde{\mathbf{U}}) = 0$ *does not necessarily imply* $\mathbf{V}(\tilde{\mathbf{U}}) \in L^{-1}\boldsymbol{\mathcal{I}'}$*.*
  - *For* $\mathbf{C}^{relax}$*, Proposition 2.1 with a strictly convex entropy provides the equivalence* $S^{relax}(\tilde{\mathbf{U}}) = 0$ *if and only if* $\mathbf{V}(\mathbf{U}) \in L^{-1}\boldsymbol{\mathcal{I}'}$*.*

  Having a vector $\mathbf{V}(\tilde{\mathbf{U}}) \in L^{-1}\boldsymbol{\mathcal{I}'}$ *implies that the mass distribution is of the form* $\tilde{U}_i = m_i = (\eta^*)'(\alpha_0 + \alpha_1 v_i + \alpha_2 v_i^2)$*.*

Such a method provides the desired properties as it mimics the trend towards equilibrium. However, it is only partially satisfactory as it fails to relate to the underlying kinetic entropy dissipation. The resulting moment equilibrium remains represented by a sum of fixed Dirac measures of the form (2.14a), which distance to a Maxwellian is not controlled. Eventually, keeping the velocities v_i constant can also affect the symmetries in the problems or in the accuracy of the approximation. In the next subsection, we focus on an approximation with free velocities, and in the next two sections, we suggest improvements of the DVM to better capture the trend towards equilibrium.

2.3.3 Quadrature-based method of moment (QMOM)

The quadrature-based moment closure is constructed by fixing:

Hypothesis: The number of Dirac measures $J = \frac{N+1}{2}$ in (2.14) is half (integer) the number of moments and the free parameters $\mathbf{p}$ are the masses $\mathbf{m}$ and the positions $\mathbf{v}$ ordered such that

$$\mathbf{p} = (m_1,\ v_1,\ \ldots,\ m_J,\ v_J)^T.$$

With this hypothesis, (2.14) is rewriten

$$\mathbf{U} = L(\mathbf{v})\mathbf{m}, \qquad L_{i,j}(\mathbf{v}) = v_j^i. \tag{2.18}$$

The matrix $L(\mathbf{v})$ is again a Vandermonde matrix, invertible for all $\mathbf{p} \in \mathcal{R}_{\mathbf{p}}$.

Proposition 2.4. *Define the set of parameters and the reduced set of moments*

$$\mathcal{R}_{\mathbf{p}} = \left\{ \mathbf{p} \in \mathbb{R}^{N+1}, \quad m_i > 0\ \forall i, \quad -\infty < v_1 < \cdots < v_J < +\infty \right\}, \quad \mathcal{R}_{app} = int(\mathcal{R}).$$

Then, (2.18) *defines a bijection from* $\mathcal{R}_{\mathbf{p}}$ *to* $\mathcal{R}_{app}$*.*

Proof. For a vector $\mathbf{p} \in \mathcal{R}_\mathbf{p}$, one construct a moment vector $\mathbf{U} \in \mathcal{R}$ using (2.18). The set $\mathcal{R}$ was studied in [81] (based on the atomic representations in [32]), and it was shown that the part of the boundary $\partial\mathcal{R} \cap \mathcal{R} =: \mathcal{R}^m$ is represented only by sums of at most N Dirac measures. Therefore, (2.18) defines an injection from $\mathcal{R}_\mathbf{p}$ into the interior $int(\mathcal{R})$.

Reciprocally, for any vector in the interior $\mathbf{U} \in int(\mathcal{R})$, there exists a measure absolutely continuous with respect to the Lebesgue measure having $\mathbf{U}$ for moments (see again [81]). The existence of the parameters $\mathbf{p}$ then follows from the existence of an exact quadrature formula with $J = \frac{N+1}{2}$ points up to degree $N+1$ (see e.g. [45]). □

The QMOM closure is generally computed using the Chebychev [31, 45] or Wheeler algorithm [98, 45] to compute these unique masses and positions $\mathbf{p}(\mathbf{U})$, and by reinjecting them in the flux definition (2.14c)

$$\mathbf{F} = \sum_{i=1}^{J} m_i(\mathbf{U}) v_i(\mathbf{U}) \mathbf{b}(v_i(\mathbf{U})). \tag{2.19}$$

We study the properties of the system (2.10) with the QMOM closure (2.19):

Proposition 2.5 (Weak hyperbolicity of QMOM)**.**

- *The Jacobian $\mathbf{F}_\mathbf{U}$ is only trigonalizable with real eigenvalues. Therefore, System* (2.10) *with the closure* (2.19) *is not symmetrizable, but only weakly hyperbolic. Its wave speeds are the velocities $(v_0, \dots, v_J) = Sp(\mathbf{F}_\mathbf{U})$.*
- *For all sets of strictly convex functions η_i, define*

$$H(\mathbf{U}) = \sum_{i=1}^{J} \eta_i(m_i(\mathbf{U})), \qquad G(\mathbf{U}) = \sum_{i=1}^{J} v_i(\mathbf{U}) \eta_i(m_i(\mathbf{U})).$$

Then, H is not strictly convex, it satisfies $G_\mathbf{U} = H_\mathbf{U} \mathbf{F}_\mathbf{U}$, but it does not symmetrize (2.10) *with the closure* (2.19)*.*

Proof. • Compute the Jacobians

$$\begin{aligned}
\mathbf{U}_\mathbf{p} &= \left(\mathbf{b}(v_1),\ m_1 \mathbf{b}'(v_1),\ \dots,\ m_J \mathbf{b}'(v_J)\right), \\
\mathbf{F}_\mathbf{p} &= \mathbf{F}_\mathbf{U} \mathbf{U}_\mathbf{p} \\
&= \left(v_1 \mathbf{b}(v_1),\ m_1(v_1 \mathbf{b}'(v_1) + \mathbf{b}(v_1)),\ \dots,\ m_J(v_J \mathbf{b}'(v_J) + \mathbf{b}(v_J))\right).
\end{aligned}$$

One verifies that $\mathbf{U}_\mathbf{p}^{-1}$ is also a Vandermonde matrix, which sends the monomials $\mathbf{b}$ on the Hermite interpolation polynomials

$$\mathbf{h} = \left(h_{1,1},\ \frac{h_{2,1}}{m_1},\ \dots,\ \frac{h_{2,J}}{m_J}\right),$$

those of the first kind being Lagrange ones squared $h_{1,i} = l_i^2$ and those of the second kind being weighted with m_i (see e.g. the computations in Theorem 5.1 in [82]). Remark that the basis $\mathbf{h}(\mathbf{U})$ now depends on $\mathbf{U}$ since the positions $v_i(\mathbf{U})$ do. This provides

$$(\mathbf{U}_\mathbf{p})^{-1} \mathbf{F}_\mathbf{p} = Diag(M^1, \dots, M^J), \qquad M^i = \begin{pmatrix} v_i & m_i \\ 0 & v_i \end{pmatrix}. \tag{2.20}$$

Especially, $\mathbf{F}_\mathbf{U}$ is similar to the matrix (2.20), and it is therefore only trigonalizable and not diagonalizable, which prevents from symmetrizability. Then, the system (2.10) with the QMOM closure (2.14) is only weakly hyperbolic and its wave speeds are the $J = \frac{N+1}{2}$ positions v_i.

- For the convexity of H, consider two vectors $\mathbf{U}, \mathbf{V} \in \mathcal{R}_{app}$ and a parameter $\theta \in [0,1]$. Since $\mathcal{R}_{app}$ and $\mathcal{R}_{\mathbf{p}}$ are in bijection, there exists $\tilde{\mathbf{V}} \in \mathcal{R}_{app}$ such that

$$\mathbf{m}(\tilde{\mathbf{V}}) = \mathbf{m}(\mathbf{V}) \qquad \text{and} \qquad \mathbf{v}(\tilde{\mathbf{V}}) = \mathbf{v}(\mathbf{U}).$$

Using (2.18),

$$\begin{aligned}\mathbf{m}(\mathbf{U} + \theta(\tilde{\mathbf{V}} - \mathbf{U})) &= L(\mathbf{v}(\mathbf{U}))^{-1} \left(\mathbf{U} + \theta(\tilde{\mathbf{V}} - \mathbf{U})\right) \\ &= \mathbf{m}(\mathbf{U}) + \theta(\mathbf{m}(\tilde{\mathbf{V}}) - \mathbf{m}(\mathbf{U})).\end{aligned}$$

Then, using the convexity of η_i

$$\begin{aligned}H(\mathbf{U} + \theta(\mathbf{V} - \mathbf{U})) &= \sum_{i=1}^{J} \eta_i \left(m_i(\mathbf{U}) + \theta(m_i(\tilde{\mathbf{V}}) - m_i(\mathbf{U}))\right) \\ &\leq \sum_{i=1}^{J} \eta_i \left(m_i(\mathbf{U})\right)(1-\theta) + \eta_i(m_i(\tilde{\mathbf{V}}))\theta \\ &= (1-\theta)H(\mathbf{U}) + \theta H(\mathbf{V}).\end{aligned}$$

The lack of strict convexity arises from the lack of variation of the entropy when $\mathbf{U}$ and $\mathbf{V}$ have the same masses.

Finally, for the entropy-entropy flux pair relation, compute the Jacobians

$$\begin{aligned}H_{\mathbf{p}} &= (\eta_1'(m_1),\ 0,\ \eta_2'(m_2),\ 0,\ \ldots,\ \eta_J'(m_J),\ 0)\,, \\ G_{\mathbf{p}} &= (v_1\eta_1'(m_1),\ \eta_1(m_1),\ v_2\eta_2'(m_2),\ \eta_2(m_2),\ \ldots,\ v_J\eta_J'(m_J),\ \eta_J(m_J))\,.\end{aligned}$$

Then, write

$$H_{\mathbf{U}}\mathbf{F}_{\mathbf{U}} = H_{\mathbf{p}}(\mathbf{U}_{\mathbf{p}})^{-1}\mathbf{F}_{\mathbf{p}}(\mathbf{U}_{\mathbf{p}})^{-1}.$$

and one verifies that $G_{\mathbf{p}} = H_{\mathbf{p}}\left((\mathbf{U}_{\mathbf{p}})^{-1}\mathbf{F}_{\mathbf{p}}\right)$ using (2.20).

□

Again, after the identification of the candidate entropy for the QMOM system, we can construct the operator

$$\mathbf{C}^{relax}(\mathbf{U}) = \frac{\mathbf{M}(\mathbf{U}) - \mathbf{U}}{\tau}, \qquad \mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} H(\mathbf{V}). \tag{2.21a}$$

The solutions of this non-strictly convex optimization problem take the form

$$\mathbf{M}(\mathbf{U}) = \sum_{i=0}^{J} \mathbf{b}(u_i)(\eta_i^*)'\left(\alpha_0 + \alpha_1 u_i + \alpha_2 u_i^2\right), \tag{2.21b}$$

for any set of velocities $-\infty < u_0 < \cdots < u_J < +\infty$, and their associated Lagrange multipliers $(\alpha_i)_{i=0,\ldots,2} \in \mathbb{R}^3$ are such that

$$\sum_{i=0}^{N} (\eta_i^*)'(\alpha_0 + \alpha_1 u_i + \alpha_2 u_i^2) \begin{pmatrix} 1 \\ u_i \\ u_i^2 \end{pmatrix} = (U_i)_{i=0,\ldots,2}.$$

These minima are not unique since H is not strictly convex (and especially independent of $\mathbf{v}$). A natural choice consists in fixing

$$u_i = v_i \tag{2.21c}$$

given by the QMOM discretization, but any other choice would dissipate the same entropy.

Proposition 2.6 (H-theorem for QMOM)**.**

- ***Entropy dissipation**:*
  - *For $\mathbf{C}^{naive}$, compute*

$$S^{naive}(\mathbf{U}) = \sum_{i=1}^{J} \int_{\mathbb{R}} l_i(\mathbf{U})(v)^2 \eta_i'(m_i(\mathbf{U})) \left(\tilde{M}(\mathbf{p}(\mathbf{U}))(v)dv - \sum_{j=0}^{J} m_j(\mathbf{U}) \delta_{v_j(\mathbf{U})}(v) \right). \tag{2.22}$$

    The criterion $S^{naive}(\mathbf{U}) \leq 0$ is not satisfied for all realizable $\mathbf{U} \in \mathcal{R}_{app}$.
  - *For $\mathbf{C}^{relax}$, Proposition 2.1 applies.*
- ***Equilibrium**:*
  - *For $\mathbf{C}^{naive}$, again only $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ implies $S^{naive}(\mathbf{U}) = 0$. But having $S^{naive}(\mathbf{U}) = 0$ does not necessarily imply $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$.*
  - *For $\mathbf{C}^{relax}$, due to the lack of strict convexity of H, only $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ implies $S^{relax}(\mathbf{U}) = 0$ and having $S^{relax}(\mathbf{U}) = 0$ does not necessarily imply $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$.*

  Having $H_{\mathbf{U}}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ implies that the mass distribution is of the form $m_i(\mathbf{U}) = (\eta^*)'(\alpha_0 + \alpha_1 v_i + \alpha_2 v_i^2)$ for some velocities $-\infty < v_i < v_{i+1} < +\infty$.

Proof. • Compute

$$S(\mathbf{U}) = H_{\mathbf{U}}(\mathbf{U})\mathbf{C}(\mathbf{U}) = H_{\mathbf{p}}(\mathbf{p}(\mathbf{U}))\mathbf{p}_{\mathbf{U}}(\mathbf{U})\mathbf{C}(\mathbf{U}).$$

Remark that $H_{\mathbf{p}} = (\eta_1'(m_1),\ 0,\ \eta_2'(m_2),\ \ldots,\ 0)$ and that $\mathbf{p}_{\mathbf{U}}$ is the Vandermonde matrix that sends the monomials $\mathbf{b}$ on the Hermite interpolation polynomials $\mathbf{h}$. This yields

$$S^{naive}(\mathbf{U}) = \sum_{i=1}^{J} \eta_i'(m_i(\mathbf{U})) \left(\int_{\mathbb{R}} l_i(\mathbf{U})(v)^2 \tilde{M}(\mathbf{U})(v)dv - m_i(\mathbf{U}) \right),$$

which provides (2.22).

• Construct the polynomials space $p \in Span(h_{1,}, \ldots, h_{1,J})$. It can be represented by $p = \mathbf{V}^T\mathbf{b}$ for some $\mathbf{V} \in Span(\mathbf{V}^1, \ldots, \mathbf{V}^J) \subset \mathbb{R}^{2J}$. Since $dim(Span(\mathbf{V}^1, \ldots, \mathbf{V}^J)) = J > 3 = dim(\boldsymbol{\mathcal{I}'})$, then $Span(\mathbf{V}^1, \ldots, \mathbf{V}^J) \subsetneq \boldsymbol{\mathcal{I}'}$ and one verifies that $H_{\mathbf{U}}(\mathbf{U}) \in Span(\mathbf{V}^1, \ldots, \mathbf{V}^J)$ also implies $S^{naive}(\mathbf{U}) = 0 = S^{relax}(\mathbf{U})$.

□

This model is certainly richer than the DVM, as the velocities v_i are no longer fixed. However, as in the last paragraph, the entropy dissipation and the equilibrium are still not related to the underlying kinetic ones. Furthermore, the equilibria may not be attainable due to the lack of strict convexity of the entropy.

2.3.4 Other quadrature methods, alternatives, and position of the problem

The quadrature-based framework has been an inspiration for the construction of a wide variety of moment closures. The reason lies in the simplicity of their construction, which also provides flexibility (e.g. in the choice of the position of the quadrature points, their weights, their number, their coupling, the shape of the distributions), and in the availability of (relatively efficient) algorithms to compute the parameters out of the moments and vice versa (see e.g. [98, 31, 93, 45]). However, remark that the condition number of the problem of computing the parameters in the construction of QMOM (masses and positions) from the moments increases rapidly with the number of moments (see [45]), and in the recent development, it is often preferred to compute directly the closure $\mathbf{F}$ from the moments $\mathbf{U}$, which is better-conditioned, without computing the parameters (see e.g. [40, 41, 39]).

Among the recent quadrature-based alternatives that have good mathematical properties, we list:

- The **multi-Gaussian closure** ([96, 29]) a.k.a. extended quadrature-based method of moments (EQMOM ; [103]): This closure consists in replacing the Dirac deltas in (2.14a) by Gaussians with a common variance. For a single peak, the variance is closely related to the physical notion of temperature, and this modification allows to accurate model more physically relevant regimes. The EQMOM system was recently analyzed in [105], and it was proved to retrieve strong hyperbolicity and entropy decay.
- The **hyperbolic quadrature-based method of moments** (HyQMOM ; [41, 39]) is at the origin of the present work. The idea originated from algorithmic considerations, by observing that the closure is entirely determined by a single parameter in an orthogonal polynomial sequence (see the precise construction in [39]). For the present purposes, we reinterpret this construction as follows: The HyQMOM closure consists in having one more parameter than necessary in (2.14), i.e. choosing $2J-1$ moments for $2J$ parameters (m_i and v_i), then the resulting set of parameters is constrained (satisfy one additional equation) such that the strong hyperbolicity is enforced. We use a similar idea below to impose symmetrizability of the resulting system. The HyQMOM system was proposed in [41, 39] and the strong hyperbolicity was shown for lower order moments. The analysis was recently completed in [104] in a perturbative framework, the strong hyperbolicity was shown in the general case, and some form of dissipativity was exhibited.
- Among the promising quadrature-inspired closures, we can also mention the **projective closure** [81, 82], which are also based on a positive combination of a Gaussian and a sum of Diracs. Such a construction allows the modeling of physical regimes involving both thermodynamic equilibria and the purely anisotropic regimes. A first attempt of analysis was initiated in [82] exhibiting interesting coupling phenomena in the entropy productions due to the Gaussian and due to the Diracs.

Eventually, the recent work on these constructions exhibited a good mathematical structure regarding the hyperbolicity and the entropy decay. The remaining mathematical difficulties that we aim to study and tackle is the lack of relations between the entropy dissipated at the hyperbolic level in (2.11) and the original kinetic entropy dissipated by (2.2a), even though the existing HyQMOM closure is surely sufficient for many physical simulations. Now, we address the question of constructing a closure of the form (2.14) dissipating a chosen entropy, or at least an approximation of it. Such a relation was also studied from another angle: The entropy-based closures (see e.g. [62] or also [75]) consist in choosing, among all the reconstructions satisfying the moment constraints, the one that minimizes the considered entropy. This leads directly to a symmetrizable moment system. At the hyperbolic level, this system dissipates an entropy corresponding to the one minimized, i.e. the kinetic one or an approximation of it (see e.g. [1, 2]). In order to compute such a closure numerically, one must solve this optimization problem, which requires the computation of integrals. These integrals are generally approximated by quadratures (see e.g. [51, 7, 6]), and bridges with the QMOM closures can be done here. These techniques also inspired part of the construction below.

3 Symmetrization by constrained over-parametrization

In order to construct entropy-dissipating moment closures based on a quadrature approximation, we have illustrated above the important role of intermediate parameters $\mathbf{p}$, i.e. the masses and positions of the quadratures. It is convenient to define the closure using these intermediate parameters. The general idea of the present construction is to have a larger number of intermediate parameters than of moments, and to impose constraints on those additional parameters in order to obtain the desired properties. This idea is more general than the case of the systems of quadrature, and it could be applied to symmetrize other types of moments systems or hyperbolic balance laws.

3.1 General constraints

In a generic framework, suppose that the parameters $\mathbf{p} \in \mathcal{R}_{\mathbf{p}} \subset \mathbb{R}^J$ live in a space of dimension $J > N+1$. And suppose that the unknown $\mathbf{U}$, the flux $\mathbf{F}$ and the source $\mathbf{C}$ are functions of $\mathbf{p} \in \mathcal{R}_{\mathbf{p}}$ into $\mathbb{R}^{N+1}$ such

that

$$\partial_t \mathbf{U}(\mathbf{p}) + \partial_x \mathbf{F}(\mathbf{p}) = \mathbf{C}(\mathbf{p}). \tag{3.1}$$

Since the number $J > N+1$ of parameters $\mathbf{p}$ is larger than that of unknowns $\mathbf{U}$, then the function $\mathbf{U}(\mathbf{p})$ can be a bijection only if we restrict $\mathcal{R}_{\mathbf{p}} \subset \mathbb{R}^J$ to a manifold of dimension $N+1$, i.e. if we add constraints on the free parameters $\mathbf{p}$. Forcing $\mathbf{p}$ to belong to a manifold also implies the existence of a bijection relating $\mathbf{p} \in \mathcal{R}_{\mathbf{p}} \subset \mathbb{R}^J$, and quantities $\mathbf{V} \in \mathbb{R}^{N+1}$, that are meant to become the entropic variables. Then the symmetrization property is rewriten as

$$\mathbf{U}_{\mathbf{p}}(\mathbf{p}(\mathbf{V}))\,\mathbf{p}_{\mathbf{V}}(\mathbf{V}) \text{ symmetric positive definite,} \tag{3.2a}$$

$$\mathbf{F}_{\mathbf{p}}(\mathbf{p}(\mathbf{V}))\,\mathbf{p}_{\mathbf{V}}(\mathbf{V}) \text{ symmetric,} \tag{3.2b}$$

where the Jacobians $\mathbf{U}_{\mathbf{p}}, \mathbf{F}_{\mathbf{p}} \in \mathbb{R}^{(N+1)\times J}$ and $\mathbf{p}_{\mathbf{V}} \in \mathbb{R}^{J\times(N+1)}$ are no longer square matrices. The Jacobians $\mathbf{U}_{\mathbf{p}}$ and $\mathbf{F}_{\mathbf{p}}$ are a priori fixed, but we aim at closing the system by imposing the coupling through the Jacobian $\mathbf{p}_{\mathbf{V}}$.

The existence of such a function $\mathbf{p}(\mathbf{V})$ can be complicated to prove, and we do not necessarily need to compute it. As an alternative, we characterize it by a (slightly) simpler algebraic criterion.

Proposition 3.1. *There exists a surjective function* $\mathbf{p}(\mathbf{V})$ *satisfying* (3.2) *if and only if there exists a matrix function* $A(\mathbf{p}) \in \mathbb{R}^{J\times(N+1)}$ *such that for all* $\mathbf{p} \in \mathcal{R}_{\mathbf{p}} \subset \mathbb{R}^J$*, then* $rank A(\mathbf{p}) = (N+1)$ *and*

$$\sum_{k=1}^{J} (U_i)_{p_k} A_{k,j} - (U_j)_{p_k} A_{k,i} \quad \textit{for all} \quad i > j = 0, \dots, N, \tag{3.3a}$$

$$\sum_{k=1}^{J} (F_i)_{p_k} A_{k,j} - (F_j)_{p_k} A_{k,i} \quad \textit{for all} \quad i > j = 0, \dots, N, \tag{3.3b}$$

$$\sum_{k=1}^{J} (A_{i,j})_{p_k} A_{k,l} - (A_{i,l})_{p_k} A_{k,j} \textit{ for all } i = 1, \dots, J \textit{ and } j > l = 0, \dots, N. \tag{3.3c}$$

Proof. Suppose there exists a surjective function $\mathfrak{p}(\mathbf{V})$ satisfying (3.2). Then its Jacobian is of maximum rank, i.e. $N+1$ and $\mathbf{V} \in Range(\mathfrak{p})$ is in bijection with $\mathbf{U} \in \mathcal{R}_{app}$. Then, defining

$$A(\mathbf{p}) = \mathfrak{p}_{\mathbf{V}}(\mathbf{V}(\mathbf{U}(\mathbf{p}))),$$

one verifies that the symmetry properties (3.2) rewrite (3.3a-3.3b) and the symmetry of the Hessian $(\mathfrak{p}_i)_{\mathbf{V},\mathbf{V}}$ rewrites (3.3c).

In the other sense, the Poincaré lemma provides the existence of a function $\mathfrak{p}_i(\mathbf{V})$ which Jacobian is $(\mathfrak{p}_i)_{\mathbf{V}}(\mathbf{V}) = \tilde{A}_{i,:}(\mathbf{V})$ if and only if $(\tilde{A}_{i,:})_{\mathbf{V}}(\mathbf{V})$ is symmetric. Assuming additionally that $\tilde{A}$ is of rank $(N+1)$ provides the surjectivity of $\mathfrak{p}$ from a non-empty set $\mathcal{R}_{\mathbf{V}} \subset \mathbb{R}^{N+1}$ on the manifold $Range(\mathfrak{p}) \subset \mathbb{R}^J$ of dimension $(N+1)$. Especially, the function $\mathbf{U}(\mathbf{p}(\mathbf{V}))$ defines a bijection from $\mathcal{R}_{\mathbf{V}} \subset \mathbb{R}^{N+1}$ into $\mathcal{R}_{app} \subset \mathbb{R}^{N+1}$ (potentially not the entire realizability set, but a non-empty part of it). Therefore, defining the matrix function

$$A(\mathbf{p}) = \tilde{A}(\mathbf{V}(\mathbf{U}(\mathbf{p}))),$$

the symmetry property on $(\tilde{A}_{i,:})_{\mathbf{V}}(\mathbf{V})$ rewrites the constraint (3.3c) on $A(\mathbf{p})$. Therefore, (3.3c) provides the existence of a function $\mathfrak{p}(\mathbf{V})$ which Jacobian is $\mathfrak{p}_{\mathbf{V}} = A(\mathbf{p})$.

The resulting function $\mathfrak{p}$ satisfies the symmetry properties on $\mathbf{U}_{\mathbf{p}}\mathfrak{p}_{\mathbf{V}}$ and $\mathbf{F}_{\mathbf{p}}\mathfrak{p}_{\mathbf{V}}$ in (3.2) if and only if (3.3a) and (3.3b) hold.

Finally, since the $rank(A(\mathbf{p}))$ is constant over the considered domain $\mathcal{R}_{\mathbf{p}}$, one can always find a clever change of unknown that imposes the positive definiteness of $\mathbf{U}_{\mathbf{p}}\mathfrak{p}_{\mathbf{V}}$ on top of the other constraints. □

This provides a technique to symmetrize the moment system (2.10) through the considered intermediate parameters. However, the constraints on dissipativity and the equilibrium

$$\mathbf{V}(\mathbf{U})^T\mathbf{C}(\mathbf{U}) \leq 0, \qquad \mathbf{V}(\mathbf{U})^T\mathbf{C}(\mathbf{U}) = 0 \;\Leftrightarrow\; \mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$$

still need to be addressed case by case.

3.2 An example in a linearized setting

It may seem difficult to construct such a matrix function satisfying (3.3). We only provide an intuition of how to construct such matrices in simpler settings.

Proposition 3.2 (Symmetric hyperbolicity in a linearized case)**.**
Suppose that $\mathbf{U}$ *and* $\mathbf{F}$ *are linear functions of* $\mathbf{p}$*:*

$$\mathbf{U} = L\mathbf{p} \quad \textit{and} \quad \mathbf{F} = M\mathbf{p}.$$

Then, there exist (constant) matrices A *that symmetrize* (3.1)*. They provide a change of unknowns:*

$$\mathbf{U} = L\mathbf{p}, \qquad \mathbf{p} = A\mathbf{V}, \qquad \mathbf{V} = (LA)^{-1}\mathbf{U}. \tag{3.4a}$$

Then, the flux take the form

$$\mathbf{F} = (MA)\mathbf{V} = (MA)(LA)^{-1}\mathbf{U}. \tag{3.4b}$$

The entropy-entropy flux pair take the form

$$H(\mathbf{U}) = \frac{\mathbf{U}^T(LA)^{-1}\mathbf{U}}{2}, \qquad G(\mathbf{U}) = \frac{\mathbf{U}^T(LA)^{-1}(MA)(LA)^{-1}\mathbf{U}}{2}. \tag{3.4c}$$

Proof. Since A is constant, the constraint (3.3c) holds.

The constraints (3.3a-3.3b) are $N(N+1)$ linear constraints over $A \in \mathbb{R}^{J\times(N+1)}$. Rewriting $A_{i,j} = a_{i+(N+1)(j-1)}$ into a vector $\mathbf{a} \in \mathbb{R}^{J(N+1)}$, this rewrites $\mathbf{a} \in Ker(P)$ with a matrix $P \in \mathbb{R}^{N(N+1)\times J(N+1)}$. Since $J > N+1$, then $dimKer(P) \geq (N+1)$ and the remaining free parameters can be chosen to enforce the positivity definiteness of $\mathbf{U_pp_V} = LA$. □

After the identification of the quadratic entropy, we can construct the operator (2.13b). The unique solution of this optimization problem is of the form

$$\mathbf{C}^{relax}(\mathbf{U}) = \frac{\mathbf{M}(\mathbf{U}) - \mathbf{U}}{\tau}, \qquad \mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} \mathbf{V}^T(LA)^{-1}\mathbf{V}.$$

This unique minimizer is the form linear combination of three vectors $\mathbf{W}^i$ with weights given by the first values U_i

$$\mathbf{M}(\mathbf{U}) = U_0\mathbf{W}^0 + U_1\mathbf{W}^1 + U_2\mathbf{W}^2. \tag{3.5}$$

Indeed, one verifies that $\mathbf{M}(U) = (LA)\boldsymbol{\lambda}$, where the only non-zero Lagrange multipliers are associated with the three-dimensional constraints $(\mathbf{M}(\mathbf{U}) - \mathbf{U})^T\boldsymbol{\phi} = 0$ for $\boldsymbol{\phi} \in \boldsymbol{\mathcal{I}}'$, i.e. $\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, 0, \ldots, 0)^T$ when $\mathbf{b}(v) = (1, v, \ldots, v^N)^T$ is the canonical basis. In that case, they are given by $\boldsymbol{\lambda} = \widetilde{(LA)^{-1}}\mathbf{U}$ where $\widetilde{(LA)^{-1}}$ is the top-left 3×3 corner of $(LA)^{-1}$.

Property 3.1 (H-theorem in a linearized framework)**.**

- ***Entropy dissipation****: The entropy dissipation takes the form:*

$$\mathbf{U}^T(LA)^{-1}\mathbf{C}(\mathbf{U}) \leq 0. \tag{3.6a}$$

– *For* $\mathbf{C}^{naive}$*, using a Cholesky decomposition* $(LA)^{-1} = BB^T$*, then*

$$S^{naive} = \frac{1}{\tau}\left(\int_{\mathbb{R}} (B\mathbf{U})^T (B\mathbf{b})(v) M(\mathbf{U}) dv - (B\mathbf{U})^2\right). \tag{3.6b}$$

The criterion $S^{naive}(\mathbf{U}) \leq 0$ *is not satisfied for all* $\mathbf{U}$*.*

– *For* $\mathbf{C}^{relax}$*, compute*

$$S^{relax} = \frac{1}{\tau}\mathbf{U}^T\left(\widetilde{(LA)^{-1}} - (LA)^{-1}\right)\mathbf{U} \leq 0, \tag{3.6c}$$

where the matrix $\widetilde{(LA)^{-1}} - (LA)^{-1}$ *is symmetric non-positive. Also Proposition 2.1 applies.*

- ***Equilibrium****: The equilibrium is represented by the equivalence:*

$$\mathbf{U}^T(LA)^{-1}\mathbf{C}(\mathbf{U}) = 0 \qquad \Leftrightarrow \qquad \mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}' \qquad \Leftrightarrow \qquad \mathbf{U} \in (LA)\boldsymbol{\mathcal{I}}'. \tag{3.6d}$$

– *For* $\mathbf{C}^{naive}$*, only* $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$ *implies* $S^{naive}(\mathbf{U}) = 0$*. But having* $S^{naive}(\mathbf{U}) = 0$ *does not necessarily imply* $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$*.*

– *For* $\mathbf{C}^{relax}$*, Proposition 2.1 with a strictly convex entropy provides the equivalence* $S^{relax}(\mathbf{U}) = 0$ *if and only if* $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$*, or equivalently if* $\mathbf{U} \in Span(\mathbf{W}^0,\ \mathbf{W}^1,\ \mathbf{W}^2)$*.*

This simplified linearized setting is sufficient for many applications. However, the quadratic entropy found in (3.4c) and the associated equilibrium in (3.5) are not yet related to the underlying kinetic entropy (2.2a). In the next two sections, we exploit this idea to construct entropy-dissipating closures for augmented DVM and augmented QMOM systems.

4 Augmented discrete velocity methods (ADVM)

First, we focus on the DVM case. The entropy property is modified by augmenting the number of quadrature nodes in the construction (2.16), i.e. $J > N + 1$.

We first focus on a direct application of the linearized case of Section 3.2. This allows for simplifications. Then, we extend the construction to mimic an underlying kinetic entropy dissipation.

4.1 Symmetrization criterion and a quadratic entropy

As in the well-parametrized case of Section 2.3.2, write the moment vectors $\mathbf{U}$ in the Lagrange interpolation polynomials associated to the first positions $v_0, \ldots, v_N$:

$$\begin{aligned} \mathfrak{l}(v) &:= \left(\prod_{k=1}^{N}\frac{(v-v_k)}{(v_0-v_k)},\ \frac{(v-v_0)}{(v_1-v_0)}\prod_{k=2}^{N}\frac{(v-v_k)}{(v_1-v_k)},\ \ldots,\ \prod_{k=1}^{N-1}\frac{(v-v_k)}{(v_N-v_k)}\right)^T, \\ \mathfrak{l}_i(v) &:= v l_i(v). \end{aligned}$$

Then, decomposing $\mathbf{p} = \mathbf{m} = (\bar{\mathbf{m}}^T, \tilde{\mathbf{m}}^T)^T$ into the two parts $\bar{\mathbf{m}} = (m_i)_{i=0,\ldots,N}$ and $\tilde{\mathbf{m}} = (m_i)_{i=N+1,\ldots,J-1}$, the Jacobians are

$$\mathbf{U}_{\mathbf{p}} = L = \left(\ Id^{N+1}\ \middle|\ \mathbf{U}_{\tilde{\mathbf{p}}}\ \right), \qquad (\mathbf{U}_{\tilde{\mathbf{p}}})_{i,j} = l_i(v_{j+N}), \tag{4.1a}$$

$$\mathbf{F}_{\mathbf{p}} = M = \left(\ D\ \middle|\ \mathbf{F}_{\tilde{\mathbf{p}}}\ \right), \quad D = Diag(v_0,\ \ldots,\ v_N), \quad (\mathbf{F}_{\tilde{\mathbf{p}}})_{i,j} = \mathfrak{l}_i(v_{j+N}), \tag{4.1b}$$

that are constants, depending only on the fixed positions v_i. In that case, the constraints (3.3a-3.3b) rewrite, for all $i > j = 0, \dots, N$

$$0 = A_{i,j} - A_{j,i} + \sum_{k=N+1}^{J-1} l_i(v_k)A_{k,j} - l_j(v_k)A_{k,i}, \tag{4.2a}$$

$$0 = v_i A_{i,j} - v_j A_{j,i} + \sum_{k=N+1}^{J-1} \mathfrak{l}_i(v_k)A_{k,j} - \mathfrak{l}_j(v_k)A_{k,i}. \tag{4.2b}$$

Then, Proposition 3.2 applies, and the system (2.10) with the flux (3.4b) is symmetric hyperbolic, associated with the entropy-entropy flux pair (3.4c) which rewrites

$$H = \frac{1}{2}\sum_{i=1}^{J}\sum_{j=0}^{N} \left(m_i q_j(v_i)\right)^2, \qquad G = \frac{1}{2}\sum_{i=1}^{J}\sum_{j=0}^{N} v_i \left(m_i q_j(v_i)\right)^2, \qquad \mathbf{V} = (LA)^{-1}\mathbf{U},$$

where $\mathbf{q} = B\mathbf{l}$ and B^T is the Cholesky decomposition of $(LA)^{-1}$. The entropic variables are still moments of the considered approximation (2.14), but written in a different polynomial basis.

After the identification of the candidate entropy, we can construct the relaxation operator (2.13b): The equilibrium $\mathbf{M}(\mathbf{U})$ in (2.13b) is the solution of the problem

$$\mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} \; H(\mathbf{V}).$$

Denoting $\mathbf{v}^j = (v_1^j, \dots, v_J^j)^T$, then this minimum is

$$\mathbf{M}(\mathbf{U}) = A(\lambda_0 \mathbf{v}^0 + \lambda_1 \mathbf{v}^1 + \lambda_2 \mathbf{v}^2), \tag{4.3}$$

where the Lagrange multipliers $(\lambda_i)_{i=0,\dots,2} \in \mathbb{R}^3$ associated with the constraints $\mathbf{V} - \mathbf{U} \in \boldsymbol{\mathcal{I}}^{\perp}$ are the only coefficients such that the moments up to order 2 of $\mathbf{M}(\mathbf{U})$ are those of $\mathbf{U}$.

Proposition 4.1 (H-theorem for the linear ADVM)**.**

- ***Entropy dissipation**:*
  - *For $\mathbf{C}^{naive}$, compute*

$$S^{naive}(\mathbf{U}) = \frac{1}{\tau}\left(\int_{\mathbb{R}} (B\mathbf{U})^T (B\mathbf{l}(v)) M(\mathbf{U})(v) dv - (B\mathbf{U})^2\right). \tag{4.4}$$

    Again this term is not signed for all $\mathbf{U} \in \mathcal{R}_{app}$.
  - *For $\mathbf{C}^{relax}$, Proposition 2.1 applies. The term S^{relax} takes the same form as in* (3.6c)*.*
- ***Equilibrium**:*
  - *For the naive operator $\mathbf{C}^{naive}$, only $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$ implies $S^{naive}(\mathbf{U}) = 0$. But having $S^{naive}(\mathbf{U}) = 0$ does not necessarily imply $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$.*
  - *For the relaxation operator $\mathbf{C}^{relax}$, Proposition 2.1 with a strictly convex entropy applies.*

  Having $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}}'$ implies that $\mathbf{U}$ is of the form (4.3)

Eventually, one can play with the size of the matrix A and the values of the v_i in order to alter the form of the entropy, but it remains a quadratic function of $\mathbf{U}$ unrelated to the kinetic entropy. As an alternative, the matrix A can be tuned to fit the kinetic entropy. It can be chosen non-constant depending on $\mathbf{p}$, but it is not obvious how to make such choices to relate to the kinetic entropy. This first linear version can still be useful when approximating linearized kinetic equations. In the next paragraph, we suggest an alternative construction to reach our goal in a non-linear setting.

4.2 Extension to a quadrature-based entropy

The aim here is to construct a closure from the considered kinetic entropy à la Levermore [62]. Choose a set of:

- Positive weights $w_i > 0$.
- Strictly convex entropies $\eta_i : \mathbb{R} \to \mathbb{R}$ such that $\eta_i' > 0$.

Define the entropy based on the intermediate parameters $\mathbf{p} = \mathbf{m}$

$$\tilde{H}(\mathbf{m}) = \sum_{i=1}^{J} w_i \eta_i(m_i), \tag{4.5a}$$

and define the entropy based on the conserved variables as

$$H(\mathbf{U}) = \min_{L\mathbf{m}=\mathbf{U}} \tilde{H}(\mathbf{m}). \tag{4.5b}$$

For all realizable $\mathbf{U} \in \mathcal{R}_{app}$, this minimum exists since this is a strictly convex problem over $\mathbf{m} \in (\mathbb{R}^+)^J$ under linear constraints.

Property 4.1. *The function H is strictly convex.*

Proof. For all pairs $\mathbf{m}^1$ and $\mathbf{m}^2$ in $(\mathbb{R}^+)^J$ and $\delta \in [0,1]$, one has

$$\tilde{H}\left(\mathbf{m}^1 + \delta(\mathbf{m}^2 - \mathbf{m}^1)\right) < \tilde{H}(\mathbf{m}^1) + \delta\left(\tilde{H}(\mathbf{m}^2) - \tilde{H}(\mathbf{m}^1)\right).$$

Taking the minimum over $(\mathbf{m}^1, \mathbf{m}^2)$ on both sides under the constraints $L\mathbf{m}^1 = \mathbf{U}^1$ and $L\mathbf{m}^2 = \mathbf{U}^2$ leads to the result. □

Let us construct its Legendre-Fenchel transform.

Property 4.2. *The Legendre-Fenchel transform H^* of H is*

$$H^*(\mathbf{V}) = \sum_{j=1}^{J} w_j \eta_j^*\left(\frac{\mathbf{V}^T\mathbf{b}(v_j)}{w_j}\right). \tag{4.6}$$

Proof. Compute the Legendre-Fenchel transform of $\tilde{H} : (\mathbb{R}^+)^J \to \mathbb{R}$

$$\tilde{H^*}(\boldsymbol{\mu}) = \sum_{j=1}^{J} w_j \eta_j^*\left(\frac{\mu_j}{w_j}\right).$$

Then, this result is a direct application of Fenchel strong duality with linear constraints $L\mathbf{m} = \mathbf{U}$ (see e.g. Corollary 3.3.11 and Exercise 3.22 in [15], see also [55, 84]). □

With this transform, one obtains the bijection relating the conserved variables $\mathbf{U}$ and the entropic variables $\mathbf{V}$:

Property 4.3. *The definition of the entropy H provides the change of unknowns:*

$$\begin{aligned} \mathbf{U}(\mathbf{V}) &= H^*_{\mathbf{V}}(\mathbf{V})^T \\ &= \sum_{i=1}^{J} \mathbf{b}(v_i)(\eta_i^*)'\left(\frac{\mathbf{V}^T\mathbf{b}(v_i)}{w_i}\right), \end{aligned} \qquad \begin{aligned} \mathbf{V}(\mathbf{U}) &= H_{\mathbf{U}}(\mathbf{U})^T = (L^TL)^{-1}L\mathbf{W}(\mathbf{U}) \\ &= \sum_{i=1}^{J} w_i \eta_i'(m_i(\mathbf{U}))\mathbf{g}(v_i), \end{aligned}$$

where $\mathbf{g}$ is a vector of polynomials, $\mathbf{W}(\mathbf{U}) \in \mathbb{R}^J$ is a vector and $L \in \mathbb{R}^{J\times(N+1)}$ is a Vandermonde matrix. They are all associated with $(v_i)_{i=1,\dots,J}$ as

$$\mathbf{g} = (LL^T)^{-1}\mathbf{b}, \qquad W(\mathbf{U})_i = w_i\eta_i'(m_i(\mathbf{U})), \qquad L_{i,j} = v_i^j.$$

Proof. Differentiating H^* gives the formula of $\mathbf{U}(\mathbf{V})$, where one identifies

$$m_j = (\eta_j^*)' \left(\frac{\mathbf{V}^T \mathbf{b}(v_j)}{w_j} \right) > 0.$$

The function H is defined implicitly as the solution of the optimization problem (4.5), and its computation is not straightforward. Instead, one identifies the masses solving the optimization problem (4.5) in the formula of $\mathbf{U}(\mathbf{V})$:

$$m_i(\mathbf{U}) = (\eta_i^*)' \left(\frac{\mathbf{V}(\mathbf{U})^T \mathbf{b}(v_i)}{w_i} \right).$$

This formula leads to an over-parametrized linear system, with a solution by invertibility of $\mathbf{U}(\mathbf{V})$. Solving this problem using the normal equation provides the given formula. This choice is also preferable for numerical purposes to avoid creating a preferred direction. □

Remark 4.1. The masses solving the optimization problem (4.5) are strictly positive. Therefore, the optimization problem (4.5) turns ill-posed along the boundary $\partial \mathcal{R}_{app}$ of the augmented DVM realizability domain and ill-conditioned in its vicinity. This domain is the polyhedral cone

$$\mathcal{R}_{app} = Cone\left(\mathbf{b}(v_i),\ i = 1, \dots, J\right).$$

This is observed identically in the construction of the classical entropy-based closures [51, 7, 6]. This is also expected for all realizable closures. However, this can be relaxed by different modifications of the entropy (see e.g. [2, 5]).

One obtains the fluxes as functions of the masses $\mathbf{m}(\mathbf{U})$, solution of the optimization problem (4.5):

Property 4.4. *The flux* $\mathbf{F}(\mathbf{U})$ *and the entropy flux* $G(\mathbf{U})$ *are*

$$\mathbf{F}(\mathbf{U}) = \sum_{i=1}^{J} m_i(\mathbf{U}) v_i \mathbf{b}(v_i) = M \mathbf{m}(\mathbf{U}), \qquad \textit{with} \qquad M_{i,j} = v_j^{i+1}, \tag{4.7a}$$

$$G(\mathbf{U}) = \sum_{i=1}^{J} w_i \left(\sum_{j=1}^{J} \eta_i(m_i(\mathbf{U})) m_j(\mathbf{U}) v_j \left(\mathbf{k}(v_i)^T \mathbf{k}(v_j) \right) \right. \tag{4.7b}$$

$$\left. - \eta_i \left(\sum_{j=1}^{J} \frac{w_j}{w_i} \eta_j(m_j(\mathbf{U})) \left(\mathbf{k}(v_j)^T \mathbf{k}(v_i) \right) \right) \right),$$

where $\mathbf{k} = B\mathbf{b}$ *and* B *is the Cholesky decomposition of* $(LL^T)^{-1} = (BB^T)$.

Proof. For the fluxes, one defines similarly

$$G^*(\mathbf{V}) = \sum_{i=1}^{J} v_i w_i \eta_i^* \left(\frac{\mathbf{V}^T \mathbf{b}(v_i)}{w_i} \right),$$

$$\mathbf{F}(\mathbf{U}(\mathbf{V})) = G^*_{\mathbf{V}}(\mathbf{V}) = \sum_{i=1}^{J} v_i \mathbf{b}(v_i) (\eta_i^*)' \left(\frac{\mathbf{V}^T \mathbf{b}(v_i)}{w_i} \right),$$

and deduce

$$\mathbf{F}(\mathbf{U}) = G^*_{\mathbf{V}}(\mathbf{V}(\mathbf{U})) = \sum_{i=1}^{J} v_i \mathbf{b}(v_i) (\eta_i^*)' \left(\frac{\mathbf{V}(\mathbf{U})^T \mathbf{b}(v_i)}{w_i} \right) = \sum_{i=1}^{J} m_i(\mathbf{U}) v_i \mathbf{b}(v_i),$$

$$G(\mathbf{U}) = \mathbf{V}(\mathbf{U})^T \mathbf{F}(\mathbf{U}) - G^*(\mathbf{V}(\mathbf{U})).$$

Reinjecting $\mathbf{V}(\mathbf{U})$ in the latter reduces to (4.7). □

Eventually, these results are summarized:

Proposition 4.2 (Symmetric hyperbolicity of ADVM with the quadrature-based entropy)**.** *The functions H, G and $\mathbf{F}$, defined in* (4.5b), (4.7b) *and* (4.7a), *satisfy*

$$G_{\mathbf{U}} = H_{\mathbf{U}}\mathbf{F}_{\mathbf{U}}.$$

Then, the system (2.10) *with the closure* (4.7a) *is symmetrizable.*

Its wave speeds are bounded by the extremum velocities

$$Sp(\mathbf{F}_{\mathbf{U}}) \subset [v_1, v_J].$$

Proof. For the bounds on the wave speeds, one has $\mathbf{F}_{\mathbf{U}} = \mathbf{F}_{\mathbf{V}}(\mathbf{U}_{\mathbf{V}})^{-1}$ with

$$\mathbf{U}_{\mathbf{V}} = \sum_{i=1}^{J} \frac{\mathbf{b}(v_i)\mathbf{b}(v_i)^T}{w_i}\eta_i''\left(\frac{\mathbf{V}^T\mathbf{b}(v_i)}{w_i}\right), \quad \mathbf{F}_{\mathbf{V}} = \sum_{i=1}^{J} v_i\frac{\mathbf{b}(v_i)\mathbf{b}(v_i)^T}{w_i}\eta_i''\left(\frac{\mathbf{V}^T\mathbf{b}(v_i)}{w_i}\right).$$

Since $\mathbf{U}_{\mathbf{V}}$ is symmetric positive definite, then it is diagonalizable with positive eigenvalues. And one observes that the matrix

$$(\mathbf{F}_{\mathbf{U}} - v_1 Id) = (\mathbf{F}_{\mathbf{V}} - v_1\mathbf{U}_{\mathbf{V}})(\mathbf{U}_{\mathbf{V}})^{-1} = \sum_{i=1}^{J}(v_i - v_1)\frac{\mathbf{b}(v_i)\mathbf{b}(v_i)^T}{w_i}\eta_i''\left(\frac{\mathbf{V}^T\mathbf{b}(v_i)}{w_i}\right)$$

is similar to a non-negative matrix. It is therefore non-negative. Similar computations holds for $(v_J Id - \mathbf{F}_{\mathbf{U}})$. □

After the identification of the candidate entropy, we construct the relaxation operator (2.13b). The equilibrium $\mathbf{M}(\mathbf{U})$ is the solution of the minimization problem (2.13b) that can be simplified:

$$\min_{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}} H(\mathbf{V}) = \min_{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}} \min_{L\mathbf{m}=\mathbf{V}} \tilde{H}(\mathbf{m}) = \min_{\tilde{L}\mathbf{m}=\tilde{\mathbf{U}}} \tilde{H}(\mathbf{m}),$$

where $\tilde{L} \in \mathbb{R}^{3\times J}$ and $\tilde{\mathbf{U}}$ correspond to the parts of L and of $\mathbf{U}$ associated with $\boldsymbol{\mathcal{I}}'$. If $\mathbf{b}(v) = (1, v, \dots, v^N)^T$, these are simply

$$\tilde{L}_{i,j}(\mathbf{v}) = v_j^i, \qquad \tilde{U}_i = U_i.$$

This simply corresponds to the minimization of entropy under the constraint that the moments $\mathbf{M}(\mathbf{U})$ are those of $\mathbf{U}$ up to order two. Therefore, this minimizer takes the form

$$\mathbf{M}(\mathbf{U}) = \sum_{i=1}^{J} \mathbf{b}(v_i)(\eta_i^*)'(\lambda_0 + \lambda_1 v_i + \lambda_2 v_i^2), \tag{4.8}$$

where the Lagrange multipliers $(\lambda_i)_{i=0,\dots,2} \in \mathbb{R}^3$ are the only coefficients such that the moments up to order 2 of $\mathbf{M}(\mathbf{U})$ are the same as those of $\mathbf{U}$.

We summarize the entropy dissipation study:

Proposition 4.3 (H-theorem for the ADVM with the quadrature-based entropy)**.**

- ***Entropy dissipation:***
  - *For $\mathbf{C}^{naive}$, compute*

$$S^{naive}(\mathbf{U}) = \frac{1}{\tau}\sum_{i=1}^{J} w_i\eta_i'(m_i(\mathbf{U}))\left(\int_{\mathbb{R}} \mathbf{k}(v_i)^T\mathbf{k}(v)M(\mathbf{U})(v)dv - \sum_{j=1}^{J} m_j(\mathbf{U})\mathbf{k}(v_i)^T\mathbf{k}(v_j)\right), \tag{4.9}$$

    where $\mathbf{k} = B\mathbf{b}$ and B is the Cholesky decomposition $BB^T = (LL^T)^{-1}$. This term is not signed for all $\mathbf{U} \in \mathcal{R}_{app}$.

– *For* $\mathbf{C}^{relax}$*, Proposition 2.1 applies.*

- ***Equilibrium****:*
  - *For* $\mathbf{C}^{naive}$*, only* $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ *implies* $S^{naive}(\mathbf{U}) = 0$*. But having* $S^{naive}(\mathbf{U}) = 0$ *does not necessarily imply* $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$*.*
  - *For* $\mathbf{C}^{relax}$*, Proposition 2.1 with a strictly convex entropy applies.*

Having $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ *implies that* $\mathbf{U}$ *is of the form* (4.8)

Remark that this is equilibrium is still not a Maxwellian. It is only an approximation of a Maxwellian by a sum of Dirac deltas. However, the number of deltas is now chosen to be $J > N + 1$.

In the next section, we extend this approach in the QMOM framework, i.e. having the velocities v_i free as well.

5 Augmented quadrature-based methods (AQMOM)

First, we write the criterion for the existence of a strictly convex entropy in the augmented QMOM framework. Then, we provide one specific choice based on the underlying kinetic entropy.

For technical reasons, in this section, the number of parameters is supposed to satisfy the following additional hypothesis:

Hypothesis: The number of moments $Card(\mathbf{b}) = N + 1 = 2M$ is even and the number of parameters $Card(\mathbf{p}) = 2J$ satisfies $M \leq J$, i.e. there are more Dirac deltas in (2.14) than moments, or more than twice as many parameters as moments.

5.1 Symmetrization criterion

In the spirit of [82] and as in Section 2.3.3, we use the Hermite interpolation polynomial at the points $(v_i)_{i=1,\dots,M}$ as basis functions $\mathbf{b}$. They give $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2)$ with

$$h_{1,i} = l_i^2, \qquad h_{2,i}(v) = (v - v_i) l_i^2(v), \qquad \text{where} \qquad l_i(v) = \prod_{\substack{j=1 \\ j \neq i}}^{M} \frac{v - v_j}{v_i - v_j}$$

are the Lagrange interpolation polynomials at the first velocities v_i for $i = 0, \dots, N$. Denote $\mathfrak{h}(v) = v\mathbf{h}(v)$. This choice provides the Jacobians

$$\mathbf{U}_{\mathbf{p}} = \left(\; Diag(B_1, \dots, B_M) \;\middle|\; \mathbf{h}(v_{M+1}),\; m_{M+1}\mathbf{h}'(v_{M+1}),\; \dots,\; m_J\mathbf{h}'(v_J) \; \right), \tag{5.1a}$$

$$\mathbf{F}_{\mathbf{p}} = \left(\; Diag(C_1, \dots, C_M) \;\middle|\; \mathfrak{h}(v_{M+1}),\; m_{M+1}\mathfrak{h}'(v_{M+1}),\; \dots,\; m_J\mathfrak{h}'(v_J) \; \right), \tag{5.1b}$$

$$B_i = \begin{pmatrix} 1 & 2m_i l_i'(v_i) \\ 0 & m_i \end{pmatrix}, \qquad C_i = \begin{pmatrix} v_i & m_i\,(1 + 2v_i l_i'(v_i)) \\ 0 & m_i v_i \end{pmatrix}, \tag{5.1c}$$

where the first part of the matrices correspond to $\mathbf{U}_{\bar{\mathbf{p}}}$ and $\mathbf{F}_{\bar{\mathbf{p}}}$ and the second to $\mathbf{U}_{\tilde{\mathbf{p}}}$ and $\mathbf{U}_{\tilde{\mathbf{p}}}$, using the decomposition $\mathbf{p} = (\bar{\mathbf{p}}^T, \tilde{\mathbf{p}}^T)^T$ with

$$\bar{\mathbf{p}} = (m_1,\; v_1,\; \dots,\; m_M,\; v_M)^T, \qquad \tilde{\mathbf{p}} = (m_{M+1},\; v_{M+1},\; \dots,\; m_J,\; v_J)^T.$$

For the sake of conciseness, the conditions (4.2) are rewritten with this choice of polynomial basis only in Appendix A.

The reformulation of these constraints in the QMOM framework leads to a system of equations that are linear over A, but polynomial over $\mathbf{p}$. Furthermore, the constraints (3.3c) are now a system of quadratic quasi-linear first-order differential conservation laws. The DVM case (4.2), i.e. by forcing the positions v_i to be constant, provides an example of parameters satisfying these constraints. However, we have not found a non-trivial modification of this closure that does not comply with the DVM framework.

Therefore, we only provide an alternative based on an entropy minimizing technique.

5.2 A tentative quadrature-based entropy based on a max-min problem

We modify the construction of Section 4.2 using the kinetic entropy to adapt it to the QMOM framework.

5.2.1 Constraints and position of the problem

First, we need to adapt the framework of Section 4.2 with varying positions v_i. Choose a set of:

- Positive weights $w_i(\mathbf{v}) > 0$, where the positions $\mathbf{v}$ are varying parameters, i.e. $\mathbf{p} = (\mathbf{m}^T, \mathbf{v}^T)^T$. Further constraints on these weights are added in the construction below.
- Strictly convex entropies $\eta_i : \mathbb{R} \to \mathbb{R}$ such that $\eta_i' > 0$.

Based on the intermediate parameters, define

$$\tilde{H}(\mathbf{p}) = \sum_{i=1}^{J} w_i(\mathbf{v})\eta_i(m_i). \tag{5.2}$$

Remark 5.1. Assume that the i-th weight $w_i(\mathbf{v}) = w_i(v_i)$ depends only on the associated position v_i. Computing the Hessian of this function leads to a block diagonal matrix $\tilde{H}_{\mathbf{p},\mathbf{p}}(\mathbf{p}) = Diag(\tilde{H}_1, \dots, \tilde{H}_J)$ with diagonal blocks of the form

$$\tilde{H}_i = \begin{pmatrix} w_i(v_i)\eta_i''(m_i) & w_i'(v_i)\eta_i'(m_i) \\ w_i'(v_i)\eta_i'(m_i) & w_i''(v_i)\eta_i(m_i) \end{pmatrix}.$$

Therefore, $\tilde{H}$ is strictly convex only under the additional condition

$$w_i(v_i)w_i''(v_i)\eta_i(m_i)\eta_i''(m_i) - w_i'(v_i)^2\eta_i'(m_i)^2 > 0$$

or equivalently

$$\frac{w_i''w_i}{(w_i')^2}(v_i) > \left(\frac{\eta_i''\eta_i}{(\eta_i')^2}(m_i)\right)^{-1}.$$

This relation between the m_i and v_i was not found meaningful in our framework, so another construction of the strictly convex entropy H is proposed.

5.2.2 Existence of a solution to a max-min problem

Define the entropy based on the conserved variables as the solution of the max-min problem (see e.g. [97, 84, 91, 36, 16]):

$$\bar{H}(\mathbf{U};\mathbf{v}) = \begin{cases} \min\limits_{\mathbf{m}\in\mathcal{R}_{\mathbf{m}}(\mathbf{U};\mathbf{v})} \tilde{H}(\mathbf{m},\mathbf{v}) & \text{if} \quad \mathcal{R}_{\mathbf{m}}(\mathbf{U};\mathbf{v}) \neq \emptyset, \\ -\infty & \text{otherwise.} \end{cases} \tag{5.3a}$$

$$H(\mathbf{U}) = \max_{\mathbf{v}\in\mathcal{R}_{\mathbf{v}}} \bar{H}(\mathbf{U};\mathbf{v}), \tag{5.3b}$$

where the sets are

$$\mathcal{R}_{\mathbf{m}}(\mathbf{U};\mathbf{v}) = \left\{ \mathbf{m} \in (\mathbb{R}^+)^J \quad \text{such that} \quad \sum_{i=1}^{J} m_i \mathbf{b}(v_i) = \mathbf{U} \right\}, \tag{5.3c}$$

$$\mathcal{R}_{\mathbf{v}} = \left\{ \mathbf{v} \in \mathbb{R}^J \quad \text{such that} \quad v_i < v_{i+1} \right\}. \tag{5.3d}$$

This can also be written as a bilevel optimization problem (see e.g. [95, 9, 8] and references therein).

Remark 5.2. This choice of couples $(\mathbf{m}, \mathbf{v})$ is motivated by two reasons:

- It provides all the mathematical properties desired for the resulting system of moments (see Section 5.2.4 below).
- The minimization part can still be interpreted as an entropy minimization, i.e. the most probable such choice ([56]) satisfying the moment constraints (5.3c). The maximization can be interpreted as a dilation from the equilibrium to observe the most accurate out-of-equilibrium representation.

We need to prove that the function H is well-defined for all realizable $\mathbf{U} \in \mathcal{R}_{app}$, and that it is strictly convex in $\mathbf{U}$.

Proposition 5.1. *Suppose that $\eta_i \in C^2(\mathbb{R}^+)$ and $w_i \in C^2(\mathcal{R}_\mathbf{v})^+$ satisfy*

$$\lim_{m\to 0} \eta_i(m) = -\infty, \qquad \forall i, \tag{5.4a}$$

$$\lim_{\mathbf{v}\to\partial\mathcal{R}_\mathbf{v}} \mathbf{w}(\mathbf{v}) = \mathbf{w}^{\lim} \notin (\mathbb{R}^{*+})^J, \tag{5.4b}$$

and at least one of its components is finite in this limit

$$w_i^{\lim} \in \mathbb{R}^{*+}. \tag{5.4c}$$

Then, for all $\mathbf{U} \in \mathcal{R}_{app}$, there exists a maximizer $\mathbf{v}^*$ to (5.3b).

Remark 5.3. The assumptions on $\mathbf{w}$ can be rewritten as:

- Every weight w_i must to be a non-negative function of the velocities $\mathbf{v} \in \mathcal{R}_\mathbf{v}$, with a C^2-regularity.
- The limit $\mathbf{v} \to \partial\mathcal{R}_\mathbf{v}$ corresponds either to the case where two velocities cross each others $v_{i+1} - v_i \to 0$, or when the extreme ones $v_0 \to -\infty$ or $v_J \to +\infty$ go to infinity.
- In such a limit, at least one of the weights w_i is assumed to go either to 0 or to $+\infty$, and at least one of them is assumed to remain strictly positive and bounded.

Proof. First, the case of uniformly distributed velocities $\mathbf{v}^{DVM}$ was studied in the last section. It was shown to have a finite entropy $\bar{H}(\mathbf{U};\mathbf{v}^{DVM}) = H^{DVM} \in \mathbb{R}$. Define

$$\mathcal{S}_\mathbf{v} = \left\{ \mathbf{v} \in \mathcal{R}_\mathbf{v} \quad \text{such that} \quad H(\mathbf{U};\mathbf{v}) \geq H^{DVM} \right\}.$$

Let us show now that this set is a compact subset of $\mathcal{R}_\mathbf{v}$.

Consider a limit $\mathbf{v} \to \partial\mathcal{R}_\mathbf{v}$, such that the two components i, j satisfy

$$\lim_{\mathbf{v}\to\partial\mathcal{R}_\mathbf{v}} w_i(\mathbf{v}) = 0, \qquad \lim_{\mathbf{v}\to\partial\mathcal{R}_\mathbf{v}} w_j(\mathbf{v}) = w_j^* \in \mathbb{R}.$$

Rewrite

$$\tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v}) = w_j(\mathbf{v}) \left[\sum_{\substack{k=1 \\ k\neq j}}^{J} \frac{w_k(\mathbf{v})}{w_j(\mathbf{v})} \eta_k(m_k) + \eta_j(m_j) \right].$$

Then, one observes that $\tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v}) \to -\infty$ in the limit $m_j \to 0$. Therefore, the minimum $\bar{H}(\mathbf{U};\mathbf{v})$ over $\mathbf{m}$ in this limit is $-\infty$. Therefore, this limit is out of $\mathcal{S}_\mathbf{v}$.

Similarly, consider a limit $\mathbf{v} \to \partial\mathcal{R}_\mathbf{v}$, such that the two components i, j satisfy

$$\lim_{\mathbf{v}\to\partial\mathcal{R}_\mathbf{v}} w_i(\mathbf{v}) = +\infty, \qquad \lim_{\mathbf{v}\to\partial\mathcal{R}_\mathbf{v}} w_j(\mathbf{v}) = w_j^* \in \mathbb{R}.$$

Rewrite

$$\tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v}) = w_i(\mathbf{v}) \left[\sum_{\substack{k=1 \\ k\neq i}}^{J} \frac{w_k(\mathbf{v})}{w_i(\mathbf{v})} \eta_k(m_k) + \eta_i(m_i) \right].$$

Then, one observes that $\tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v}) \to -\infty$ in the limit $m_i \to 0$. Therefore, the minimum $\bar{H}(\mathbf{U};\mathbf{v})$ over $\mathbf{m}$ is $-\infty$ in this limit. Therefore, this limit is out of $\mathcal{S}_{\mathbf{v}}$.

This implies that $\mathcal{S}_{\mathbf{v}}$ is a compact subset of $\mathcal{R}_{\mathbf{v}}$ and the function $\bar{H}(\mathbf{U};\cdot)$ is smooth on this set. Therefore, it has a maximum. □

5.2.3 Discussions on the uniqueness of the max-min and the choice of the weights

We have not been able to prove the uniqueness of the solution to (5.3). As an alternative, we show that the present framework is still favorable in this direction and highlight the difficulties for the proof. And we show the uniqueness under stronger (unsatisfied) assumptions.

First, only a few constraints on the quadrature weights as functions of $\mathbf{v}$ were imposed to obtain the existence of a max-min. But further constraints can be imposed to ease this problem. A first step in this direction is the following strict concave-convex property:

Proposition 5.2. *Denote*

$$\tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}}(\mathbf{U};\mathbf{v}) = \tilde{\mathcal{R}}_{\mathbf{m}}(\mathbf{U};\mathbf{v}) \bigcap \prod_i (\eta_i')^{-1}\left([\epsilon,+\infty)\right).$$

Suppose that the weights w_i are concave functions of $\mathbf{v} \in \mathcal{R}_{\mathbf{v}}$ such that, for all $\mathbf{u} \in (\mathbb{R}^*)^N$, at least one of them satisfies

$$\mathbf{u}^T (w_i)_{\mathbf{vv}}(\mathbf{v})\mathbf{u} < 0.$$

Then, for all realizable $\mathbf{U} \in \mathcal{R}_{app}$, the function $(\mathbf{m},\mathbf{v}) \mapsto \tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v})$ is strictly concave-convex over $\tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}} \times \mathcal{R}_{\mathbf{v}}$.

Proof. For all $\mathbf{m} \in \tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}}$, one has $\eta_i(m_i) > 0$ and $\tilde{H}(\mathbf{U};\mathbf{m},\mathbf{v})$ becomes a positive combination of the weights w_i and the strict concave-convex property follows. □

Remark 5.4.

- In the case of the Boltzmann entropy $\eta_i(m) = m\log m - m$, the constraint $\mathbf{m} \in \mathcal{R}^{\epsilon}_{\mathbf{m}}$ implies that $\log m_i > \epsilon$, and therefore $m_i > 1$.
- We provide below examples of sets of weights w_i satisfying the constraints of positivity, concavity and (5.4): For all $1 < i < J$

$$w_i(\mathbf{v}) = (v_{i+1} - v_i)(v_i - v_{i-1}) \quad \text{or} \quad w_i(\mathbf{v}) = \frac{(v_{i+1} - v_i)(v_i - v_{i-1})}{v_{i+1} - v_{i-1}}$$

and the extreme points

$$\begin{aligned} & w_1(\mathbf{v}) = \sqrt{v_2 - v_1}, \qquad w_J(\mathbf{v}) = \sqrt{v_J - v_{J-1}}, \\ \text{or} \qquad & w_1(\mathbf{v}) = \frac{v_2 - v_1}{2}, \qquad w_J(\mathbf{v}) = \frac{v_J - v_{J-1}}{2}. \end{aligned}$$

One verifies that all these weight functions w_i are non-strictly concave functions of $\mathbf{v}$, but any strictly positive combination of them is strictly concave. Furthermore, they are all strictly positive and at least one of them has a zero or infinity limit when $v_i = v_{i+1}$, when $v_1 \to -\infty$ or when $v_J \to +\infty$.

This strict concave-convex property provides a good framework for the problem (5.3), but the main difficulties arise from the non-linear moment constraints over $(\mathbf{m},\mathbf{v})$.

Proposition 5.3. *Assume (falsely) that the set $\tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}}(\mathbf{U};\mathbf{v}) = \tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}}$ is independent of $\mathbf{v}$, i.e. consider* (5.3) *without the moment constraints* (5.3d)*. Consider that $\tilde{H}$ is strictly concave-convex over $\tilde{\mathcal{R}}_{\mathbf{v}} \times \tilde{\mathcal{R}}^{\epsilon}_{\mathbf{m}}$.*

Then (5.3) *has a unique solution.*

Proof. Consider two solutions $(\mathbf{m}, \mathbf{v})$ and $(\mathbf{n}, \mathbf{u})$ of (5.3). On one side,

$$\tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{n}, \mathbf{u}) \leq \tilde{H}(\mathbf{m}, \mathbf{u}), \qquad \tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{n}, \mathbf{u}) \leq \tilde{H}(\mathbf{n}, \mathbf{v}),$$

since $\mathbf{m}$, resp. $\mathbf{n}$, minimizes $\tilde{H}$ at given $\mathbf{v}$, resp. $\mathbf{u}$. On the other side, concave-convexity was shown to provide strong duality e.g. in [84, 36, 91, 16] (under various formulations of this constraints over $\tilde{H}$), i.e.

$$\max_{\mathcal{R}_\mathbf{v}} \min_{\mathcal{R}_\mathbf{m}} \tilde{H} = \min_{\mathcal{R}_\mathbf{m}} \max_{\mathcal{R}_\mathbf{v}} \tilde{H}.$$

And one obtains

$$\tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{n}, \mathbf{u}) \geq \tilde{H}(\mathbf{m}, \mathbf{u}), \qquad \tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{n}, \mathbf{u}) \geq \tilde{H}(\mathbf{n}, \mathbf{v}),$$

since $\mathbf{v}$, resp. $\mathbf{u}$, maximizes $\tilde{H}$ at given $\mathbf{m}$, resp. $\mathbf{n}$. This provides especially the equalities $\tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{n}, \mathbf{v})$ and $\tilde{H}(\mathbf{m}, \mathbf{v}) = \tilde{H}(\mathbf{m}, \mathbf{u})$, which can hold together only if $\mathbf{m} = \mathbf{n}$ and $\mathbf{v} = \mathbf{u}$ due to strict concave-convexity of $\tilde{H}$. □

In reality, $\mathcal{R}_\mathbf{m}(\mathbf{U}; \mathbf{v})$ depends non-linearly on $\mathbf{v}$, and this manifold is a non-convex.

5.2.4 Properties of the resulting system of quadrature

Eventually, we exhibit the properties of the system of moments resulting from the closure (2.14) constructed with the solution of the max-min problem (5.3).

Property 5.1. *The function H defined by the max-min problem* (5.3) *is convex. If this solution is unique for all $\mathbf{U} \in E \subset \mathcal{R}_{app}$, then H is strictly convex over E.*

Proof. The function $\bar{H}(\mathbf{U}; \mathbf{v})$ was shown to be strictly convex in $\mathbf{U}$ in Property 4.1 for all $\mathbf{v} \in \mathcal{R}_\mathbf{v}$. Then, we apply Danskin's theorem ([33, 10]): Compute for all $\mathbf{v}$

$$\bar{H}\left(\mathbf{U}^1 + \delta(\mathbf{U}^2 - \mathbf{U}^1); \mathbf{v}\right) < \bar{H}\left(\mathbf{U}^1; \mathbf{v}\right) + \delta\left(\bar{H}(\mathbf{U}^2; \mathbf{v}) - \bar{H}(\mathbf{U}^1; \mathbf{v})\right).$$

Maximizing over $\mathbf{v}$ on both sides gives

$$\begin{aligned} H\left(\mathbf{U}^1 + \delta(\mathbf{U}^2 - \mathbf{U}^1)\right) &\leq \max_{\mathbf{v} \in \mathcal{R}_\mathbf{v}} \bar{H}\left(\mathbf{U}^1; \mathbf{v}\right) + \delta\left(\bar{H}(\mathbf{U}^2; \mathbf{v}) - \bar{H}(\mathbf{U}^1; \mathbf{v})\right) \\ &\leq H(\mathbf{U}^1) + \delta\left(H(\mathbf{U}^2) - H(\mathbf{U}^1)\right). \end{aligned}$$

When the maximum is unique, the inequality becomes strict. □

The Legendre-Fenchel transform of the entropy can be expressed using the solution $\mathbf{v}^*$ of the maximization problem (5.3b):

Property 5.2. *Suppose that the maximization problem* (5.3b) *has a unique solution $\mathbf{v}^* \in \mathcal{R}_\mathbf{v}$ for all $\mathbf{U} \in \mathcal{R}_{app}$.*

Then, the Legendre-Fenchel transform H^* of H is

$$H^*(\mathbf{V}) = \sum_{i=1}^{J} w_i(\mathbf{v}^*) \eta_i^* \left(\frac{\mathbf{V}^T \mathbf{b}(v_i^*)}{w_i(\mathbf{v}^*)} \right). \tag{5.5}$$

Proof. Following Danskin's theorem ([33, 10]), the derivative of the maximum H of $\bar{H}$ satisfies

$$H_\mathbf{U}(\mathbf{U}) = \bar{H}_\mathbf{U}(\mathbf{U}; \mathbf{v}^*).$$

Inverting this formula with respect to $\mathbf{U}$ provides

$$H^*_\mathbf{V}(\mathbf{V}) = (H_\mathbf{U})^{-1}(\mathbf{V}) = (\bar{H}_\mathbf{U})^{-1}(\mathbf{V}; \mathbf{v}^*) = \sum_{i=1}^{J} \mathbf{b}(v_i^*)(\eta_i^*)' \left(\frac{\mathbf{V}^T \mathbf{b}(v_i^*)}{w_i(\mathbf{v}^*)} \right),$$

as in (4.6). This is indeed the derivative of (5.5). □

Remark 5.5. If the maximum $\mathbf{v}^*$ is not unique, the directional derivative $H_{\mathbf{U}}(\mathbf{U})\mathbf{d}$ in the direction $\mathbf{d}$ takes the maximum value over all maxima $\mathbf{v}^*$, and one can still find a formula for H^*.

Property 5.3. *Suppose that the maximization problem* (5.3b) *has a unique solution* $\mathbf{v}^* \in \mathcal{R}_{\mathbf{v}}$ *for all* $\mathbf{U} \in \mathcal{R}_{app}$.

Then, the definition of the entropy H *provides the change of unknowns*

$$\begin{aligned} \mathbf{U}(\mathbf{V}) &= H^*_{\mathbf{V}}(\mathbf{V}) \\ &= \sum_{i=1}^{J} \mathbf{b}(v_i^*)(\eta_i^*)' \left(\frac{\mathbf{V}^T \mathbf{b}(v_i^*)}{w_i(\mathbf{v}^*)} \right), \end{aligned} \qquad \begin{aligned} \mathbf{V}(\mathbf{U}) &= H_{\mathbf{U}}(\mathbf{U}) \\ &= \sum_{i=1}^{J} w_i(\mathbf{v}^*) \eta_i' \left(m_i^*(\mathbf{U}) \right) \mathbf{g}(v_i^*), \end{aligned}$$

where $(\mathbf{m}^*(\mathbf{U}), \mathbf{v}^*)$ *is the solution of the max-min problem* (5.3). *This provides the flux and the entropy flux*

$$\mathbf{F}(\mathbf{U}) = \sum_{i=1}^{J} m_i^*(\mathbf{U}) v_i^* \mathbf{b}(v_i^*), \tag{5.6a}$$

$$\begin{aligned} G(\mathbf{U}) = \sum_{i=1}^{J} w_i(\mathbf{v}^*) \Bigg(& \sum_{j=1}^{J} \eta_i(m_i^*(\mathbf{U})) m_j^*(\mathbf{U}) v_j^* \left(\mathbf{k}(v_i^*)^T \mathbf{k}(v_j^*) \right) \\ & - \eta_i \left(\sum_{j=1}^{J} \frac{w_j(\mathbf{v}^*)}{w_i(\mathbf{v}^*)} \eta_j(m_j^*(\mathbf{U})) \left(\mathbf{k}(v_j^*)^T \mathbf{k}(v_i^*) \right) \right) \Bigg). \end{aligned} \tag{5.6b}$$

Proof. This follows from the evaluation of the functions and their derivatives at the value $(\mathbf{m}^*, \mathbf{v}^*)$, and using the computations of Section 4.2. □

We summarize the hyperbolicity results into the following proposition:

Proposition 5.4 (Symmetric hyperbolicity of AQMOM with the quadrature-based entropy)**.** *The functions* H, G *and* $\mathbf{F}$, *defined in* (5.3b), (5.6a) *and* (5.6b) *satisfy*

$$G_{\mathbf{U}} = H_{\mathbf{U}} \mathbf{F}_{\mathbf{U}}.$$

Suppose additionally that the maximum (5.3b) *is unique for all* $\mathbf{U} \in \mathcal{R}_{app}$. *Then, the system* (2.10) *with the closure* (5.6a) *is symmetrizable.*

Its wave speeds are bounded by the extremum velocities

$$Sp(\mathbf{F}_{\mathbf{U}}) \subset [v_1, v_J].$$

Proof. This is a straightforward adaptation of the proof of Proposition 4.2. □

After the identification of the candidate entropy, we construct the relaxation operator (2.13b) in the AQMOM case. The equilibrium $\mathbf{M}(\mathbf{U})$ in (2.13b) is the solution of the problem

$$\min_{\mathbf{V}-\mathbf{U} \in \boldsymbol{\mathcal{I}}^{\perp}} H(\mathbf{V}) = \min_{\mathbf{V}-\mathbf{U} \in \boldsymbol{\mathcal{I}}^{\perp}} \max_{\mathbf{v} \in \mathcal{R}_{\mathbf{v}}} \min_{\mathbf{m} \in \mathcal{R}_{\mathbf{m}}(\mathbf{V};\mathbf{v})} \tilde{H}(\mathbf{V}; \mathbf{m}, \mathbf{v}).$$

Since $\tilde{H}$ is convex in $\mathbf{m}$ and concave in $\mathbf{v}$, and the constraints on $\mathbf{V}$ and on $\mathbf{v}$ are independent, then the min-max theorem applies ([36, 91]) and provides strong duality. Therefore, the optimization problem (2.13b) reduces to

$$\begin{aligned} \min_{\mathbf{V}-\mathbf{U} \in \boldsymbol{\mathcal{I}}^{\perp}} \max_{\mathbf{v} \in \mathcal{R}_{\mathbf{v}}} \min_{\mathbf{m} \in \mathcal{R}_{\mathbf{m}}(\mathbf{V};\mathbf{v})} \tilde{H}(\mathbf{V}; \mathbf{m}, \mathbf{v}) &= \max_{\mathbf{v} \in \mathcal{R}_{\mathbf{v}}} \min_{\mathbf{V}-\mathbf{U} \in \boldsymbol{\mathcal{I}}^{\perp}} \min_{\mathbf{m} \in \mathcal{R}_{\mathbf{m}}(\mathbf{V};\mathbf{v})} \tilde{H}(\mathbf{V}; \mathbf{m}, \mathbf{v}) \\ &= \max_{\mathbf{v} \in \mathcal{R}_{\mathbf{v}}} \min_{\tilde{L}(\mathbf{v})\mathbf{m} = \tilde{\mathbf{U}}} \tilde{H}(\mathbf{V}; \mathbf{m}, \mathbf{v}), \end{aligned} \tag{5.7}$$

where $\tilde{L}(\mathbf{v}) \in \mathbb{R}^{3\times J}$ and $\tilde{\mathbf{U}}$ correspond to the parts of $L(\mathbf{v})$ and of $\mathbf{U}$ associated with $\boldsymbol{\mathcal{I}'}$. If $\mathbf{b}(v) = (1, v, \dots, v^N)^T$, these are

$$\tilde{L}_{i,j}(\mathbf{v}) = v_j^i, \qquad \tilde{U}_i = U_i.$$

This simply corresponds to the max-min problem (5.3) under the (reduced) constraint that the moments $\mathbf{M}(\mathbf{U})$ are those of $\mathbf{U}$ up to order two. The solution to (5.7) takes the form

$$\mathbf{M}(\mathbf{U}) = \sum_{i=1}^{J} \mathbf{b}(u_i^*)(\eta_i^*)' \left(\lambda_0 + \lambda_1 u_i^* + \lambda_2 (u_i^*)^2\right), \tag{5.8}$$

where the Lagrange multipliers $(\lambda_i)_{i=0,\dots,2} \in \mathbb{R}^3$ are the only coefficients such that the moments up to order 2 of $\mathbf{M}(\mathbf{U})$ are the same as those of $\mathbf{U}$. And the velocities $\mathbf{u}^*$ solve of the upper maximization problem (5.7). Since the constraints are different, then $\mathbf{u}^*$ is a priori different from $\mathbf{v}^*$ computed for the closure problem (5.3), they are therefore denoted differently.

We summarize the entropy dissipation in the following proposition:

Proposition 5.5 (H-theorem for the AQMOM with the quadrature-based entropy)**.**

- ***Entropy dissipation**:*
  - *For $\mathbf{C}^{naive}$, compute*

$$S^{naive}(\mathbf{U}) = \frac{1}{\tau} \sum_{i=1}^{J} w_i(\mathbf{v}^*) \eta_i'(m_i(\mathbf{U})) \left(\int_{\mathbb{R}} \left(\mathbf{k}(v_i^*)^T \mathbf{k}(v)\right) M(\mathbf{U})(v) dv - \sum_{j=1}^{J} m_i(\mathbf{U}) \left(\mathbf{k}(v_i^*)^T \mathbf{k}(v_j^*)\right) \right),$$

    where $\mathbf{k} = B\mathbf{b}$ and B is the Cholesky decomposition $BB^T = (LL^T)^{-1}$. This term is not signed for all $\mathbf{U} \in \mathcal{R}_{app}$.
  - *For $\mathbf{C}^{relax}$, Proposition 2.1 applies.*
- ***Equilibrium**:*
  - *For $\mathbf{C}^{naive}$, only $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ implies $S^{naive}(\mathbf{U}) = 0$. But having $S^{naive}(\mathbf{U}) = 0$ does not necessarily imply $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$.*
  - *For $\mathbf{C}^{relax}$, Proposition 2.1 with a strictly convex entropy applies.*

  Having $\mathbf{V}(\mathbf{U}) \in \boldsymbol{\mathcal{I}'}$ implies that $\mathbf{U}$ is of the form (5.8)

6 An entropy-dissipating numerical scheme

Eventually, we provide an entropy-dissipative and realizability-preserving discretization of the moment system (2.10) adapted to the different quadrature-based closures defined in the last sections.

Write the generic system to solve

$$\partial_t \mathbf{U} + \partial_x \mathbf{F}(\mathbf{U}) = \frac{\mathbf{M}(\mathbf{U}) - \mathbf{U}}{\tau}, \tag{6.1a}$$

where the unknown $\mathbf{U}$ and the flux $\mathbf{F}(\mathbf{U})$ satisfy

$$\mathbf{U} = \sum_{i=1}^{J} m_i(\mathbf{U}) \mathbf{b}(v_i(\mathbf{U})), \qquad \mathbf{F}(\mathbf{U}) = \sum_{i=1}^{J} m_i(\mathbf{U}) v_i(\mathbf{U}) \mathbf{b}(v_i(\mathbf{U})), \tag{6.1b}$$

where the number J of quadrature points, the size $N = Card(\mathbf{b})$ and the dependencies of v_i and m_i with respect to $\mathbf{U}$ depend on the considered closure. The equilibrium $\mathbf{M}(\mathbf{U})$ is assumed to dissipate the entropy H, which provides a symmetrization of (6.1), i.e. assume there exists H strictly convex and G such that

$$G_{\mathbf{U}} = H_{\mathbf{U}} F_{\mathbf{U}}, \qquad \mathbf{M}(\mathbf{U}) = \underset{\mathbf{V}-\mathbf{U}\in\boldsymbol{\mathcal{I}}^{\perp}}{\operatorname{argmin}} H(\mathbf{V}).$$

Consider a finite volume scheme for (6.1) of the form

$$\frac{\mathbf{U}_j^{n+1} - \mathbf{U}_j^n}{\Delta t} - \frac{\mathbf{F}_{j+1/2}^n - \mathbf{F}_{j-1/2}^n}{\Delta x} = \frac{\mathbf{M}(\mathbf{U}_j^{n+1}) - \mathbf{U}_j^{n+1}}{\tau}, \tag{6.2a}$$

with explicit fluxes and a fully implicit collision term.

Define kinetic fluxes (see e.g. [17, 18, 19, 80]) for the explicit numerical fluxes:

$$\begin{aligned}\mathbf{F}_{j+1/2}^n = \boldsymbol{\mathcal{F}}^+(\mathbf{U}_j^n) + \boldsymbol{\mathcal{F}}^-(\mathbf{U}_{j+1}^n), \qquad \boldsymbol{\mathcal{F}}^{\pm}(\mathbf{U}) &= \sum_{i=1}^{J} m_i(\mathbf{U}) v_i^{\pm}(\mathbf{U}) \mathbf{b}(v_i(\mathbf{U})) \\ &= \int_{\mathbb{R}^{\pm}} v\mathbf{b}(v) \left(\sum_{i=1}^{J} m_i(\mathbf{U}) \delta_{v_i(\mathbf{U})}(v) \right). \end{aligned} \tag{6.2b}$$

Since the Maxwellian $\mathbf{M}(\mathbf{U})$ depends only on the components of $\mathbf{U}$ corresponding to the collision invariants $\boldsymbol{\mathcal{I}}'$, i.e. $(U_k)_{k=0,\dots,2}$ if $\mathbf{b}(v) = (1, v, \dots, v^N)^T$, then (6.2a) can be rewritten as an explicit predictor-corrector scheme:

$$\mathbf{U}_j^{n+1,-} = \mathbf{U}_j^n + \Delta t \frac{\mathbf{F}_{j+1/2}^n - \mathbf{F}_{j-1/2}^n}{\Delta x}, \tag{6.2c}$$

$$\begin{aligned}\mathbf{U}_j^{n+1} &= \mathbf{U}_j^{n+1,-} + \frac{\Delta t}{\tau} \left(\mathbf{M}(\mathbf{U}_j^{n+1}) - \mathbf{U}_j^{n+1} \right) \\ &= \frac{\tau \mathbf{U}_j^{n+1,-} + \Delta t \mathbf{M}(\mathbf{U}_j^{n+1,-})}{\tau + \Delta t}, \end{aligned} \tag{6.2d}$$

by observing that $\mathbf{M}(\mathbf{U}_j^{n+1}) = \mathbf{M}(\mathbf{U}_j^{n+1,-})$. One verifies that reinjecting (6.2c) in (6.2d) provides indeed (6.2a).

Consider a CFL condition of the form

$$\Delta t \leq \frac{\Delta x}{\max_{i,j} |v_i(\mathbf{U}_j^n)|}. \tag{6.3}$$

Proposition 6.1. *Suppose that the CFL condition* (6.3) *holds.*

Then, the numerical scheme (6.2) *preserves the realizability from one time step to another, i.e. if* $\mathbf{U}_i^n \in \mathcal{R}_{app}$ *for all* i*, then* $\mathbf{U}_i^{n+1} \in \mathcal{R}_{app}$ *for all* i*.*

Suppose furthermore that for all $\mathbf{U} \in \mathcal{R}_{app}$*,*

$$S(\mathbf{U}) = H_{\mathbf{U}}(\mathbf{U}) \frac{(\mathbf{M}(\mathbf{U}) - \mathbf{U})}{\tau} \leq 0.$$

Then

$$H(\mathbf{U}_i^{n+1}) \leq H(\mathbf{U}_i^n) + \frac{\Delta t}{\Delta x}(G_{i+1/2} - G_{i-1/2}),$$

where

$$G_{i+1/2} = \mathcal{G}^+(\mathbf{U}_i^n) + \mathcal{G}^-(\mathbf{U}_{i+1}^n), \qquad \mathcal{G}_{\mathbf{U}}^{\pm} = H_{\mathbf{U}} \mathcal{F}_{\mathbf{U}}^{\pm}.$$

Proof. For the realizability, rewrite the updated value $\mathbf{U}_i^{n+1}$

$$\mathbf{U}_i^{n+1,-} = \sum_{j=1}^{J} \left(1 - \frac{v_j(\mathbf{U}_i^n)\Delta t}{\Delta x}\right) m_j(\mathbf{U}_i^n)\mathbf{b}(v_j(\mathbf{U}_i^n)) + \sum_{\pm}\sum_{j=1}^{J} \pm \frac{v_j(\mathbf{U}_{i\pm1}^n)^{\mp}\Delta t}{\Delta x} m_j(\mathbf{U}_{i\pm1}^n)\mathbf{b}(v_j(\mathbf{U}_{i\pm1}^n)).$$

This is a positive (under the CFL condition) combination of realizable vectors. Therefore, $\mathbf{U}_i^{n+1,-} \in \mathcal{R}_{app}$ is realizable. Similarly,

$$\mathbf{U}_i^{n+1} = \frac{\tau}{\Delta t + \tau}\mathbf{U}_i^{n+1,-} + \frac{\Delta t}{\Delta t + \tau}\mathbf{M}(\mathbf{U}_i^{n+1,-})$$

is a positive combination of two realizable vectors. Then, $\mathbf{U}_i^{n+1} \in \mathcal{R}_{app}$ is realizable.

For the entropy dissipation, the predictor step

$$H(\mathbf{U}_i^{n+1,-}) \leq H(\mathbf{U}_i^n) + \frac{\Delta t}{\Delta x}(G_{i+1/2}^n - G_{i-1/2}^n)$$

is a direct application of Theorem 3.1 in [18]. For the corrector step, the equilibrium $\mathbf{M}(\mathbf{U})$ minimizes the entropy H under constraints which are also satisfied by $\mathbf{U}$. Therefore, $H(\mathbf{M}(\mathbf{U})) \leq H(\mathbf{U})$. Then, using the convexity of the entropy H,

$$\begin{aligned} H(\mathbf{U}_i^{n+1}) &= H\left(\frac{\tau}{\Delta t + \tau}\mathbf{U}_i^{n+1,-} + \frac{\Delta t}{\Delta t + \tau}\mathbf{M}(\mathbf{U}_i^{n+1,-})\right) \\ &\leq H(\mathbf{U}_i^{n+1,-})\frac{\tau}{\Delta t + \tau} + H\left(\mathbf{M}(\mathbf{U}_i^{n+1,-})\right)\frac{\Delta t}{\Delta t + \tau} \leq H(\mathbf{U}_i^{n+1,-}). \end{aligned}$$

□

7 Conclusion

We have proposed several constructions of augmented systems of quadratures adapted to specific entropies. The most advanced one, developed in the last section, is based on a quadrature formula for the kinetic entropy, and it suffers from two drawbacks that need to be addressed before the numerical implementation of a solver and that are left for future work:

7.1 Discussion on the uniqueness of the entropy-based AQMOM fluxes

The lack of uniqueness of the solution to the max-min problem (5.3) implies the possibility of having multiple sets of quadrature points and weights representing the same realizable $\mathbf{U}$. If such a configuration exists, the uniqueness can be obtained locally in $\mathbf{U}$ by adding a simple selection criterion. Note that $\tilde{H}$ is C^∞ in all its variables in its domain of definition. Then, if such configurations giving non-uniqueness of $(\mathbf{m}, \mathbf{v})$ exist for $\mathbf{U}$ over a manifold in $\mathcal{R}_{app}$, a selection criterion would naturally follow the same regularity. A similar idea was followed in [82]. A more serious problem would be the existence of a realizable $\mathbf{U} \in \mathcal{R}$ where the parameters $(\mathbf{m}, \mathbf{v})$ solving (5.3) are not unique, and jump from one side to to the other. By construction, such a jump in the parameters $(\mathbf{m}, \mathbf{v})$ would still imply the continuity of the entropy $H(\mathbf{U}) = \tilde{H}(\mathbf{U}; \mathbf{m}, \mathbf{v})$. We have not been able yet to show whether or not such configurations exist.

7.2 Discussion on the algorithms used to compute the closures

The construction of these closures ultimately boils down to solving optimization problems:

- For ADVM: A strictly convex minimization problem (4.5) under the linear moment constraints $L\mathbf{m} = \mathbf{U}$ on the masses $\mathbf{m}$.
- For AQMOM: A strictly concave-convex (in part of the domain) max-min problem (5.3) under the (non-linear) polynomial moment constraints $\mathbf{m} \in \mathcal{R}_{\mathbf{m}}(\mathbf{U}; \mathbf{v})$ on the masses and positions $(\mathbf{m}, \mathbf{v})$.

The first case is closely related to the common entropy-minimizing closure [62]. Several algorithms adapted to this problem have been studied in the literature (see e.g. [51, 7, 6, 5, 12]). Preliminary tests on this problem (and the related dynamical problem (6.1)) with standard numerical optimization tools show comparable results with these references. However, this problem is known to be ill-conditioned in certain regimes, and the development of numerical tools adapted to both (potentially ill-conditioned) optimization problems is left for future work. This convex minimization problem also corresponds to the inner (or lower-level) problem (5.3a) of the max-min problem (5.3). Therefore, it must be solved for AQMOM as well.

The second case is much more difficult to solve. Preliminary tests were made with a naive approach consisting of maximizing the outer problem (5.3b) agnostic to the inner problem (5.3a), again using standard numerical optimization tools. However, such a naive approach requires a rather high computational cost, and is not well adapted to all regimes $\mathbf{U} \in \mathcal{R}$. The development of a numerical tool that is more stable and needs less computational cost is deferred as it requires further consideration.

A Numerical constraints for the symmetrization of AQMOM

Similar to the computations for the symmetrization of ADVM, compute

$$
\begin{aligned}
(\mathbf{U}_{\mathbf{p}}A)_{2i+1,j} &= A_{2i+1,j} + 2m_i l_i'(v_i) A_{2i+2,j} \\
&\quad + \sum_{k=M+1}^{J} \left[\mathbf{h}_{2i+1}(v_k) A_{2k+1,j} + m_k \mathbf{h}_{2i+1}'(v_k) A_{2k+2,j} \right], \\
(\mathbf{U}_{\mathbf{p}}A)_{2i+2,j} &= m_i A_{2i+2,j} + \sum_{k=M+1}^{J} \left[\mathbf{h}_{2i+2}(v_k) A_{2k+1,j} + m_k \mathbf{h}_{2i+2}'(v_k) A_{2k+2,j} \right], \\
(\mathbf{F}_{\mathbf{p}}A)_{2i+1,j} &= v_i A_{2i+1,j} + m_i \left(1 + 2v_i l_i'(v_i)\right) A_{2i+2,j} \\
&\quad + \sum_{k=M+1}^{J} \left[\mathfrak{h}_{2i+1}(v_k) A_{2k+1,j} + m_k \mathfrak{h}_{2i+1}'(v_k) A_{2k+2,j} \right], \\
(\mathbf{F}_{\mathbf{p}}A)_{2i+2,j} &= m_i v_i A_{2i+2,j} + \sum_{k=M+1}^{J} \left[\mathfrak{h}_{2i+2}(v_k) A_{2k+1,j} + m_k \mathfrak{h}_{2i+2}'(v_k) A_{2k+2,j} \right].
\end{aligned}
$$

Then, the conditions (3.3a-3.3b) are rewritten in the basis $\mathbf{h} = (\mathbf{h}_1, \mathbf{h}_2)$ of Hermite interpolation polyno-

mials as: For all $1 \leq i < j \leq M$:

$$
\begin{aligned}
0 = &(A_{2i+1,2j+1} - A_{2j+1,2i+1}) \\
&+ 2\left(m_i l_i'(v_i) A_{2i+2,2j+1} - m_j l_j'(v_j) A_{2j+2,2i+1}\right) \\
&\quad + \sum_{k=M+1}^{J} \big[(\mathbf{h}_{2i+1}(v_k) A_{2k+1,2j+1} - \mathbf{h}_{2j+1}(v_k) A_{2k+1,2i+1}) \\
&\qquad + m_k\left(\mathbf{h}_{2i+1}'(v_k) A_{2k+2,2j+1} - \mathbf{h}_{2j+1}'(v_k) A_{2k+2,2i+1}\right)\big], \qquad \text{(A.1a)} \\
0 = &(m_i A_{2i+2,2j+2} - m_j A_{2j+2,2i+2}) \\
&\quad + \sum_{k=M+1}^{J} \big[(\mathbf{h}_{2i+2}(v_k) A_{2k+1,2j+2} - \mathbf{h}_{2j+2}(v_k) A_{2k+1,2i+2}) \\
&\qquad + m_k\left(\mathbf{h}_{2i+2}'(v_k) A_{2k+2,2j+2} - \mathbf{h}_{2j+2}'(v_k) A_{2k+2,2i+2}\right)\big], \qquad \text{(A.1b)} \\
0 = &(v_i A_{2i+1,2j+1} - v_j A_{2j+1,2i+1}) \\
&+ \left(m_i\left(1 + 2v_i l_i'(v_i)\right) A_{2i+2,2j+1} - m_j\left(1 + 2v_j l_j'(v_j)\right) A_{2j+2,2i+1}\right) \\
&\quad + \sum_{k=M+1}^{J} \big[(\mathfrak{h}_{2i+1}(v_k) A_{2k+1,2j+1} - \mathfrak{h}_{2j+1}(v_k) A_{2k+1,2i+1}) \\
&\qquad + m_k\left(\mathfrak{h}_{2i+1}'(v_k) A_{2k+2,2j+1} - \mathfrak{h}_{2j+1}'(v_k) A_{2k+2,2i+1}\right)\big], \qquad \text{(A.1c)} \\
0 = &(m_i v_i A_{2i+2,2j+2} - m_j v_j A_{2j+2,2i+2}) \\
&\quad + \sum_{k=M+1}^{J} \big[(\mathfrak{h}_{2i+2}(v_k) A_{2k+1,2j+2} - \mathfrak{h}_{2j+2}(v_k) A_{2k+1,2i+2}) \\
&\qquad + m_k\left(\mathfrak{h}_{2i+2}'(v_k) A_{2k+2,2j+2} - \mathfrak{h}_{2j+2}'(v_k) A_{2k+2,2i+2}\right)\big], \qquad \text{(A.1d)}
\end{aligned}
$$

and for all $1 \leq i \leq j \leq M$:

$$
\begin{aligned}
0 = &(A_{2i+1,2j+2} - m_j A_{2j+2,2i+1}) + (2m_i l_i'(v_i) A_{2i+2,2j+2}) \\
&\quad + \sum_{k=M+1}^{J} \big[(\mathbf{h}_{2i+1}(v_k) A_{2k+1,2j+2} - \mathbf{h}_{2j+2}(v_k) A_{2k+1,2i+1}) \\
&\qquad + m_k\left(\mathbf{h}_{2i+1}'(v_k) A_{2k+2,2j+2} - \mathbf{h}_{2j+2}'(v_k) A_{2k+2,2i+1}\right)\big], \qquad \text{(A.1e)} \\
0 = &(v_i A_{2i+1,2j+2} - m_j v_j A_{2j+2,2i+1}) + (m_i(1 + 2v_i l_i'(v_i)) A_{2i+2,2j+2}) \\
&\quad + \sum_{k=M+1}^{J} \big[\left(\mathfrak{h}_{2i+1}(v_k) A_{2k+1,2j+2} - \mathfrak{h}_{2j+2}(v_k) A_{2k+1,2i+1}\right) \\
&\qquad + m_k\left(\mathfrak{h}_{2i+1}'(v_k) A_{2k+2,2j+2} - \mathfrak{h}_{2j+2}'(v_k) A_{2k+2,2i+1}\right)\big]. \qquad \text{(A.1f)}
\end{aligned}
$$

References

[1] M. Abdelmalik, *Adaptive algorithms for optimal multiscale model hierarchies of the boltzmann equation: Galerkin methods for kinetic theory*, Ph.D. thesis, Mechanical Engineering, 2017.

[2] M. Abdelmalik and H. van Brummelen, *Moment closure approximations of the Boltzmann equation based on φ-divergences*, J. Stat. Phys. **164** (2016), no. 1, 77–104.

[3] N. I. Akhiezer, *The classical moment problem*, Hafner Publ. Co., 1965.

[4] N. I. Akhiezer and M. Krein, *Theory of moments*, vol. 2, AMS Trans. Math. Monographs, 1962.

[5] G. W. Alldredge, M. Frank, and C. D. Hauck, *A regularized entropy-based moment method for kinetic equations*, SIAM J. Appl. Math. **79** (2019), no. 5, 1627–1653.

[6] G. W. Alldredge, C. D. Hauck, D. P. O'Leary, and A. L. Tits, *Adaptive change of basis in entropy-based moment closures for linear kinetic equations*, J. Comp. Phys. **74** (2014), no. 4, 489–508.

[7] G. W. Alldredge, C. D. Hauck, and A. L. Tits, *High-order entropy-based closures for linear transport in slab geometry II: A comutational study of the optimization problem*, SIAM J. Sci. Comput. **34** (2012), no. 4, 361–391.

[8] J. F. Bard, *Practical bilevel optimization: Algorithms and applications*, Springer, 1998.

[9] Y. Beck and M. Schmidt, *A gentle and incomplete introduction to bilevel optimization*, Tech. report, Trier Uni., 2021.

[10] D. P. Bertsekas, *Nonlinear programming*, Athena Scientific, 1999.

[11] G. A. Bird, *Molecular gas dynamics and the direct numerical simulation of gas flows*, Oxford, 1976.

[12] N. Böhmer and M. Torrilhon, *Entropic quadrature for moment approximations of the Boltzmann-BGK equation*, J. Comput. Phys. **401** (2022), 108992.

[13] J. Borwein and A. Lewis, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim. **29** (1991), 325–338.

[14] ———, *Partially finite convex programming, part I: Quasi relative interiors and duality theory*, Math. Program. **57** (1992), 15–48.

[15] ———, *Convex analysis and nonlinear optimization*, Springer, 2006.

[16] J. Borwein and D. Zhuang, *On fan's minmax theorem*, Math. Programming **34** (1986), 232–234.

[17] F. Bouchut, *On zero pressure gas dynamics*, Advances in kinetic theory and computing, Ser. Adv. Math. Appl. Sci., vol. 22, World Sci. Publ., 1994, pp. 171–190.

[18] ———, *Entropy satisfying flux vector splittings and kinetic bgk models*, Numer. Math. **94** (2003), 623–672.

[19] F. Bouchut, S. Jin, and X. Li, *Numerical approximations of pressureless and isothermal gas dynamics*, SIAM J. Numer. Anal. **41** (2003), 135–158.

[20] I. D. Boyd and T. E. Schwartzentruber, *Nonequilibrium gas dynamics and molecular simulation*, Cambridge, 2017.

[21] J. E. Broadwell, *Shock structure in a simple discrete velocity gas*, Phys. Fluid **7** (1964), 1243–1247.

[22] ———, *Study of rarefied shear flow by the discrete velocity method*, J. Fluid Mech. **19** (1964), 401–414.

[23] S. Brull, V. Pavan, and J. Schneider, *Derivation and analysis of relaxation operators in kinetic theory*, https://hal.science/hal-04561642 (2024), 1–49.

[24] Z. Cai, *Moment method as a numerical solver: Challenge from shock structure problems*, J. Comput. Phys. **444** (2021), 110593.

[25] Z. Cai, Y. Fan, and R. Li, *Globally hyperbolic regularization of grad's moment system in one dimensional space*, Commun. Math. Sci. **11** (2013), no. 2, 547–571.

[26] Z. Cai, B. Lin, and M. Lin, *A positive and moment-preserving fourier spectral method*, SIAM J. Numer. Anal. **62** (2024), 273–294.

[27] C. Cercignani, *The boltzmann equation and its applications*, Springer, 1987.

[28] C. Cercignani, R. Illner, and M. Pulvirenti, *The mathematical theory of dilute gases*, Springer, 1994.

[29] C. Chalons, R. O. Fox, F. Laurent, M. Massot, and A. Vié, *Multivariate gaussian extended quadrature method of moments for turbulent disperse multiphase flow*, SIAM Multiscale Mod. Sim. **15** (2017), 1553–1583.

[30] C. Chalons, D. Kah, and M. Massot, *Beyond pressureless gas dynamics: quadrature-based velocity moment models*, Commun. Math. Sci. **10** (2012), no. 4, 1241–1272.

[31] P. L. Chebyshev, *Sur l'interpolation par la méthode des moindres carrés*, Mém. Acad. Impér. Sci. St. Petersbourg **1** (1988), 1–24.

[32] R. Curto and L. A. Fialkow, *Recusiveness, positivity, and truncated moment problems*, Houston J. Math. **17** (1991), no. 4, 603–634.

[33] J. M. Danskin, *The theory of max-min and its application to weapons allocation problems*, Springer, 1967.

[34] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, Elsevier, 1984.

[35] B. C. Eu, *Kinetic theory of nonequilibrium ensembles, irreversible thermodynamics, and generalized hydrodynamics*, vol. 1: Nonrelativistic theories, Springer, 2016.

[36] K. Fan, *Minimax theorems*, Proc. Nat. Acad. Sci. **39** (1953), 42–47.

[37] Y. Fan, J. Koellermeier, J. Li, R. Li, and M. Torrilhon, *Model reduction of kinetic equations by operator projection*, J. Stat. Phys. **162** (2016), 457–486.

[38] F. Filbet and C. Mouhot, *Analysis of spectral methods for the homogeneous boltzmann equation*, Trans. AMS **363** (2011), 1947–1980.

[39] R. O. Fox and F. Laurent, *Hyperbolic quadrature method of moments for the one-dimensional kinetic equation*, SIAM J. Appl. Math. **82** (2022), 750–771.

[40] R. O. Fox, F. Laurent, and A. Passalacqua, *The generalized quadrature method of moments*, J. Aerosol Sci. **167** (2023), 106096.

[41] R. O. Fox, F. Laurent, and A. Vié, *Conditional hyperbolic quadrature method of moments for kinetic equations*, J. Comput. Phys. **365** (2018), 269–293.

[42] K. O. Friedrichs and P. D. Lax, *Systems of conservation equations with a convex extension*, Proc. Nat. Acad. Sci. **68** (1971), no. 8, 1686–1688.

[43] I. Gamba and S. Tharkabhushanam, *Spectral-lagrangian methods for collisional models of non-equilibrium statistical states*, J. Comput. Phys. **228** (2009), 2012–2036.

[44] A. L. Garcia, *Numerical methods for physics*, Prentice Hall, 1994.

[45] W. Gautschi, *Orthogonal polynomials*, Oxford press, 2004.

[46] E. Godlewski and P.-A. Raviart, *Numerical approximation of hyperbolic systems of conservation laws*, Springer, 1996.

[47] S. K. Godunov, *An interesting class of quasilinear systems*, Dokl. Akad. Nauk SSSR **139** (1961), no. 3, 521–523.

[48] H. Grad, *On the kinetic theory of rarefied gases*, Commun. Pure Appl. Math. **2** (1949), no. 4, 331–407.

[49] C. P. T. Groth, *Moment closure methods for kinetic equations of complex transport phenomena and their numerical solution*, Tech. report, École polytechnique, 2020.

[50] N. Hale and L. N. Trefethen, *Chebfun and numerical quadrature*, Sci. China Math. **55** (2012), 1749–1760.

[51] C. D. Hauck, *High-order entropy-based closures for linear transport in slab geometry*, Commun. Math. Sci **9** (2011), no. 1, 187–205.

[52] C. D. Hauck, C. D. Levermore, and A. L. Tits, *Convex duality and entropy-based moment closures: Characterizing degenerate densities*, SIAM J. Control Optim. **47** (2008), 1977–2015.

[53] L. H. Holway, *New statistical models for kinetic theory: Methods of construction*, Phys. Fluids **9** (1966), 1658–1673.

[54] Q. Huang, S. Li, and W.-A. Yong, *Stability analysis of quadrature-based moment methods for kinetic equations*, SIAM J. Appl. Math. **80** (2020), no. 1, 206–231.

[55] A. D. Ioffe and V. M. Tichomirov, *Theorie der extremalaufgaben*, Deutscher Verlag der Wssenschaften, 1979.

[56] E. T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. **106** (1957), no. 4, 620–630.

[57] M. Junk, *Maximum entropy for reduced moment problems*, Math. Mod. Meth. Appl. Sci. **10** (2000), no. 7, 1001–1028.

[58] S. Kawashima, *Systems of a hyperbolic-parabolic composite type, with applications to the equations of magnetohydrodynamics*, Ph.D. thesis, Kyoto University, 1984.

[59] S. Kawashima and W.-A. Yong, *Dissipative structure and entropy for hyperbolic systems of balance laws*, Arch. Rational Mech. Anal. **174** (2004), 345–364.

[60] J. Köllermeier, R. P. Schaerer, and M. Torrilhon, *A framework for hyperbolic approximation of kinetic equations using quadrature based projection methods*, Kin. Rel. Mod. **7** (2014), 531–549.

[61] J.-B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College press optimization series, 2010.

[62] C. D. Levermore, *Moment closure hierarchies for kinetic theories*, J. Stat. Phys. **83** (1996), no. 5-6, 1021–1065.

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

[64] ———, *Computational models for polydisperse particulate and multiphase systems*, Cambridge University Press, 2013.

[65] D. L. Marchisio, J. T. Pikturna, R. O. Fox, R. D. Vigil, and A. A. Barresi, *Quadrature method of moments for population-balance equations*, AIChE J. **49** (2003), 1266–1276.

[66] J. McDonald and M. Torrilhon, *Affordable robust moment closures for CFD based on the maximum-entropy hierarchy*, J. Comput. Phys. **251** (2013), 500–523.

[67] R. McGraw, *Description of aerosol dynamics by the quadrature method of moments*, Aerosol S. and Tech. **27** (1997), 255–265.

[68] L. R. Mead and N. Papanicolaou, *Maximum entropy in the problem of moments*, J. Math. Phys. **25** (1984), 2404–2417.

[69] L. Mieussens, *Discrete velocity models and numerical schemes for the boltzmann-bgk equation in plane and axisymmetric geometries*, J. Comput. Phys. **162** (200), 429–466.

[70] M. S. Mock, *Systems of conservation laws of mixed type*, J. Diff. Eq. **37** (1980), no. 1, 70–88.

[71] P. Monreal, *Moment realizability and kershaw closures in radiative transfer*, Ph.D. thesis, RWTH Aachen University, 2012.

[72] P. Monreal and M. Frank, *Higher order minimum entropy approximations in radiative transfer*, arXiv:0812.3063 (2008), 1–18.

[73] W. Morin and J. McDonald, *Applications of orthogonal-polynomial-based one-dimensional moment closures to plasma flows*, J. Comput. Phys. **523** (2025), 113659.

[74] ———, *Development of globally hyperbolic one-dimensional moment closures based on orthogonal polynomials*, - (2025), –.

[75] I. Müller and T. Ruggeri, *Rational extended thermodynamics*, Springer, 1998.

[76] E. P. Gross et M. Krook P. L. Bhatnagar, *A model for collision processes in gases. i. small amplitude processes in charged and neutral one-component systems*, Phys. Rev. **94** (1954), no. 3, 541–561.

[77] L. Pareschi and B. Perthame, *A spectral method for the homogeneous boltzmann equation*, Transport Theo. Stat. Phys. **25** (1996), 369–383.

[78] L. Pareschi and T. Rey, *Moment preserving fourier-galerkin spectral methods and application to the boltzmann equation*, SIAM J. Num. Anal. **60** (2022), 1947–1980.

[79] L. Pareschi and G. Russo, *An introduction to Monte Carlo methods for the Boltzmann equation*, ESAIM Proc. **10** (1999), 1–38.

[80] B. Perthame and C. Simeoni, *A kinetic scheme for the saint-venant system with a source term*, Calcolo **38** (2001), 201–231.

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

[82] ———, *A moment closure based on a projection on the boundary of the realizability domain: Analysis and extension*, Kin. rel. mod. **15** (2022), 793–822.

[83] ———, *Some recent advances on the method of moments in kinetic theory*, ESAIM: Proc **75** (2023), 86–95.

[84] R. T. Rockafellar, *Convex analysis*, Princeton Landmarks, 1970.

[85] F. Rogier and J. Schneider, *A direct method for solving the Boltzmann equation*, Transport Th. Stat. Phys. **23** (1994), no. 1-3, 313–338.

[86] K. Schmüdgen, *The moment problem*, Springer, 2017.

[87] F. Schneider, *Kershaw closures for linear transport equations in slab geometry i: model derivation*, J. Comput. Phys. **322** (2016), 905–919.

[88] J. Schneider, *Entropic approximation in kinetic theory*, ESAIM: M2AN **38** (2004), no. 3, 541–561.

[89] Y. Shizuta and S. Kawashima, *Systems of equations of hyperbolic-parabolic type with applications to the discrete boltzmann equation*, Hokkaido Math. J. **14** (1985), 249–275.

[90] J. A. Shohat and J. D. Tamarkin, *The problem of moments*, AMS, 1943.

[91] M. Sion, *On general minimax theorems*, Pacific J. Math. **8** (1958), 171–176.

[92] H. Struchtrup and M. Torrilhon, *Regularization of grad's 13 moment equations: Derivation and linear analysis*, Phys. Fluids **15** (2003), no. 9, 2668–2680.

[93] G. Szegö, *Orthogonal polynomials*, vol. XXIII, AMS colloquium publications, 1939.

[94] C. Truesdell, *Rational thermodynamics*, Springer, 1984.

[95] L. N. Vicente and P. H. Calamai, *Bilevel and multilevel programming: A bibliography review*, J. Global Optim. **5** (1994), 291–306.

[96] A. Vié, C. Chalons, R. O. Fox, F. Laurent, and M. Massot, *A multi-gaussian quadrature method of moments for simulating high stokes number turbulent two-phase flows*, Annual Research Briefs of the CTR-Stanford University, CTR-Stanford University, 2012, pp. 309–320.

[97] J. von Neumann, *Zur theorie der gesellschaftsspiele*, Math. Annalen **100** (1928), 295–320.

[98] J. C. Wheeler, *Modified moments and gaussian quadratures*, Rocky Mt. J. Math. **4** (1974), 287–296.

[99] E. Yilmaz, G. Oblapenko, and M. Torrilhon, *On nonlinear closures for moment equations based on orthogonal polynomials*, arXiv preprint:2407.05894 (2024), 1–27.

[100] W.-A. Yong, *Singular perturbations of first-order hyperbolic systems with stiff source terms*, J. Diff. Eq. **155** (1999), 89–132.

[101] ———, *Entropy and global existence for hyperbolic balance laws*, Arch. Rational Mech. Anal. **172** (2004), 247–266.

[102] C. Yuan and R. Fox, *Conditional quadrature method of moments for kinetic equations*, J. Comput. Phys. **230** (2011), no. 22, 8216–8246.

[103] C. Yuan, F. Laurent, and R. O. Fox, *An extended quadrature method of moments for population balance equations*, J. Aerosol Sci. **51** (2012), 1–23.

[104] R. Zhang, Y. Chen, Q. Huang, and W.-A. Yong, *Dissipativeness of the hyperbolic quadrature method of moments for kinetic equations*, arXiv:2406.13931v1 **155** (2024), 1–32.

[105] R. Zhang, Q. Huang, and W.-A. Yong, *Stability analysis of an extended quadrature method of moments for kinetic equations*, SIAM J. Appl. Math. **56** (2024), 4687–4711.