Journal of Scientific Computing (2023) 95:19
https://doi.org/10.1007/s10915-023-02117-5

Positivity-Preserving Lax–Wendroff Discontinuous Galerkin Schemes for Quadrature-Based Moment-Closure Approximations of Kinetic Models

Erica R. Johnson[1] · James A. Rossmanith[2] · Christine Vaughan[3]

Received: 5 November 2021 / Revised: 23 October 2022 / Accepted: 12 January 2023 /
Published online: 22 February 2023

Abstract
The quadrature-based method of moments (QMOM) offers a promising class of approximation techniques for reducing kinetic equations to fluid equations that are valid beyond thermodynamic equilibrium. In this work, we study a particular five-moment variant of QMOM known as HyQMOM and establish that this system is moment-invertible over a convex region in solution space. We then develop a high-order discontinuous Galerkin (DG) scheme for solving the resulting fluid system. The scheme is based on a predictor–corrector approach, where the prediction is a localized space-time DG scheme. The nonlinear algebraic system in this prediction is solved using a Picard iteration. The correction is a straightforward explicit update based on the time-integral of the evolution equation, where the space-time prediction replaces all instances of the exact solution. In the absence of limiters, the high-order scheme does not guarantee that solutions remain in the convex set over which HyQMOM is moment-realizable. To overcome this, we introduce novel limiters that rigorously guarantee that the computed solution does not leave the convex set of realizable solutions, thus guaranteeing the hyperbolicity of the system. We develop positivity-preserving limiters in both the prediction and correction steps and an oscillation limiter that damps unphysical oscillations near shocks. We also develop a novel extension of this scheme to include a BGK collision operator; the proposed method is shown to be asymptotic-preserving in the high-collision limit. The HyQMOM and the HyQMOM-BGK solvers are verified on several test cases,

This research was partially funded by NSF Grants DMS–1620128 and DMS–2012699.

✉ James A. Rossmanith
rossmani@iastate.edu

Erica R. Johnson
erica.johnson@swri.edu

Christine Vaughan
cvaughan880@gmail.com

[1] Space Science Department, Southwest Research Institute, 6220 Culebra Road, San Antonio, TX 78238, USA

[2] Department of Mathematics, Iowa State University, 411 Morrill Road, Ames, IA 50011, USA

[3] Physicians Health Plan, 1400 East Michigan Ave, Lansing, MI 48912, USA

demonstrating high-order accuracy on smooth problems and shock-capturing capability on problems with shocks. The asymptotic-preserving property of the HyQMOM-BGK solver is also numerically verified.

Keywords Discontinuous Galerkin · Hyperbolic conservation laws · Moment closure · Positivity-preserving limiters

Mathematics Subject Classification 65M60 · 82C80

Contents

1 Introduction

Kinetic Boltzmann equations model the non-equilibrium dynamics of a wide variety of fluids, including gases, multiphase flows, and plasma. These equations have the following general form:

$$f_{,t} + \underline{v} \cdot \underline{\nabla}_{\underline{x}} f + \underline{\mathcal{F}} \cdot \underline{\nabla}_{\underline{v}} f = \mathbb{C}(f), \tag{1.1}$$

where $f\left(t, \underline{x}, \underline{v}\right) : \mathbb{R}_{\geq 0} \times \mathbb{R}^D \times \mathbb{R}^V \mapsto \mathbb{R}_{\geq 0}$ is the distribution function that describes the state of the fluid, $t \in \mathbb{R}_{\geq 0}$ is time, $\underline{x} \in \mathbb{R}^D$ is the spatial coordinate, and $\underline{v} \in \mathbb{R}^V$ is the velocity coordinate. Additionally, $\underline{\mathcal{F}}\left(t, \underline{x}, \underline{v}\right) \in \mathbb{R}_{\geq 0} \times \mathbb{R}^D \times \mathbb{R}^V \mapsto \mathbb{R}^V$ is the forcing term that could include lift, drag, gravity, and other forces acting on the particles, and $\mathbb{C}(f) : \mathbb{R}_{\geq 0} \mapsto \mathbb{R}$ is the collision term that describes direct particle-particle interactions.

Kinetic models of the form (1.1) offer two desirable features: (1) the evolution equations have a relatively simple form (i.e., advection in phase space); and (2) the models are capable of accurately describing a large class of physical phenomena that are important in application problems. However, the main difficulty with kinetic models is that their solutions live in high-dimensional phase space, which means that high fidelity numerical computations are very expensive.

1.1 Fluid Models and the Moment-Closure Problem

One approach for reducing the computational complexity of kinetic models is to replace them with so-called *fluid models*, which means that instead of evolving the distribution function directly, one evolves a finite set of *moments* of the distribution function. For example the (ℓ_1, ℓ_2, ℓ_3) moment of the distribution function, f, is defined as follows:

$$\mathsf{M}_{(\ell_1,\ell_2,\ell_3)} := \int_{\mathbb{R}^3} v_1^{\ell_1} v_2^{\ell_2} v_3^{\ell_3} f\left(t, \underline{x}, \underline{v}\right) dv_1\, dv_2\, dv_3, \tag{1.2}$$

where $\underline{v} = (v_1, v_2, v_3)$. If moments of the kinetic equation (1.1) are computed, we arrive at three-dimensional evolution equations of the following form:

$$\mathsf{M}_{(\ell_1,\ell_2,\ell_3),t} + \mathsf{M}_{(\ell_1+1,\ell_2,\ell_3),x_1} + \mathsf{M}_{(\ell_1,\ell_2+1,\ell_3),x_2} + \mathsf{M}_{(\ell_1,\ell_2,\ell_3+1),x_3} = 0, \tag{1.3}$$

where for simplicity, we have set the forcing and collision operator to zero: $\underline{\mathcal{F}} \equiv \underline{0}$ and $\mathbb{C} \equiv 0$.

The key benefit of considering a finite set of evolution equations of the form (1.3) over the fully kinetic equation (1.1) is the reduction in the number of independent variables from $1 + D + V \leq 7$ to $1 + D \leq 4$. However, we observe from (1.3) the key challenge in fluid model approximations of the kinetic equation: to evolve the moment $\mathsf{M}_{(\ell_1,\ell_2,\ell_3)}$, we need to know higher-order moments: $\mathsf{M}_{(\ell_1+1,\ell_2,\ell_3)}$, $\mathsf{M}_{(\ell_1,\ell_2+1,\ell_3)}$, and $\mathsf{M}_{(\ell_1+1,\ell_2,\ell_3+1)}$. This issue is known simply as the *moment-closure problem*. And in particular, to obtain a closed fluid system, one needs to somehow approximate the highest moments in the system. Furthermore, different choices lead to systems of differential (or integro-differential) equations with vastly different mathematical properties.

1.2 Moment-Closure Methods

A standard approach for developing moment-closure approximation for (1.1) is to assume a specific ansatz for the distribution function:

$$f(t,\underline{x},\underline{v}) \approx \sum_{\ell=1}^{M} \psi_\ell\left(\underline{v},\beta_\ell\left(t,\underline{x}\right)\right) \quad \text{or} \quad f(t,\underline{x},\underline{v}) \approx \prod_{\ell=1}^{M} \psi_\ell\left(\underline{v},\beta_\ell\left(t,\underline{x}\right)\right). \tag{1.4}$$

The first systematic attempt at developing a moment-closure approach is due to the seminal work of Grad [25], in which he proposed a moment-closure that assumed the distribution was a Maxwell–Boltzmann distribution multiplied by a polynomial in $\underline{v}$. Since Grad's work, a vast body of literature has developed on various moment closures, including modern modifications of Grad's closure (e.g., [9, 10, 34–36]), maximum entropy [17, 40, 44] and its numerous variants (e.g., [1, 5]), and quadrature-based moment closures (e.g., [16, 20–22, 41, 45]). A full review of all methods is well beyond the scope of the current work, but can be found in the paper of Torrilhon [55] and the references therein.

1.3 Scope of this Work

In this work, we study a particular five-moment moment-closure known as the hyperbolic quadrature-based moment closure (HyQMOM), originally due to Fox, Laurent, and Vié [22]. Our focus here is only on the one-dimensional version of kinetic equation (1.1) (i.e., 1D1V). We begin in Sect. 2 with a brief review of the moment-closure problem and a few strategies for producing fluid approximations with desirable mathematical properties. In Sect. 3 we provide a brief review of the classical quadrature-based moment closure (QMOM), show its shortcomings, and then establish that the HyQMOM system is moment-invertible over a convex set in solution space. In Sect. 4 we introduce a novel high-order Lax–Wendroff discontinuous Galerkin scheme for solving the HyQMOM fluid system. The scheme is based on a predictor–corrector approach, where the prediction step is based on a localized space-time discontinuous Galerkin scheme. The nonlinear algebraic system that arises in this prediction step is solved using a Picard iteration. The correction is a straightforward explicit update based on the time-integral of the evolution equation, where the space-time prediction replaces all instances of the exact solution. In the absence of additional limiters, the proposed high-order scheme does not guarantee that the numerical solution remains in the convex set over which HyQMOM is moment-realizable. To overcome this challenge, we introduce in Sect. 5 novel limiters that rigorously guarantee that the computed solution does not leave the convex set over which moment-invertible and hyperbolicity of the fluid system is guaranteed. We develop positivity-preserving limiters in both the prediction and correction steps and an oscillation limiter that damps unphysical oscillations near shocks. In Sect. 6 we perform convergence tests to verify the order of accuracy of the scheme and test the scheme on Riemann data to demonstrate the shock-capturing and robustness of the method. In Sect. 7 we develop an asymptotic-preserving [29] extension of the proposed scheme that allows us to solve a five-moment fluid model with a Bhatnagar–Gross–Krook (BGK) [4] collision operator. Finally, in Sect. 8 we perform convergence tests to verify the order of accuracy of the scheme and verify the method on Riemann data with different Knudsen numbers. Conclusions are provided in Sect. 9.

2 Brief Review of Moment-Realizability

For the remainder of the present work, we focus exclusively on the one-dimensional version of (1.1). In particular, in this section, we focus on the transport portion of the kinetic equation, and thus restrict ourselves to the 1D1V collisionless Boltzmann equation (aka Vlasov equation):

$$f_{,t} + v f_{,x} = 0, \tag{2.1}$$

where $f(t, x, v) : \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_{\geq 0}$ is the probability distribution function that describes the state of the fluid.

The moments of f are defined as follows:

$$\mathsf{M}_\ell := \int_{\mathbb{R}} v^\ell \, f \, dv, \quad \text{for} \quad \ell \in \mathbb{Z}_{\geq 0}. \tag{2.2}$$

A simple calculation reveals that for each $\ell \in \mathbb{Z}_{\geq 0}$, the moments satisfy the following equation:

$$\mathsf{M}_{\ell,t} + \mathsf{M}_{\ell+1,x} = 0. \tag{2.3}$$

The key difficulty is that the evolution of the ℓ^{th} moment depends on the $(\ell+1)^{\text{st}}$ moment, meaning that the moment expansion does not produce a closed system. Therefore, the key challenge for developing fluid approximations of kinetic models is the following question: How does one close the moment hierarchy?

Definition 2.1 *(Univariate moment-closure problem)* Let $S \in \mathbb{Z}_{\geq 0}$. Given only the first $S+1$ moments of a univariate distribution function $f(v)$:

$$\mathsf{M}_\ell = \int_{-\infty}^{\infty} v^\ell f(v) \, dv, \quad \text{for} \quad \ell = 0, 1, \ldots, S, \tag{2.4}$$

find an approximation of the next moment, M_{S+1}, in terms of the given moments.

The basic strategy in most moment-closure approaches is as follows: (1) start with a finite set of moments (e.g., $\ell = 0, 1, \ldots, S$); (2) assume a form of the distribution function with several free parameters (typically, the number of free parameters is the same as the number of moments that will be tracked); (3) determine the free parameters in the assumed distribution function so that its moments match all of the known moments (this part is called *moment-inversion*); and (4) compute the next moment of the assumed distribution function, which is then used to provide the flux in the evolution equation for M_S:

$$\mathsf{M}_{S,t} + \overline{\mathsf{M}}_{S+1,x} = 0, \tag{2.5}$$

where $\overline{\mathsf{M}}_{S+1} = \overline{\mathsf{M}}_{S+1}(\mathsf{M}_0, \mathsf{M}_1, \ldots, \mathsf{M}_S)$.

2.1 Existence: Moment-Realizability

Before considering specific strategies for approximating the missing moment, M_{S+1}, it is worthwhile to discuss the general existence problem first. We begin by defining some important quantities relevant throughout this work, namely the mass density, macroscopic velocity,

pressure, heat flux, and modified kurtosis:

$$\rho := \int_{-\infty}^{\infty} f\,dv, \qquad u := \frac{1}{\rho}\int_{-\infty}^{\infty} vf\,dv, \qquad p := \int_{-\infty}^{\infty} (v-u)^2 f\,dv,$$
$$h := \int_{-\infty}^{\infty} (v-u)^3 f\,dv, \qquad \text{and} \qquad k := \int_{-\infty}^{\infty} (v-u)^4 f\,dv - \left(\frac{p^3+\rho q^2}{\rho\, p}\right), \tag{2.6}$$

where we assume that $\rho,\ p > 0$. These *primitive variables* are directly linked to the moments M_ℓ:

$$\rho = \mathsf{M}_0, \qquad u = \frac{\mathsf{M}_1}{\mathsf{M}_0}, \qquad p = \mathsf{M}_2 - \frac{\mathsf{M}_1^2}{\mathsf{M}_0}, \qquad h = \mathsf{M}_3 - \frac{3\mathsf{M}_1\mathsf{M}_2}{\mathsf{M}_0} + \frac{2\mathsf{M}_1^3}{\mathsf{M}_0^2},$$
$$\text{and} \qquad k = \frac{\mathsf{M}_2^3 - 2\mathsf{M}_1\mathsf{M}_2\mathsf{M}_3 + \mathsf{M}_0\mathsf{M}_3^2 + \mathsf{M}_1^2\mathsf{M}_4 - \mathsf{M}_0\mathsf{M}_2\mathsf{M}_4}{\mathsf{M}_1^2 - \mathsf{M}_0\mathsf{M}_2}. \tag{2.7}$$

Using these we define the *normalized velocity variable*, s, and the *normalized moments*: $\widetilde{\mathsf{M}}_j$:

$$s := \frac{v-u}{\sqrt{T}} \qquad \text{and} \qquad \widetilde{\mathsf{M}}_j := \frac{\sqrt{T}}{\rho}\int_{-\infty}^{\infty} s^j\, f\,(v(s))\,ds = \int_{-\infty}^{\infty} s^j\,\widetilde{f}\,(s)\,ds, \tag{2.8}$$

where $T = p/\rho$ is the temperature. The moments and the normalized moments are related as follows:

$$\mathsf{M}_\ell = \rho \sum_{j=0}^{\ell} \binom{\ell}{j} T^{\frac{j}{2}}\, u^{\ell-j}\, \widetilde{\mathsf{M}}_j \qquad \text{and} \qquad \widetilde{\mathsf{M}}_j = \rho^{-1}\, T^{-\frac{j}{2}} \sum_{\ell=0}^{j} \binom{j}{\ell} (-u)^{j-\ell}\, \mathsf{M}_\ell. \tag{2.9}$$

By construction, the normalized moments have the following property:

$$\widetilde{\mathsf{M}}_0 = 1, \qquad \widetilde{\mathsf{M}}_1 = 0, \qquad \text{and} \qquad \widetilde{\mathsf{M}}_2 = 1. \tag{2.10}$$

Definition 2.2 *(Realizable moments)* The following rescaled moments:

$$\widetilde{\mathsf{M}}_0 = 1, \quad \widetilde{\mathsf{M}}_1 = 0, \quad \widetilde{\mathsf{M}}_2 = 1, \quad \widetilde{\mathsf{M}}_3, \quad \ldots, \quad \widetilde{\mathsf{M}}_S,$$

where $S \in \mathbb{Z}_{\geq 3}$ and $\left|\widetilde{\mathsf{M}}_j\right| < \infty\ \forall\, j \in \mathbb{Z}_{\geq 0}$, are called **realizable** if there exists a probability density function, $\widetilde{f}(s) : \mathbb{R} \mapsto \mathbb{R}_{\geq 0}$, such that

$$\widetilde{\mathsf{M}}_j = \int_{-\infty}^{\infty} s^j\, \widetilde{f}(s)\, ds \qquad \text{for} \quad j = 0, 1, \ldots, S.$$

This leads to an obvious question: for a given $S \in \mathbb{Z}_{\geq 3}$, under what conditions is the set of moments, $\{\widetilde{\mathsf{M}}_0 = 1,\ \widetilde{\mathsf{M}}_1 = 0,\ \widetilde{\mathsf{M}}_2 = 1,\ \widetilde{\mathsf{M}}_3, \ldots,\ \widetilde{\mathsf{M}}_S\}$, realizable? This question is the celebrated **truncated Hamburger moment problem** (e.g., see Chapter 9 of [51]) for which we can state the following result. Note that we state only the case where S is even, although a similar result also exists when S is odd [51].

Theorem 2.1 (Truncated Hamburger moment problem (adapted from Theorem 9.27 of [51])) *Let $S \in \mathbb{Z}_{>3}$ be an even integer. The set of moments:*

$$\{\widetilde{M}_0 = 1,\ \widetilde{M}_1 = 0,\ \widetilde{M}_2 = 1,\ \widetilde{M}_3, \ldots,\ \widetilde{M}_S\}$$

is realizable if all of the Hankel determinants for $m = 0, 1, \ldots, S/2$ are positive:

$$D_m := \begin{vmatrix} 1 & 0 & 1 & \widetilde{M}_3 & \cdots & \widetilde{M}_m \\ 0 & 1 & \widetilde{M}_3 & \widetilde{M}_4 & \cdots & \widetilde{M}_{m+1} \\ 1 & \widetilde{M}_3 & \widetilde{M}_4 & \widetilde{M}_5 & \cdots & \widetilde{M}_{m+2} \\ \widetilde{M}_3 & \widetilde{M}_4 & \widetilde{M}_5 & \widetilde{M}_6 & \cdots & \widetilde{M}_{m+3} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ \widetilde{M}_m & \widetilde{M}_{m+1} & \widetilde{M}_{m+2} & \widetilde{M}_{m+3} & \cdots & \widetilde{M}_{2m} \end{vmatrix} > 0.$$

2.2 Example: $S = 4$ Case

As an example, which will become relevant later in this work, consider the case $S = 4$, where the three relevant Hankel determinants are

$$D_0 = 1, \quad D_1 = 1, \quad D_2 = \begin{vmatrix} 1 & 0 & 1 \\ 0 & 1 & \widetilde{\mathrm{M}}_3 \\ 1 & \widetilde{\mathrm{M}}_3 & \widetilde{\mathrm{M}}_4 \end{vmatrix} = \widetilde{\mathrm{M}}_4 - \widetilde{\mathrm{M}}_3^2 - 1 > 0. \tag{2.11}$$

Thus, the realizability condition in the univariate $S = 4$ case is

$$\widetilde{\mathrm{M}}_4 > \widetilde{\mathrm{M}}_3^2 + 1, \tag{2.12}$$

which is depicted in Fig. 1. Also shown in this figure is the location of the distribution in thermodynamic equilibrium (i.e., the Maxwellian distribution):

$$f(v) = \frac{\rho}{\sqrt{2\pi T}} e^{-\frac{(v-u)^2}{2T}} \implies \widetilde{f}(s) = \frac{1}{\sqrt{2\pi}} e^{-\frac{s^2}{2}} \implies \begin{cases} \widetilde{\mathrm{M}}_3 = 0 \\ \widetilde{\mathrm{M}}_4 = 3 \end{cases}. \tag{2.13}$$

Finally, we note that the realizability condition (2.12) in terms of the primitive variables can be written as follows:

$$\frac{\rho\, k}{p^2} > 0, \tag{2.14}$$

which is satisfied if $\rho > 0$, $p > 0$, and $k > 0$.

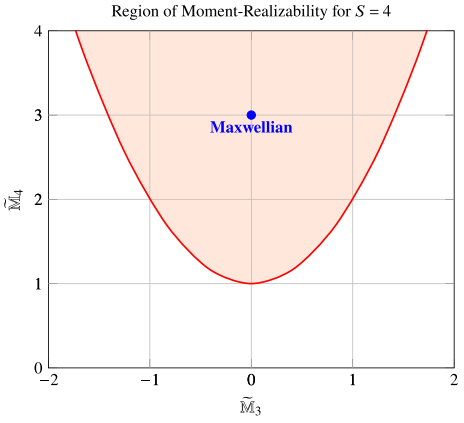

Fig. 1 Region of the moment-realizability when $S = 4$. In other words, given the five rescaled moments: $\widetilde{\mathrm{M}}_0 = 1$, $\widetilde{\mathrm{M}}_1 = 0$, $\widetilde{\mathrm{M}}_2 = 1$, $\widetilde{\mathrm{M}}_3$, and $\widetilde{\mathrm{M}}_4$, there exists a positive distribution function matching the given moments provided that these moments lie within the shaded pink region of the above graph. Thermodynamic equilibrium (i.e., the Maxwellian distribution) occurs at $\widetilde{\mathrm{M}}_3 = 0$ and $\widetilde{\mathrm{M}}_4 = 3$, which is in the interior of the shaded pink region (color figure online)

2.3 Techniques for Moment-Closure

There are many different moment closure methods for approximating the final moment, $\overline{\mathsf{M}}_{S+1}$, and each method has its own merits and challenges. A full review of all methods is well beyond the scope of the current work, but can be found in the paper of Torrilhon [55] and the references therein. In this work, we settle for a brief summary of three broad classes of the most commonly used closures.

Grad closure: This approach was originally developed by Grad [25] in 1949, but variants with improved hyperbolicity properties have been introduced more recently [9, 10, 34–36]. The basic idea is that the distribution function is approximated as a Maxwellian multiplied by a polynomial in v:

$$\widetilde{f}(s) \sim \frac{e^{-\frac{s^2}{2}}}{\sqrt{2\pi T}} \left[\rho + \sum_{\ell=3}^{S} T^{-\frac{\ell}{2}} \beta_\ell \psi_\ell(s) \right], \tag{2.15}$$

where $\psi_\ell(s)$ is the Hermite polynomial of degree ℓ and $\underline{\beta}$ are coefficients chosen so that $\widetilde{f}(s)$ matches the first $S+1$ input moments. A related approach widely used in applications is the R13 model, which regularizes the 13-moment Grad closure through additional terms from the Chapman–Enskog expansion; an excellent review of the R13 model can be found in Torrilhon [56].

Maximum entropy closure: The maximum entropy closure [17, 40, 44] and its numerous variants (e.g., see [1, 5]) formulate the moment-inversion problem as an optimization problem to maximize the entropy under some assumed form of the distribution function. In the original formulation, the distribution function is approximated as an exponential of a polynomial:

$$\widetilde{f}(s) \sim e^{\underline{\beta} \cdot \underline{\Phi}(s)} \qquad \text{where} \qquad \underline{\Phi}(s) = [1, s, s^2, \dots]^T, \tag{2.16}$$

and the coefficients, $\underline{\beta}$, are chosen so that $\widetilde{f}(s)$ matches the first $S+1$ input moments.

Quadrature-based moment closures: In the quadrature-based method of moments (QMOM) (e.g., [16, 20–22, 41, 45]), the distribution function is represented as a sum of Dirac delta functions. This closure will be discussed in detail in Sect. 3.

3 HyQMOM: Hyperbolic Quadrature-Based Moment Closure

In this section, we review the quadrature-based moment-closure (QMOM) approach and describe in detail the five-moment hyperbolic regularization of QMOM (HyQMOM), which is the main focus of the current work.

3.1 Classical QMOM Approach

The classical quadrature-based moment-closure (QMOM) approach is widely used in modeling multiphase flows; key developments in this methodology have been developed over the course of the past several years, e.g., see [12, 16, 19–21, 41, 57].

The key idea is to assume that the distribution is a sum of Dirac delta functions whose locations (abscissas) and strengths (weights) are free parameters:

$$f(t,x,v) \approx f^\star(t,x,v) := \sum_{j=1}^{N} \omega_j \delta\left(v-\mu_j\right), \tag{3.1}$$

where $\delta(v)$ is the Dirac delta and the quadrature weights, ω_j, and abscissas, μ_j, are all functions of t and x. This approach is reminiscent of other discrete velocity models such as the Broadwell model [6, 7, 47]; however, a key difference is that the discrete velocities, μ_ℓ, change with the solution.

The moment inversion problem requires us to find (ω_j, μ_j) for $j = 1, \ldots, N$ by matching the first $2N$ moments of $f(t,x,v)$:

$$\mathsf{M}_\ell = \sum_{j=1}^{N} \omega_j \mu_j^\ell \quad \text{for} \quad \ell = 0, 1, \ldots, 2N-1. \tag{3.2}$$

The closure then comes from taking the next moment as follows:

$$\mathsf{M}_{2N}^\star = \sum_{j=1}^{N} \omega_j \mu_j^{2N}. \tag{3.3}$$

3.1.1 Example: $N = 2$ Case

As a simple example, let us consider the $N = 2$ case. The first four moments of $f^\star$ in (3.1) with $N = 2$ are

$$\mathsf{M}_\ell = \omega_1 \mu_1^\ell + \omega_2 \mu_2^\ell \quad \text{for} \quad \ell = 0, 1, 2, 3. \tag{3.4}$$

The moment inversion problem is then this: given $(\mathsf{M}_0, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3)$, find the parameters $(\mu_1, \mu_2, \omega_1, \omega_2)$ such that (3.4) is satisfied.

This inversion problem is equivalent to finding the quadrature points and weights for the following weighted Gaussian quadrature rule:

$$\int_{-\infty}^{\infty} g(v)\, f(v)\, dv \;\approx\; \omega_1\, g\left(\mu_1\right) + \omega_2\, g\left(\mu_2\right), \tag{3.5}$$

where $f(v)$ is a probability density function with moments $(\mathsf{M}_0, \mathsf{M}_1, \mathsf{M}_2, \mathsf{M}_3)$. If we attempt to make this quadrature rule exact with $g(v) = 1, v, v^2$, and v^3, we again arrive at (3.4).

To find the correct Gaussian quadrature rule, we invoke results from classical numerical analysis and look for polynomials of degree up to two that are orthogonal in the weighted $L^2(-\infty, \infty)$ inner product: $\langle \cdot, \cdot \rangle_f$. Such polynomials are easily obtained via Gram-Schmidt, and indeed the relevant one here is the quadratic polynomial:

$$\psi_2(v) = p^2 - p\rho\,(v-u)^2 + \rho h\,(v-u), \tag{3.6}$$

where ρ, u, p, and h are defined by (2.6)–(2.7). The abscissas are the two distinct real roots of $\psi_2(v)$ and the weights can easily be obtained by enforcing (3.4):

$$\mu_1,\ \mu_2 = u + \frac{h}{2p} \mp \sqrt{\frac{p}{\rho} + \left(\frac{h}{2p}\right)^2}, \quad \omega_1,\ \omega_2 = \frac{\rho}{2}\left[1 \pm \frac{\left(\frac{h}{2p}\right)}{\sqrt{\frac{p}{\rho} + \left(\frac{h}{2p}\right)^2}}\right]. \tag{3.7}$$

3.1.2 Weak Hyperbolicity and Linear Degeneracy of QMOM for all $N \geq 1$

While the above-described process can be used for any $N \geq 1$, it turns out that the resulting fluid equations are always only *weakly hyperbolic*. Furthermore, all of the waves in the Riemann problem solution are *linearly degenerate*. We state these facts in the form of Theorem A.1 in "Appendix A"; for completeness, we also provide a full proof of this theorem in "Appendix A". The theorem results are well-known in the QMOM literature (e.g., see Chalons et al. [12]), although previously, no theorem had been presented to rigorously show both the weak hyperbolicity and linear degeneracy of all the waves for all $N \geq 1$.

3.2 Pressure Regularized QMOM

To overcome the weak hyperbolicity present in classical QMOM, Chalons, Fox, and Massot [11] proposed to replace the delta function ansatz (3.1) with a multi-Gaussian ansatz of the form:

$$f(t,x,v) \approx f^{\star}(t,x,v) := \frac{1}{\sqrt{2\pi\sigma}} \sum_{j=1}^{N} \omega_j \exp\left[-\frac{\left(v-\mu_j\right)^2}{2\sigma} \right], \tag{3.8}$$

where the free parameters are now the quadrature weights, ω_k, the abscissas, μ_k, and the additional parameter σ. A similar approach using B-splines was also considered by Cheng and Rossmanith [13]. The additional parameter σ allows this closure to match an additional moment (i.e., a total of $2N+1$ moments can now be matched), but more importantly, it provides a pressure regularization that restores strong hyperbolicity. Unfortunately, this closure exhibits a singularity in the limit of thermodynamic equilibrium, since in that limit, all the quadrature points collapse to the macroscopic velocity:

$$\sigma \to T, \qquad \sum_{j=1}^{N} \omega_j = \rho, \qquad \text{and} \qquad \mu_j \to u \quad \forall\, j \in [1,N]\,. \tag{3.9}$$

This type of singularity is also evident in other closures, most notably the *maximum entropy* closure [32].

3.3 HyQMOM: Density Regularized QMOM

As an alternative to the above-described pressure regularized QMOM, Fox et al. [22, 45] developed a density regularized version, which they refer to as HyQMOM (hyperbolic quadrature-based method of moments). This approach was also studied by Johnson [31] and Wiersma [59]. The five-moment HyQMOM system is the subject of the current work, and we briefly review it in this section.

The idea of HyQMOM is as follows: approximate the distribution as a sum of delta functions (as in classical QMOM), but place one (or more) of these delta functions at known locations. This converts the quadrature rule from classical Gaussian quadrature to something akin to Gauss–Radau quadrature. The version of this idea relevant to the current work is the case of three delta functions:

$$f \approx f^{\star} = \omega_1 \delta(v-\mu_1) + \omega_2 \delta(v-u) + \omega_3 \delta(v-\mu_3), \tag{3.10}$$

where two delta distributions are at unknown locations μ_1, μ_3 and the last delta distribution

is fixed at the velocity, u. Each of the distributions is weighted by $\omega_1, \omega_2, \omega_3$. This results in the following moment-inversion problem:

$$\begin{aligned} \omega_1 + \omega_3 &= \widetilde{\rho}, \\ \omega_1\mu_1 + \omega_3\mu_3 &= \widetilde{\rho}u, \\ \omega_1\mu_1^2 + \omega_3\mu_3^2 &= \widetilde{\rho}u^2 + p, \\ \omega_1\mu_1^3 + \omega_3\mu_3^3 &= \widetilde{\rho}u^3 + 3pu + h, \end{aligned} \tag{3.11}$$

and

$$\widetilde{\rho} := \rho - \omega_2 = \omega_1\left(\frac{\mu_1}{u}\right)^4 + \omega_3\left(\frac{\mu_3}{u}\right)^4 - \frac{6\rho p^2u^2 + 4\rho phu + p^3 + \rho h^2 + \rho pk}{\rho pu^4}. \tag{3.12}$$

System (3.11) can be solved in the same way that the $N = 2$ classical QMOM system was solved, namely by constructing a quadratic polynomial:

$$\psi_2(v) = p^2 - p\widetilde{\rho}\,(v-u)^2 + h\widetilde{\rho}\,(v-u)\,, \tag{3.13}$$

which is just (3.6) with $\rho \to \widetilde{\rho}$. The roots of (3.13) provide μ_1 and μ_3, and the corresponding weights, ω_1 and ω_3, can easily be computed from (3.11) once μ_1 and μ_3 are known:

$$\mu_1,\ \mu_3 = u + \frac{h}{2p} \mp \sqrt{\frac{p}{\widetilde{\rho}} + \left(\frac{h}{2p}\right)^2}, \quad \omega_1,\ \omega_3 = \frac{\widetilde{\rho}}{2}\left[1 \pm \frac{\left(\frac{h}{2p}\right)}{\sqrt{\frac{p}{\widetilde{\rho}} + \left(\frac{h}{2p}\right)^2}}\right]. \tag{3.14}$$

Finally, we can obtain ω_2 and fully solve the moment inversion problem by inserting the above expressions for ω_1, ω_3, μ_1, and μ_3 into (3.12):

$$\widetilde{\rho} = \rho - \omega_2 = \frac{p^3}{p\left(k + \frac{p^2}{\rho} + \frac{h^2}{p}\right) - h^2}. \tag{3.15}$$

Putting all of these results together yields:

$$\mathsf{M}_5^\star = \rho u^5 + 10pu^3 + 10hu^2 + 5ru + \frac{2hr}{p} - \frac{h^3}{p^2}, \quad \text{where} \quad r := \frac{p^2}{\rho} + \frac{h^2}{p} + k. \tag{3.16}$$

From this, we can now assemble the full fluid approximation implied by the 5-moment HyQMOM approximation (3.10).

Definition 3.1 *(5-moment HyQMOM fluid approximation)* The 5-moment HyQMOM approximation can be written either in conservative or primitive form:

$$\underline{q}_{,t} + \underline{f}\left(\underline{q}\right)_{,x} = \underline{0} \qquad \text{or} \qquad \underline{\alpha}_{,t} + \underline{\underline{B}}\left(\underline{\alpha}\right)\underline{\alpha}_{,x} = \underline{0}\,, \tag{3.17}$$

respectively, where

$$\underline{q} = \begin{bmatrix} \rho \\ \rho u \\ \rho u^2 + p \\ \rho u^3 + 3pu + h \\ \rho u^4 + 6pu^2 + 4hu + r \end{bmatrix}, \quad \underline{f}\left(\underline{q}\right) = \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u^3 + 3pu + h \\ \rho u^4 + 6pu^2 + 4hu + r \\ \mathsf{M}_5^\star \end{bmatrix}, \tag{3.18}$$

where r and $\mathsf{M}_5^\star$ are defined by(3.16), and

$$\underline{\alpha} = \begin{bmatrix} \rho \\ u \\ p \\ h \\ k \end{bmatrix}, \quad \underline{\underline{B}}\left(\underline{\alpha}\right) = \begin{bmatrix} u & \rho & 0 & 0 & 0 \\ 0 & u & \frac{1}{\rho} & 0 & 0 \\ 0 & 3p & u & 1 & 0 \\ -\frac{p^2}{\rho^2} & 4h & -\frac{h^2}{p^2} - \frac{p}{\rho} & u + \frac{2h}{p} & 1 \\ 0 & 5k & -\frac{2kh}{p^2} & \frac{2k}{p} & u \end{bmatrix}. \tag{3.19}$$

Furthermore, we note that that conservative flux Jacobian has the following form:

$$\underline{\underline{A}}\left(\underline{q}\right) := \underline{f}\left(\underline{q}\right)_{,\underline{q}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \frac{\partial \mathsf{M}_5^\star}{\partial \mathsf{M}_0} & \frac{\partial \mathsf{M}_5^\star}{\partial \mathsf{M}_1} & \frac{\partial \mathsf{M}_5^\star}{\partial \mathsf{M}_2} & \frac{\partial \mathsf{M}_5^\star}{\partial \mathsf{M}_3} & \frac{\partial \mathsf{M}_5^\star}{\partial \mathsf{M}_4} \end{bmatrix}, \tag{3.20}$$

where the details of the last row have been omitted for brevity. The matrices $\underline{\underline{A}}$ and $\underline{\underline{B}}$ as defined in (3.20) and (3.19), respectively, are *similar matrices*, meaning that they share the same eigenvalues.

Proposition 3.1 (Hyperbolicity of HyQMOM) *The HyQMOM fluid model as expressed either in conservative or primitive form* (3.17)–(3.19) *is strictly hyperbolic for all* $\rho,\ p,\ k > 0$. *Note that* $\rho,\ p,\ k > 0$ *defines a convex set in the solution space:* $\underline{q} \in \mathbb{R}^5$. *The resulting wave structure includes one linearly degenerate wave and four nonlinear waves.*

Note that the essential contents of this theorem were already known in Fox, Laurent, and Vié [22], but never presented as a formal proposition and proved. Therefore, we include the proof here.

Proof The eigenvalues of the Jacobian matrix, $\underline{\underline{B}}\left(\underline{\alpha}\right)$, in (3.19) can be computed explicitly:

$$\begin{aligned} \lambda_1 = u + \frac{h}{2p} - \sqrt{a+b}, \quad \lambda_2 = u + \frac{h}{2p} - \sqrt{a-b}, \quad \lambda_3 = u, \\ \lambda_4 = u + \frac{h}{2p} + \sqrt{a-b}, \quad \lambda_5 = u + \frac{h}{2p} + \sqrt{a+b}, \end{aligned} \tag{3.21}$$

where

$$a = \frac{p}{\rho} + \frac{k}{p} + \left(\frac{h}{2p}\right)^2 \quad \text{and} \quad b = \sqrt{\frac{k^2}{p^2} + \frac{k}{\rho}}. \tag{3.22}$$

One can show via simple calculations that for all $\rho,\ p, k > 0$:

$$a > b > 0, \quad \sqrt{a+b} > \frac{|h|}{2p}, \quad \text{and} \quad \sqrt{a-b} > \frac{|h|}{2p}. \tag{3.23}$$

Therefore, all five eigenvalues shown in (3.21)–(3.22) are real and distinct for all $\rho,\ p, k > 0$, which is sufficient to show that system (3.17)–(3.19) is strictly hyperbolic.

The eigenvectors of the flux Jacobian (3.20) can be written as follows:

$$\underline{\mathcal{R}_\ell} = \left[1, \lambda_\ell, \lambda_\ell^2, \lambda_\ell^3, \lambda_\ell^4\right]^T \quad \text{for} \quad \ell = 1, 2, 3, 4, 5. \tag{3.24}$$

To determine whether the corresponding waves are linearly degenerate or not, we need to compute the quantities:

$$\frac{\partial \lambda_\ell}{\partial \underline{q}} \cdot \underline{\mathcal{R}_\ell} = \frac{\partial \lambda_\ell}{\partial \underline{\alpha}} \cdot \left(\frac{\partial \underline{q}}{\partial \underline{\alpha}} \right)^{-1} \cdot \underline{\mathcal{R}_\ell}, \quad \forall \ell = 1, 2, 3, 4, 5, \tag{3.25}$$

where

$$\frac{\partial \lambda_\ell}{\partial \underline{q}} := \left[\frac{\partial \lambda_\ell}{\partial q_1}, \frac{\partial \lambda_\ell}{\partial q_2}, \frac{\partial \lambda_\ell}{\partial q_3}, \frac{\partial \lambda_\ell}{\partial q_4}, \frac{\partial \lambda_\ell}{\partial q_5} \right], \tag{3.26}$$

and

$$\frac{\partial \underline{q}}{\partial \underline{\alpha}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ u & \rho & 0 & 0 & 0 \\ u^2 & 2\rho u & 1 & 0 & 0 \\ u^3 & 3\left(p + \rho u^2\right) & 3u & 1 & 0 \\ \frac{\rho^2 u^4 - p^2}{\rho^2} & 4\left(h + 3pu + \rho u^3\right) & \frac{6\rho p^2 u^2 + 2p^3 - \rho h^2}{\rho p^2} & \frac{2h + 4pu}{p} & 1 \end{bmatrix}. \tag{3.27}$$

We note that one of the waves is linearly degenerate while the remaining are nonlinear:

$$\frac{\partial \lambda_3}{\partial \underline{q}} \cdot \underline{\mathcal{R}_3} \equiv 0 \quad \text{and} \quad \frac{\partial \lambda_\ell}{\partial \underline{q}} \cdot \underline{\mathcal{R}_\ell} \not\equiv 0 \quad \text{for} \quad \ell = 1, 2, 4, 5. \tag{3.28}$$

□

4 Locally-Implicit Lax–Wendroff Discontinuous Galerkin

We consider generic one-dimensional conservation laws of the form:

$$\underline{q}_{,t} + \underline{f}\left(\underline{q}\right)_{,x} = \underline{0}, \tag{4.1}$$

where t is time, x is space, $q(t, x) :\in \mathbb{R}^+ \times \mathbb{R} \mapsto \mathbb{R}^{M_{\text{eqn}}}$ is the vector of conserved variables, M_{eqn} is the number of equations, and $\underline{f}\left(\underline{q}\right) : \mathbb{R}^{M_{\text{eqn}}} \mapsto \mathbb{R}^{M_{\text{eqn}}}$ is the flux function. We assume that this system is hyperbolic, meaning that the flux Jacobian:

$$\underline{f}\left(\underline{q}\right)_{,\underline{q}} : \mathbb{R}^{M_{\text{eqn}}} \mapsto \mathbb{R}^{M_{\text{eqn}} \times M_{\text{eqn}}}, \tag{4.2}$$

has real eigenvalues and a complete set of eigenvectors over some convex region $\mathcal{D} \subset \mathbb{R}^{M_{\text{eqn}}}$ in solution space inside of which we are interested in solving the equation.

The Lax–Wendroff method [37] is a time discretization for hyperbolic conservation laws based on the the Cauchy–Kovalevskaya [58] procedure to convert temporal derivatives into spatial derivatives. We begin with a Taylor series in time:

$$\underline{q}(t + \Delta t, x) = \underline{q}(t, x) + \Delta t \underline{q}_{,t}(t, x) + \frac{1}{2} \Delta t^2 \underline{q}_{,t,t}(t, x) + \ldots, \tag{4.3}$$

and then replace all time derivatives by spatial derivatives:

$$\underline{q}_{,t} = -\underline{f}\left(\underline{q}\right)_{,x}, \quad \underline{q}_{,t,t} = -\underline{f}\left(\underline{q}\right)_{,t,x} = -\left[\underline{\underline{f'}}\left(\underline{q}\right) \underline{q}_{,t}\right]_{,x} = \left[\underline{\underline{f'}}\left(\underline{q}\right) \underline{f}\left(\underline{q}\right)_{,x}\right]_{,x}, \quad \ldots, \tag{4.4}$$

which results in the following:

$$\underline{q}(t + \Delta t, x) = \underline{q} - \Delta t \underline{f}\left(\underline{q}\right)_{,x} + \frac{1}{2} \Delta t^2 \left[\underline{\underline{f'}}\left(\underline{q}\right) \underline{f}\left(\underline{q}\right)_{,x}\right]_{,x} + \ldots, \tag{4.5}$$

where on the right-hand side, we have suppressed the evaluation at (t, x). The final step is to truncate the Taylor series at some finite number of terms, and then replace all spatial derivatives by some discrete spatial derivative operators. The above Lax–Wendroff formalism [37] has been used in conjunction with with a variety of spatial discretizations, including finite volume [38], weighted essentially non-oscillatory (WENO) [54], and discontinuous Galerkin [48] operators.

In this work, we are concerned with the discontinuous Galerkin version of Lax–Wendroff [48]; and in particular, we make use of the reformulation of Gassner et al. [24] of the Lax–Wendroff discontinuous Galerkin (LxW-DG) scheme in terms of a locally-implicit prediction step, followed by an explicit correction step. The key advantage of this formulation is that we do not need to explicitly compute the partial derivatives as shown in (4.4); and instead, the locally-implicit solver automatically produces discrete versions of these derivatives. The next challenge is to efficiently solve the nonlinear algebraic equations arising from the locally-implicit prediction step; we solve these equations by following Gassner et al. [24] and making use of a Picard fixed point iteration. We will follow the notational conventions of Guthrey and Rossmanith [26] developed for locally-implicit and regionally-implicit LxW-DG schemes. Note that the predictor-correct method is equivalent to the Lax–Wendroff DG method for linear constant-coefficient hyperbolic systems. For nonlinear systems, the predictor-correct method differs slightly from Lax–Wendroff DG in that the Picard iteration inside the prediction step approximates the direct computation of the nonlinear Taylor series expansion.

4.1 Discontinuous Galerkin Finite Elements

To discretize Eq. (4.1) in space, we use the discontinuous Galerkin (DG) finite element method, which was first introduced by Reed and Hill [49]. It was fully developed for time-dependent hyperbolic conservation laws in a series of papers by Bernardo Cockburn, Chi-Wang Shu, and collaborators (see [14] and references therein for details).

We define the broken finite element space

$$\mathcal{W}^{\Delta x} := \left\{ w^{\Delta x} \in [L^\infty(\Omega)]^{M_{\text{eqn}}} : w^{\Delta x}|_{\mathcal{T}_i} \in [\mathbb{P}(M_{\text{deg}})]^{M_{\text{eqn}}} \quad \forall \mathcal{T}_i \right\}, \tag{4.6}$$

where $\Delta x = (x_{\text{high}} - x_{\text{low}})/M_{\text{elem}}$ is a uniform grid spacing with M_{elem} being the number of elements. Additionally, M_{eqn} is the number of conserved variables, $\mathbb{P}(M_{\text{deg}})$ is the set of all polynomials of degree at most M_{deg}, and the computational mesh is described by non-overlapping elements of width Δx centered at the points x_i:

$$\mathcal{T}_i = \left[x_i - \frac{\Delta x}{2},\, x_i + \frac{\Delta x}{2} \right] \qquad \text{for } i = 1, \ldots, M_{\text{elem}}. \tag{4.7}$$

On each element, we define the local spatial variable, ξ:

$$x = x_i + \left(\frac{\Delta x}{2} \right) \xi \quad \text{for} \quad \xi \in [-1, 1]. \tag{4.8}$$

On each element, we approximate the solution by a finite expansion in terms of the following orthonormal Legendre polynomial basis functions:

$$\underline{\Phi} = \left(1,\ \sqrt{3}\xi,\ \frac{\sqrt{5}}{2}\left(3\xi^2-1\right),\ \frac{\sqrt{7}}{2}\left(5\xi^3-3\xi\right),\ \frac{\sqrt{9}}{8}\left(35\xi^4-30\xi^2+3\right),\ \ldots\right), \tag{4.9}$$

with the orthonormality property:

$$\frac{1}{2}\int_{-1}^{1} \underline{\Phi}\,\underline{\Phi}^T\, d\xi = \underline{\underline{\mathbb{I}}}, \tag{4.10}$$

where $\underline{\underline{\mathbb{I}}}$ is the identity matrix. On each element at time $t = t^n$, we approximate the solution as follows:

$$\underline{q}^h\left(t^n, x_i + \frac{\Delta x}{2}\xi\right) := \underline{\Phi}(\xi)^T\, \underline{\underline{Q_i^n}} \quad \text{for} \quad \xi \in [-1,1], \tag{4.11}$$

where

$$\underline{\Phi}(\xi) : [-1,1] \mapsto \mathbb{R}^{M_C} \quad \text{and} \quad \underline{\underline{Q_i^n}} \in \mathbb{R}^{M_C \times M_{\text{eqn}}}. \tag{4.12}$$

The number of basis functions in 1D is $M_C = M_{\text{deg}} + 1$, and the order of accuracy is $M_O = M_{\text{deg}} + 1$.

4.2 Prediction Step

The prediction step is completely local to each element, and therefore, the prediction step is inconsistent with the underlying conservation law. This inconsistency allows us to freely choose updated variables; we use the primitive variables for this step: $\underline{\alpha} = (\rho, u, p, h, k)$.

The prediction step is local on each space-element $\left[t^n, t^{n+1}\right] \times \mathcal{T}_i$, where $t^{n+1} = t^n + \Delta t$. Let $t = t^n + \frac{\Delta t}{2}(1+\tau)$, for $\tau \in [-1,1]$ and

$$\underline{\alpha}_{,\tau} = \underline{\Theta}(\underline{\alpha}) := -\frac{\Delta t}{\Delta x}\,\underline{\underline{B}}\left(\underline{\alpha}\right)\underline{\alpha}_{,\xi}, \tag{4.13}$$

where $\underline{\underline{B}}(\underline{\alpha})$ is the (primitive variable) flux Jacobian matrix defined by (3.19). We introduce a space-time Legendre basis on each element:

$$\Psi_\ell(\tau,\xi) = \Phi_{\ell_1}(\tau)\Phi_{\ell_2}(\xi), \quad \text{for} \quad \ell_1 = 1,\ldots,M_O, \quad \ell_2 = 1,\ldots,M_O+1-\ell_1, \tag{4.14}$$

where

$$\ell = M_O\left(\ell_1 - 1\right) - \frac{\left(\ell_1-1\right)\left(\ell_1-2\right)}{2} + \ell_2 \quad \text{such that} \quad \ell = 1,\ldots,M_P, \tag{4.15}$$

and $M_P = M_O(M_O+1)/2$ is the number of space-time basis functions. These space-time basis functions are orthonormal on $[-1,1]^2$:

$$\frac{1}{4}\iint_{-1}^{1} \underline{\Psi}\,\underline{\Psi}^T\, d\tau\, d\xi = \underline{\underline{\mathbb{I}}}. \tag{4.16}$$

We write the predicted solution as follows:

$$\underline{\alpha}^{\text{ST}}\left(t^n + \frac{\Delta t}{2}(1+\tau),\ x_i + \frac{\Delta x}{2}\,\xi\right) := \underline{\Psi}\,(\tau,\xi)^T\ \underline{\underline{W_i^{n+\frac{1}{2}}}}, \qquad \underline{\underline{W_i^{n+\frac{1}{2}}}} \in \mathbb{R}^{M_P \times M_{\text{eqn}}}, \tag{4.17}$$

for $(\tau, \xi) \in [-1, 1]^2$, where $\underline{\underline{W}}$ represents the matrix of unknown coefficients.

We proceed by multiplying (4.13) by the test function $\underline{\Psi}$, then integrating over $(\tau, \xi) \in [-1, 1]^2$:

$$\frac{1}{4} \iint_{-1}^{1} \alpha_{m,\tau} \, \underline{\Psi} \, d\tau \, d\xi = \frac{1}{4} \iint_{-1}^{1} \Theta_m(\underline{\alpha}) \, \underline{\Psi} \, d\tau \, d\xi, \tag{4.18}$$

and applying integration-by-parts in τ only:

$$\begin{aligned} &\frac{1}{4} \int_{-1}^{1} \alpha_m^\star \, (\tau = 1, \xi) \, \underline{\Psi} \, (\tau = 1, \xi) \, d\xi - \frac{1}{4} \int_{-1}^{1} \alpha_m^\star \, (\tau = -1, \xi) \, \underline{\Psi} \, (\tau = -1, \xi) \, d\xi \\ &\quad - \frac{1}{4} \iint_{-1}^{1} \alpha_m \, \underline{\Psi}_{,\tau} \, d\tau \, d\xi = \frac{1}{4} \iint_{-1}^{1} \Theta_m(\underline{\alpha}) \, \underline{\Psi} \, d\tau \, d\xi, \end{aligned} \tag{4.19}$$

where the choice of values for $\alpha_m^\star$ at $\tau = 1$ and $\tau = -1$ still remains to be made. Before making this choice, however, let us reverse integrate-by-parts, such that the newly introduced boundary terms are always the internal values of the current space-time element:

$$\begin{aligned} &\frac{1}{4} \int_{-1}^{1} \Big[\alpha_m^\star \, (\tau = 1, \xi) - \alpha_m \, (\tau = 1, \xi) \Big] \underline{\Psi} \, (\tau = 1, \xi) \, d\xi \\ &\quad - \frac{1}{4} \int_{-1}^{1} \Big[\alpha_m^\star \, (\tau = -1, \xi) - \alpha_m \, (\tau = -1, \xi) \Big] \underline{\Psi} \, (\tau = -1, \xi) \, d\xi \\ &\quad + \frac{1}{4} \iint_{-1}^{1} \underline{\Psi} \, \alpha_{m,\tau} \, d\tau \, d\xi = \frac{1}{4} \iint_{-1}^{1} \Theta_m(\underline{\alpha}) \, \underline{\Psi} \, d\tau \, d\xi. \end{aligned}$$

Now we plug-in the following values for $\alpha_m^\star$ at $\tau = 1$ and $\tau = -1$:

$$\alpha_m^\star \, (\tau = 1, \xi) = \alpha_m \, (\tau = 1, \xi) \quad \text{and} \quad \alpha_m^\star \, (\tau = -1, \xi) = \underline{\Phi} \, (\xi)^T \, \underline{A_{i\,(:,m)}^n}, \tag{4.20}$$

as well as the following values for α_m on the interior of the space-time element:

$$\alpha_m \, (\tau, \xi) = \underline{\Psi} \, (\tau, \xi)^T \, \underline{W_{i\,(:,m)}^{n+\frac{1}{2}}}. \tag{4.21}$$

The result is

$$\begin{aligned} \underline{\underline{L}} \, \underline{W_{i\,(:,m)}^{n+\frac{1}{2}}} &= \frac{1}{4} \iint_{-1}^{1} \Theta_m \left(\underline{\Psi}^T \, \underline{\underline{W_i^{n+\frac{1}{2}}}} \right) \underline{\Psi} \, d\tau \, d\xi \\ &\quad + \left[\frac{1}{4} \int_{-1}^{1} \underline{\Psi} \, (-1, \xi) \, \underline{\Phi} \, (\xi)^T \, d\xi \right] \underline{A_{i\,(:,m)}^n}, \end{aligned} \tag{4.22}$$

for each equation $m = 1, \ldots, M_{\text{eqn}}$, where

$$\underline{\underline{L}} = \frac{1}{4} \iint_{-1}^{1} \underline{\Psi} \, \underline{\Psi}_{,\tau}^T \, d\tau \, d\xi + \frac{1}{4} \int_{-1}^{1} \underline{\Psi}_{|_{\tau=-1}} \underline{\Psi}_{|_{\tau=-1}}^T \, d\xi, \tag{4.23}$$

$$\underline{\underline{A_i^n}} = \frac{1}{2} \sum_{a=1}^{M_{\text{O}}} \omega_a \, \underline{\Phi} \, (\mu_a) \left[\underline{\alpha} \left(\underline{\Phi} \, (\mu_a)^T \, \underline{\underline{Q_i^n}} \right) \right]^T, \tag{4.24}$$

and $\underline{\alpha}(\underline{q})$ gives the relationship between conservative and primitive variables. Equation (4.22) is a nonlinear algebraic equation that must be solved on each space-time element for the matrix of unknown coefficients: $\underline{\underline{W}}$.

Following Gassner et al. [24], instead of using Newton's method to solve the resulting non-linear equation, which involves inverting a Jacobian matrix at every step, we used the

much simpler Picard iteration. There are two key advantages of the Picard iteration over Newton's method. First, since $\underline{\underline{L}}$ is independent of the solution and the same on each space-time element, we only invert this relatively small matrix once. Second, the Picard iteration converges to sufficiently high order accuracy after $M_{\mathrm{O}} - 1$ iterations, so the need to compute residuals is eliminated (see justification in [24]). We can write the Picard iteration as

$$\begin{aligned}\underline{W_{i\,(:,m)}^{n+\frac{1}{2}(j)}} =& \frac{1}{4}\sum_{a=1}^{M_{\mathrm{O}}}\sum_{b=1}^{M_{\mathrm{O}}} \omega_a\,\omega_b\,\underline{\hat{\Psi}}\,(\mu_b,\,\mu_a)\;\Theta_m\left(\underline{\Psi}\,(\mu_b,\,\mu_a)^T\;\underline{\underline{W_i^{n+\frac{1}{2}(j-1)}}}\right)\\ &+\frac{1}{4}\sum_{b=1}^{M_{\mathrm{O}}}\omega_b\,\underline{\hat{\Psi}}\,(-1,\,\xi_b)\;\underline{\Phi}\,(\xi_b)^T\;\underline{A_{i\,(:,m)}^n},\end{aligned} \tag{4.25}$$

where $j = 1, \ldots, M_{\mathrm{O}} - 1$ is the iteration count, $m = 1, \ldots, M_{\mathrm{eqn}}$ is the equation index, $\underline{\hat{\Psi}} = \underline{\underline{L}}^{-1}\underline{\Psi}$, and ω_a and μ_a for $a = 1, \ldots, M_{\mathrm{O}}$ are the weights and abscissas of the M_{O}-point Gauss–Legendre quadrature rule. This gives a solution for the prediction step, which we know is inconsistent with the conservation law. To make the final solution consistent (and high-order) with the original conservation law, we next need to add a correction step.

4.3 Correction Step

The correction step is designed to work like a single forward Euler-like step that uses the predicted solution. To perform this step, we begin with the hyperbolic conservation law (4.1) and multiply by the spatial basis functions defined in (4.9). Next, we integrate over the space-time element $(\tau, \xi) \in [-1, 1]^2$:

$$\frac{1}{2}\iint_{-1}^{1}\left[\underline{\Phi}(\xi)\,\underline{q}_{,\tau} + \frac{\Delta t}{\Delta x}\,\underline{\Phi}(\xi)\,\underline{f}\left(\underline{q}\right)_{,\xi}\right]d\xi\,d\tau = \underline{0}, \tag{4.26}$$

which can be written as

$$\begin{aligned}\frac{1}{2}\int_{-1}^{1}\underline{\Phi}(\xi)\,\underline{q}\left(t^{n+1}, x_i + \frac{\Delta x}{2}\xi\right)\,d\xi =& \frac{1}{2}\int_{-1}^{1}\underline{\Phi}(\xi)\,\underline{q}\left(t^{n}, x_i + \frac{\Delta x}{2}\xi\right)\,d\xi\\ &- \frac{\Delta t}{2\Delta x}\iint_{-1}^{1}\underline{\Phi}(\xi)\,\underline{f}\left(\underline{q}\right)_{,\xi}\,d\xi\,d\tau.\end{aligned} \tag{4.27}$$

We approximate $\underline{q}\left(t^{n+1}, \cdot\right)$ and $\underline{q}\left(t^{n}, \cdot\right)$ in (4.27) via appropriate versions of ansatz (4.11). For the remaining term, we first apply integration-by-parts in space, then replace the true solution q by its space-time predicted solution: (4.17), and replace exact integration by numerical quadrature. This results in the following expression:

$$\begin{aligned}\underline{\underline{Q_i^{n+1}}} =& \underline{\underline{Q_i^n}} + \frac{\Delta t}{2\Delta x}\sum_{a=1}^{M_{\mathrm{O}}}\sum_{b=1}^{M_{\mathrm{O}}}\omega_a\,\omega_b\,\underline{\Phi}_{,\xi}\,(\mu_a)\left[\underline{f}\left(\underline{\Psi}\,(\mu_b,\,\mu_a)^T\;\underline{\underline{W_i^{n+\frac{1}{2}}}}\right)\right]^T\\ &- \frac{\Delta t}{\Delta x}\left(\underline{\Phi}(1)\left[\underline{\mathcal{F}_{i+\frac{1}{2}}^{n+\frac{1}{2}}}\right]^T - \underline{\Phi}(-1)\left[\underline{\mathcal{F}_{i-\frac{1}{2}}^{n+\frac{1}{2}}}\right]^T\right).\end{aligned} \tag{4.28}$$

In the expressions after the approximation symbol, $\approx$, we replaced all exact integration by a Gauss–Legendre quadrature, where ω_a and μ_a for $a = 1, \ldots, M_{\mathrm{O}}$ are the weights and abscissas of the M_{O}-point Gauss–Legendre quadrature rule.

Table 1 The CFL numbers used in all simulations as a function of the method order

Order	$M_{\mathrm{O}} = 1$	$M_{\mathrm{O}} = 2$	$M_{\mathrm{O}} = 3$	$M_{\mathrm{O}} = 4$
CFL # used in practice	0.90	0.30	0.14	0.09

The time-integrated numerical fluxes are defined using the predicted solution and the Rusanov [50] time-averaged flux:

$$\underline{\mathcal{F}}_{i-\frac{1}{2}}^{n+\frac{1}{2}} := \frac{1}{2}\sum_{a=1}^{M_{\mathrm{O}}} \omega_a \, \underline{\mathcal{F}}\left(\mu_a\right), \tag{4.29}$$

where the numerical flux at each temporal quadrature point is given by

$$\underline{\mathcal{F}}(\tau) := \frac{1}{2}\left\{\underline{f}\left(\underline{W}_{\mathrm{R}}(\tau)\right) + \underline{f}\left(\underline{W}_{\mathrm{L}}(\tau)\right)\right\} - \frac{1}{2}\lambda_{\max}(\tau)\left\{\underline{q}\left(\underline{W}_{\mathrm{R}}(\tau)\right) - \underline{q}\left(\underline{W}_{\mathrm{L}}(\tau)\right)\right\}, \tag{4.30}$$

where

$$\underline{W}_{\mathrm{L}}(\tau) := \underline{\Psi}(\tau, 1)^T \, \underline{\underline{W}}_{i-1}^{n+\frac{1}{2}}, \qquad \underline{W}_{\mathrm{R}}(\tau) := \underline{\Psi}(\tau, -1)^T \, \underline{\underline{W}}_{i}^{n+\frac{1}{2}}, \tag{4.31}$$

and $\lambda_{\max}(\tau)$ is a local bound on the spectral radius of the flux Jacobian, $\underline{f}(\underline{q})_{,\underline{q}} = \underline{\underline{A}}\left(\underline{q}\right)$, in the neighborhood of interface $x = x_{i-\frac{1}{2}}$ and at time $t = t^n + (\tau + 1)\Delta t/2$.

These steps are all it takes to regain the coupling neglected in the prediction step. We now have a solution that is not only consistent with the conservation law but is also high order. The Courant–Friedrichs–Lewy (CFL) number we use for each order is given in Table 1. However, we still have some work to do to ensure that the solution is physical. We must be careful to maintain the positivity of the primitive variables ρ, p, k, as was necessary for the system's hyperbolicity and moment-realizability. We address the limiters utilized to accomplish this in the next section.

5 HyQMOM Limiters

The high-order numerical method as described in Sect. 4 does not guarantee that density, pressure, and modified kurtosis remain positive throughout a time-step:

$$\rho^n > 0, \quad p^n > 0, \quad k^n > 0 \quad \Longrightarrow \quad \rho^{n+1} > 0, \quad p^{n+1} > 0, \quad k^{n+1} > 0. \tag{5.1}$$

Recall that positivity of these quantities is needed to guarantee the moment-realizability of the moment-closure and strict hyperbolicity of the resulting evolution equations. If we want positivity over a time step, we will need to introduce *positivity-preserving limiters*. Additionally, if we want to control unphysical oscillations near large gradients, shocks, and rarefactions, we will also need *non-oscillatory limiters*. In this section, we derive all of these limiters. In particular, we first need to establish that a simple first-order scheme is positivity-preserving under some appropriate time-step restriction; this is done in Sect. 5.1. Using this result we derive a suite of limiters that ensure positivity: Sect. 5.2 (positivity at select points in the prediction step), Sect. 5.3 (positivity of the average density, pressure, and modified kurtosis in each element in the correction step), and Sect. 5.4 (positivity at select points in the correction step). We then develop an unphysical oscillation limiter in Sect. 5.5.

5.1 Positivity of the Rusanov Scheme

Before considering positivity limiters for the high-order method, we must first establish that simple first-order schemes, in this case, we consider the Rusanov (aka local Lax–Friedrichs) scheme [50], are positivity-preserving under some appropriate time-step restriction. This is established in the theorem below, which is an extension of the result of Zhang and Shu [62] for the compressible Euler equations.

Theorem 5.1 *Let $\underline{Q_i^n}$ be some approximation of the element averages of the conserved variables on element $\mathcal{T}_i$ at time t^n, and let $\underline{Q_i^{n+1}}$ be the element averages produced by the Rusanov scheme [50] at time $t^{n+1} = t^n + \Delta t$. Then*

$$\rho_i^n > 0, \quad p_i^n > 0, \quad k_i^n > 0 \quad \forall i \quad \Longrightarrow \quad \rho_i^{n+1} > 0, \quad p_i^{n+1} > 0, \quad k_i^{n+1} > 0 \quad \forall i, \tag{5.2}$$

under the CFL condition:

$$\frac{\Delta t}{\Delta x} \max_i \left(\lambda_{i+\frac{1}{2}} \right) < 1, \tag{5.3}$$

where

$$\lambda_{i\pm\frac{1}{2}} = \max \left\{ \lambda_{\max} \left(\underline{Q_i^n} \right), \lambda_{\max} \left(\underline{Q_{i\pm1}^n} \right), \lambda_{\max} \left(\frac{1}{2} \left(\underline{Q_i^n} + \underline{Q_{i\pm1}^n} \right) \right) \right\}, \tag{5.4}$$

and $\lambda_{\max} \left(\underline{Q_i^n} \right)$ is a bound on the spectral radius of the flux Jacobian (3.20) at state $\underline{Q_i^n}$.

Proof Recall that the Rusanov scheme can be written as

$$\underline{Q_i^{n+1}} = \underline{Q_i^n} - \frac{\Delta t}{\Delta x} \left(\underline{\mathcal{F}_{i+\frac{1}{2}}} - \underline{\mathcal{F}_{i-\frac{1}{2}}} \right), \tag{5.5}$$

where the numerical fluxes are given by

$$\underline{\mathcal{F}_{i\pm\frac{1}{2}}} = \frac{1}{2} \left[\underline{f}(\underline{Q_i^n}) + \underline{f}(\underline{Q_{i\pm1}^n}) \right] - \frac{1}{2} \lambda_{i\pm\frac{1}{2}} \left(\underline{Q_i^n} - \underline{Q_{i\pm1}^n} \right), \tag{5.6}$$

where the flux function, $\underline{f} \left(\underline{q} \right)$, is defined by (3.18) and the local wave speed, $\lambda_{i\pm\frac{1}{2}}$, is defined by (5.4). We can rewrite the above expression into the following numerical update:

$$\underline{Q_i^{n+1}} = \left[1 - \frac{\Delta t}{2\Delta x} \left(\lambda_{i+\frac{1}{2}} + \lambda_{i-\frac{1}{2}} \right) \right] \underline{Q_i^n} + \left[\frac{\Delta t \lambda_{i+\frac{1}{2}}}{2\Delta x} \right] \underline{M_i^+} + \left[\frac{\Delta t \lambda_{i-\frac{1}{2}}}{2\Delta x} \right] \underline{M_i^-}, \tag{5.7}$$

where

$$\underline{M_i^+} = \underline{Q_{i+1}^n} - \left(\lambda_{i+\frac{1}{2}} \right)^{-1} \underline{f} \left(\underline{Q_{i+1}^n} \right) \quad \text{and} \quad \underline{M_i^-} = \underline{Q_{i-1}^n} + \left(\lambda_{i-\frac{1}{2}} \right)^{-1} \underline{f} \left(\underline{Q_{i-1}^n} \right). \tag{5.8}$$

Under the CFL condition (5.3), we note that

$$1 - \frac{\Delta t}{2\Delta x} \left(\lambda_{i+\frac{1}{2}} + \lambda_{i-\frac{1}{2}} \right) \geq 0, \quad \frac{\Delta t \lambda_{i+\frac{1}{2}}}{2\Delta x} \geq 0, \quad \text{and} \quad \frac{\Delta t \lambda_{i-\frac{1}{2}}}{2\Delta x} \geq 0. \tag{5.9}$$

Additionally, note that the coefficients in (5.7) sum to unity:

$$\left[1-\frac{\Delta t}{2\Delta x}\left(\lambda_{i+\frac{1}{2}}+\lambda_{i-\frac{1}{2}}\right)\right]+\left[\frac{\Delta t\lambda_{i+\frac{1}{2}}}{2\Delta x}\right]+\left[\frac{\Delta t\lambda_{i-\frac{1}{2}}}{2\Delta x}\right]=1. \tag{5.10}$$

Now let $\mathcal{C}_\alpha$ be any convex function of the conserved variables: $\underline{q}=(\mathsf{M}_0,\mathsf{M}_1,\mathsf{M}_2,\mathsf{M}_3,\mathsf{M}_4)$. Then, applying the convex function to both sides of (5.7) we see that

$$\begin{aligned}\mathcal{C}_\alpha\left(\underline{Q}_i^{n+1}\right)\le&\left[1-\frac{\Delta t}{2\Delta x}\left(\lambda_{i+\frac{1}{2}}+\lambda_{i-\frac{1}{2}}\right)\right]\mathcal{C}_\alpha\left(\underline{Q}_i^{n}\right)\\&+\left[\frac{\Delta t\lambda_{i+\frac{1}{2}}}{2\Delta x}\right]\mathcal{C}_\alpha\left(\underline{M}_i^{+}\right)+\left[\frac{\Delta t\lambda_{i-\frac{1}{2}}}{2\Delta x}\right]\mathcal{C}_\alpha\left(\underline{M}_i^{-}\right),\end{aligned} \tag{5.11}$$

which follows from conditions (5.9) and (5.10), as well as the property of convex functions shown in Lemma A.4.

Furthermore, we note that density, pressure, and modified kurtosis (ρ, p, and k) are all convex functions of $\underline{q}=(\mathsf{M}_0,\mathsf{M}_1,\mathsf{M}_2,\mathsf{M}_3,\mathsf{M}_4)$; the density is trivially convex, while the pressure is convex if $\rho>0$, and the modified kurtosis is convex if $\rho>0$ and $p>0$ (see definitions (2.7)). Therefore, to prove result (5.2), all we need to show is that

$$\mathcal{C}_\alpha\left(\underline{M}_i^{+}\right)>0\qquad\text{and}\qquad\mathcal{C}_\alpha\left(\underline{M}_i^{-}\right)>0, \tag{5.12}$$

with $\mathcal{C}_\alpha$ chosen as the density, pressure, and modified kurtosis:

$$\begin{aligned}\mathcal{C}_\rho\left(\underline{q}\right):=q_1,\quad\mathcal{C}_p\left(\underline{q}\right):=q_3-\frac{q_2^2}{q_1},\\ \mathcal{C}_k\left(\underline{q}\right):=\frac{q_3^3-2q_2q_3q_4+q_1q_4^2+q_2^2q_5-q_1q_3q_5}{q_2^2-q_1q_3}.\end{aligned} \tag{5.13}$$

1. **Density:** We take $\alpha\equiv\rho$ and note that

$$\mathcal{C}_\rho\left(\underline{Q}\pm\lambda^{-1}\underline{f}(\underline{Q})\right)=\frac{(\lambda\pm u)\,\rho}{\lambda}. \tag{5.14}$$

   Positivity of (5.14) follows from the fact that the wave speed, λ, always exceeds the local fluid speed, $|u|$.
2. **Pressure:** We take $\alpha\equiv p$ and note that

$$\mathcal{C}_p\left(\underline{Q}\pm\lambda^{-1}\underline{f}(\underline{Q})\right)=\frac{p\rho\,(\lambda\pm u)^2+h\rho\,(u\pm\lambda)-p^2}{\lambda\,(\lambda\pm u)\,\rho}, \tag{5.15}$$

   for which we note that the numerator is a quadratic polynomial in λ. The roots of this quadratic polynomial can easily be computed:

$$z_1=\mp\left(u+\frac{h}{2p}\right)-\sqrt{\frac{p}{\rho}+\left(\frac{h}{2p}\right)^2},\quad z_2=\mp\left(u+\frac{h}{2p}\right)+\sqrt{\frac{p}{\rho}+\left(\frac{h}{2p}\right)^2}.$$

   Positivity of (5.15) follows from the fact that $\lambda>\max\{z_1,z_2\}$ (i.e., λ is always to the right of the roots) and $p\rho>0$ (i.e., the quadratic is concave up).

3. **Modified kurtosis:** We take $\alpha \equiv k$ and note that

$$\mathcal{C}_k\left(\underline{Q} \pm \lambda^{-1}\underline{f}(\underline{Q})\right) = \left(\frac{k(\lambda \pm u)}{\lambda}\right)\left(\frac{p\rho\,(\lambda \pm u)^2 + h\rho\,(u \pm \lambda) - p^2 - k\rho}{p\rho\,(\lambda \pm u)^2 + h\rho\,(u \pm \lambda) - p^2}\right), \tag{5.16}$$

where the numerator of the second fraction is again a quadratic polynomial in λ. Showing that this quadratic is positive is sufficient to show that the whole expression is positive since the remaining pieces are already positive due to the previously established positivity of (5.14) and (5.15). The roots of the quadratic are:

$$z_1 = \mp\left(u + \frac{h}{2p}\right) - \sqrt{\frac{k}{p} + \frac{p}{\rho} + \left(\frac{h}{2p}\right)^2}, \quad z_2 = \mp\left(u + \frac{h}{2p}\right) + \sqrt{\frac{k}{p} + \frac{p}{\rho} + \left(\frac{h}{2p}\right)^2}.$$

Positivity of (5.16) follows from the fact that $\lambda > \max\{z_1, z_2\}$ (i.e., λ is always to the right of the roots) and $p\rho > 0$ (i.e., the quadratic is concave up).

□

We have now shown that the first-order method will maintain positivity from one time step to the next under an appropriate CFL condition. However, higher-order methods will not automatically guarantee positivity; we address this issue in the subsequent subsections.

5.2 Limiter I: Positivity-at-Points in the Prediction Step

The prediction step, as described in (4.25) requires numerical quadrature in space-time in each Picard iteration. Furthermore, once the predicted solution has been computed, it will again be integrated in space-time in the correction step [i.e., see (4.28)–(4.31)]. To guarantee that all of the numerical quadratures in both the prediction and correction steps only use positive values of density, pressure, and modified kurtosis, we introduce a prediction-step positivity limiter.

Let the 1D Gauss–Legendre points internal to each element, augmented by the element end-points (i.e., the location of the element faces), be defined as follows:

$$\mathbb{X}_{M_O} := \Big\{-1, 1\Big\} \cup \Big\{\text{roots of the } M_O^{\text{th}} \text{ degree Legendre polynomial}\Big\}, \tag{5.17}$$

where M_O is the desired order of accuracy. Note that $\mathbb{X}_{M_O}$ contains a total of $M_O + 2$ points. We note that all of the quadratures in the prediction update (4.25) and correction update (4.28)–(4.31) only depends on the predicted solution at a small number of quadrature points, which are fully contained in the Cartesian product of $\mathbb{X}_{M_O}$ with itself:

$$\mathbb{X}^2_{M_O} := \mathbb{X}_{M_O} \otimes \mathbb{X}_{M_O}. \tag{5.18}$$

Therefore, $\mathbb{X}^2_{M_O}$ contains a total of $(M_O + 2)^2$ points. Our goal is thus to enforce positivity at all the space-time points $(\tau, \xi) \in \mathbb{X}^2_{M_O}$.

Following the strategy developed by Zhang and Shu [61] for the Runge–Kutta discontinuous Galerkin scheme, we apply the following procedure, which is applied element-by-element.

Step 1. On the current space-time element defined over

$$(t, x) \in \left[t^n,\, t^n + \Delta t\right] \times \left[x_i - \frac{\Delta x}{2},\, x_i + \frac{\Delta x}{2}\right],$$

the solution is given by (4.17). Find the minimum density, pressure, and modified kurtosis of this solution over the points $(\tau, \xi) \in \mathbb{X}^2_{M_O}$:

$$\alpha_i^{(m)} := \min_{(\tau,\xi)\in\mathbb{X}^2_{M_O}} \left\{ \underline{\Psi}\,(\tau,\xi)^T \, \underline{W_{i\,(:,m)}^{n+\frac{1}{2}}} \right\}, \tag{5.19}$$

for $m = 1$ (density), $m = 3$ (pressure), and $m = 5$ (modified kurtosis).

Step 2. Rewrite the solution as

$$\underline{\alpha}^{\mathrm{ST}} \left(t^n + \frac{\Delta t}{2}(1+\tau),\; x_i + \frac{\Delta x}{2}\,\xi;\; \theta \right) := (1-\theta)\, \underline{W_{i\,(1,:)}^{n+\frac{1}{2}}} + \theta\, \underline{\Psi}\,(\tau,\xi)^T \, \underline{\underline{W_i^{n+\frac{1}{2}}}}, \tag{5.20}$$

where $\theta \in [0,1]$, such that $\theta = 1$ recovers the original solution (4.17) and $\theta = 0$ results in reducing the entire solution on to its space-time average. We now choose the largest possible $\theta \in [0,1]$ so that (5.20) is positive for components $m = 1$ (density), 3 (pressure), 5 (modified kurtosis) at all the space-time points in $\mathbb{X}^2_{M_O}$. This is achieved by taking

$$\theta = \min_{m\in\{1,3,5\}} \min \left\{ 1, \frac{W_{i\,(1,m)}^{n+\frac{1}{2}} - \varepsilon}{W_{i\,(1,m)}^{n+\frac{1}{2}} - \alpha_i^{(m)}} \right\}, \tag{5.21}$$

where $\varepsilon > 0$ is a preselected small constant (e.g., in this work we select $\varepsilon = 10^{-14}$).

5.3 Limiter II: Positivity-in-the-Mean in the Correction Step

One of the key challenges in the correction step, as described by (4.28)–(4.31) is to make sure that the element averages of the density, pressure, and modified kurtosis remain positive at the end of the time-step: $\bar{\rho}_i^{n+1} > 0$, $\bar{p}_i^{n+1} > 0$, and $\bar{k}_i^{n+1} > 0$, where the bar over each variable refers to the element average. The prediction step limiter described in the previous Sect. 5.2, helps with this positivity-in-the-mean but cannot guarantee it. Furthermore, without positivity-in-the-mean, we cannot achieve positivity of the higher-order polynomial inside the element (i.e., if the polynomial average is negative, a significant portion of the polynomial must be negative inside the element). To overcome this challenge, we extend the approach developed by Moe et al. [43], which, for the element averages, blends the high-order update described by (4.28)–(4.31) with a first-order Rusanov scheme. We have already proved in Theorem 5.1 that the Rusanov scheme is guaranteed to preserve positivity.

We begin by defining the Rusanov [50] (aka local Lax–Friedrichs) update based on the element averages at $t = t^n$:

$$\underline{Q_i^{\mathrm{Rus}}} := \underline{Q_{i(1,:)}^n} - \frac{\Delta t}{\Delta x} \left(\underline{\mathcal{F}_{i+\frac{1}{2}}^{\mathrm{Rus}}} - \underline{\mathcal{F}_{i-\frac{1}{2}}^{\mathrm{Rus}}} \right), \tag{5.22}$$

where the numerical flux is given by

$$\underline{\mathcal{F}_{i-\frac{1}{2}}^{\mathrm{Rus}}} := \frac{1}{2}\left[\underline{f}\left(\underline{Q_{i(1,:)}^n} \right) + \underline{f}\left(\underline{Q_{i-1(1,:)}^n} \right) \right] - \frac{1}{2}\left| \lambda_{i-\frac{1}{2}} \right| \left(\underline{Q_{i(1,:)}^n} - \underline{Q_{i-1(1,:)}^n} \right), \tag{5.23}$$

and $\left|\lambda_{i-\frac{1}{2}}\right|$ is a local bound of the flux Jacobian spectral radius. Recall that $\underline{Q_i^{\mathrm{Rus}}}$ is guaranteed to have positive density, pressure, and modified kurtosis under a time-step restriction (see Theorem 5.1).

Next, we write the update for the element averages of the high-order method in terms of the low-order update (5.22):

$$\underline{Q^{n+1}_{i\,(1,:)}} = \underline{Q^{\text{Rus}}_{i\,(1,:)}} - \frac{\Delta t}{\Delta x}\left(\theta_{i+\frac{1}{2}}\,\underline{\Delta\mathcal{F}_{i+\frac{1}{2}}} - \theta_{i-\frac{1}{2}}\,\underline{\Delta\mathcal{F}_{i-\frac{1}{2}}}\right), \tag{5.24}$$

where the difference between the high and low-order fluxes is given by

$$\underline{\Delta\mathcal{F}_{i-\frac{1}{2}}} := \underline{\mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}}} - \underline{\mathcal{F}^{\text{Rus}}_{i-\frac{1}{2}}}, \tag{5.25}$$

and $\theta_{i+\frac{1}{2}} \in [0,\,1]$ measures the amount of flux limiting, where $\theta_{i+\frac{1}{2}} = 0$ represents maximal limiting (i.e., no high-order flux contributions) and $\theta_{i+\frac{1}{2}} = 1$ represents no limiting (i.e., no low-order flux contributions).

The strategy for the positivity-in-the-mean limiter is then to find the maximum $\theta_{i+\frac{1}{2}} \in [0,\,1]$ such that $\forall i$

$$\mathcal{C}_\alpha\left(\underline{Q^{n+1}_{i\,(1,:)}}\right) > 0, \tag{5.26}$$

for $\alpha \equiv \rho$ (density), $\alpha \equiv p$ (pressure), and $\alpha \equiv k$ (modified kurtosis), as defined in (5.13). The strategy for achieving this is outlined below and is applied element-by-element. The process begins by initializing $\theta_{i+\frac{1}{2}} = 1\ \forall i$.

Step 1: (density) Define

$$\Gamma := \frac{\Delta x}{\Delta t}\left(Q^{\text{Rus}}_{i(1)} - \varepsilon\right). \tag{5.27}$$

Set $\Lambda_{\text{left}} = \Lambda_{\text{right}} = 1$ (i.e., full high-order flux), but modify these if there is any potential for the density to decrease below zero.

Case 1. If $\Delta\mathcal{F}_{i-\frac{1}{2}(1)} < 0$ and $\Delta\mathcal{F}_{i+\frac{1}{2}(1)} < 0$, then

$$\Lambda_{\text{left}} = \Lambda_{\text{right}} = \min\left\{1,\ \frac{\Gamma}{\left|\Delta\mathcal{F}_{i-\frac{1}{2}(1)}\right| + \left|\Delta\mathcal{F}_{i+\frac{1}{2}(1)}\right|}\right\}. \tag{5.28}$$

Case 2. If $\Delta\mathcal{F}_{i-\frac{1}{2}(1)} < 0$ and $\Delta\mathcal{F}_{i+\frac{1}{2}(1)} > 0$ then

$$\Lambda_{\text{left}} = \min\left\{1,\ \frac{\Gamma}{\left|\Delta\mathcal{F}_{i-\frac{1}{2}(1)}\right|}\right\}. \tag{5.29}$$

Case 3. If $\Delta\mathcal{F}_{i+\frac{1}{2}(1)} < 0$ and $\Delta\mathcal{F}_{i-\frac{1}{2}(1)} > 0$ then

$$\Lambda_{\text{right}} = \min\left\{1,\ \frac{\Gamma}{\left|\Delta\mathcal{F}_{i+\frac{1}{2}(1)}\right|}\right\}. \tag{5.30}$$

Step 2: (pressure) Compute the Rusanov pressure (which is guaranteed to be positive):

$$p^{\text{Rus}} := \mathcal{C}_p\left(\underline{Q^{\text{Rus}}_i}\right). \tag{5.31}$$

Set $\mu_{11} = \mu_{10} = \mu_{01} = 1$, but modify these if there is any potential for the pressure to decrease below zero.

Part 2A. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} - \frac{\Delta t}{\Delta x}\left(\Lambda_{\text{right}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i+\frac{1}{2}}} - \Lambda_{\text{left}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}}}\right), \quad p^\star = \mathcal{C}_p\left(\underline{Q^\star}\right). \tag{5.32}$$

If $p^\star < \varepsilon$, then we set $\mu_{11} = \left(p^{\text{Rus}} - \varepsilon\right) / \left(p^{\text{Rus}} - p^\star\right)$.
Part 2B. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} + \frac{\Delta t}{\Delta x}\Lambda_{\text{left}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}}}, \quad p^\star = \mathcal{C}_p\left(\underline{Q^\star}\right). \tag{5.33}$$

If $p^\star < \varepsilon$, then $\mu_{10} = \left(p^{\text{Rus}} - \varepsilon\right) / \left(p^{\text{Rus}} - p^\star\right)$.
Part 2C. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} - \frac{\Delta t}{\Delta x}\Lambda_{\text{right}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i+\frac{1}{2}}}, \quad p^\star = \mathcal{C}_p\left(\underline{Q^\star}\right). \tag{5.34}$$

If $p^\star < \varepsilon$, then $\mu_{01} = \left(p^{\text{Rus}} - \varepsilon\right) / \left(p^{\text{Rus}} - p^\star\right)$.
Part 2D. Set

$$\mu = \min\left\{\mu_{11}, \mu_{10}, \mu_{01}\right\}, \quad \Lambda_{\text{left}} \leftarrow \mu\,\Lambda_{\text{left}}, \quad \Lambda_{\text{right}} \leftarrow \mu\,\Lambda_{\text{right}}. \tag{5.35}$$

Step 3: (modified kurtosis) Compute the Rusanov modified kurtosis (which is guaranteed to be positive):

$$k^{\text{Rus}} := \mathcal{C}_k\left(\underline{Q^{\text{Rus}}_i}\right). \tag{5.36}$$

Set $\mu_{11} = \mu_{10} = \mu_{01} = 1$, but modify these if there is any potential for the pressure to decrease below zero.

Part 3A. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} - \frac{\Delta t}{\Delta x}\left(\Lambda_{\text{right}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i+\frac{1}{2}}} - \Lambda_{\text{left}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}}}\right), \quad k^\star = \mathcal{C}_k\left(\underline{Q^\star}\right). \tag{5.37}$$

If $k^\star < \varepsilon$, then we set $\mu_{11} = \left(k^{\text{Rus}} - \varepsilon\right) / \left(k^{\text{Rus}} - k^\star\right)$.
Part 3B. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} + \frac{\Delta t}{\Delta x}\Lambda_{\text{left}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}}}, \quad k^\star = \mathcal{C}_k\left(\underline{Q^\star}\right). \tag{5.38}$$

If $k^\star < \varepsilon$, then $\mu_{10} = \left(k^{\text{Rus}} - \varepsilon\right) / \left(k^{\text{Rus}} - k^\star\right)$.
Part 3C. Set

$$\underline{Q^\star} = \underline{Q^n_{i\,(1,:)}} - \frac{\Delta t}{\Delta x}\Lambda_{\text{right}}\,\underline{\mathcal{F}^{n+\frac{1}{2}}_{i+\frac{1}{2}}}, \quad k^\star = \mathcal{C}_k\left(\underline{Q^\star}\right). \tag{5.39}$$

If $k^\star < \varepsilon$, then $\mu_{01} = \left(k^{\text{Rus}} - \varepsilon\right) / \left(k^{\text{Rus}} - k^\star\right)$.
Part 3D. Set

$$\mu = \min\left\{\mu_{11}, \mu_{10}, \mu_{01}\right\}, \quad \Lambda_{\text{left}} \leftarrow \mu\,\Lambda_{\text{left}}, \quad \Lambda_{\text{right}} \leftarrow \mu\,\Lambda_{\text{right}}. \tag{5.40}$$

Step 4: Set

$$\theta_{i-\frac{1}{2}} \leftarrow \min\left\{\theta_{i-\frac{1}{2}},\ \Lambda_{\text{left}}\right\} \quad \text{and} \quad \theta_{i+\frac{1}{2}} \leftarrow \min\left\{\theta_{i+\frac{1}{2}},\ \Lambda_{\text{right}}\right\}. \tag{5.41}$$

In all of the above formulas, we select in this work: $\varepsilon = 10^{-14}$.

5.4 Limiter III: Positivity-at-Points in the Correction Step

Once we have ensured that the element averages are positive, we then look to enforce positivity of the corrected solution at spatial quadrature points: $\xi \in \mathbb{X}_{M_O}$ as defined by (5.17). Following the ideas developed by Zhang and Shu [61] for the Runge–Kutta discontinuous Galerkin scheme, we aim to find the maximum $\theta \in [0, 1]$ such that

$$\underline{q}^h\left(t^{n+1},\, x_i + \frac{\Delta x}{2}\,\xi; \theta\right) := (1-\theta)\ \underline{Q}_{i\,(1,:)}^{n+1} + \theta\ \underline{\Phi}(\xi)^T\ \underline{\underline{Q}}_i^{n+1} \tag{5.42}$$

is positive at all points $\xi \in \mathbb{X}_{M_O}$ for every space element $\mathcal{T}_i$. As in the prediction step limiter from Sect. 5.2, $\theta = 0$ means that the solution is limited fully down to its element average, while $\theta = 1$ means that no limiting is needed and the full high-order approximation can be used. We apply the following procedure element-by-element.

Step 1. On the current element defined over

$$x \in \left[x_i - \frac{\Delta x}{2},\, x_i + \frac{\Delta x}{2}\right],$$

the solution is given by (4.11). Find the minimum density $\forall \xi \in \mathbb{X}_{M_O}$ [see (5.17)] and compute the corresponding damping parameter (θ):

$$\rho_i^{\min} := \min_{\xi \in \mathbb{X}_{M_O}} \left\{\underline{\Phi}\,(\xi)^T\ \underline{Q}_{i\,(:,1)}^{n+1}\right\}, \quad \theta = \min\left\{1,\, \frac{Q_{i\,(1,1)}^{n+1} - \varepsilon}{Q_{i\,(1,1)}^{n+1} - \rho_i^{\min}}\right\}. \tag{5.43}$$

Finally, rescale the higher-order coefficients using the above calculated damping parameter (θ):

$$\underline{Q}_{i\,(:,\ell)}^{n+1} \leftarrow \theta\ \underline{Q}_{i\,(:,\ell)}^{n+1} \quad \forall \ell = 2, \ldots, M_C. \tag{5.44}$$

Step 2. Now that the density is positive $\forall \xi \in \mathbb{X}_{M_O}$, we repeat **Step 1** for the pressure. That is, we find the average pressure and the minimum pressure $\forall \xi \in \mathbb{X}_{M_O}$:

$$\overline{P}_i := \mathcal{C}_p\left(\underline{Q}_{i\,(1,:)}^{n+1}\right), \quad p_i^{\min} := \min_{\xi \in \mathbb{X}_{M_O}} \left\{\mathcal{C}_p\left(\underline{\Phi}\,(\xi)^T\ \underline{\underline{Q}}_i^{n+1}\right)\right\}, \tag{5.45}$$

where $\mathcal{C}_p\left(\underline{q}\right)$ is defined in (5.13). From here, we compute the corresponding damping parameter and rescale the higher-order coefficients:

$$\theta = \min\left\{1,\, \frac{\overline{P}_i - \varepsilon}{\overline{P}_i - p_i^{\min}}\right\}, \quad \underline{Q}_{i\,(:,\ell)}^{n+1} \leftarrow \theta\ \underline{Q}_{i\,(:,\ell)}^{n+1} \quad \forall \ell = 2, \ldots, M_C. \tag{5.46}$$

Step 3. Now that both density and pressure are positive $\forall \xi \in \mathbb{X}_{M_O}$, we repeat **Step 2** for the modified kurtosis. That is, we find the average modified kurtosis and the minimum modified kurtosis $\forall \xi \in \mathbb{X}_{M_O}$:

$$\overline{K}_i := \mathcal{C}_k\left(\underline{Q}_{i\,(1,:)}^{n+1}\right), \quad k_i^{\min} := \min_{\xi \in \mathbb{X}_{M_O}} \left\{\mathcal{C}_k\left(\underline{\Phi}\,(\xi)^T\ \underline{\underline{Q}}_i^{n+1}\right)\right\}, \tag{5.47}$$

where $\mathcal{C}_k\left(\underline{q}\right)$ is defined in (5.13). From here, we compute the corresponding damping parameter and rescale the higher-order coefficients:

$$\theta = \min\left\{1, \frac{\overline{K}_i - \varepsilon}{\overline{K}_i - k_i^{\min}}\right\}, \quad \underline{Q_{i\,(:,\ell)}^{n+1}} \leftarrow \theta\, \underline{Q_{i\,(:,\ell)}^{n+1}} \quad \forall \ell = 2, \ldots, M_{\mathrm{C}}. \tag{5.48}$$

In all the formulas presented above we use $\varepsilon = 10^{-14}$.

5.5 Limiter IV: Unphysical Oscillation Limiter

The previously described limiters guarantee positivity for ρ, p, and k, but there still may be unphysical oscillations near shocks, rarefactions, or large gradients. We augment the method with one more limiter to eliminate these oscillations: a variant of the strategy developed in Moe et al. [42]. This limiter is applied once per time step and can remove unphysical oscillations without overly diffusing the numerical solution. We apply the following procedure.

Step 1. Loop over each element $\mathcal{T}_i$ and compute the minimum and maximum values of all of the following variables: $w^\ell \in \{\rho, u, p, h, r\}$:

$$w_{M_i}^\ell := \max_{\xi \in \mathbb{X}_{M_{\mathrm{O}}}} \left\{ w^\ell \left(\underline{q^{\Delta x}}(\xi) \right) \Big|_{\mathcal{T}_i} \right\}, \quad w_{m_i}^\ell := \min_{\xi \in \mathbb{X}_{M_{\mathrm{O}}}} \left\{ w^\ell \left(\underline{q^{\Delta x}}(\xi) \right) \Big|_{\mathcal{T}_i} \right\}, \tag{5.49}$$

for all $\ell = 1, 2, 3, 4, 5$. Here $\mathbb{X}_{M_{\mathrm{O}}}$ is taken to be the M_{O} roots of the $M_{\mathrm{O}}{}^{\mathrm{th}}$ Legendre polynomial (i.e., Gauss–Legendre points) plus the element ends points (see Eq. 5.17).
Step 2. Compute upper and lower bounds over all neighborhoods, $N_{\mathcal{T}_i} := \{\mathcal{T}_{i-1}, \mathcal{T}_i, \mathcal{T}_{i+1}\}$:

$$\begin{aligned} M_i^\ell &= \max\left\{ \overline{w}_i^\ell + \mathcal{A}_0 h^{1.5}, \max_{j \in N_{\mathcal{T}_i}} \left\{ w_{M_j}^\ell \right\} \right\}, \\ m_i^\ell &= \min\left\{ \overline{w}_i^\ell - \mathcal{A}_0 h^{1.5}, \min_{j \in N_{\mathcal{T}_i}} \left\{ w_{m_j}^\ell \right\} \right\}, \end{aligned} \tag{5.50}$$

where $\overline{w}_i^\ell$ are the element averages of each variable, and $\mathcal{A}_0 h^{1.5}$ is used to offset these averages to recover high-order accuracy for smooth solutions in the limit $h \to 0$ (see Moe et al. [42] for more details).
Step 3. On each element $\mathcal{T}_i$, compute the largest damping parameters between [0, 1] that guarantee that the high-order solution in $\mathcal{T}_i$ does not violate the maximum and minimum bounds defined by (5.50):

$$\theta = \min\left\{1, \ \mu \cdot \min_\ell \left\{ \frac{M_i^\ell - \bar{w}_i^\ell}{w_{M_i}^\ell - \bar{w}_i^\ell} \right\}, \ \mu \cdot \min_\ell \left\{ \frac{m_i^\ell - \bar{w}_i^\ell}{w_{m_i}^\ell - \bar{w}_i^\ell} \right\} \right\}, \tag{5.51}$$

where the factor $\mu = 10/11$ is introduced to slightly increase the aggressiveness of the limiter (again, see Moe et al. [42] for more details).
Step 4. On each element $\mathcal{T}_i$, limit the conserved variables:

$$\underline{Q_{i\,(\ell,:)}^{n+1}} \leftarrow \theta\, \underline{Q_{i\,(\ell,:)}^{n+1}} \quad \forall \ell = 2, \ldots, M_{\mathrm{C}}. \tag{5.52}$$

6 Collisionless HyQMOM Numerical Examples

In this section, we apply the proposed scheme and the corresponding limiters to several test cases. In Sect. 6.1, we verify the claimed orders of accuracy on a smooth exact solution. In Sects. 6.2 and 6.3 we apply the scheme to *shock tube* initial data. These results demonstrate the ability of the non-oscillatory limiter to control unphysical oscillations. Finally, in Sect. 6.4, we fully validate the positivity limiters by applying the scheme to piecewise constant initial data that lead to the formation of a vacuum. This example demonstrates the ability of the positivity limiters to prevent negative states in density, pressure, and modified kurtosis, both on the element average and on the solution values internal to the element. In all the cases presented in Sects. 6.2, 6.3, and 6.4, we compare the high-order scheme against a highly-resolved first-order Rusanov scheme that is guaranteed to be positivity-preserving without the need for any limiters.

6.1 Smooth Solution Convergence Test

Consider the following exact solution to the 1D HyQMOM system (3.17)–(3.18) with periodic boundary conditions on $x \in [-1, 1]$:

$$\begin{aligned} &\rho(t,x) = 2 + \sin(2\pi(x-t)), \\ u(t,x) = 1, \quad &p(t,x) = 2, \quad h(t,x) = 4, \quad k(t,x) = 8 - 4\left[\rho(t,x)\right]^{-1}. \end{aligned} \tag{6.1}$$

The numerical solution is computed with grid resolutions of

$$M_{\text{elem}} = 10 \times 2^{\ell}, \quad \text{for} \quad \ell = 0, 1, 2, 3, 4, 5, \tag{6.2}$$

up to a final time of $t = 1$. We verify the order of accuracy for the schemes with orders of accuracy $M_{\text{O}} = 2, 3, 4$.

The errors we report are based on the following error measure:

$$\sum_{\ell=1}^{M_{\text{eqn}}} \frac{\left\| f_\ell - g_\ell \right\|_{L^2[-1,1]}}{\left\| g_\ell \right\|_{L^2[-1,1]}} = \sum_{\ell=1}^{M_{\text{eqn}}} \sqrt{\frac{\int_{-1}^{1} \left| f_\ell(x) - g_\ell(x) \right|^2 dx}{\int_{-1}^{1} \left| g_\ell(x) \right|^2 dx}}, \tag{6.3}$$

where $\underline{f}(x) : [-1, 1] \mapsto \mathbb{R}^{M_{\text{eqn}}}$ is the approximate solution and $\underline{g}(x) : [-1, 1] \mapsto \mathbb{R}^{M_{\text{eqn}}}$ is the exact solution. In practice, however, we replace the exact solution with a piecewise Legendre polynomial approximation of degree $M_{\text{O}} + 1$ on the computational mesh. Repeated use of the orthonormality of the Legendre basis functions yields the following (approximate) relative error on a mesh with N elements and a numerical method of order M_{O}:

$$e_N := \sum_{\ell=1}^{M_{\text{eqn}}} \sqrt{\frac{\sum_{i=1}^{N} \left\{ \sum_{j=1}^{M_C} \left(Q_{i(j,\ell)} - Q^{\star}_{i(j,\ell)} \right)^2 + \left(Q^{\star}_{i(M_C+1,\ell)} \right)^2 \right\}}{\sum_{i=1}^{N} \sum_{j=1}^{M_C+1} \left(Q^{\star}_{i(j,\ell)} \right)^2}}, \tag{6.4}$$

where Q and $Q^{\star}$ are the Legendre coefficients of the numerical and exact solutions at the final

time, respectively. The exact solution coefficients are computed using Gaussian quadrature with 20 quadrature points per element:

$$\underline{Q^{\star}_{i(k,:)}} := \frac{1}{2}\sum_{a=1}^{20} \omega_a^{\star}\, \phi_k\left(\mu_a^{\star}\right) \underline{q}^{\star}\left(t=1,\; x_i + \frac{\Delta x}{2}\mu_a^{\star}\right), \tag{6.5}$$

where $\omega_a^{\star}$ and $\mu_a^{\star}$ for $a = 1, \ldots, 20$ are the weights and abscissas of the 20 point quadrature rule, and $\underline{q}^{\star}$ is the exact solution. Gaussian quadrature rules have been tabulated in many books and websites; we obtained our data from [33].

The errors as defined by (6.4), as well as the base-2 logarithms of the ratio of consecutive errors,

$$\log_2\left(\frac{e_{N/2}}{e_N}\right) \approx \log_2\left(\frac{(N/2)^{-M_{\mathrm{O}}}}{N^{-M_{\mathrm{O}}}}\right) = \log_2\left(2^{M_{\mathrm{O}}}\right) = M_{\mathrm{O}}, \tag{6.6}$$

are shown in Tables 2(all limiters are turned off) and 3(all limiters are turned on). For the simulations that result in Table 3, none of the three positivity limiters (i.e., Limiters I, II, and III) are active because the solution is far away from positivity violations. In Table 3, the values affected by Limiter IV are highlighted in red. Note that at low resolutions, Limiter IV is active, and the results in Tables 2 and 3 differ slightly, but that at higher resolutions, the effect of the limiter disappears. Note that for all the simulations with limiters turned on, we used the value of $\mathcal{A}_0 = 5$ in formula Eq. 5.50.

Table 2 Section 6.1: smooth solution convergence test with limiters turned off

N	$M_{\mathrm{O}} = 2$	Eq. (6.6)	$M_{\mathrm{O}} = 3$	Eq. (6.6)	$M_{\mathrm{O}} = 4$	Eq. (6.6)
10	1.143e−01	–	1.171e−02	–	4.924e−03	–
20	2.005e−02	2.511	2.260e−03	2.374	4.617e−04	3.415
40	3.759e−03	2.415	4.032e−04	2.487	5.337e−06	6.435
80	8.802e−04	2.095	6.077e−05	2.730	1.962e−07	4.765
160	2.192e−04	2.006	8.127e−06	2.903	1.203e−08	4.028
320	5.485e−05	1.999	1.040e−06	2.966	7.500e−10	4.004

Relative L^2 errors for the HyQMOM equations with variable density, constant fluid velocity, pressure, heat flux, and fourth primitive moment, and periodic boundary conditions. In these simulations all four limiters were turned off

Table 3 Section 6.1: smooth solution convergence test with limiters turned on

N	$M_{\mathrm{O}} = 2$	Eq. (6.6)	$M_{\mathrm{O}} = 3$	Eq. (6.6)	$M_{\mathrm{O}} = 4$	Eq. (6.6)
10	3.154e−01	–	5.360e−02	–	4.924e−03	–
20	4.887e−02	2.690	2.260e−03	4.568	4.617e−04	3.415
40	3.759e−03	3.700	4.032e−04	2.487	5.337e−06	6.435
80	8.802e−04	2.095	6.077e−05	2.730	1.962e−07	4.765
160	2.192e−04	2.006	8.127e−06	2.903	1.203e−08	4.028
320	5.485e−05	1.999	1.040e−06	2.966	7.500e−10	4.004

Relative L^2 errors for the one-dimensional HyQMOM equations with variable density, constant fluid velocity, pressure, heat flux, and fourth primitive moment, and periodic boundary conditions. In these simulations, all four limiters were turned on

6.2 Shock Tube Problem #1

Consider the Riemann problem for (3.17)–(3.19) with the following initial data at $t = 0$:

$$\big(\rho, u, p, h, k\big)(t = 0, x) = \begin{cases} \big(1.5,\ -0.5,\ 1.5,\ 1.0,\ 2.3\bar{3}\big) & x < 0, \\ \big(1.0,\ -0.5,\ 1.0,\ 0.5,\ 1.75\big) & x > 0, \end{cases} \tag{6.7}$$

on $x \in [-1.2, 1.2]$ with extrapolation boundary conditions.

Shown in Fig. 2 are results from a simulation run with two distinct methods: (1) the $M_{\mathrm{O}} = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_{\mathrm{O}} = 4$

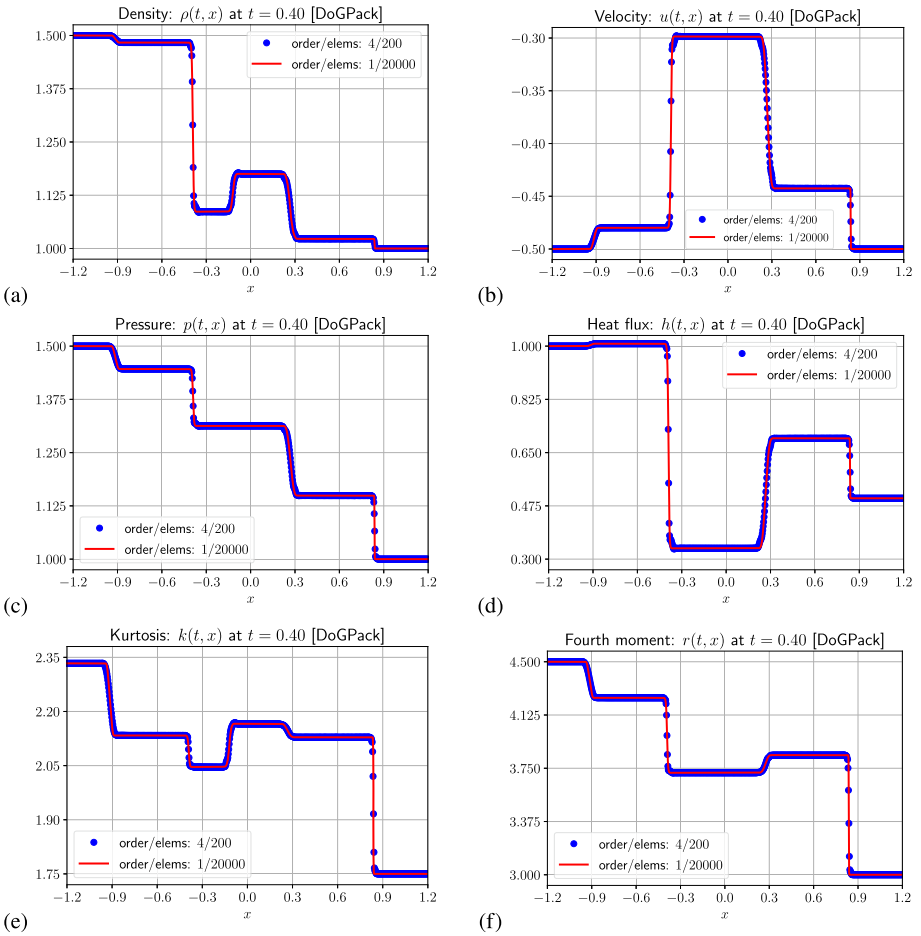

Fig. 2 (Section 6.2: shock tube problem #1) Numerical solution of shock tube problem #1 on $x \in [-1.2, 1.2]$ with initial conditions given by (6.7). Shown are the results from a simulation run with two distinct methods: (1) the $M_{\mathrm{O}} = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t, x)$, **b** macroscopic velocity: $u(t, x)$, **c** pressure: $p(t, x)$, **d** heat flux: $h(t, x)$, **e** modified kurtosis: $k(t, x)$, and **f** primitive fourth-moment: $r(t, x)$ (color figure online)

scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: (a) density: $\rho(t,x)$, (b) macroscopic velocity: $u(t,x)$, (c) pressure: $p(t,x)$, (d) heat flux: $h(t,x)$, (e) modified kurtosis: $k(t,x)$, and (f) primitive fourth-moment: $r(t,x)$. Note that we used the value of $\mathcal{A}_0 = 5$ in formula Eq. 5.50. These results clearly demonstrate the non-oscillatory limiters' ability to adequately control unphysical oscillations and produce accurate solutions.

6.3 Shock Tube Problem #2

Consider the Riemann problem for (3.17)–(3.19) with the following initial data at $t = 0$:

$$\big(\rho, u, p, h, k\big)(t=0,x) = \begin{cases} \big(1.0,\ -0.7,\ 1.5,\ 1.5,\ 1.75\big) & x < 0, \\ \big(0.5,\ -0.9,\ 1.0,\ 1.0,\ 1.0\big) & x > 0, \end{cases} \tag{6.8}$$

on $x \in [-1.2, 1.2]$ with extrapolation boundary conditions.

Shown in Fig. 3 are results from a simulation run with two distinct methods: (1) the $M_{\mathrm{O}} = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points-per-element in order to show the intra-element solution structure. The panels show the primitive variables: (a) density: $\rho(t,x)$, (b) macroscopic velocity: $u(t,x)$, (c) pressure: $p(t,x)$, (d) heat flux: $h(t,x)$, (e) modified kurtosis: $k(t,x)$, and (f) primitive fourth-moment: $r(t,x)$. Note that we used the value of $\mathcal{A}_0 = 5$ in formula Eq. 5.50.

Again, just as in the previous example, these results demonstrate the ability of the non-oscillatory limiters to adequately control unphysical oscillations and produce accurate solutions.

6.4 Double Rarefaction Vacuum Problem

In the final example, we solve a vacuum problem where the right and left initial velocities are large and opposite, creating a vacuum state in the center of the solution domain. The initial states are

$$\big(\rho, u, p, h, k\big)(t=0,x) = \begin{cases} \big(1.0,\ -2.0,\ 1.0,\ 0.0,\ 2.0\big) & x < 0, \\ \big(1.0,\ +2.0,\ 1.0,\ 0.0,\ 2.0\big) & x > 0. \end{cases} \tag{6.9}$$

The computational domain is $x \in [-1.2, 1.2]$ with extrapolation boundary conditions.

Shown in Fig. 4 are results from a simulation run with two distinct methods: (1) the $M_{\mathrm{O}} = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: (a) density: $\rho(t,x)$, (b) macroscopic velocity: $u(t,x)$, (c) pressure: $p(t,x)$, (d) heat flux: $h(t,x)$, (e) modified kurtosis: $k(t,x)$, and (f) primitive fourth-moment: $r(t,x)$. Note that we used the value of $\mathcal{A}_0 = 5$ in formula Eq. 5.50.

We comment on two important findings from this simulation. First, this example demonstrates the ability of the positivity limiters to prevent negative states in density, pressure, and modified kurtosis, both on the element average and the solution values internal to the element. In this simulation, all three variables, ρ, p, and k, become very small, but all stay

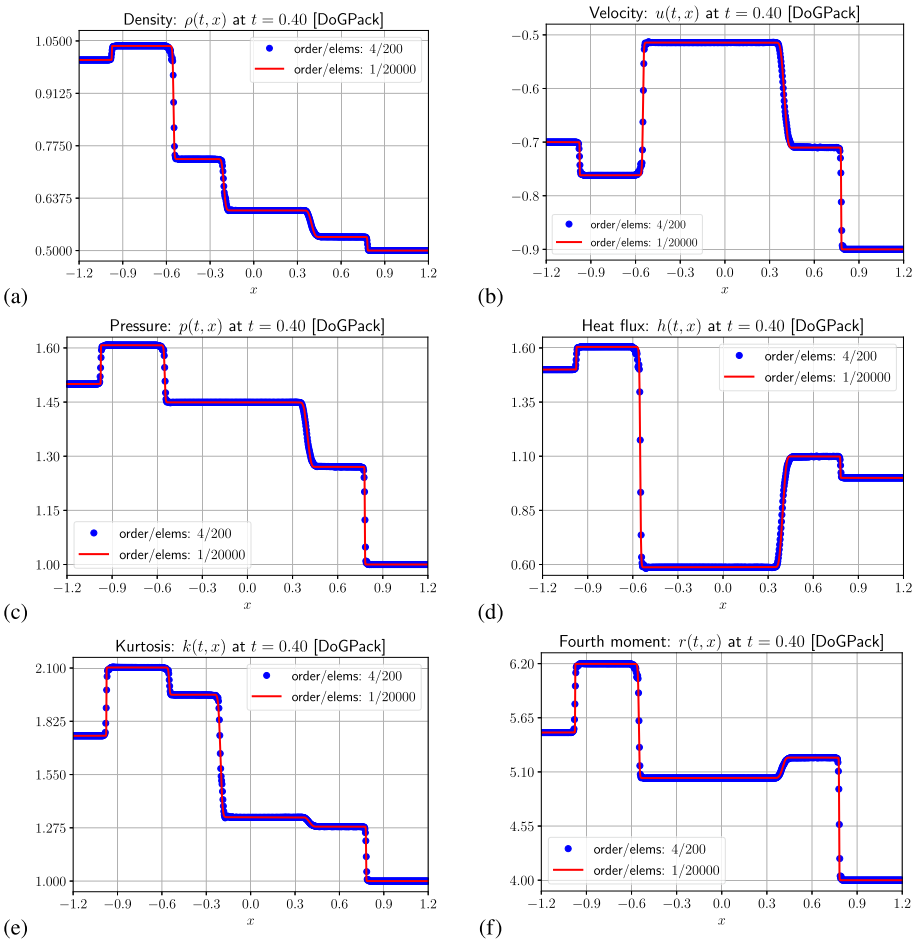

Fig. 3 (Section 6.3: shock tube problem #2) Numerical solution of shock tube problem #2 on $x \in [-1.2, 1.2]$ with initial conditions given by (6.8). Shown are the results from a simulation run with two distinct methods: (1) the $M_O = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_O = 4$ scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t, x)$, **b** macroscopic velocity: $u(t, x)$, **c** pressure: $p(t, x)$, **d** heat flux: $h(t, x)$, **e** modified kurtosis: $k(t, x)$, and **f** primitive fourth-moment: $r(t, x)$ (color figure online)

strictly above zero. Because all three remain strictly positive, the moments remain realizable, and the numerical simulation remains nonlinear stable. Second, while the simulation results from the $M_O = 4$ scheme with 200 elements do show some differences in the vacuum region with the highly resolved Rusanov solution, especially in the density plot shown in Fig. 4a, the solution remains qualitatively correct. We can investigate this further by increasing the grid resolution; in Fig. 5we show the density plots at different grid resolutions: (a) $N = 200$, (b) $N = 400$, (c) $N = 800$, and (b) $N = 1600$. These results verify that the differences between the $M_O = 4$ scheme and the highly resolved Rusanov scheme disappear at higher resolutions.

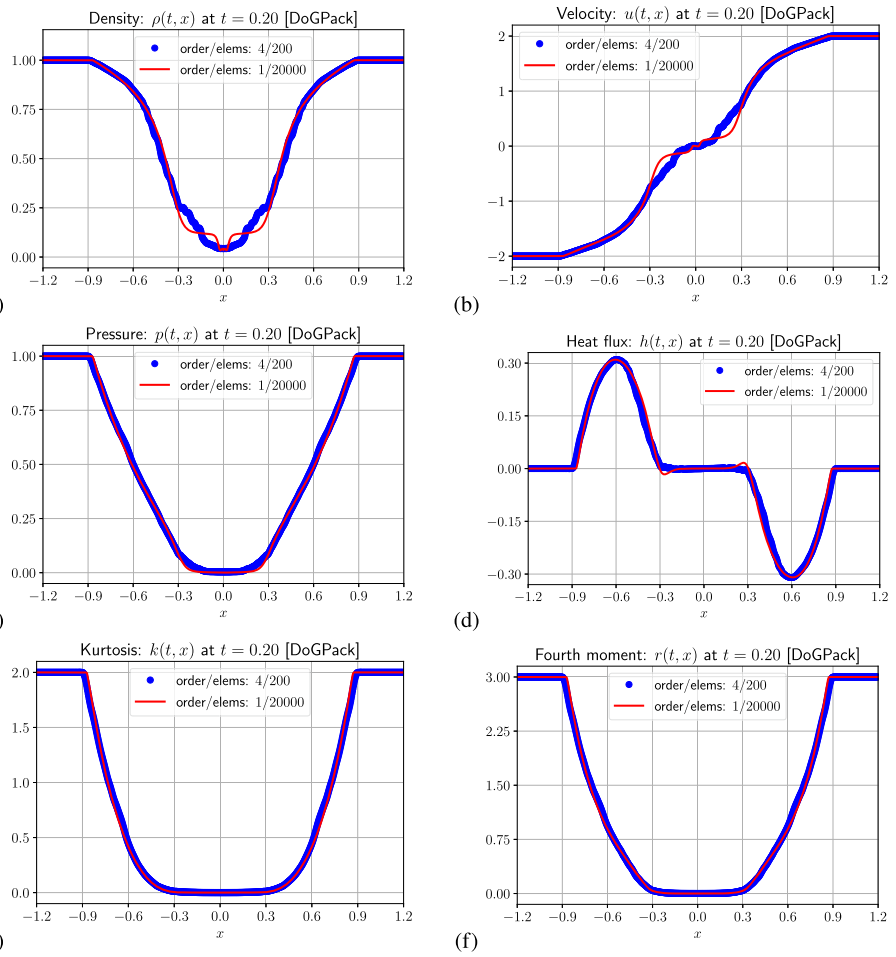

Fig. 4 (Section 6.4: double rarefaction vacuum problem) Numerical solution of the double rarefaction vacuum problem on $x \in [-1.2, 1.2]$ with initial conditions given by (6.9). Shown are the results from a simulation run with two distinct methods: (1) the $M_{\mathrm{O}} = 4$ scheme with 200 elements and full limiters (shown as blue dots), and (2) the first-order Rusanov scheme with 20,000 elements (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t,x)$, **b** macroscopic velocity: $u(t,x)$, **c** pressure: $p(t,x)$, **d** heat flux: $h(t,x)$, **e** modified kurtosis: $k(t,x)$, and **f** primitive fourth-moment: $r(t,x)$ (color figure online)

7 Extension to HyQMOM-BGK

Up to this point, we have only considered the HyQMOM approximation applied to the Vlasov model Eq. 2.1; this allowed us to study the mathematical structure of HyQMOM and to develop accurate high-order methods and limiters. On the other hand, the practicality of HyQMOM is not for solving collisionless kinetic models since, in this regime, it would be far better to directly solve the Vlasov equation with Lagrangian or semi-Lagrangian approaches. Instead, the true benefit of the HyQMOM approximation is in the approximation of kinetic systems near thermodynamic equilibrium—a regime we study in this section.

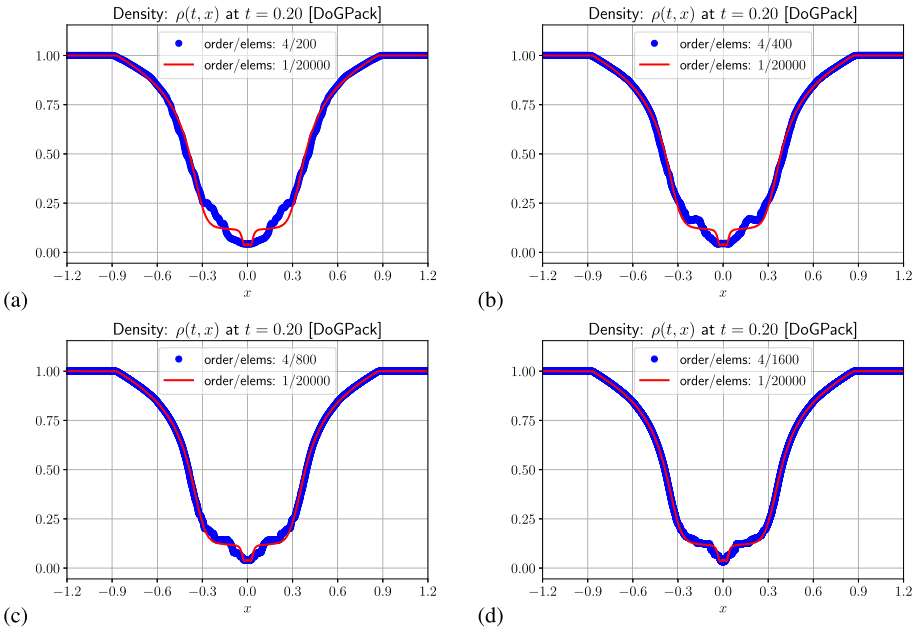

Fig. 5 (Section 6.4: double rarefaction vacuum problem) Numerical solution of the double rarefaction vacuum problem on $x \in [-1.2, 1.2]$ with initial conditions given by (6.9). Shown are the densities at various grid resolutions: **a** $N = 200$, **b** $N = 400$, **c** $N = 800$, and **d** $N = 1600$. In each panel, we compare the $M_{\mathrm{O}} = 4$ scheme with full limiters (shown as blue dots) with the first-order Rusanov scheme with 20,000 elements (shown as a solid red line) (color figure online)

In this section, we extend the previously developed numerical method to HyQMOM with a BGK collision operator. Importantly, we develop this extension so that the resulting HyQMOM solver adheres to the following two key design parameters:

1. The method should remain high-order accurate irrespective of the Knudsen number: $\varepsilon > 0$.
2. For fixed mesh parameters (i.e., fixed Δx and Δt), the method should remain stable in the singular limit: $\varepsilon \to 0^+$. This property is often referred to as the *asymptotic-preserving* (AP) property, and a variety of schemes with this property can be found in the literature (e.g., see [3, 8, 15, 23, 27–30, 46, 60]).

The specific approach we detail in this section is novel and directly relies on the prediction and correction format of the method developed in Sect. 4.

7.1 1D1V Boltzmann–BGK Equation

Consider the 1D1V Boltzmann–BGK equation [4]:

$$f_{,t} + v f_{,x} = \frac{1}{\varepsilon} (\mathcal{M} - f) , \tag{7.1}$$

where $\varepsilon > 0$ is the Knudsen number, which is a non-dimensional ratio of the particle mean-free path to a characteristic length scale, and $\mathcal{M}(t, x, v) : \mathbb{R}_{\geq 0} \times \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}_{\geq 0}$ is the Maxwell–Boltzmann distribution:

$$\mathcal{M}(t,x,v) := \frac{\rho}{\sqrt{2\pi T}} e^{-\frac{(v-u)^2}{2T}}. \tag{7.2}$$

In this expression, ρ is density, p is pressure, and $T = p/\rho$ is temperature (e.g., see Definitions (2.6)). For $\varepsilon \gg 1$, and for a fixed t and x, the collision operator is weak, and the solution behaves similarly to Vlasov equation (2.1). For $\varepsilon \ll 1$, and for a fixed t and x, the BGK collision operator forces f towards the Maxwell–Boltzmann distribution (i.e., thermodynamic equilibrium):

$$f(t,x,v) \to \mathcal{M}(t,x,v) + \mathcal{O}(\varepsilon). \tag{7.3}$$

7.2 HyQMOM-BGK and the Asymptotic-Preserving Property

Relevant in this work are the first five moments of (7.1) with the HyQMOM moment-closure (3.16):

$$\underline{q}_{,t} + \underline{f}\left(\underline{q}\right)_{,x} = \frac{1}{\varepsilon}\underline{S^{\text{cons}}}\left(\underline{q}\right), \qquad \underline{\alpha}_{,t} + \underline{\underline{B}}\left(\underline{\alpha}\right)\underline{\alpha}_{,x} = \frac{1}{\varepsilon}\underline{S^{\text{prim}}}\left(\underline{\alpha}\right), \tag{7.4}$$

where only the fourth and fifth components of the source terms are nonzero:

$$S_4^{\text{cons}} = S_4^{\text{prim}} = -h, \quad S_5^{\text{cons}} = -k + \frac{2p^2}{\rho} - \frac{4puh + h^2}{p}, \quad S_5^{\text{prim}} = -k + \frac{2p^2}{\rho} + \frac{h^2}{p}. \tag{7.5}$$

In Eq. 7.4 we are using definitions (3.16), (3.18), and (3.19). For $\varepsilon \ll 1$, and for a fixed t and x, the BGK collision operator forces the heat flux, h, and modified kurtosis, k, towards their Maxwell–Boltzmann values:

$$h(t,x) = 0 + \mathcal{O}(\varepsilon) \qquad \text{and} \qquad k(t,x) = \frac{2p(t,x)^2}{\rho(t,x)} + \mathcal{O}(\varepsilon). \tag{7.6}$$

In particular, in the $\varepsilon \to 0^+$ limit, solutions of the HyQMOM-BGK system converge to solutions of the 1D compressible Euler equations at a convergence rate of $\mathcal{O}(\varepsilon)$:

$$\begin{bmatrix} \rho \\ \rho u \\ \rho u^2 + p \end{bmatrix}_{,t} + \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u^3 + 3pu \end{bmatrix}_{,x} = \underline{0}, \tag{7.7}$$

where $h \equiv 0$ and $k \equiv \frac{2p}{\rho}$. Furthermore, by including the next order term in the Chapman–Enskog expansion, one can show that solutions to HyQMOM-BGK converge to solutions of the 1D Navier-Stokes equations at a convergence rate of $\mathcal{O}(\varepsilon^2)$ [2]:

$$\begin{bmatrix} \rho \\ \rho u \\ \rho u^2 + p \end{bmatrix}_{,t} + \begin{bmatrix} \rho u \\ \rho u^2 + p \\ \rho u^3 + 3pu \end{bmatrix}_{,x} = \begin{bmatrix} 0 \\ 0 \\ \frac{3}{2}\varepsilon p T_{,x} \end{bmatrix}_{,x}, \tag{7.8}$$

where $T = \frac{p}{\rho}$ and $q \equiv -\frac{3}{2}\varepsilon p T_{,x}$.

Definition 7.1 *(Asymptotic-preserving (AP) property* [29]*)* Let $\underline{q^{(\Delta t,\Delta x)}}(t,x;\varepsilon)$ be an approximation to the exact solution of Eqs. 7.4, 7.5 as computed by a numerical method with mesh parameters Δt, $\Delta x > 0$. We assume that for a fixed $\varepsilon > 0$, this method is convergent to the exact solution of Eqs. 7.4, 7.5. This numerical method is said to be *asymptotic-preserving*

(AP) provided that the vanishing mesh parameter limit, $\Delta t, \Delta x \to 0^+$, and the vanishing Knudsen number limit, $\varepsilon \to 0^+$, commute:

$$\lim_{\varepsilon \to 0^+} \left[\lim_{\Delta t, \Delta x \to 0^+} \underline{q^{(\Delta t, \Delta x)}}(t, x; \varepsilon) \right] = \lim_{\Delta t, \Delta x \to 0^+} \left[\lim_{\varepsilon \to 0^+} \underline{q^{(\Delta t, \Delta x)}}(t, x; \varepsilon) \right].$$

Practically, this means that an AP scheme remains stable and accurate for a fixed mesh, Δt, and Δx, for all $\varepsilon > 0$, including in the limit $\varepsilon \to 0^+$.

The goal of this section is to develop an extension of the Lax–Wendroff DG scheme developed in Sects. 4 and 5 for the HyQMOM-BGK system Eqs. 7.4, 7.5 that behaves, on the discrete level, as a consistent and stable numerical method for Eqs. 7.7 and 7.8 in the singular limit $\varepsilon \to 0^+$. The key innovation in this work is that we make use of the prediction-correction formulation of Lax–Wendroff DG to incorporate the collision operator.

7.3 Prediction

To describe the HyQMOM-BGK prediction step, it is first useful to define the following matrices that allow us to map Legendre coefficients to nodal space-time Gauss–Legendre quadrature points and back again:

$$C^1_{(a,b)} := \Psi_b(\tau_a, \xi_a), \qquad \underline{\underline{C^1}} \in \mathbb{R}^{M_{\mathrm{O}}^2 \times M_{\mathrm{P}}}, \tag{7.9}$$

$$C^2_{(b,a)} := \frac{\omega_a}{4} \Psi_b(\tau_a, \xi_a), \qquad \underline{\underline{C^2}} \in \mathbb{R}^{M_{\mathrm{P}} \times M_{\mathrm{O}}^2}, \tag{7.10}$$

where ω_a and (τ_a, ξ_a) for $a = 1, \ldots, M_{\mathrm{O}}^2$ are tensor product Gauss–Legendre weights and abscissas. These two matrices satisfy

$$\underline{\underline{C^2}}\,\underline{\underline{C^1}} = \underline{\underline{\mathbb{I}}} \in \mathbb{R}^{M_{\mathrm{P}} \times M_{\mathrm{P}}}. \tag{7.11}$$

HyQMOM consists of five evolution equations, and the first three are unaffected by the collision operator; therefore, the update inside the Picard iteration for the three collision invariants (i.e., density, macroscopic velocity, and pressure) remains the same as in the collisionless case: Eq. 4.25. On the other hand, the update for the heat flux, h, has a non-zero BGK contribution; however, the BGK term is linear in the heat flux (see Eqs. 7.4 and 7.5), which allows for simple treatment. The strategy we pursue here is to include the BGK source term in the implicit portion of the Picard update to remain uniformly stable in $\varepsilon > 0$. After simple algebra, we arrive at the following update for the heat flux, h:

$$\begin{aligned} \left(\frac{\Delta t}{2} \underline{\underline{I}} + \varepsilon \underline{\underline{L}} \right) \underline{W^{n+\frac{1}{2}(j)}_{i\,(:,4)}} = {} & \frac{\varepsilon}{4} \sum_{a=1}^{M_{\mathrm{O}}} \sum_{b=1}^{M_{\mathrm{O}}} \omega_a \omega_b \underline{\Psi}(\mu_b, \mu_a) \Theta_{(a,b,4)} \\ & + \frac{\varepsilon}{4} \sum_{b=1}^{M_{\mathrm{O}}} \omega_b \underline{\Psi}(-1, \xi_b) \underline{\Phi}(\xi_b)^T \underline{A^n_{i\,(:,4)}}, \end{aligned} \tag{7.12}$$

where we are using the short-hand:

$$\Theta_{(a,b,m)} := \Theta_m \left(\underline{\Psi}(\mu_b, \mu_a)^T \underline{\underline{W^{n+\frac{1}{2}(j-1)}_i}} \right). \tag{7.13}$$

The update for the final primitive variable, k (modified kurtosis), requires more work. The source term shown in Eqs. 7.4 and 7.5 is linear in k, but it also includes nonlinear terms from three previously updated quantities: ρ, p, and h. To construct these nonlinear quantities, we

first apply the mapping from Legendre to nodal values via Eq. 7.9, and then evaluate the nonlinear portion of the source:

$$\begin{aligned} m = 1,3,4: \quad & \underline{\widehat{W}_{(:,m)}} = \underline{\underline{C}}^1 \, \underline{W^{n+\frac{1}{2}(j)}_{i(:,m)}} \in \mathbb{R}^{M_{\mathrm{O}}^2}, \\ a = 1,\ldots,M_{\mathrm{O}}^2: \quad & \{\rho_a, p_a, h_a\} = \widehat{W}_{(a,\{1,3,4\})}, \quad \widehat{\mathcal{S}}^{(j)}_{ia} = \frac{2p_a^2}{\rho_a} + \frac{h_a^2}{p_a}. \end{aligned} \tag{7.14}$$

From here, the update for the modified kurtosis inside the Picard iteration looks very similar to the update for heat flux (see Eq. 7.12), but with the additional nonlinear terms computed from Eq. 7.14, which now need to be mapped back to Legendre coefficients via Eq. 7.10. After some simple algebra, the update takes the following form:

$$\begin{aligned} \left(\frac{\Delta t}{2} \underline{\underline{I}} + \varepsilon \, \underline{\underline{L}} \right) \underline{W^{n+\frac{1}{2}(j)}_{i\,(:,5)}} = {} & \frac{\varepsilon}{4} \sum_{a=1}^{M_{\mathrm{O}}} \sum_{b=1}^{M_{\mathrm{O}}} \omega_a \, \omega_b \, \underline{\Psi} \left(\mu_b, \, \mu_a \right) \, \Theta_{(a,b,5)} \\ & + \frac{\varepsilon}{4} \sum_{b=1}^{M_{\mathrm{O}}} \omega_b \, \underline{\Psi} \left(-1, \xi_b \right) \, \underline{\Phi} \left(\xi_b \right)^T \, \underline{A^n_{i\,(:,5)}} + \frac{\Delta t}{2} \, \underline{\underline{C}}^2 \, \underline{\widehat{\mathcal{S}}^{(j)}_i}. \end{aligned} \tag{7.15}$$

7.4 Post-prediction BGK Source Evaluation

Once the Picard iterations are complete and all five primitive variables have been predicted, there is one final computation that must be completed to prepare us for the correction step: we need to evaluate and project the BGK source term, $\underline{S^{\mathrm{cons}}}$, from Eq. 7.5. This is done similar to Eq. 7.14 by first mapping the predicted solution from Legendre to nodal values via Eq. 7.9, then evaluating the source components at nodal values, and finally mapping back to Legendre coefficients via Eq. 7.10:

$$\begin{aligned} m = 1,\ldots,5: \quad & \underline{\widehat{W}_{(:,m)}} = \underline{\underline{C}}^1 \, \underline{W^{n+\frac{1}{2}}_{i(:,m)}} \in \mathbb{R}^{M_{\mathrm{O}}^2}, \\ a = 1,\ldots,M_{\mathrm{O}}^2: \quad & \{\rho_a, \, u_a, \, p_a, \, h_a, \, k_a\} = \widehat{W}_{(a,1:5)}, \\ & \widehat{\Delta\mathcal{M}}_{(a,4)} = -h_a, \quad \widehat{\Delta\mathcal{M}}_{(a,5)} = -k_a + \frac{2p_a^2}{\rho_a} - \frac{4p_a u_a h_a + h_a^2}{p_a}, \\ m = 4,5: \quad & \underline{\Delta\mathcal{M}_{i(:,m)}} = \underline{\underline{C}}^2 \, \underline{\widehat{\Delta\mathcal{M}}_{(:,m)}} \in \mathbb{R}^{M_{\mathrm{P}}}. \end{aligned} \tag{7.16}$$

In the above expressions, we use the notation $\Delta\mathcal{M}$ to signify that these BGK source terms are, in fact, measuring the deviations in the heat flux, h, and the modified kurtosis, k, from their Maxwell–Boltzmann values [e.g., see (7.6)].

We choose to do the above BGK source evaluation and projection, Eq. 7.16, as a separate step rather than just as part of the correction update since we need to be extra careful in assuring that the final update is asymptotic-preserving. Indeed, we show in the next section how to obtain a fully asymptotic-preserving scheme.

7.5 Correction

Just as in the prediction step, we begin by defining matrices that allow us to map Legendre coefficients to nodal space Gauss–Legendre quadrature points and back again:

$$C^3_{(a,b)} := \Phi_b \left(\xi_a \right), \qquad \underline{\underline{C}}^3 \in \mathbb{R}^{M_{\mathrm{O}} \times M_{\mathrm{C}}}, \tag{7.17}$$

$$C^4_{(b,a)} := \frac{\omega_a}{2}\,\Phi_b\left(\xi_a\right), \qquad \underline{\underline{C}}^4 \in \mathbb{R}^{M_\mathrm{C}\times M_\mathrm{O}}, \tag{7.18}$$

where ω_a and ξ_a for $a = 1, \ldots, M_\mathrm{O}$ are Gauss–Legendre weights and abscissas. These two matrices satisfy

$$\underline{\underline{C}}^4\,\underline{\underline{C}}^3 = \underline{\underline{\mathbb{I}}} \in \mathbb{R}^{M_\mathrm{C}\times M_\mathrm{C}}. \tag{7.19}$$

As far as the correction step is concerned, the only difference between the collisionless update, as shown through Eqs. 4.28, 4.29, 4.30 and 4.31 and the BGK version is the additional BGK source integral needed in Eq. 4.28:

$$\underline{\underline{Q}}_i^{n+1} = \underline{\underline{Q}}_i^{n} + \ldots + \underbrace{\frac{\frac{\Delta t}{2}}{2\varepsilon}\iint_{-1}^{1} \underline{\Phi}\left[\underline{S}^{\mathrm{cons}}\right]^T d\tau\, d\xi}_{\text{BGK source}}. \tag{7.20}$$

To eventually achieve the asymptotic-preserving (AP) property, we introduce the following Legendre-in-space-Radau-in-time quadrature:

$$\iint_{-1}^{1} g\left(\tau,\xi\right)\, d\tau\, d\xi \approx \sum_{k=1}^{M_\mathrm{O}}\sum_{\ell=1}^{M_\mathrm{O}} \omega_k\,\omega_\ell^R\, g\left(\tau_\ell^R, \xi_k\right), \tag{7.21}$$

where for $a = 1, \ldots, M_\mathrm{O}$, (ω_a, ξ_a) are again 1D Gauss–Legendre weights/abscissas, while (ω_a^R, ξ_a^R) are 1D Gauss–Radau weights/abscissas. In particular, what we aim to do here is to handle the BGK source in Eq. 7.20 using a strategy that replaces the actual Legendre-in-space-Radau-in-time quadrature shown via Eq. 7.21, by a version where the function values at the $\tau = 1$ quadrature points are replaced by the unknown solution Q^{n+1}:

$$m = 4, 5: \quad \frac{1}{2\varepsilon}\iint_{-1}^{1} \underline{\Phi}\, S_m^{\mathrm{cons}}\, d\tau\, d\xi \approx \underbrace{\frac{1}{\varepsilon}\,\underline{\underline{R}}\,\underline{\Delta\mathcal{M}_{i(:,m)}}}_{\text{explicit}} + \underbrace{\frac{1}{r\varepsilon}\left(\underline{S_{(:,m)}} - \underline{Q^{n+1}_{i(:,m)}}\right)}_{\text{implicit}}, \tag{7.22}$$

where $\Delta\mathcal{M}$ is defined by Eq. 7.16, S are Maxwell–Boltzmann moments (the precise definition is provided below in Eq. 7.24), and

$$\underline{\underline{R}} = \frac{1}{2}\sum_{k=1}^{M_\mathrm{O}}\sum_{\ell=1}^{M_\mathrm{O}-1} \omega_k\,\omega_\ell^R\,\underline{\Phi}\left(\xi_k\right)\underline{\Psi}\left(\tau_\ell^R, \xi_k\right)^T \in \mathbb{R}^{M_\mathrm{C}\times M_\mathrm{P}}, \quad r = 1/\omega^R_{M_\mathrm{O}}. \tag{7.23}$$

This quadrature provides a strategy for implicitly handling the BGK collision term, which is critical for achieving the asymptotic-preserving (AP) property. We illustrate the modified Gauss–Radau quadrature strategy in Fig. 6.

The full correction update is detailed below. The first three moments are collision invariants and thus updated via Eqs. 4.28, 4.29, 4.30 and 4.31. From these updated moments, we compute the Maxwell–Boltzmann moments, S, that are required in Eq. 7.22:

$$\begin{aligned}
m = 1, 2, 3: \quad & \underline{\widehat{M}_{(:,m)}} = \underline{\underline{C}}^3\,\underline{Q^{n+1}_{i(:,m)}} \in \mathbb{R}^{M_\mathrm{O}},\\
a = 1, \ldots, M_\mathrm{O}: \quad & \{\rho,\, u,\, p\} = \left\{\widehat{M}_{(a,1)},\ \frac{\widehat{M}_{(a,2)}}{\widehat{M}_{(a,1)}},\ \widehat{M}_{(a,3)} - \frac{\widehat{M}^2_{(a,2)}}{\widehat{M}_{(a,1)}}\right\},\\
& \left\{\widehat{S}_{i(a,4)},\ \widehat{S}_{i(a,5)}\right\} = \left\{\rho u^3 + 3pu,\ \ \rho u^4 + 6pu^2 + \frac{3p^2}{\rho}\right\}.
\end{aligned} \tag{7.24}$$

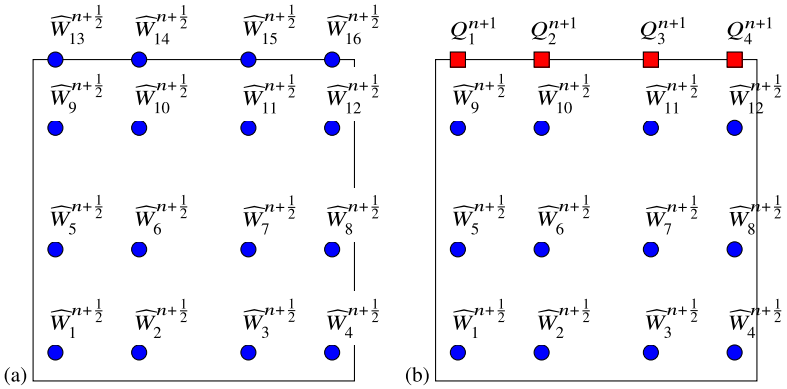

Fig. 6 Numerical quadrature on the canonical space-time element $(\tau, \xi) \in [-1, 1]^2$ using a tensor product between 1D Gauss–Legendre points in ξ and Gauss–Radau points in τ. **a** Shows the Legendre-in-space-Radau-in-time points in the case $M_{\mathrm{O}} = 4$ and the known solution values from the prediction step at those quadrature points. **b** Shows the same thing, but the function values at the $\tau = 1$ quadrature points are replaced by the unknown solution Q^{n+1}. This strategy of replacing the $\tau = 1$ quadrature point function values with the unknown solution provides a strategy for implicitly handling the BGK collision term (color figure online)

We then update the final two moments in a two-step process, where the first step is to apply a collisionless update:

$$m = 4, 5: \quad \underline{\widetilde{Q}^{n+1}_{i(:,m)}} = \underline{Q^{n}_{i(:,m)}} - \frac{\Delta t}{\Delta x}\left(\underline{\Phi}(1)\, \mathcal{F}^{n+\frac{1}{2}}_{i+\frac{1}{2}(m)} - \underline{\Phi}(-1)\, \mathcal{F}^{n+\frac{1}{2}}_{i-\frac{1}{2}(m)} \right) + \frac{\Delta t}{2\Delta x} \sum_{a=1}^{M_{\mathrm{O}}} \sum_{b=1}^{M_{\mathrm{O}}} \omega_a\, \omega_b\, \underline{\Phi}_{,\xi}\,(\mu_a)\, f_m\left(\underline{\underline{W^{n+\frac{1}{2}}_{i}}}\, \underline{\Psi}\,(\mu_b, \mu_a) \right), \tag{7.25}$$

followed by a collision step:

$$\underline{Q^{n+1}_{i(:,4:5)}} = \left(\frac{r\varepsilon}{\frac{\Delta t}{2} + r\varepsilon} \right) \underline{\widetilde{Q}^{n+1}_{i(:,4:5)}} + \left(\frac{\frac{\Delta t}{2}}{\frac{\Delta t}{2} + r\varepsilon} \right) \left(r\, \underline{\underline{R}}\, \underline{\Delta \mathcal{M}_{i(:,4:5)}} + \underline{\underline{C}}^4\, \underline{\widehat{S}_{i(:,4:5)}} \right). \tag{7.26}$$

Note 7.1 All the limiters described in Sect. 5 can still be applied to the HyQMOM-BGK solver described in this section.

7.6 Asymptotic-Preserving Condition

The advantage of the above-proposed scheme for the HyQMOM-BGK is that it remains high-order accurate uniformly in $\varepsilon > 0$ and is asymptotic-preserving in the $\varepsilon \to 0^+$ limit. The first claim is demonstrated via numerical examples in the next section; the second claim is easily demonstrated in this section.

Lemma 7.1 *The method LxW-DG method for HyQMOM-BGK described in Sects. 7.3, 7.4, and 7.5 is asymptotic-preserving in the $\varepsilon \to 0^+$ limit.*

Proof In the prediction step, the only updates directly affected by the Knudsen number, $\varepsilon > 0$, are the updates for the heat flux, h, and the modified kurtosis, k. Taking the $\varepsilon \to 0^+$

limit of both Eqs. 7.12 and 7.15 yields:

$$\underline{W_{i\,(:,4)}^{n+\frac{1}{2}(j)}} \to \underline{0} \quad \text{and} \quad \underline{W_{i\,(:,5)}^{n+\frac{1}{2}(j)}} \to \underline{\underline{C^2}}\ \underline{\widehat{\mathcal{S}}_i^{(j)}}. \tag{7.27}$$

This is precisely the desired effect: moments converge to their Maxwell–Boltzmann values.

In the correction step, the only update directly affected by the Knudsen number, $\varepsilon > 0$, is Eq. 7.26. Taking the $\varepsilon \to 0^+$ limit of this update yields:

$$\underline{\Delta\mathcal{M}_{i\,(:,4:5)}} \to \underline{0} \quad \text{and} \quad \underline{Q_{i\,(:,4:5)}^{n+1}} \to \underline{\underline{C^4}}\ \underline{\widehat{S}_{i\,(:,4:5)}}. \tag{7.28}$$

Again, this is precisely the desired effect: moments converge to their Maxwell–Boltzmann values. □

8 HyQMOM-BGK Numerical Examples

In this section, we apply the proposed HyQMOM-BGK scheme to several test cases. In Section 8.1 we verify the claimed orders of accuracy on a smooth manufactured solution with different Knudsen numbers. These tests also show the scheme's uniform accuracy and order of accuracy as a function of the Knudsen number. In Section 8.2 we apply the scheme to *shock tube* initial data with different Knudsen numbers. These results demonstrate the ability of the non-oscillatory limiter to control unphysical oscillations. Also shown by these results is the asymptotic-preserving (AP) property of the scheme for small $\varepsilon > 0$; in particular, we include the exact Riemann solution for the compressible Euler equations as a point of comparison.

8.1 Manufactured Solution Convergence Test

We consider the following manufactured solution:

$$\begin{aligned} \underline{\alpha^{\mathrm{ms}}}(t,x) &:= \left(\rho^{\mathrm{ms}},\ u^{\mathrm{ms}},\ p^{\mathrm{ms}},\ h^{\mathrm{ms}},\ k^{\mathrm{ms}}\right)(t,x), \\ \left(\rho^{\mathrm{ms}},\ p^{\mathrm{ms}},\ h^{\mathrm{ms}},\ k^{\mathrm{ms}}\right)(t,x) &:= \left(\rho_\varepsilon,\ p_\varepsilon,\ h_\varepsilon,\ k_\varepsilon\right)\sqrt{\pi}\left(2-\cos\left(2\pi\,(t-x)\right)\right), \\ u^{\mathrm{ms}}(t,x) &:= \frac{1-3\varepsilon}{4+8\varepsilon}, \end{aligned} \tag{8.1}$$

where

$$\begin{aligned} \rho_\varepsilon &= \frac{1+2\varepsilon}{2+2\varepsilon}, \quad p_\varepsilon = \frac{2+33\left(\varepsilon+\varepsilon^2\right)}{32(1+\varepsilon)(1+2\varepsilon)}, \quad h_\varepsilon = -\frac{125\varepsilon}{128\,(1+2\varepsilon)^2}, \\ r_\varepsilon &= \frac{12+\left(\varepsilon+\varepsilon^2\right)\left(1021+2017\left(\varepsilon+\varepsilon^2\right)\right)}{512(1+\varepsilon)(1+2\varepsilon)^3}, \quad k_\varepsilon = r_\varepsilon - \frac{p_\varepsilon^2}{\rho_\varepsilon} - \frac{h_\varepsilon^2}{p_\varepsilon}. \end{aligned} \tag{8.2}$$

Note that this solution is ε-dependent and well-defined for all $0 < \varepsilon < \infty$. Since this is

not an exact solution to HyQMOM-BGK, we need to augment Eq. 7.4 with an additional manufactured solution source term:

$$\underline{q}_{,t} + \underline{f}\left(\underline{q}\right)_{,x} = \frac{1}{\varepsilon}\underline{S^{\text{cons}}}\left(\underline{q}\right) + \underline{s_q}, \quad \underline{\alpha}_{,t} + \underline{\underline{B}}\left(\underline{\alpha}\right)\underline{\alpha}_{,x} = \frac{1}{\varepsilon}\underline{S^{\text{prim}}}\left(\underline{\alpha}\right) + \underline{q}_{,\underline{\alpha}}^{-1}\left(\underline{\alpha^{\text{ms}}}\right)\underline{s_q}, \tag{8.3}$$

where

$$\begin{aligned} \underline{s_q} &= \left[\pi^{3/2}\sin\left(2\pi(t-x)\right)\right]\underline{v_1} + \left[\pi^{1/2}\left(2-\cos\left(2\pi(t-x)\right)\right)\right]\underline{v_2}, \\ \underline{v_1} &= \left(A_1,\ A_2,\ A_3,\ A_4,\ A_7\right), \quad \underline{v_2} = \left(0,\ 0,\ 0,\ A_5,\ A_6\right). \end{aligned} \tag{8.4}$$

with

$$\begin{aligned} A_1 = \frac{3+11\varepsilon}{4(1+\varepsilon)}, \quad A_2 = \frac{1-33\varepsilon}{16(1+\varepsilon)}, \quad A_3 = \frac{5(1+33\varepsilon)}{64(1+\varepsilon)}, \quad A_4 = \frac{3-809\varepsilon}{256(1+\varepsilon)}, \\ A_5 = \frac{-125}{128(1+2\varepsilon)^2}, \quad A_6 = \frac{125(1+2\varepsilon-10\varepsilon^2)}{512(1+2\varepsilon)^3}, \\ A_7 = \frac{76+3620\varepsilon+521895\varepsilon^2+5285445\varepsilon^3+9544425\varepsilon^4+4794867\varepsilon^5}{1024(1+\varepsilon)(2+33\varepsilon(1+\varepsilon))^2}. \end{aligned} \tag{8.5}$$

Convergence tables for the $M_{\text{O}} = 4$ scheme are shown in Table 4. Importantly, we consider various values of the Knudsen number that span ten orders of magnitude: $\varepsilon = 10^4,\ 10^2,\ 10^0,\ 10^{-2},\ 10^{-4}$, and 10^{-6}, and in each case, we achieve optimal convergence. These results confirm the asymptotic-preserving (AP) property for small $\varepsilon > 0$.

Table 4 Section 8.1: HyQMOM-BGK manufactured solution problem)

N	$\varepsilon = 10^4$	$\log_2 \frac{e_{N/2}}{e_N}$	$\varepsilon = 10^2$	$\log_2 \frac{e_{N/2}}{e_N}$	$\varepsilon = 10^0$	$\log_2 \frac{e_{N/2}}{e_N}$
10	1.181e−03	–	1.193e−03	–	1.426e−03	–
20	5.809e−05	4.346	5.897e−05	4.339	6.321e−05	4.496
40	3.541e−06	4.036	3.515e−06	4.068	3.655e−06	4.112
80	2.212e−07	4.001	2.622e−07	3.745	2.601e−07	3.813
160	1.376e−08	4.007	1.622e−08	4.015	1.529e−08	4.088
320	8.592e−10	4.001	9.983e−10	4.022	8.986e−10	4.089
N	$\varepsilon = 10^{-2}$	$\log_2 \frac{e_{N/2}}{e_N}$	$\varepsilon = 10^{-4}$	$\log_2 \frac{e_{N/2}}{e_N}$	$\varepsilon = 10^{-6}$	$\log_2 \frac{e_{N/2}}{e_N}$
10	1.327e−03	–	1.644e−03	–	1.660e−03	–
20	6.608e−05	4.328	6.826e−05	4.590	6.861e−05	4.597
40	4.040e−06	4.032	4.222e−06	4.015	4.148e−06	4.048
80	2.537e−07	3.993	2.575e−07	4.035	2.557e−07	4.020
160	1.572e−08	4.012	1.622e−08	3.989	1.601e−08	3.998
320	9.929e−10	3.985	1.006e−09	4.011	1.008e−09	3.989

Relative L^2 errors for a manufactured solution example for the one-dimensional HyQMOM equations with a BGK collision operator. The errors are computed for various values of the Knudsen number: $\varepsilon = 10^4,\ 10^2,\ 10^0,\ 10^{-2},\ 10^{-4}$, and 10^{-6}. In each case, we use the fourth order method: $M_{\text{O}} = 4$

8.2 BGK Shock Tube Problem

Consider the Riemann problem for (7.4)–(7.5) with the following initial data at $t = 0$:

$$(\rho, u, p, h, k)(t = 0, x) = \begin{cases} (1.0,\ 0.0,\ 1.0,\ 0.0,\ 2.0) & x < 0, \\ (0.125,\ 0.0,\ 0.1,\ 0.0,\ 0.16) & x > 0, \end{cases} \tag{8.6}$$

on $x \in [-1, 1]$ with extrapolation boundary conditions. This is the standard Sod shock tube problem [52], which is ubiquitous in shock-capturing literature, and is also often found as a standard test for Boltzmann–BGK solvers (e.g., see [3]).

We consider three different values of the Knudsen number: (a) $\varepsilon = 10^{-2}$, (b) $\varepsilon = 10^{-3}$, and (c) $\varepsilon = 10^{-4}$. In each case we run the $M_{\mathrm{O}} = 4$ scheme with $M_{\mathrm{elem}} = 200$; we also compare in each case the HyQMOM-BGK solution to the exact solution for the compressible Euler equations Eq. 7.7 (e.g., see Chapter 14 of LeVeque [39] for a derivation). We used the following values of $\mathcal{A}_0$ in formula Eq. 5.50: (a) $\mathcal{A}_0 = 50$ for $\varepsilon = 10^{-2}$, (b) $\mathcal{A}_0 = 50$ for $\varepsilon = 10^{-3}$, and (c) $\mathcal{A}_0 = 350$ for $\varepsilon = 10^{-4}$.

Figure 7displays the $\varepsilon = 10^{-2}$ numerical simulation at $t = 0.28$ showing the (a) density: $\rho(t, x)$, (b) macroscopic velocity: $u(t, x)$, (c) pressure: $p(t, x)$, and (d) heat flux: $h(t, x)$. At this Knudsen number, the solution is still significantly different than the compressible Euler

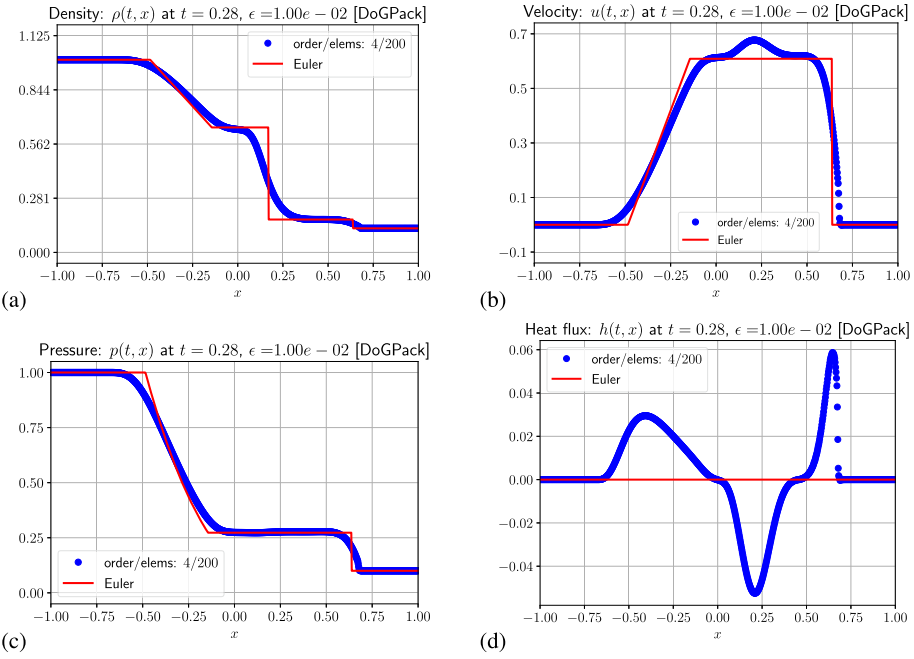

Fig. 7 (Section 8.2: BGK shock tube problem with $\varepsilon = 10^{-2}$) Numerical solution of shock tube problem on $x \in [-1, 1]$ with initial conditions given by (8.6). Shown are the results from a simulation run with the proposed $M_{\mathrm{O}} = 4$ scheme with 200 elements (shown as blue dots) and the exact solution of the compressible Euler equations with the same initial conditions (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points-per-element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t, x)$, **b** macroscopic velocity: $u(t, x)$, **c** pressure: $p(t, x)$, and **d** heat flux: $h(t, x)$ (color figure online)

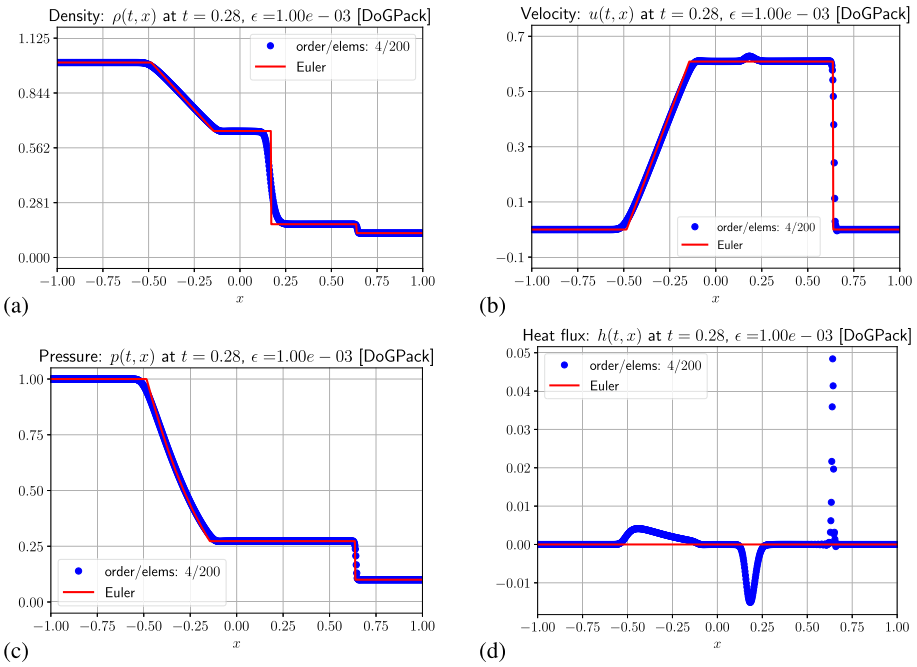

Fig. 8 (Section 8.2: BGK shock tube problem with $\varepsilon = 10^{-3}$) Numerical solution of shock tube problem on $x \in [-1, 1]$ with initial conditions given by (8.6). Shown are the results from a simulation run with the proposed $M_{\mathrm{O}} = 4$ scheme with 200 elements (shown as blue dots) and the exact solution of the compressible Euler equations with the same initial conditions (shown as a solid red line). For the $M_{\mathrm{O}} = 4$ scheme, we are plotting four points per element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t, x)$, **b** macroscopic velocity: $u(t, x)$, **c** pressure: $p(t, x)$, and **d** heat flux: $h(t, x)$ (color figure online)

solution, which is also shown in each panel. The results are consistent with fully kinetic solutions [3].

Figure 8 displays the $\varepsilon = 10^{-3}$ numerical simulation at $t = 0.28$ showing the (a) density: $\rho(t, x)$, (b) macroscopic velocity: $u(t, x)$, (c) pressure: $p(t, x)$, and (d) heat flux: $h(t, x)$. At this Knudsen number, the solution looks closer to the compressible Euler solution, which is also shown in each panel. The results are again consistent with fully kinetic solutions [3].

Figure 9 displays the $\varepsilon = 10^{-4}$ numerical simulation at $t = 0.28$ showing the (a) density: $\rho(t, x)$, (b) macroscopic velocity: $u(t, x)$, (c) pressure: $p(t, x)$, and (d) heat flux: $h(t, x)$. At this Knudsen number, the solution is very close to the compressible Euler solution, which is also shown in each panel. The results are again consistent with fully kinetic solutions [3].

9 Conclusions

In this work, we considered a particular moment closure called HyQMOM (the hyperbolic quadrature-based method of moments), which was originally introduced by Fox, Laurent, Vie [22] and further studied by Johnson [31] and Wiersma [59]. Quadrature-based method of moments (QMOM), including the HyQMOM variant, are a promising class of approximation techniques for reducing kinetic equations to fluid equations that are valid beyond thermodynamic equilibrium. In particular, the goal of the present work was to develop high-order

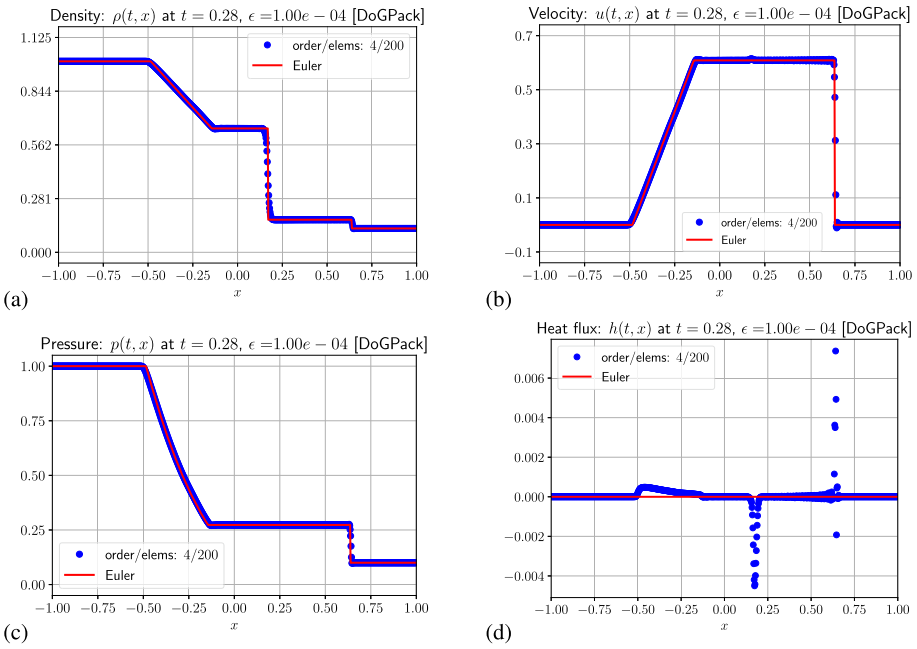

Fig. 9 (Section 8.2: BGK shock tube problem with $\varepsilon = 10^{-4}$) Numerical solution of shock tube problem on $x \in [-1, 1]$ with initial conditions given by (8.6). Shown are the results from a simulation run with the proposed $M_O = 4$ scheme with 200 elements (shown as blue dots) and the exact solution of the compressible Euler equations with the same initial conditions (shown as a solid red line). For the $M_O = 4$ scheme, we are plotting four points-per-element in order to show the intra-element solution structure. The panels show the primitive variables: **a** density: $\rho(t, x)$, **b** macroscopic velocity: $u(t, x)$, **c** pressure: $p(t, x)$, and **d** heat flux: $h(t, x)$ (color figure online)

discontinuous Galerkin schemes and corresponding limiters that control both unphysical oscillations and eliminate positivity violations.

The numerical scheme developed is based on the Lax–Wendroff discontinuous Galerkin scheme introduced by Qiu, Dumbser, and Shu [48], with the predictor–corrector interpretation developed by Gassner et al. [24], and further refinements developed by Felton et al. [18]. The resulting numerical method is performed in two phases at each time step.

Prediction step. The equation and numerical solution are written in the primitive variables in this phase. In the space-time DG approximation, which is applied to each element, integration-by-parts is only performed on the time variable. The result is a system of local nonlinear equations on each element. These equations are solved using a Picard iteration, which provides a sufficiently accurate solution after $M_O - 1$ iterations, where M_O is the order of accuracy of the method.

Correction step. The correction is a straightforward explicit update based on the time-integral of the evolution equation in conservation form, where the space-time prediction replaces all instances of the exact solution.

Several limiters were applied to the scheme to guarantee positivity and achieve solutions without unphysical oscillations.

Limiter I: Prediction step positivity limiter. This limiter is completely local and minimally damps high-order corrections to the primitive variables to get pointwise positivity

of the predicted density, pressure, and modified kurtosis on all space-time quadrature points (Gauss–Legendre + edges). The limiter is applied once after each Picard iteration, meaning it is applied a total of $M_O - 1$ times per time step.

Limiter II: Correction step positivity limiter on cell average. This limiter is applied once per time step and blends high-order numerical fluxes with positivity preserving low-order fluxes in such a way as to preserve the positivity of the corrected element averages of density, pressure, and modified kurtosis. This limiter is applied once per time step.

Limiter III: Correction step positivity limiter on quadrature points. This limiter is similar to Limiter I and involves minimally damping the high-order corrections to preserve the positivity of the corrected density, pressure, and modified kurtosis, on all spatial quadrature points (Gauss–Legendre + edges). This limiter is applied once per time step at the end of the step.

Limiter IV: Oscillation Limiter. This limiter damps the solution if the primitive solution variables on the current element significantly exceed the primitive solution variables on neighboring elements. This limiter is applied once per time step at the end of the time step.

In the collisionless regime, the proposed high-order method and the limiting strategy were tested on both smooth and Riemann problems. The smooth solution was used to perform convergence tests that demonstrated the expected orders of accuracy. The Riemann data tests clearly showed that the limiters were successful in damping unphysical oscillations without adversely diffusing the solution and preserving the positivity of the relevant variables.

Once the collisionless method is fully developed, we propose a version of the scheme for HyQMOM with a BGK collision operator. We carefully show how to handle the collision operator in both the prediction and collision steps to achieve an asymptotic-preserving (AP) property in the high-collision limit. Several numerical examples are provided to validate the scheme both for smooth solutions and Riemann initial data. The asymptotic-preserving property is validated for smooth solutions and Riemann initial data.

Future work will focus on extending this work to higher dimensions; perhaps using the conditional moment strategy of [22, 45], or some other higher-dimensional extension.

Acknowledgements We would like to thank the anonymous reviewers for their thoughtful comments and suggestions that helped to improve this paper.

Funding This research was funded by Iowa State University in Ames, Iowa, USA and US National Science Foundation Grants DMS–1620128 and DMS–2012699.

Data Availability Data sharing does not apply to this article as no datasets were generated or analyzed during the current study.

Declarations

Competing interest The authors have no conflicts of interest to disclose.

A Appendix

Lemma A.1 (Hermite interpolation) *Consider the Hermite interpolation problem of interpolating the function* $f(v) = v^{2N}$ *with a polynomial of degree* $2N - 1$*:*

$$P_{2N-1}(v) = a_0 + a_1 v + a_2 v^2 + \cdots + a_{2N-1} v^{2N-1} = \sum_{j=0}^{2N-1} a_j v^j, \tag{A.1}$$

with interpolating conditions for $\ell = 1, 2, \ldots, N$*:*

$$P_{2N-1}(\mu_\ell) = f(\mu_\ell) = \mu_\ell^{2N} \quad \text{and} \quad P'_{2N-1}(\mu_\ell) = f'(\mu_\ell) = 2N\mu_\ell^{2N-1}, \tag{A.2}$$

where

$$\mu_1 < \mu_2 < \cdots < \mu_{N-1} < \mu_N. \tag{A.3}$$

Applying these conditions yields the following formula for the polynomial coefficients:

$$\begin{bmatrix} a_0 \\ \vdots \\ a_{N-1} \\ a_N \\ \vdots \\ a_{2N-1} \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & \mu_1^2 & \mu_1^3 & \cdots & \mu_1^{2N-1} \\ \vdots & \vdots & \vdots & & & \vdots \\ 1 & \mu_N & \mu_N^2 & \mu_N^3 & \cdots & \mu_N^{2N-1} \\ 0 & 1 & 2\mu_1 & 3\mu_1^2 & \cdots & (2N-1)\mu_1^{2N-2} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 2\mu_N & 3\mu_N^2 & \cdots & (2N-1)\mu_N^{2N-2} \end{bmatrix}^{-1} \begin{bmatrix} \mu_1^{2N} \\ \vdots \\ \mu_N^{2N} \\ 2N\mu_1^{2N-1} \\ \vdots \\ 2N\mu_N^{2N-1} \end{bmatrix}. \tag{A.4}$$

Proof The claimed result follows directly from applying the interpolating conditions to the polynomial $P_{2N-1}(v)$. □

Lemma A.2 (Moment gradient operator I) *Let* $f\left(\underline{\omega}, \underline{\mu}\right) : \mathbb{R}^N \times \mathbb{R}^N \mapsto \mathbb{R}$ *be a continuously differentiable function, where* $\underline{\omega}, \underline{\mu} \in \mathbb{R}^N$ *satisfy the moment condition* (3.2)*. The gradient of* f *with respect to the moments,* $\underline{M} = (M_0, M_1, \ldots, M_{2N-1})$*, is given by*

$$\nabla_{\underline{M}} f\left(\underline{\omega}, \underline{\mu}\right) = \underline{\underline{B}}^{-1} \nabla_{\left(\underline{\omega}, \underline{\mu}\right)} f\left(\underline{\omega}, \underline{\mu}\right), \tag{A.5}$$

where

$$\nabla_{\underline{M}} := \left(\frac{\partial}{\partial M_0}, \frac{\partial}{\partial M_1}, \ldots, \frac{\partial}{\partial M_{2N-1}}\right), \tag{A.6}$$

$$\nabla_{\left(\underline{\omega}, \underline{\mu}\right)} := \left(\frac{\partial}{\partial \omega_1}, \ldots, \frac{\partial}{\partial \omega_N}, \frac{\partial}{\partial \mu_1}, \ldots, \frac{\partial}{\partial \mu_N}\right), \tag{A.7}$$

$$\underline{\underline{B}} := \frac{\partial \underline{M}}{\partial \left(\underline{\omega}, \underline{\mu}\right)} = \begin{bmatrix} 1 & \mu_1 & \mu_1^2 & \cdots & \mu_1^{2N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \mu_N & \mu_N^2 & \cdots & \mu_N^{2N-1} \\ 0 & \omega_1 & 2\omega_1\mu_1 & \cdots & (2N-1)\omega_1\mu_1^{2N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \omega_N & 2\omega_N\mu_N & \cdots & (2N-1)\omega_N\mu_N^{2N-2} \end{bmatrix}. \tag{A.8}$$

Proof The results follows directly from the chain rule applied to the moment condition: (3.2). □

Lemma A.3 *(*Moment gradient operator II) *The moment gradient as defined by* (A.5)–(A.8) *applied to the function* $f\left(\underline{\omega}, \underline{\mu}\right) = \mu_\ell$ *for some* $\ell = 1, 2, \ldots, N$ *is the following vector:*

$$\mathbb{R}^{2N} \ni \underline{b} := \left[b_0,\ b_1,\ \ldots, b_{2N-1}\right]^T = \nabla_{\underline{M}}\, \mu_\ell = \underline{\underline{B}}^{-1}\, \nabla_{\left(\underline{\omega},\underline{\mu}\right)}\, \mu_\ell = \underline{\underline{B}}^{-1}\, \underline{e_{N+\ell}}, \tag{A.9}$$

where $\underline{e_{N+\ell}} \in \mathbb{R}^{2N}$ *is a vector with a value of one in component* $N + \ell$ *and a value of zero in all other components.*

Furthermore, $\underline{b}$ *as defined above can be interpreted as the vector of coefficients of the following polynomial:*

$$Q_{2N-1}(v) := b_0 + b_1 v + b_2 v^2 + \cdots + b_{2N-1} v^{2N-1} = \sum_{j=0}^{2N-1} b_j v^j, \tag{A.10}$$

which satisfies all of the following conditions:

$$Q_{2N-1}\left(\mu_m\right) = 0 \quad \text{and} \quad Q'_{2N-1}\left(\mu_m\right) = \frac{1}{\omega_\ell}\, \delta^\ell_m \quad \text{for} \quad m = 1, 2, \ldots, N. \tag{A.11}$$

The polynomial Q_{2N-1} *can be explicitly written as follows:*

$$Q_{2N-1}(v) = \frac{(v - \mu_\ell)}{\omega_\ell} \left[\prod_{\substack{j=1 \\ j\neq\ell}}^{N} \left(v - \mu_j\right)^2 \Bigg/ \prod_{\substack{j=1 \\ j\neq\ell}}^{N} \left(\mu_\ell - \mu_j\right)^2 \right]. \tag{A.12}$$

Finally, the dot product between the vector $\underline{b} \in \mathbb{R}^{2N}$ *defined by* (A.9) *and the following vector:*

$$\underline{\mathcal{R}}(s) := \left[1,\ s,\ s^2,\ \ldots,\ s^{2N-1}\right]^T, \qquad s \in \mathbb{R}, \tag{A.13}$$

can be written as

$$\underline{b} \cdot \underline{\mathcal{R}}(s) = Q_{2N-1}(s) = \frac{(s - \mu_\ell)}{\omega_\ell} \left[\prod_{\substack{j=1 \\ j\neq\ell}}^{N} \left(s - \mu_j\right)^2 \Bigg/ \prod_{\substack{j=1 \\ j\neq\ell}}^{N} \left(\mu_\ell - \mu_j\right)^2 \right]. \tag{A.14}$$

Proof Equation (A.9) follows directly from definitions (A.5)–(A.8). Polynomial (A.10) with Hermite interpolation conditions (A.11) follows from an argument similar to the one provided in Lemma (A.1). Equation (A.12) follows from invoking the Lagrange form of the interpolating polynomial that satisfies conditions (A.11).

Finally, dot product (A.14) follows from the simple observation that

$$Q_{2N-1}(sv) = \sum_{j=0}^{2N-1} b_j s^j v^j = b_0 + (b_1 s)\, v + \cdots + \left(b_{2N-1} s^{2N-1}\right) v^{2N-1}$$

$$\implies \quad Q_{2N-1}(s) = \sum_{j=0}^{2N-1} b_j \mathcal{R}_j(s) = \underline{b} \cdot \underline{\mathcal{R}}(s),$$

which when combined with (A.12) gives the desired result. □

Theorem A.1 (Weak hyperbolicity and linear degeneracy of QMOM) *The classical quadrature-based moment (QMOM) closure for a fixed $N \in \mathbb{N}_{\geq 1}$, denoted by Eq. 3.1, leads to a system of partial differential equations that has the following quasilinear form:*

$$\underline{q}_{,t} + \underline{\underline{A}}\left(\underline{q}\right) \underline{q}_{,x}, \qquad \text{where} \quad \underline{q} = \left[M_0,\, M_1,\, \cdots,\, M_{2N-2},\, M_{2N-1}\right], \tag{A.15}$$

where the flux Jacobian matrix is given by

$$\underline{\underline{A}}\left(\underline{q}\right) = \begin{bmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ \frac{\partial M^{\star}_{2N}}{\partial M_0} & \frac{\partial M^{\star}_{2N}}{\partial M_1} & \cdots & \frac{\partial M^{\star}_{2N}}{\partial M^{\star}_{2N-2}} & \frac{\partial M^{\star}_{2N}}{\partial M_{2N-1}} \end{bmatrix}, \tag{A.16}$$

where $M^{\star}_{2N}$ is given by (3.2). System (A.15) and (A.16) is weakly hyperbolic for any integer $N \geq 1$ with the following properties:

1. *The eigenvalues of (A.16) are $\lambda_\ell = \mu_\ell$ for $\ell = 1, 2, 3, \ldots, N$, where μ_ℓ are the abscissas in (3.1);*
2. *Every eigenvalue has algebraic multiplicity exactly two;*
3. *Every eigenvalue has geometric multiplicity exactly one; and*
4. *Every wave in the system is linearly degenerate: $\frac{\partial \lambda_\ell}{\partial \underline{q}} \cdot \underline{\mathcal{R}^\ell} = 0$ for $\ell = 1, 2, 3, \ldots, N$, where $\left(\lambda_\ell, \underline{\mathcal{R}^\ell}\right)$ is an eigenvalue-eigenvector pair of flux Jacobian (A.16).*

Proof The key to understanding the eigenvalues of flux Jacobian (A.16) is to understand the last row. To this end, consider:

$$\mathsf{M}^{\star}_{2N} = \sum_{j=1}^{N} \omega_j \mu_j^{2N} \quad \Longrightarrow \quad \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_\ell} = \sum_{j=1}^{N} \left[\mu_j^{2N} \frac{\partial \omega_j}{\partial \mathsf{M}_\ell} + 2N \omega_j \mu_j^{2N-1} \frac{\partial \mu_j}{\partial \mathsf{M}_\ell} \right]. \tag{A.17}$$

To make sense of this we need to obtain expressions for the partial derivatives of the quadrature weights and abscissas with respect to the moments. To this end, we compute the related quantities:

$$\mathsf{M}_s = \sum_{j=1}^{N} \omega_j \mu_j^s \quad \Longrightarrow \quad \frac{\partial \mathsf{M}_s}{\partial \mathsf{M}_\ell} = \sum_{j=1}^{N} \left[\mu_j^s \frac{\partial \omega_j}{\partial \mathsf{M}_\ell} + s \omega_j \mu_j^{s-1} \frac{\partial \mu_j}{\partial \mathsf{M}_\ell} \right] = \delta^s_\ell, \tag{A.18}$$

where $s, \ell = 0, 1, \ldots, 2N-1$ and δ^s_ℓ is the Kronecker delta, which arises due to the fact that M_ℓ and M_s are independent variables if $s \neq \ell$. The expression in (A.18) can be written in matrix form to obtain the following result:

$$\begin{bmatrix} \frac{\partial \omega_1}{\partial \mathsf{M}_0} & \cdots & \frac{\partial \omega_1}{\partial \mathsf{M}_{2N-1}} \\ \vdots & & \vdots \\ \frac{\partial \omega_N}{\partial \mathsf{M}_0} & \cdots & \frac{\partial \omega_N}{\partial \mathsf{M}_{2N-1}} \\ \omega_1 \frac{\partial \mu_1}{\partial \mathsf{M}_0} & \cdots & \omega_1 \frac{\partial \mu_1}{\partial \mathsf{M}_{2N-1}} \\ \vdots & & \vdots \\ \omega_N \frac{\partial \mu_N}{\partial \mathsf{M}_0} & \cdots & \omega_N \frac{\partial \mu_N}{\partial \mathsf{M}_{2N-1}} \end{bmatrix}^T = \begin{bmatrix} 1 & \mu_1 & \mu_1^2 & \cdots & \mu_1^{2N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & \mu_N & \mu_N^2 & \cdots & \mu_N^{2N-1} \\ 0 & 1 & 2\mu_1 & \cdots & (2N-1)\mu_1^{2N-2} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 1 & 2\mu_N & \cdots & (2N-1)\mu_N^{2N-2} \end{bmatrix}^{-1}. \tag{A.19}$$

Using this result in (A.17) produces expressions for the last row of flux Jacobian (A.16):

$$\begin{bmatrix} \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_0} \\ \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_1} \\ \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_2} \\ \vdots \\ \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_{2N-2}} \\ \frac{\partial \mathsf{M}^{\star}_{2N}}{\partial \mathsf{M}_{2N-1}} \end{bmatrix} = \begin{bmatrix} 1 & \mu_1 & \cdots & \mu_1^{2N-1} \\ \vdots & \vdots & & \vdots \\ 1 & \mu_N & \cdots & \mu_N^{2N-1} \\ 0 & 1 & \cdots & (2N-1)\mu_1^{2N-2} \\ \vdots & \vdots & & \vdots \\ 0 & 1 & \cdots & (2N-1)\mu_N^{2N-2} \end{bmatrix}^{-1} \begin{bmatrix} \mu_1^{2N} \\ \vdots \\ \mu_N^{2N} \\ 2N\mu_1^{2N-1} \\ \vdots \\ 2N\mu_N^{2N-1} \end{bmatrix} = \begin{bmatrix} a_0 \\ \vdots \\ a_{N-1} \\ a_N \\ \vdots \\ a_{2N-1} \end{bmatrix}, \tag{A.20}$$

where the last equality follows from Lemma A.1 and a_j for $j = 0, 1, \ldots, 2N-1$ are the coefficients of the Hermite interpolating polynomial defined through (A.1) and (A.2).

Next we attempt to directly compute the eigenvalues of the flux Jacobian:

$$\left| \underline{\underline{A}} - v \underline{\underline{\mathbb{I}}} \right| = \begin{vmatrix} [1.5] - v & 1 & & \\ & \ddots & \ddots & \\ & & -v & 1 \\ a_0 & \cdots & a_{2N-2} & (a_{2N-1} - v) \end{vmatrix} = v^{2N} - \sum_{j=0}^{2N-1} a_j v^j. \tag{A.21}$$

Using a classical result from Hermite polynomial interpolation, we can write the right-most term in the above expression as follows (e.g., see Theorem 6.4 on page 190 of Süli and Mayer [53]):

$$\left| \underline{\underline{A}} - v \underline{\underline{\mathbb{I}}} \right| = (v - \mu_1)^2 (v - \mu_2)^2 \cdots (v - \mu_N)^2. \tag{A.22}$$

This proves the first two claims of the theorem: (1) the eigenvalues are the quadrature abscissas, and (2) each eigenvalue has algebraic multiplicity exactly two.

Next we look at the eigenvectors. For example, the ℓ^{th} eigenvector for each $\ell = 1, 2, \ldots, N$ satisfies the relationship:

$$\left(\underline{\underline{A}} - \mu_\ell \underline{\underline{\mathbb{I}}} \right) \underline{\mathcal{R}^\ell} = \underline{0}. \tag{A.23}$$

By inspection, we see that $\underline{\mathcal{R}^\ell} \not\equiv \underline{0}$ if and only if the first component of $\underline{\mathcal{R}^\ell}$ is not zero. Without loss of generality the first component is taken to be unity, and then by inspection we note that the only eigenvector associated to eigenvalue $v = \mu_\ell$ must be

$$\underline{\mathcal{R}^\ell} = \left(1, \ \mu_\ell, \ \mu_\ell^2, \ \ldots, \ \mu_\ell^{2N-1} \right)^T. \tag{A.24}$$

This proves the third claim of the theorem: (3) each eigenvalue has geometric multiplicity exactly one. Since the geometric multiplicity for each eigenvalue is strictly less than the algebraic multiplicity, system (A.15) and (A.16) is weakly hyperbolic for any integer $N \geq 1$.

The final claim in the theorem is that each wave is linearly degenerate. Proving this requires us to investigate the following dot product

$$\frac{\partial \mu_\ell}{\partial \underline{q}} \cdot \underline{\mathcal{R}^\ell}. \tag{A.25}$$

Invoking Lemmas (A.2) and (A.3) shows that for every $\ell = 1, 2, \ldots, N$:

$$\frac{\partial \mu_\ell}{\partial \underline{q}} \cdot \underline{\mathcal{R}^\ell} = \lim_{s \to \mu_\ell} \left\{ \frac{(s - \mu_\ell)}{\omega_\ell} \left[\prod_{\substack{j=1 \\ j \neq \ell}}^{N} (s - \mu_j)^2 \Big/ \prod_{\substack{j=1 \\ j \neq \ell}}^{N} (\mu_\ell - \mu_j)^2 \right] \right\} = 0, \tag{A.26}$$

which proves that every wave for $\ell = 1, 2, \ldots, N$ is linearly degenerate. □

Lemma A.4 (Convexity property) *Let $\mathcal{C}_\alpha\left(\underline{Q}\right) : \mathbb{R}^5 \mapsto \mathbb{R}$ be a convex function. Then for all $a, b \in \mathbb{R}$ such that*

$$a \geq 0, \quad b \geq 0, \quad 1 - a - b \geq 0, \tag{A.27}$$

the following inequality holds:

$$\mathcal{C}_\alpha\Big((1 - a - b)\underline{P} + a\underline{Q} + b\underline{R}\Big) \leq (1 - a - b)\mathcal{C}_\alpha\Big(\underline{P}\Big) + a\mathcal{C}_\alpha\Big(\underline{Q}\Big) + b\mathcal{C}_\alpha\Big(\underline{R}\Big) \quad \forall \underline{P}, \underline{Q}, \underline{R} \in \mathbb{R}^5. \tag{A.28}$$

Proof By definition, the function $\mathcal{C}_\alpha$ is convex *if and only if* the following is true for all $\theta \in [0, 1]$:

$$\mathcal{C}_\alpha\Big(\big(1 - \theta\big)\underline{P} + \theta\underline{Q}\Big) \leq \big(1 - \theta\big)\,\mathcal{C}_\alpha\Big(\underline{P}\Big) + \theta\,\mathcal{C}_\alpha\Big(\underline{Q}\Big) \quad \forall \underline{P}, \underline{Q} \in \mathbb{R}^5. \tag{A.29}$$

Consider the convex function applied to a sum of three vectors of the following form:

$$\mathcal{C}_\alpha\Big((1 - a - b)\underline{P} + a\underline{Q} + b\underline{R}\Big) \quad \text{where} \quad a, b, (1 - a - b) \geq 0. \tag{A.30}$$

We can temporarily define the following vector:

$$\underline{S} := \left(\frac{a}{a + b}\right)\underline{Q} + \left(\frac{b}{a + b}\right)\underline{R} \quad \Longrightarrow \quad a\underline{Q} + b\underline{R} = (a + b)\underline{S}, \tag{A.31}$$

such that

$$\mathcal{C}_\alpha\Big((1 - a - b)\underline{P} + a\underline{Q} + b\underline{R}\Big) = \mathcal{C}_\alpha\Big((1 - a - b)\underline{P} + (a + b)\underline{S}\Big). \tag{A.32}$$

Invoking the convexity Definition (A.29) with $\theta = a + b$, which by assumption satisfies $\theta \in [0, 1]$, we have that

$$\mathcal{C}_\alpha\Big((1 - a - b)\underline{P} + a\underline{Q} + b\underline{R}\Big) \leq (1 - a - b)\mathcal{C}_\alpha\Big(\underline{P}\Big) + (a + b)\mathcal{C}_\alpha\left(\left(\frac{a}{a + b}\right)\underline{Q} + \left(\frac{b}{a + b}\right)\underline{R}\right). \tag{A.33}$$

We then again invoke Definition (A.29), this time with $\theta = \frac{b}{a+b}$, which also satisfies $\theta \in [0, 1]$, to get that

$$\mathcal{C}_\alpha\left(\left(\frac{a}{a + b}\right)\underline{Q} + \left(\frac{b}{a + b}\right)\underline{R}\right) \leq \left(\frac{a}{a + b}\right)\mathcal{C}_\alpha\Big(\underline{Q}\Big) + \left(\frac{b}{a + b}\right)\mathcal{C}_\alpha\Big(\underline{R}\Big). \tag{A.34}$$

Combining the last two inequalities, (A.33) and (A.34), results in desired inequality: (A.28). □

References

1. Abdelmalik, M., van Brummelen, E.: Moment closure approximations of the Boltzmann equation based on φ-divergences. J. Stat. Phys. **164**, 77–104 (2016)
2. Bardos, C., Golse, F., Levermore, D.: Fluid dynamic limits of kinetic equations. I. Formal derivations. J. Stat. Phys. **63**(1–2), 323–344 (1991)
3. Bennoune, M., Lemou, M., Mieussens, L.: Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier–Stokes asymptotics. J. Comput. Phys. **227**, 3781–3803 (2008)
4. Bhatnagar, P., Gross, E., Krook, M.: A model for collision processes in gases I. Small amplitude processes in charged and neutral one-component systems. Phys. Rev. Lett. **94**, 511–525 (1954)
5. Böhmer, N., Torrilhon, M.: Entropic quadrature for moment approximations of the Boltzmann–BGK equation. J. Comput. Phys. **401**, 108992 (2020)
6. Broadwell, J.: Study of rarefied shear flow by the discrete velocity method. J. Fluid Mech. **19**, 401–414 (1964)
7. Broadwell, J.E.: Shock structure in a simple discrete velocity gas. Phys. Fluids **7**, 1243–1247 (1964)
8. Caflisch, R., Jin, S., Russo, G.: Uniformly accurate schemes for hyperbolic systems with relaxation. SIAM J. Numer. Anal. **34**, 246–281 (1997)
9. Cai, Z., Fan, Y., Li, R.: Globally hyperbolic regularization of grad's moment system in one dimensional space. Commun. Math. Sci. **11**, 547–571 (2013)
10. Cai, Z., Fan, Y., Li, R.: Globally hyperbolic regularization of grad's moment system. Commun. Pure Appl. Math. **32**, 464–518 (2014)
11. Chalons, C., Fox, R., Massot, M.: A multi-Gaussian quadrature method of moments for gas-particle fows in a LES framework. In: Proceedings of the Summer Program, pp. 347–358. Center for Turbulence Research (2010)
12. Chalons, C., Kah, D., Massot, M.: Beyond pressureless gas dynamics: quadrature-based velocity moment models. Commun. Math. Sci. **10**, 1241–1272 (2012)
13. Cheng, Y., Rossmanith, J.: A class of quadrature-based moment-closure methods with application to the Vlasov–Poisson–Fokker–Planck system in the high-field limit. J. Comput. Appl. Math. **262**, 384–398 (2014)
14. Cockburn, B., Shu, C.W.: The Runge–Kutta discontinuous Galerkin method for conservation laws V. J. Comput. Phys. **141**(2), 199–224 (1998). https://doi.org/10.1006/jcph.1998.5892
15. Coron, F., Perhame, B.: Numerical passage from kinetic to fluid equations. SIAM J. Numer. Anal. **28**, 26–42 (1991)
16. Desjardins, O., Fox, R., Villedieu, P.: A quadrature-based moment method for dilute fluid-particle flows. J. Comput. Phys. **227**, 2514–2539 (2008)
17. Dreyer, W.: Maximisation of the entropy in non-equilibrium. J. Phys. A Math. **20**, 6505–6517 (1987)
18. Felton, C., Harris, M., Logemann, C., Nelson, S., Pelakh, I., Rossmanith, J.: A positivity-preserving limiting strategy for locally-implicit Lax–Wendroff discontinuous Galerkin methods. arXiv:1806.06756 (2018)
19. Fox, R.: Computational Models for Turbulent Flows. Cambridge University Press, Cambridge (2003)
20. Fox, R.: A quadrature-based third-order moment method for dilute gas-particle flows. J. Comput. Phys. **227**, 6313–6350 (2008)
21. Fox, R.: Higher-order quadrature-based moment methods for kinetic equations. J. Comput. Phys. **228**, 7771–7791 (2009)
22. Fox, R., Laurent, F., Vié, A.: Conditional hyperbolic quadrature method of moments for kinetic equations. J. Comput. Phys. **365**, 269–293 (2018)
23. Gabetta, E., Pareschi, L., Toscani, G.: Relaxation schemes for nonlinear kinetic equations. SIAM J. Numer. Anal. **34**, 2168–2194 (1997)
24. Gassner, G., Dumbser, M., Hindenlang, F., Munz, C.D.: Explicit one-step time discretizations for discontinuous Galerkin and finite volume schemes based on local predictors. J. Comput. Phys. **230**, 4232–4247 (2011)
25. Grad, H.: On the kinetic theory of rarefied gases. Commun. Pure Appl. Math. **2**, 331–407 (1949)
26. Guthrey, P., Rossmanith, J.: The regionally-implicit discontinuous Galerkin method: improving the stability of DG-FEM. SIAM J. Num. Anal. **57**, 1263–1288 (2019)
27. Jin, S.: Runge–Kutta methods for hyperbolic conservation laws with stiff relaxation terms. J. Comput. Phys. **122**, 51–67 (1995)
28. Jin, S.: Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. SIAM J. Sci. Comput. **21**, 441–454 (1999)

29. Jin, S.: Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. Riv. Mat. Della Univ. Parma **3**, 177–216 (2012)
30. Jin, S., Levermore, C.: Numerical schemes for hyperbolic conservation laws with stiff relaxation terms. J. Comput. Phys. **126**, 449–467 (1996)
31. Johnson, E.R.: A high-order discontinuous Galerkin finite element method for a quadrature-based moment-closure model. Master's thesis, Iowa State University (2017)
32. Junk, M.: Domain of definition of Levermore's five-moment system. J. Stat. Phys. **93**, 1143–1167 (1998)
33. Kamermans, M.: Gaussian quadrature weights and abscissae. https://pomax.github.io/bezierinfo/legendre-gauss.html
34. Koellermeier, J., Castro, M.: High-order non-conservative simulation of hyperbolic moment models in partially conservative form. East Asian J. Appl. Math. **11**, 435–467 (2021)
35. Koellermeier, J., Torrilhon, M.: Numerical solution of hyperbolic moment models for the Boltzmann equation. Eur. J. Mech. B/Fluids **64**, 41–46 (2017)
36. Koellermeier, J., Schaerer, R., Torrilhon, M.: A framework for hyperbolic approximation of kinetic equations using quadrature-based projection methods. Kinet. Relat. Models **7**(3), 531–549 (2014)
37. Lax, P., Wendroff, B.: Systems of conservation laws. Commun. Pure Appl. Math. **13**, 217–237 (1960)
38. LeVeque, R.: Wave propagation algorithms for multi-dimensional hyperbolic systems. J. Comput. Phys. **131**, 327–335 (1997)
39. LeVeque, R.: Finite Volume Methods for Hyperbolic Problems. Cambridge University Press, Cambridge (2002)
40. Levermore, C.: Moment closure hierarchies for kinetic theories. J. Stat. Phys. **83**, 1021–1065 (1996)
41. Marchisio, D., Fox, R.: Solution of population balance equations using the direct quadrature method of moments. J. Aerosol. Sci. **36**, 43–73 (2005)
42. Moe, S., Rossmanith, J., Seal, D.: A simple and effective high-order shock-capturing limiter for discontinuous Galerkin methods (2015). arXiv:1507.03024
43. Moe, S., Rossmanith, J., Seal, D.: Positivity-preserving discontinuous Galerkin methods with Lax–Wendroff time discretizations. J. Sci. Comput. **71**, 44–70 (2017)
44. Müller, I., Ruggeri, T.: Extended Thermodynamics. Springer, New York (1993)
45. Patel, R., Desjardins, O., Fox, R.: Three-dimensional conditional hyperbolic quadrature method of moments. J. Comput. Phys. X **1**, 100006 (2019)
46. Pieraccini, S., Puppo, G.: Implicit-explicit schemes for BGK kinetic equations. J. Sci. Comput. **32**, 1–28 (2007)
47. Platkowski, T., Illner, R.: Discrete velocity models of the Boltzmann equation: a survey on the mathematical aspects of the theory. SIAM Rev. **30**(2), 213–255 (1988)
48. Qiu, J., Dumbser, M., Shu, C.W.: The discontinuous Galerkin method with Lax–Wendroff type time discretizations. Comput. Methods Appl. Mech. Eng. **194**, 4528–4543 (2005)
49. Reed, W., Hill, T.: Triangular mesh methods for the neutron transport equation. Tech. Rep. LA-UR-73-479, Los Alamos Scientific Laboratory (1973)
50. Rusanov, V.: Calculation of interaction of non-steady shock waves with obstacles. J. Comp. Math. Phys. USSR **1**, 267–279 (1961)
51. Schmüdgen, K.: The Moment Problem. Graduate Texts in Mathematics. Springer, New York (2017)
52. Sod, G.: A survey of several finite difference methods for systems of nonlinear hyperbolic conservation laws. J. Comput. Phys. **27**, 1–31 (1978)
53. Süli, E., Mayers, D.: An Introduction to Numerical Analysis. Cambridge University Press, Cambridge (2003)
54. Titarev, V., Toro, E.: ADER: arbitrary high order Godunov approach. J. Sci. Comput. **17**, 609–618 (2002)
55. Torrilhon, M.: Editorial: special issue on moment methods in kinetic gas theory. Contin. Mech. Thermodyn. **21**, 341–343 (2009)
56. Torrilhon, M.: Modeling nonequilibrium gas flows based on moment equations. Ann. Rev. Fluid Mech. **48**(1), 429–458 (2016)
57. Vikas, V., Wang, Z., Passalacqua, A., Fox, R.: Realizable high-order finite-volume schemes for quadrature-based moment methods. J. Comput. Phys. **230**, 5328–5352 (2011)
58. von Kowalesky, S.: Zur Theorie der partiallen Differentialgleichungen. J. Reine Angew. Math. **80**, 1–32 (1875)
59. Wiersma, C.: A locally-implicit Lax–Wendroff discontinuous Galerkin scheme with limiters that guarantees moment-realizability for quadrature-based moment closures. Master's thesis, Iowa State University (2019)
60. Xiong, T., Qiu, J.M.: A hierarchical uniformly high order DG-IMEX scheme for the 1D BGK equation. J. Comput. Phys. **336**, 164–191 (2017)

61. Zhang, X., Shu, C.W.: Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: survey and new developments. Proc. R. Soc. A **467**, 2752–2776 (2011)
62. Zhang, X., Shu, C.W.: Positivity-preserving high order discontinuous Galerkin schemes for compressible Euler equations with source terms. J. Comput. Phys. **230**, 1238–1248 (2011)