A POSITIVE AND STABLE L2-MINIMIZATION BASED MOMENT METHOD FOR THE BOLTZMANN EQUATION OF GAS DYNAMICS*

NEERAJ SARNA†

Abstract. We consider the method-of-moments approach to solve the Boltzmann equation of rarefied gas dynamics, which results in the following moment-closure problem. Given a set of moments, find the underlying probability density function. The moment-closure problem has infinitely many solutions and requires an additional optimality criterion to single-out a unique solution. Motivated from a discontinuous Galerkin velocity discretization, we consider an optimality criterion based upon L2-minimization. To ensure a positive solution to the moment-closure problem, we enforce positivity constraints on L2-minimization. This results in a quadratic optimization problem with moments and positivity constraints. We show that a (Courant-Friedrichs-Lewy) CFL-type condition ensures both the feasibility of the optimization problem and the L2-stability of the space-time discrete moment approximation. We provide an extension of our method to multi-dimensional space-velocity domains and perform several numerical experiments to showcase its accuracy.

1 Introduction

Due to modeling assumptions, the Euler and the Navier-Stokes equations become inaccurate as a flow deviates significantly from a thermodynamic equilibrium. This motivates one to consider mathematical models that can approximate flows in all regimes of thermodynamic non-equilibrium. One such model is the Boltzmann equation (BE) that govern the evolution of a probability density function (pdf) $f(x,t,\xi) \in \mathbb{R}^+$ and reads

$$\mathcal{L}(f) = 0 \quad \text{where} \quad \mathcal{L} := \partial_t + \xi \cdot \nabla - Q. \tag{1.1}$$

Above, $\xi \in \mathbb{R}^{d_\xi}$ is the molecular velocity with $1 \leq d_\xi \leq 3$ being the velocity-dimension, $D := [0,T]$ is the temporal domain with $T > 0$ being the final time, and ∇ represents a gradient in the spatial domain $\Omega \subseteq \mathbb{R}^d$ with $1 \leq d \leq 3$ being the space-dimension. The BE signifies the fact that the pdf changes due to the free-streaming of the gas molecules and the inter-molecular collisions—the collision operator Q models the inter-particle collisions, and the transport operator $\partial_t + \xi \cdot \nabla$ models the free-streaming of the gas molecules.

In practical applications, one is not interested in the fine details of a pdf but in the macroscopic quantities like density, velocity, temperature, etc. These quantities can be recovered by taking the velocity-moments of the pdf. This motivates the method-of-moments (MOM) approach, where, rather than directly solving the BE, we solve for a finite number of moments of the pdf. The velocity-moments of the BE provide the governing equation for the moments of $f(x,t,\cdot)$, or the so-called moment equations. However, a finite set of moment equations is not closed—the flux term $(\xi \cdot \nabla f)$ results in a moment of degree higher than that included in the moment set. Nevertheless, one can close the moment equations by solving the following moment-closure problem.

(1.2) Moment-closure problem: given a set of moments, find the underlying pdf.

There are infinitely many solutions to the moment-closure problem [21]. To single-out a unique solution, one can introduce an optimality criterion by minimizing a

*Submitted to the editors xxxx

Funding: N.S is supported by the German Federal Ministry for Economic Affairs and Energy (BMWi) in the joint project "MathEnergy - Mathematical Key Technologies for Evolving Energy Grids", sub-project: Model Order Reduction (Grant number: 0324019B).

†Corresponding author, Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr 1, 39106, Magdeburg, Germany, sarna@mpi-magdeburg.mpg.de.

strictly convex functional of the pdf. We choose this functional to be the L2-norm of the pdf. Our choice is motivated by a discontinuous Galerkin (DG) discretization of the velocity domain that we interpret as an L2-minimization problem with moment constraints—later sections provide further clarification.

A major drawback of L2-minimization (and so of a DG-discretization [3]) is that it does not penalize the negativity of a solution—for the same reason, a Hermite spectral method proposed in [15] does not ensure positivity. This is undesirable given that we are approximating a pdf that is positive by definition. There is ample numerical and theoretical evidence supporting the claim that a positive solution to a moment-closure problem better approximates a pdf—see the different works on positive moment-methods [1, 8, 10, 18, 20, 22, 30, 37]. Furthermore, in theory, a negative solution to a moment-closure problem can result in a negative density and temperature, resulting in a breakdown of the solution algorithm. For this reason, we enforce positivity constraints on our L2-minimization problem. This results in a quadratic optimization problem with moments and positivity constraints.

For the robustness of the algorithm, the feasibility of the quadratic optimization problem is imperative. We show that a CFL-type condition ensures (i) the feasibility of the optimization problem and (ii) the L2-stability of the moment approximation—we insist that stability is crucial in analyzing the convergence of a moment approximation [24, 26]. A proof for both these properties hinges on relating our moment approximation to a discrete-velocity-method (DVM). We emphasize that our proof is general in the sense that it is independent of the objective functional being minimized to single-out a unique solution to the moment-closure problem.

Other than L2-minimization (with positivity constraints) one can consider entropy-minimization. Despite the many favourable properties (see [16, 19, 20, 29, 31]), it is challenging to compute an entropy-minimization based closure. A few reasons for this are as follows. Firstly, to perform entropy-minimization one usually performs Newton iterations where in every iteration one inverts the Hessian of the objective functional. This Hessian (despite the adaptivity of basis proposed in [5]) can become severely ill-conditioned—particularly inside shocks and for large moment sets—leading to a slow (or no) convergence of the Newton solver [29, 30]. Secondly, in every Newton iteration, one needs to compute integrals over the d_ξ-dimensional velocity domain. An analytical expression for these integrals is usually unavailable and one seeks a numerical approximation via some quadrature routine. The number of these quadrature points can grow drastically with d_ξ, making the solution algorithm expensive for multi-dimensional applications [8, 29, 30]. For instance, the number of tensorized Gauss-Legendre quadrature points grow as $\mathcal{O}(N^{d_\xi})$, where N is the number of quadrature points in one direction.

Replacing entropy minimization by L2-minimization (with positivity constraints) does not necessarily solves the two problems mentioned above. We use the interior-convex-set algorithm to perform L2-minimization and even for problems with strong shocks and large moment sets, we did not encounter issues with the conditioning of the Hessian. Our results suggest that L2-minimization could be an alternative to entropy-minimization for flow regimes where entropy-minimization losses robustness. Furthermore, since L2-minimization is robust for large moment sets, it is appealing for an adaptive approach where depending upon the accuracy requirements, the moment set can change locally in the space-time domain [2, 38].

We note that although our L2-minimization procedure is robust, to approximate the integrals, we use tensorized Gauss-Legendre quadrature points in the velocity domain, which, we expect, makes L2-minimization expensive. Specialized quadrature points can make both L2 and entropy minimization efficient [8]. However, these quadrature points do not guarantee the feasibility of the minimization problem—a property we consider crucial for the robustness of the solution algorithm. To tackle infeasible problems, one can try regularizing the minimization problem by relaxing the moment constraints [4]. The use of a specialized quadrature with a regularized minimization problem is an interesting direction to pursue and we plan to consider it in the future.

We acknowledge that our work draws inspiration from the positive PN closure proposed in [18] for the radiative transport equation. Indeed, we solve a similar optimization problem as that solved by the positive PN closure. Nevertheless, our work differs from [18] in the following ways. Firstly, unlike the linear isotropic collision operator considered in [18], we consider the non-linear Boltzmann-BGK operator that we discretize using entropy-minimization to ensure mass, momentum and energy conservation. Secondly, using the solution of the optimization problem, to close the moment system, authors in [18] perform a spherical harmonics based velocity reconstruction of the pdf. Our framework suggests that such a reconstruction is not needed if one uses the same quadrature points to compute moments in the moment equations and to solve the minimization problem. Thirdly, through numerical experiments, we study the convergence of our moment approximation and compare it to the DVM. These studies were not performed in [18]. Lastly, we establish robust (under vanishing Knudsen limit) L2-stability estimates for our moment approximation. Let us emphasize that to the best of our knowledge, for gas transport applications, none of the previous works consider L2-minimization based moment-closures with positivity constraints.

We have organized the rest of the article as follows. In Section 2 we discuss our moment approximation and the details of the BE. In Section 3 we discuss the space-time discretization of the moment equations, the feasibility of the optimization problem, and the stability of the moment approximation. In Section 4 we extend our framework to multi-dimensional problems, and in Section 5 we perform numerical experiments.

2 Moment Approximation

Throughout this section we consider a one dimensional space-velocity domain i.e., $d = d_\xi = 1$ in (1.1). An extension to multi-dimensions is straightforward and is discussed in Section 4. We start by discussing a positive L2-minimization based moment-closure and use it later to define a moment approximation for the BE.

2.1 A positive L2-method-of-moments (pos-L2-MOM)

Consider a m-th order polynomial in ξ given as $p_m(\xi) := \xi^m$. Collect all the different $p_m(\xi)$ upto some order $(M-1) \in \mathbb{N}$ in a vector $P_M(\xi)$ given as

$$P_M(\xi) := (p_0(\xi), \ldots, p_{M-1}(\xi))^T, \tag{2.1}$$

where $(\cdot)^T$ represents the transpose of a vector. For a function $\xi \mapsto g(\xi) \in \mathbb{R}$, we introduce the shorthand notation

$$\langle g \rangle := \int_{\mathbb{R}} g(\xi) d\xi. \tag{2.2}$$

Note that the definition of P_M implies that the vector $\langle P_M g \rangle$ contains the first M moments of g.

For some moment vector $\lambda \in \mathbb{R}^M$, consider the mathematical formulation of the moment-closure problem described earlier in the introduction

$$\text{Find } g_M \ : \ \langle P_M g_M \rangle = \lambda. \tag{2.3}$$

Even for a realizable moment vector λ (i.e., there exists a $g^* > 0$ such that $\lambda = \langle P_M g^* \rangle$), the above problem can have infinitely many solutions [21]. To single-out a unique solution, we use L2-minimization as an optimality criterion. Since L2-minimization does not penalize negativity and since we prefer a positive solution to the moment-closure problem, we explicitly enforce a positivity constraint. This result in an optimization problem given as

$$g_M := \underset{g^* \in L^2(\mathbb{R})}{\arg\min} \frac{1}{2} \|g^*\|^2_{L^2(\mathbb{R})} : \ \langle P_M g^* \rangle = \lambda, \ g^* > 0. \tag{2.4}$$

In the above minimization problem, as yet, it is unclear how to enforce the positivity constraint almost everywhere on $\mathbb{R}$. To tackle this problem, we consider the following two steps—we refer to [18, 29, 30] for similar steps related to the minimum-entropy closure and the positive PN closure.

1. *Truncate the velocity domain:* We truncate the velocity domain $\mathbb{R}$ to $\Omega_\xi := [\xi_{\min}, \xi_{\max}]$. A decent estimate for $\xi_{\max/\min}$ follows from the velocity and the temperature field of the gas and is discussed later in Subsection 2.4. The same sub-section discusses the pros and cons associated with truncating the velocity domain.
2. *Positivity constraints on quadrature points:* To perform the integrals in the minimization problem, we use some quadrature points defined over Ω_ξ. We enforce the positivity constraints only over these quadrature points. Although our framework is valid for any set of space-time-independent quadrature points, for completeness, we consider N Gauss-Legendre quadrature points and we denote their weights and abscissas by $\{\omega_i\}_i$ and $\{\xi_i\}_i$, respectively. Using the quadrature points, for some function $\xi \mapsto g(\xi) \in \mathbb{R}$, we define

$$\langle g \rangle \approx \langle g \rangle_N := \sum_{i=1}^{N} \omega_i g(\xi_i). \tag{2.5}$$

For convenience, with $W(g) \in \mathbb{R}^N$ we represent a vector that collects all the values of g at the quadrature points i.e.,

$$(W(g))_i := g(\xi_i), \quad \forall i \in \{1, \dots, N\}. \tag{2.6}$$

With the above two simplifications, the optimization problem in (2.4) transforms to an optimization problem for $W(g_M)$ given as

$$W(g_M) = \underset{W^* \in \mathbb{R}^N}{\arg\min} \frac{1}{2} \|W^*\|^2_{l^2} : \ \underline{ALW^* = \lambda}, \ W^* > 0. \tag{2.7}$$

To write down the moment constraint (the underlined term) in the above problem, we have used the relation

$$\langle P_M g^* \rangle \approx \langle P_M g^* \rangle_N = ALW(g^*), \tag{2.8}$$

where the matrices $A \in \mathbb{R}^{M\times N}$ and $L \in \mathbb{R}^{N\times N}$ are given as

$$A := (P_M(\xi_1), \ldots, P_M(\xi_N)), \quad L_{ij} = \begin{cases} \omega_i, & i = j \\ 0, & i \neq j \end{cases}. \tag{2.9}$$

Thus, L is a diagonal matrix containing the quadrature weights $\{\omega_i\}$ at its diagonal, and A is a Vandermonde matrix. Note that in (2.7), for notational simplicity, we defined $W^* = W(g^*)$.

Remark 1 (A DG discretization). *To see the similarity between the pos-L2-MOM and a DG velocity space discretization and understand our motivation behind considering L2-minimization, consider the optimization problem*

$$g_M^{DG} := \underset{g^* \in L^2(\Omega_\xi)}{\arg\min} \frac{1}{2}\|g^*\|^2_{L^2(\Omega_\xi)} : \int_{\Omega_\xi} P_M g^* d\xi = \lambda.$$

The above problem is a continuous-in-velocity analogue of (2.4) *but without positivity constraints. Using the first order-optimality conditions, one can conclude that a solution to the above problem is given as (see page-2611 of [18] for a proof related to the PN closure)*

$$g_M^{DG} = \alpha^T P_M,$$

where α is a vector of expansion coefficients related linearly to the moment vector λ—the exact form of α is not important here. The above expansion is the same as the DG velocity discretization proposed in [3]. Thus, one can interpret pos-L2-MOM as a DG velocity discretization with positivity constraints. Note that the optimization problem corresponding to g_M^{DG} does not penalize the negativity of a solution therefore, g_M^{DG} is not necessarily positive on the quadrature points.

Remark 2 (A Hermite expansion). *One can also interpret a Hermite approximation to a pdf as a solution to a weighted L2-minimization problem—we refer to [15, 43] for an exhaustive discussion on Hermite expansions. Let $p_m(\xi)$ denote the m-th order Hermite polynomial $He_m(\xi)$. Normalize the Hermite polynomials such that they are orthogonal under the inner-product of the weighted L2-space $L^2(\mathbb{R}, \exp(-\xi^2/2))$. Let P_M be as defined in* (2.1)*. Note that instead of monomials, the vector $P_M(\xi)$ now contains Hermite polynomials.*

Consider a weighted L2-minimization problem given as

$$g_M^H := \underset{g^* \in L^2(\mathbb{R}, \exp(\xi^2/2))}{\arg\min} \frac{1}{2}\|g^*\|^2_{L^2(\mathbb{R}, \exp(\xi^2/2))} : \langle P_M g^* \rangle = \lambda.$$

Note that as compared to a DG approximation, in the above optimization problem, we did not truncate the velocity domain. One can show that the solution to the above minimization problem is given as

$$g_M^H = \lambda^T P_M \exp(-\xi^2/2),$$

which is similar to the Hermite spectral method proposed in [15]. Using the same methodology as for the L2-minimization, one can impose positivity constraints in the above minimization problem and enforce them on a set of Gauss-Hermite quadrature points. We leave the development of a positive weighted L2-minimization based moment method as a part of our future work.

2.1.1 Feasibility of the positive L2-minimization If there exists a $z > 0$ such that $\lambda = ALz$ then the optimization problem in (2.7) is feasible with the feasible point $W^* = z$. We collect this simple, but noteworthy, result as follows. We first define a set of realizable moments

$$R := \{\lambda \ : \ \lambda \in \mathbb{R}^M,\ \lambda = ALz,\ z > 0\}. \tag{2.10}$$

Using R, we collect our statement related to the feasibility of the optimization problem.

LEMMA 2.1 (Feasibility of the optimization problem). *The optimization problem in* (2.7) *is feasible if* $\lambda \in R$.

Note that for a given $\lambda \in R$, the number of feasible points of the optimization problem vary depending upon the value of N relative to M. Let $z > 0$ be such that $\lambda = ALz$. A feasible point W^* of the optimization problem (2.7) is a positive solution of the linear system

$$ALW^* = ALz.$$

Since AL is a full-rank matrix (A is a Vandermonde matrix and the Gauss-Legendre quadrature weights are positive), the above linear system has a unique solution $W^* = z$ for $N \leq M$. Thus, the optimization problem has a single feasible point for $N \leq M$. In contrast, the above linear system has infinitely many positive solutions for $N > M$, resulting in infinitely many feasible points.[1]

The above discussion indicates that for $N \leq M$, we do not need to perform L2-minimization. A unique positive $W(g_M)$ can be recovered by solving the moment constraint $ALW(g_M) = \lambda$. However, for $N \leq M$, a moment-based approach is meaningless because we can directly compute $W(g_M)$ using a discrete-velocity-method (DVM). Since $N \leq M$, this would be less expensive than first computing λ and then computing $W(g_M)$ using the optimization problem. Therefore, in the following discussion we only consider $N > M$. The discussion here becomes clearer when we later relate our moment approximation to a DVM.

REMARK 3 (Practical considerations while choosing N). *Practical considerations suggest a compromise between small and large values of N. We use an inter-convex-set algorithm to solve the minimization problem in* (2.7). *A crude estimate for the complexity of this algorithm is* $\mathcal{O}(N^3)$ *[42]. Thus, choosing a large value of N increases the computational cost of solving the optimization problem, which, as we discuss later, is the most expensive part of our moment approximation. On the contrary, we do not want N to be so small that the error (measured in some norm) in our moment approximation is dominated by the error in our quadrature approximation. Numerical experiments suggest that choosing N between $2M$ and $5M$ is a good compromise between accuracy and efficiency.*

2.2 The Boltzmann Equation (BE) Equipping the BE with initial and boundary data provides

$$\begin{aligned} \mathcal{L}(f) = 0 \text{ on } \Omega \times D \times \mathbb{R}, \quad f(\cdot, t = 0, \cdot) = f_0 \text{ on } \Omega \times \mathbb{R}, \\ f = f_{in} \text{ on } \partial\Omega_- \times D. \end{aligned} \tag{2.11}$$

[1] Let W^* be a solution to $ALW^* = ALz$. Let v be an element of the null-space of AL—since AL is a flat matrix, its null-space is non-empty. Then, for all β such that $\min_i(\beta v_i) > -\min_i(w_i)$, we find that $W^* + \beta v$ is also a feasible point.

Above, the spatial domain is given as $\Omega := [x_{\min}, x_{\max}]$, and $\partial\Omega_-$ is the inflow part of the boundary that reads

$$\partial\Omega_- := \{(x, \xi) \ : \ \xi \cdot n(x) \leq 0, \ x \in \partial\Omega\}, \tag{2.12}$$

where $n(x)$ is a unit normal at $x \in \partial\Omega$ that points out of the domain. For simplicity, we consider only inflow type boundary conditions and not wall boundary conditions i.e., f_{in} is the given data and is independent of the solution f [12]. An inflow type boundary simplifies our result related to the stability of the moment approximation discussed later. With some additional technical details, one can extend our stability result to solid-wall boundaries—see [28] for stability results related to a solid-wall boundary for a Grad's moment method.

We normalise f such that the density ρ, the velocity v and the temperature θ (in energy units) reads

$$\begin{pmatrix} \rho(x,t) \\ \rho(x,t)v(x,t) \\ \rho(x,t)\left(\theta(x,t) + v(x,t)^2\right) \end{pmatrix} := \langle P_{\text{cons}} f(x,t,\cdot)\rangle, \quad P_{\text{cons}}(\xi) := \begin{pmatrix} 1 \\ \xi \\ \xi^2 \end{pmatrix}. \tag{2.13}$$

Note that for $M \geq 3$, $P_{\text{cons}}(\xi)$ is nothing but the first three entries of $P_M(\xi)$.

We consider a Boltzmann-BGK collision operator given as

$$Q(f(x,t,\xi)) := \frac{1}{\tau(x,t)}(f_{\mathcal{M}}(x,t,\xi) - f(x,t,\xi)), \tag{2.14}$$

where the collision frequency $\tau(x,t)^{-1}$ reads $\tau(x,t)^{-1} := C\rho(x,t)\theta(x,t)^{1-\omega}$ with ω begin the exponent in the viscosity law of the gas [13]. The collision operator represents the fact that the pdf $f(x,t,\cdot)$ is pushed towards the Maxwell-Boltzmann pdf $f_{\mathcal{M}}(x,t,\cdot)$ given as

$$f_{\mathcal{M}}(x,t,\xi) := \frac{\rho(x,t)}{\sqrt{2\pi\theta(x,t)}} \exp\left(-\frac{(\xi - v(x,t))^2}{2\theta(x,t)}\right). \tag{2.15}$$

We can also interpret $f_{\mathcal{M}}$ as a solution to an entropy-minimization problem. Out of all the pdfs that have the same mass, momentum and energy as $f(x,t,\cdot)$, the pdf $f_{\mathcal{M}}(x,t,\cdot)$ is the one that minimizes the Boltzmann's entropy. Equivalently,

$$f_{\mathcal{M}}(x,t,\cdot) = \underset{f^*(\xi)\geq 0}{\arg\min} \left\{\langle f^* \log(f^*)\rangle \ : \ \langle P_{\text{cons}} f^*\rangle = \langle P_{\text{cons}} f(x,t,\cdot)\rangle\right\}. \tag{2.16}$$

Later, we use the above interpretation of $f_{\mathcal{M}}$ to discretize it on a velocity grid. A noteworthy property of $Q(f)$ is its collision invariance i.e., $\langle P_{\text{cons}} Q(f)\rangle = 0$ for all f in the domain of Q. This ensures that the BE conserves mass, momentum and energy. By considering $M \geq 3$, which ensures that $P_{\text{cons}}(\xi)$ is contained in the vector $P_M(\xi)$, and by carefully discretizing the collision operator as in [22], we will ensure that our moment system also conserves these quantities.

2.3 Moment equations We present a moment approximation to the BE based upon the pos-L2-MOM described in Subsection 2.1. To derive a governing equation for the moments $\langle P_M f(x,t,\cdot)\rangle$, we take (discrete) velocity moments of the BE given in (2.11) to find

$$\partial_t \langle P_M f(x,t,\cdot)\rangle_N + \partial_x \underline{\langle P_M \xi f(x,t,\cdot)\rangle_N} = \langle P_M Q(f(x,t,\cdot))\rangle_N. \tag{2.17}$$

Recall that $\langle\cdot\rangle_N$ is as defined in (2.5) and is a numerical approximation to the integral $\langle\cdot\rangle$.

The above system of equations is not closed—the underlined flux-term contains a M-order moment that is not contained in the moment vector $\langle P_M f(x,t,\cdot)\rangle_N$. To close the system of equations, using the moments $\langle P_M f(x,t,\cdot)\rangle_N$, we need to approximate the values of $f(x,t,\cdot)$ at the quadrature points i.e., we need to approximate the vector $W(f(x,t,\cdot))$ using the moments $\langle P_M f(x,t,\cdot)\rangle_N$. We approximate $W(f(x,t,\cdot))$ by $W(f_M(x,t,\cdot))$. To compute $W(f_M(x,t,\cdot))$, we use the L2-minimization problem given in (2.7) with the moment vector λ set to $\langle P_M f(x,t,\cdot)\rangle_N$. This results in the following closed set of moment equations

$$\begin{aligned} \partial_t \langle P_M f_M\rangle_N + \partial_x \langle P_M \xi f_M\rangle_N &= \frac{1}{\tau}(\langle P_M f_{\mathcal{M},N}\rangle_N - \langle P_M f_M\rangle_N) \text{ on } \Omega\times D,\\ \langle P_M f_M(t=0)\rangle_N &= \langle P_M f_0\rangle_N \text{ on } \Omega. \end{aligned} \tag{2.18}$$

Our space-time discretization Subsection 3.2 will discuss the boundary discretization. Let us emphasis again that to compute the flux term $\langle P_M \xi f_M(x,t,\cdot)\rangle_N$, we only need the value of $W(f_M(x,t,\cdot))$, which are available after solving the L2-minimization problem.

The pdf $f_{\mathcal{M},N}$ is an approximation to the Maxwell-Boltzmann pdf $f_{\mathcal{M}}$ and is such that $W(f_{\mathcal{M},N})$ is a solution to an entropy-minimization problem given as [22]

$$W(f_{\mathcal{M},N}(x,t,\cdot)) = \underset{W^*\in\mathbb{R}^N_{>0}}{\arg\min}\left\{\sum_i w_i^* \log(w_i^*)\omega_i \;:\; A_{\text{cons}} L W^* = \langle P_{\text{cons}} f_M(x,t,\cdot)\rangle_N\right\}. \tag{2.19}$$

The above problem is a discrete-in-velocity analogue of the entropy minimization problem given in (2.16). Furthermore, the moment constraints in the minimization problem ensure that the moment system (2.18) conserves mass, momentum and energy.

2.4 Computing the velocity cut-off Recall that we truncate the velocity domain $\mathbb{R}$ to $\Omega_\xi = [\xi_{\min}, \xi_{\max}]$. We use the same technique as a DVM to compute the velocity cut-off $\xi_{\max/\min}$. The technique is summarised as follows—for further details, we refer to [7, 22] and the references therein. Estimating $\xi_{\max/\min}$ using the velocity and the temperature of the gas provides

$$\begin{aligned} \xi_{\min} &:= \inf_{(x,t)\in\Omega\times D}\left(v(x,t) - c\sqrt{\theta(x,t)}\right),\\ \xi_{\max} &:= \sup_{(x,t)\in\Omega\times D}\left(v(x,t) + c\sqrt{\theta(x,t)}\right). \end{aligned} \tag{2.20}$$

From arguments in statistical mechanics, a value of c between 3 and 4 is desirable. Choosing $c = 3.5$ balances accuracy and computational cost. During numerical experiments, we compute a reference solution using a DVM. To ensure that the DVM solution is sufficiently refined, we perform a convergence study by first estimating $\xi_{\max/\min}$ using the initial data and the above formulae and then increasing $\xi_{\max}$ (and decreasing $\xi_{\min}$) till the relative error between two subsequent refinements drops below an acceptable value. We use $\xi_{\max}$ from the last refinement cycle for both the DVM and the pos-L2-MOM—Section 5 provides further details. In practical applications, one can estimate $v(x,t)$ and $\theta(x,t)$ using a Navier-Stokes solver, which is usually much cheaper than a BE solver [7].

REMARK 4 (Pros and cons of a space-time-independent $\xi_{\max}$). *Our choice of* $\xi_{\max}$ *(and* $\xi_{\min}$*) is space-time-independent, which has both positive and negative consequences. Such a velocity cut-off can be accurate only if, on the entire space-time domain,* $f(x,t,\cdot)$ *is sufficiently small outside of* Ω_ξ*. In terms of the macroscopic quantities, we can expect to be accurate only for flows with a velocity and a temperature inside a certain range [22]. Let us mention that we share these negative consequences of truncating a velocity domain with the DVM and the entropy-minimization based closures [22, 30]. On the positive side, as we discuss later, with a space-time-independent* $\xi_{\max}$ *it is straightforward to ensure the feasibility of the optimization problem in* (2.7)*. Furthermore, the stability of the moment equations that we establish later can also be attributed to* $\xi_{\max}$ *being fixed in space-time.*

REMARK 5 (A space-time-dependent $\xi_{\max}$). *To overcome the limitations mentioned in the previous remark, similar to [9], one can introduce space-time-dependence in* $\xi_{\max}$*. We failed to introduce this dependence without sacrificing the feasibility of the optimization problem* (2.7) *and the stability result discussed later. To overcome the feasibility issue, one can try modifying the optimization problem by regularizing it [4]. The regularization adds the moment constraint as a penalty term and tries to minimize both the L2-norm of the pdf and the error in satisfying the moment constraint. As for the stability, it is unclear how one can ensure it with a space-time-dependent* $\xi_{\max}$*. We leave the development of pos-L2-MOM with space-time adaptive* $\xi_{\max}$ *as a part of our future work.*

3 Space-time discretization

3.1 Preliminaries We partition $\Omega = [x_{\min}, x_{\max}]$ into N_x intervals given as

$$\Omega = \bigcup_{i=1}^{N_x} \mathcal{I}_i, \quad \mathcal{I}_i = [x_{i-1/2}, x_{i+1/2}], \tag{3.1}$$

where $x_{1/2} = x_{\min}$ and $x_{N_x+1/2} = x_{\max}$. With $\{t_i\}_{i=1,\ldots,K} \subset D$ we represent a set of discrete time instances such that $0 = t_1 < t_2 \cdots < t_K = T$. For simplicity of notation, we assume that all the space and the time intervals are of the same size Δx and Δt, respectively. An extension to non-uniform space-time grids is straightforward. We denote the finite volume (FV) approximation of $\langle P_M f_M(x,t,\cdot)\rangle$ and $\langle P_M f_{\mathcal{M},N}(x,t,\cdot)\rangle$ in the i-th cell and at the k-th time instance by

$$\begin{aligned} \left\langle P_M f_i^k \right\rangle_N &\approx \frac{1}{\Delta x}\int_{\mathcal{I}_i} \langle P_M f_M(x,t_k,\cdot)\rangle_N \, dx, \\ \left\langle P_M f_{\mathcal{M},i}^k \right\rangle_N &\approx \frac{1}{\Delta x}\int_{\mathcal{I}_i} \langle P_M f_{\mathcal{M},N}(x,t_k,\cdot)\rangle_N \, dx. \end{aligned} \tag{3.2}$$

Above, $f_{\mathcal{M},N}$ is the discretization of the Maxwell-Boltzmann distribution introduced in (2.16) and for notational simplicity, we have suppressed the M dependence in f_i^k.

Using the matrix A and L given in (2.9), we can express the space-time discrete moments in a matrix-vector product form as

$$\left\langle P_M f_i^k \right\rangle_N = ALW(f_i^k), \quad \left\langle P_M f_{\mathcal{M},i}^k \right\rangle = ALW(f_{\mathcal{M},i}^k), \tag{3.3}$$

where $W(f_i^k)$ and $W(f_{\mathcal{M},i}^k)$ are the FV-approximations to $W(f_M(x,t_k,\cdot))$ and $W(f_{\mathcal{M}}(x,t_k,\cdot))$, respectively, in the i-th cell and at the k-th time step. For later convenience, with f_{M,N_x} we represent an FV approximation to f_M defined as

$$f_{M,N_x}(x,t_k,\xi) = f_i^k(\xi), \quad \forall x \in \mathcal{I}_i, k \in \{1,\ldots,K\}, \xi \in \{\xi_i\}_i. \tag{3.4}$$

3.2 Evolution scheme The evolution scheme consists of four steps outlined below. We present these steps for some representative $t = t_k$. Each step is repeated from $k = 1$ to $k = K - 1$. For $k = 1$, we initialize with

$$\left\langle P_M f_i^k \right\rangle_N = \frac{1}{\Delta x} \int_{\mathcal{I}_i} \langle P_M f_0(x, \cdot) \rangle_N \, dx, \quad \forall i \in \{1, \dots, N_x\}, \tag{3.5}$$

where f_0 is the initial data in (2.11). We approximate the above space integral with 10 Gauss-Legendre quadrature points in each cell.

1. *Entropy-minimization step:* Using the conserved moments $\{\left\langle P_{\text{cons}} f_i^k \right\rangle_N\}_i$, solve the entropy minimization problem in (2.19). This provides the discrete Maxwell-Boltzmann pdf $\{f_{\mathcal{M},i}^k\}_i$.
2. *Collision step:* With the output of the previous step, perform collisions with an implicit Euler time-stepping scheme. At some intermediate $t_{k^*} \in (t_k, t_{k+1})$ and for all $i \in \{1, \dots, N_x\}$, this provides [14]

$$\frac{\left\langle P_M f_i^{k^*} \right\rangle_N - \left\langle P_M f_i^k \right\rangle_N}{\Delta t} = \frac{1}{\tau(x_i, t_{k^*})} \left(\left\langle P_M f_{\mathcal{M},i}^{k^*} \right\rangle_N - \left\langle P_M f_i^{k^*} \right\rangle_N \right). \tag{3.6}$$

There is an explicit solution to the above implicit collision step. Since the collision step preserves mass, moment and energy and since the solution of the entropy-minimization problem (2.19) is unique for a given set of conserved moments, we find $W(f_{\mathcal{M},i}^{k^*}) = W(f_{\mathcal{M},i}^k)$. This implies that $\left\langle P_M f_{\mathcal{M},i}^{k^*} \right\rangle_N = \left\langle P_M f_{\mathcal{M},i}^k \right\rangle_N$, which provides

$$\begin{aligned} \left\langle P_M f_i^{k^*} \right\rangle_N =& \frac{1}{1 + \Delta t/\tau(x_i, t_{k^*})} \left\langle P_M f_i^k \right\rangle_N \\ &+ \frac{\Delta t/\tau(x_i, t_{k^*})}{1 + \Delta t/\tau(x_i, t_{k^*})} \left\langle P_M f_{\mathcal{M},i}^k \right\rangle_N . \end{aligned} \tag{3.7}$$

3. *Optimization step:* Using the moments $\{\left\langle P_M f_i^{k^*} \right\rangle_N\}_i$, compute the weights $\{W(f_i^{k^*})\}_i$ by solving the optimization problem in (2.7).
4. *Transport step:* Using the output of the previous step, perform the transport step given as

$$\begin{aligned} \frac{\left\langle P_M f_i^{k+1} \right\rangle_N - \left\langle P_M f_i^{k^*} \right\rangle_N}{\Delta t} = -\frac{1}{\Delta x} \Big(& \mathcal{F}(W(f_{i+1}^{k,*}), W(f_i^{k,*})) \\ & - \mathcal{F}(W(f_i^{k,*}), W(f_{i-1}^{k,*})) \Big). \end{aligned} \tag{3.8}$$

To impose boundary conditions, for $i = 1$, set $W(f_{i-1}^{k,*}) = W(f_{in}(t, \cdot))$ and for $i = N_x$, set $W(f_{i+1}^{k,*}) = W(f_{in,N}(t, \cdot))$, where f_{in} is the boundary data given in (2.11). Above, $\mathcal{F} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}^M$ is the numerical flux and since we consider a kinetic upwind numerical flux, it reads [1]

$$\mathcal{F}(W_1, W_2) := \frac{1}{2} \left(AL(\Xi - |\Xi|)W_1 + AL(\Xi + |\Xi|)W_2 \right). \tag{3.9}$$

Above, A and L are the two matrices defined in (2.9). The matrix Ξ is a diagonal matrix with the locations of the quadrature points $\{\xi_i\}_i$ at its diagonal. Furthermore, $|\Xi|$ is a matrix representing the absolute value of Ξ in the sense that $(|\Xi|)_{ij} = |\Xi_{ij}|$. For clarity, to express $\mathcal{F}$ in a standard kinetic upwind flux form, note that $AL(\Xi \pm |\Xi|)W_1 = \langle P_M(\xi \pm |\xi|) f_1 \rangle_N$.

REMARK 6 (Space-time locality of the optimization step). *The optimization step (and also the entropy minimization step) is a local in space-time operation. We loop over each spatial cell, solve the optimization problem, add the local contributions to the numerical flux and move over to the next cell. Therefore, at any given point in time, we store only the moments in all the spatial cells and not the values of the pdf at the quadrature points. This results in a drastic reduction in memory consumption since, in practice, the number of quadrature points are much larger than the number of moments—see [29, 30] for a similar comment related to a maximum-entropy closure. Let us emphasis that in comparison, a DVM stores the values of the pdf at the quadrature points in all the cells, which, particularly for multi-dimensional velocity domain, results in a memory intensive algorithm [7].*

3.3 Properties of the evolution scheme The entropy-minimization problem in (2.19) ensures that our moment approximation conserves mass, moment and energy. In addition to being conservative, the following discussion establishes that our space-time discrete moment approximation (i) under a CFL-type condition, results in a feasible optimization problem; and (ii) is L2 stable in the sense that the L2-energy $\sum_{i=1}^{N_x}\|\langle P_M f_i^k\rangle_N\|_{l^2}^2$ has an upper-bound that depends solely on the initial data f_0 and the boundary data f_{in}.

We start with making the following assumptions on the initial and the boundary data. We assume that the first M-moments of f_0 and f_{in} belong to the realizability set R defined in (2.10) i.e.,

$$\langle P_M f_{in}(x,t,\cdot)\rangle_N \in R, \quad \langle P_M f_0(x,\cdot)\rangle_N \in R, \quad \forall (x,t)\in \Omega\times D. \tag{3.10}$$

The above assumption will be helpful in establishing the feasibility of the optimization problem in the optimization step. For the boundary data, we also assume that

$$\begin{gathered}|f_{in}(\cdot,t,\cdot)|_{\partial\Omega,N} < \infty, \quad \forall t\in D,\\ \text{where } |f_{in}(\cdot,t,\cdot)|^2_{\partial\Omega,N} := \sum_{\xi_i\cdot n(x)\leq 0}\oint_{\partial\Omega}|\xi_i\cdot n(x)| f_{in}(x,t,\xi_i)^2\omega_i ds.\end{gathered} \tag{3.11}$$

Above, the unit vector $n(x)$ is as given in (2.12), and $\{\xi_i\}$ and $\{\omega_i\}_i$ are the abscissas and the weights of the quadrature points, respectively. Note that the assumption on $|f_{in}(\cdot,t,\cdot)|_{\partial\Omega,N}$ is a discrete-in-velocity analogue of a standard assumption that $f_{in}(\cdot,t,\cdot)\in L^2(\partial\Omega_-,|\xi\cdot n(x)|)$—see [40] for further details. Here, $L^2(\partial\Omega_-,|\xi\cdot n(x)|)$ represents a L^2 space over $\partial\Omega_-$ with the Lebesgue measure $|\xi\cdot n(x)|$, and the set $\partial\Omega_-$ contains all the incoming velocities and is as defined in (2.12). Intuitively, the above assumption states that the total L2-energy flux associated with f_{in} should be bounded. We insist that the above assumptions are valid for most applications of practical relevance.

3.4 Feasibility of the optimization problem We show that under a CFL-condition, the moments resulting from the collision step and the transport step belong to the realizability set R given in (2.10) i.e., both the steps are realizability preserving. The feasibility of the optimization problem then follows from Lemma 2.1. The details are as follows.

Our result is a straightforward extension of the proof for the realizability preserving space-time discretization of radiative transport equations considered in [6]. Using the definition of R given in (2.10), we find

$$a_1\lambda_1 + a_2\lambda_2 \in R, \quad \forall a_1,a_2\geq 0,\ \lambda_1,\lambda_2\in R. \tag{3.12}$$

We consider the collision step given in (3.7). Suppose that $\langle P_M f_i^k \rangle \in R$, which implies that entropy-minimization step is well-posed (see [22]) and that $\langle P_M f_{\mathcal{M},i}^k \rangle \in R$. Then, the above relation implies that for any $\Delta t, \tau(x_i, t_k) > 0$, the collision step is realizability preserving i.e., for all $i \in \{1, \ldots, N_x\}$, we have $\langle P_M f_i^{k^*} \rangle \in R$.

We show that under a CFL-condition, the transport step in (3.8) is also realizability preserving. Replacing the numerical flux function from (3.9) in the transport step given in (3.8) and re-arranging a few terms provides

$$\begin{aligned} \langle P_M f_i^{k+1} \rangle_N = & AL(1 - \Lambda|\Xi|)W(f_i^{k^*}) \\ & + \underline{\frac{\Lambda}{2} AL(|\Xi| - \Xi)W(f_{i+1}^{k^*})} + \underline{\frac{\Lambda}{2} AL(|\Xi| + \Xi)W(f_{i-1}^{k^*})}. \end{aligned} \tag{3.13}$$

where $\Lambda := \frac{\Delta t}{\Delta x}$. For all $i \in \{1, \ldots, N_x\}$, due to the positivity constraints in the optimization problem (2.7), we have $W(f_i^{k^*}) > 0$, which, for $\Lambda > 0$, implies that the underlined terms are in R. To ensure that the first term on the right is in R, we choose

$$0 < \Lambda \leq \min\{|\xi_{\text{max}}^{-1}|, |\xi_{\text{min}}^{-1}|\}. \tag{3.14}$$

The above range of Λ, the relation in (3.12) and the assumption on the initial and the boundary data (3.10) provides $\langle P_M f_i^{k+1} \rangle_N \in R$. We collect our findings in the result below.

LEMMA 3.1. *Consider the evolution scheme outlined in Subsection 3.2 and define $\Lambda = \Delta t/\Delta x$. Assume that the initial and the boundary data satisfies (3.10), then the quadratic optimization problem in the evolution scheme is feasible if $\Lambda \in (0, \min\{|\xi_{\text{max}}^{-1}|, |\xi_{\text{min}}^{-1}|\}]$.*

3.5 $L2$ stability of the scheme Define the total L2-energy at $t = t_{k+1}$ as

$$\mathcal{E}_{k+1} := \sum_{i=1}^{N_x} \| \langle P_M f_i^{k+1} \rangle_N \|_{l^2}^2. \tag{3.15}$$

We establish that $\mathcal{E}_{k+1}$ is bounded by the L2-energy of the previous time-step $\mathcal{E}_k$ and $|f_{in}(\cdot, t_k, \cdot)|_{\partial\Omega, N}$. Recursion then implies that $\mathcal{E}_{k+1}$ is bounded solely by the initial and the boundary data.

For convenience, we define a few objects. For a vector $z \in \mathbb{R}^N$, with $\|z\|_L$ we represent the norm

$$\|z\|_L := \sqrt{z^T L z}.$$

Interpreting z as a vector that contains the value of a function $g : \Omega_\xi \to \mathbb{R}$ at the quadrature points and recalling that L is a diagonal matrix with the quadrature weights on its diagonal, we conclude that $\|z\|_L$ represent an approximation to $\|g\|_{L^2(\Omega_\xi)}$. We bound the l^2-norm of a moment vector $\lambda = ALz$ as

$$\|\lambda\|_{l^2} \geq \sigma_{\min}(A\sqrt{L})\|z\|_L, \quad \|\lambda\|_{l^2} \leq \sigma_{\max}(A\sqrt{L})\|z\|_L, \tag{3.16}$$

where $\sigma_{\min/\max}(A\sqrt{L})$ represent the minimum/maximum singular value of the matrix $A\sqrt{L}$. We will use the above two bounds to convert stability results for the DVM to stability results for the moment approximation.

3.5.1 Collision step We start with the collision step given in (3.6). Applying triangle's inequality to the collision step we find

$$\mathcal{E}_{k^*} \leq \frac{2}{(1+\Delta t/\tau)^2}\mathcal{E}_k + 2\left(\frac{\Delta t/\tau}{1+\Delta t/\tau}\right)^2 \sum_{i=1}^{N_x} \underbrace{\|\langle P_M f_{\mathcal{M},i}^k\rangle_N\|_{l^2}^2}_{\leq \sigma_{\max}(A\sqrt{L})^2 \|W(f_{\mathcal{M},i}^k)\|_L^2} . \tag{3.17}$$

The bound on the right hand side follows from the inequalities in (3.16). From page-92 of [23] we know that the solution to the entropy-minimization problem (2.16) satisfies

$$\|W(f_{\mathcal{M},i}^k)\|_L^2 \leq N^3 \exp(2Nt_k). \tag{3.18}$$

The above relation and the bound on $\mathcal{E}_{k^*}$ given in (3.17) provides

$$\mathcal{E}_{k^*} \leq \frac{2}{(1+\Delta t/\tau)^2}\mathcal{E}_k + 2\left(\frac{\Delta t/\tau}{1+\Delta t/\tau}\right)^2 N_x \sigma_{\max}(A\sqrt{L})^2 N^3 \exp(2Nt_k). \tag{3.19}$$

3.5.2 Transport step With the following three steps, we establish the stability of the transport step given in (3.8). (i) We recover a DVM underlying the transport step in (3.8). (ii) Using stability properties of an upwind scheme, we establish the stability of the DVM. (iii) Finally, relating the discrete velocity solution to the moment solution, we establish the stability of the moment scheme. The details of these three steps is as follows.

We consider the reformulated transport step given in (3.13). Let $\mathcal{N}(AL)$ represent the null-space of the matrix AL, where A and L are as given in (2.8) and (2.9), respectively. Then, the transport step provides

$$\begin{aligned} W(f_i^{k+1}) =& (1-\Lambda|\Xi|)W(f_i^{k^*}) \\ &+ \frac{\Lambda}{2}(|\Xi|-\Xi)W(f_{i+1}^{k^*}) + \frac{\Lambda}{2}(|\Xi|+\Xi)W(f_{i-1}^{k^*}) + v, \end{aligned} \tag{3.20}$$

where v belongs to $\mathcal{N}(AL)$. Since the moments at time step t_{k+1}—given as $ALW(f_i^{k+1})$—are invariant under the choice of v, we choose $v = 0$. This makes the above evolution equation a space-time discretization of a system of decoupled linear advection equations given as $\partial_t W(f) + \Xi\partial_x W(f) = 0$. The discretization uses an explicit Euler and an upwind FV scheme to discretize the space and the time domain, respectively. From Example-7.2 of [33] we know that such a discretization is L2-stable under the CFL condition

$$0 < \Lambda \leq \min\{|\xi_{\max}^{-1}|, |\xi_{\min}^{-1}|\}/2. \tag{3.21}$$

This provides

$$\sum_{i=1}^{N_x}\|W(f_i^{k+1})\|_L^2 \leq \sum_{i=1}^{N_x}\|W(f_i^{k^*})\|_L^2 + |f_{in}(\cdot,t_k,\cdot)|_{\partial\Omega,N}^2. \tag{3.22}$$

Above, $|\cdot|_{\partial\Omega,N}$ is as defined in (3.11). Using the bounds in (3.16) , we express the above bound in terms of moments to find

$$\mathcal{E}_{k+1} \leq \kappa(A\sqrt{L})^2\mathcal{E}_{k^*} + \sigma_{\max}(A\sqrt{L})^2 |f_{in}(\cdot,t_k,\cdot)|_{\partial\Omega,N}^2. \tag{3.23}$$

Above, $\kappa(A\sqrt{L})$ represents the condition number of the matrix $A\sqrt{L}$. We collect our stability estimate in the result below.

THEOREM 3.2. *Consider the evolution scheme given in Subsection 3.2 and let $\mathcal{E}_k$ be the L2-energy defined in (5.5). Assume that the boundary data satisfies (3.10) and that the ratio $\Lambda = \Delta t/\Delta x$ satisfies $\Lambda \in (0, \min\{|\xi_{\max}^{-1}|, |\xi_{\min}^{-1}|\}/2]$. Then, $\mathcal{E}_{k+1}$ is bounded as*

$$\mathcal{E}_{k+1} \leq \mathcal{B}_k + \mathcal{B}_{\mathcal{M}} + \mathcal{B}_{in} \tag{3.24}$$

where

$$\begin{aligned}
\mathcal{B}_k :=& \kappa(A\sqrt{L})^2 \frac{2}{(1+\Delta t/\tau)^2}\mathcal{E}_k, \\
\mathcal{B}_{\mathcal{M}} :=& 2\kappa(A\sqrt{L})^2 \sigma_{\max}(A\sqrt{L})^2 \left(\frac{\Delta t/\tau}{1+\Delta t/\tau}\right)^2 N_x N^3 \exp(2Nt_k), \\
\mathcal{B}_{in} :=& \sigma_{\max}(A\sqrt{L})^2 |f_{in}(\cdot, t_k, \cdot)|^2_{\partial\Omega, N}.
\end{aligned} \tag{3.25}$$

We make the following remarks related to the above theorem.

1. The terms $\mathcal{B}_k$, $\mathcal{B}_{\mathcal{M}}$ and $\mathcal{B}_{in}$ appearing in (3.24) represent the contribution from the previous time step, the discrete Maxwell-Boltzmann distribution function and the boundary data, respectively, into bound for the L2-energy at time t_{k+1}. Note that out of all these three terms, only $\mathcal{B}_k$ depends upon the solution of the previous time-step.
2. For the limit $\tau \to 0$, at least formally, the BE results in the Euler equations [13]. Under this limit, the bound in (3.24) is robust, which is a result of performing the collision step implicitly.
3. The DVM corresponding to the transport step given in (3.20) is a space-time discretization of a linear hyperbolic PDE. As a result, the L2-bound for the transport step (given in (3.23)) is linear in time. In contrast, since the collision operator is non-linear, the collision step is non-linear. One can attribute this non-linearity to the exponential-in-time growth in $\mathcal{B}_{\mathcal{M}}$.
4. For a fixed truncated velocity domain Ω_ξ, consider the limit $N, M \to \infty$ with $N > M$. Under this limit, the bound on $\mathcal{E}_{k+1}$ is not robust because—at least heuristically—both $\kappa(A\sqrt{L})$ and $\sigma_{\max}(A\sqrt{L})$ are almost independent of N and grow polynomially with M. To derive bounds that are independent of $\kappa(A\sqrt{L})$ and $\sigma_{\max}(A\sqrt{L})$, one should directly consider the moment approximation without accessing the underlying DVM. As yet, it is unclear how to proceed with such a technique.
5. Nowhere in the proof of the above theorem we used the fact that we minimize the L2-norm in the moment-closure problem given in (2.7). Therefore, the bound on $\mathcal{E}_{k+1}$ holds for any other objective functional and specifically for the minimum-entropy closure considered in [16, 30].

3.6 Computational costs We study the cost of evolution scheme outlined in Subsection 3.2. We consider the cost of a single time-step performed in a single spatial cell.

1. *Entropy-minimization step:* We use Newton iteration to solve the entropy-minimization problem where we compute and invert a Hessian $H(x,t)$ given as

$$(H(x,t))_{kl} := \sum_i \left(P_{\text{cons}}(\xi_i)\right)_k \left(P_{\text{cons}}(\xi_i)\right)_l \exp(P_{\text{cons}} \cdot \alpha(x,t))\omega_i, \tag{3.26}$$

where $\alpha(x,t)$ are the Lagrange multipliers in $\mathbb{R}^3$. Computing the Hessian is an $\mathcal{O}(N)$ operation. As a stopping criterion to the Newton solver, we consider a user-defined tolerance of `TOL` in the moment constraints. Suppose we need $m_{\texttt{TOL}}$ Newton iterations to reach this tolerance then, the total cost of entropy-minimization is given as

$$C_{\text{entropy}} = \mathcal{O}(N m_{\texttt{TOL}}).$$

In all our numerical examples, we choose $\texttt{TOL} = 10^{-8}$.

2. *Collision step:* Computing the M-moments of the discrete Maxwell-Boltzmann pdf is an $\mathcal{O}(NM)$ operation and updating the moments in the collision step is an $\mathcal{O}(M)$ operation. Thus, the total cost of the collision step is given as

$$C_{\text{col}} = \mathcal{O}(MN).$$

3. *Optimization step:* We use the `quadprog` routine from matlab to solve the optimization problem in (2.7) and we use the default interior-point-convex solver with all the parameters set to their default values. Usually, it is difficult to estimate the complexity of this algorithm but a crude estimate gives [42]

$$C_{\text{opt}} = \mathcal{O}(N^3). \tag{3.27}$$

4. *Transport step:* Flux computation is an $\mathcal{O}(MN)$ operation and the time update of the moments is an $\mathcal{O}(M)$ operation. Thus the cost of the transport step is

$$C_{\text{tran}} = \mathcal{O}(MN).$$

Summing up the above costs, the total cost of our evolution scheme is given as

$$C_{\text{total}} = \mathcal{O}(N m_{\texttt{TOL}}) + \mathcal{O}(MN) + \mathcal{O}(N^3).$$

REMARK 7 (Efficiency of the optimization step). *For $N > M$ (the values of N that interest us, see Remark 3) and a sufficiently small $m_{\texttt{TOL}}$, solving the quadratic optimization problem is the most expensive part of the algorithm. A possible way to overcome this high cost is to train an auto-encoder/gaussian-regression to replace the quadratic optimization problem [17, 25]. We plan to consider this direction in the future.*

4 Extension to multi-dimensions Maintaining consistency with our numerical experiments, we propose an extension of our method to two-dimensional planar flows. An extension to three-dimensional problems is similar and is not discussed for brevity. For 2D problems, we reduce the storage requirements by solving for the reduced pdfs h_1 and h_2 given as [41]

$$\begin{aligned} h_1(x,t,\xi_1,\xi_2) &:= \int_{\mathbb{R}} f(x,t,\xi_1,\xi_2,\xi_3)\,d\xi_3, \\ h_2(x,t,\xi_1,\xi_2) &:= \int_{\mathbb{R}} \xi_3^2 f(x,t,\xi_1,\xi_2,\xi_3)\,d\xi_3. \end{aligned} \tag{4.1}$$

In the coming discussion, ξ will represent a velocity vector in $\mathbb{R}^2$ and with $\langle g \rangle$ we will represent the integral of a function $\xi \mapsto g(\xi)$ over $\mathbb{R}^2$. To derive the governing equation for h_1 and h_2, we multiply the BE given in (1.1) by 1 and ξ_3^2 and integrate over $\mathbb{R}$ with respect to ξ_3 to find

$$\partial_t h_i + \xi_1 \partial_{x_1} h_i + \xi_2 \partial_{x_2} h_i = \frac{1}{\tau}\left(h_{i,\mathcal{M}} - h_i\right). \tag{4.2}$$

Above, $h_{i,\mathcal{M}}$ represents the reduced Maxwell-Boltzmann pdf and is given as

$$h_{1,\mathcal{M}} = \frac{\rho}{2\pi\theta}\exp\left(-\frac{|\xi - v|^2}{2\theta}\right), \quad h_{2,\mathcal{M}} = \frac{\rho}{2\pi}\exp\left(-\frac{|\xi - v|^2}{2\theta}\right), \tag{4.3}$$

where, $|\cdot|$ is the Eucledian norm of a vector. Note that the mass ρ, the momentum ρv and the temperature θ can be recovered from h_1 and h_2 via

$$\rho = \langle h_1\rangle, \quad \rho v = \langle \xi h_1\rangle, \quad \rho\theta = \frac{1}{3}\left(\langle |\xi|^2 h_1\rangle - \rho|v|^2 + \langle h_2\rangle\right). \tag{4.4}$$

4.1 Moment equations The moment approximation we discuss below is the same for both h_1 and h_2. Therefore, for the simplicity of notation, we present our approximation for some representative h. Similar to the 1D case, we truncate the velocity domain to $\mathbb{R}^2 \supset \Omega_\xi = [\xi_{1,\min}, \xi_{1,\max}] \times [\xi_{2,\min}, \xi_{2,\max}]$. To compute $\xi_{i,\max/\min}$, we adopt the same methodology as that outlined in Subsection 2.4. We consider tensorized $N \times N$ Gauss-Legendre quadrature points inside Ω_ξ. Using these quadrature points, we approximate $\langle\cdot\rangle$ by $\langle\cdot\rangle_{N,N}$.

To derive a governing equation for the moments of h, we first define a polynomial in ξ. With $\beta_M := (\beta_1^M, \beta_2^M) \in \mathbb{R}^2$ we represent a multi-index with each entry being a natural number and the l^1-norm of β_M being equal to M. Using β_M, we define a M-th order polynomial in ξ via $p_{\beta_M} = \xi_1^{\beta_1^M}\xi_2^{\beta_1^M}$. Note that for a given M, β_M is non-unique—for $M = 1$, β_M could either be $(0,1)$ or $(1,0)$. In a vector $P_M(\xi)$, we collect all the polynomials p_{β_M} upto order $M-1$. For completeness, we present the entries in $P_M(\xi)$ for $M = 3$ and $M = 5$.

$$\begin{aligned} &\text{M=3: } P_M(\xi) = \left(1, \xi_1, \xi_2, \xi_1^2, \xi_1\xi_2, \xi_2^2\right)^T; \\ &\text{M=5: } P_M(\xi) = \left(1, \xi_1, \xi_2, \xi_1^2, \xi_1\xi_2, \xi_2^2, \xi_1^3,\right. \\ &\qquad\qquad \left.\xi_1^2\xi_2, \xi_1\xi_2^2, \xi_2^3, \xi_1^4, \xi_1^3\xi_2^1, \xi_1^2\xi_2^2, \xi_1\xi_2^3, \xi_2^4\right)^T. \end{aligned} \tag{4.5}$$

Note that for $M = 3$ and $M = 3$, $P_M(\xi)$ contains 6 and 15 entries, respectively.

For some $M \in \mathbb{N}$, we approximate h by h_M where we compute h_M (more precisely $W(h_M)$) using the L2-minimization problem given in (2.7). To evolve the moments of h_M, we use a multi-dimensional version of the moment system given in (2.18), which reads

$$\begin{aligned} \partial_t \langle P_M h_M\rangle_{N,N} &+ \partial_{x_1}\langle P_M \xi_1 h_M\rangle_{N,N} + \partial_{x_2}\langle P_M \xi_2 h_M\rangle_{N,N} \\ &= \frac{1}{\tau}(\langle P_M h_{\mathcal{M},N}\rangle_{N,N} - \langle P_M h_M\rangle_{N,N}) \text{ on } \Omega \times D, \\ \langle P_M h_M(t=0)\rangle_{N,N} &= \langle P_M h_0\rangle_{N,N} \text{ on } \Omega. \end{aligned} \tag{4.6}$$

Above, $h_{\mathcal{M},N}$ is a discretization of the Maxwell-Boltzmann pdf that results from solving a multi-dimensional version of the optimization problem given in (2.16)—see [22] for an explicit form of this optimization problem. The treatment of boundary conditions is the same as that for the 1D case and is not discussed for brevity.

4.2 Space-time discretization For simplicity, we consider a square spatial domain $\Omega = [x_{1,\min}, x_{1,\max}] \times [x_{2,\min}, x_{2,\max}]$. We discretize Ω with N_x number of uniform elements in each spatial dimension and with Δx we represent the grid spacing. With some additional technical details, it is straightforward to extend our framework to curved domain discretized with unstructured meshes. For simplicity, we consider a fixed time-step of size Δt.

We index a spatial cell with (i, j) where $i, j \in \{1, \ldots, N_x\}$. With $\left\langle P_M h_{i,j}^k \right\rangle_{N,N}$ we represent a FV approximation to $\langle P_M h_M(x, t_k, \cdot)\rangle_{N,N}$ in the cell $\mathcal{I}_{i,j}$. Given $\left\langle P_M h_{i,j}^k \right\rangle_{N,N}$, we want to compute the FV approximation at the next time instance. To this end, we follow the same four steps as those outlined for the 1D-case in Subsection 3.2. The entropy-minimization step, the collision step and the optimization step are very similar to the 1D case and, for brevity, we do not repeat them here. The transport step is slightly different and is given as

$$\begin{aligned}\frac{\left\langle P_M f_{i,j}^{k+1} \right\rangle_{N,N} - \left\langle P_M f_{i,j}^{k^*} \right\rangle_{N,N}}{\Delta t} = &- \frac{1}{\Delta x}\Big(\mathcal{F}_1(W(f_{i+1,j}^{k,*}), W(f_{i,j}^{k,*})) \\ &-\mathcal{F}(W(f_{i,j}^{k,*}), W(f_{i-1,j}^{k,*}))\Big) \\ &- \frac{1}{\Delta x}\Big(\mathcal{F}_2(W(f_{i,j+1}^{k,*}), W(f_{i,j}^{k,*})) \\ &-\mathcal{F}_2(W(f_{i,j}^{k,*}), W(f_{i,j-1}^{k,*}))\Big) .\end{aligned} \tag{4.7}$$

Above, $\{W(f_{i,j}^{k,*})\}_{i,j}$ results from the optimization step and $\mathcal{F}_1(W_1, W_2)$ and $\mathcal{F}_2(W_1, W_2)$ are the numerical fluxes given as

$$\mathcal{F}_i(W_1, W_2) := \frac{1}{2}\left(AL(\Xi_i - |\Xi_i|)W_1 + AL(\Xi_i + |\Xi_i|)W_2\right). \tag{4.8}$$

Above, A and L are multi-dimensional versions of the matrices given in (2.8) and Ξ_i is a diagonal matrix with all the i-th components of the quadrature point's locations at its diagonal.

Assuming that the initial and the boundary data satisfies (3.10), one can show that the space-time discretization results in a feasible optimization if the ratio $\Lambda = \Delta t/\Delta x$ satisfies

$$0 < \Lambda \leq \frac{1}{2}\min_i \left\{\min\left\{|\xi_{i,\max}^{-1}|, |\xi_{i,\min}^{-1}|\right\}\right\}. \tag{4.9}$$

Similarly, one can show that the space-time discretization is L2-stable if Λ satisfies

$$0 < \Lambda \leq \frac{1}{4}\min_i \left\{\min\left\{|\xi_{i,\max}^{-1}|, |\xi_{i,\min}^{-1}|\right\}\right\}. \tag{4.10}$$

A proof of the above two results uses the exact same technique as that for the 1D case and is not repeated for brevity.

5 Numerical Results For simplicity, we non-dimensionalize the BE and all the macroscopic quantities with appropriate powers of some reference density ρ_0, temperature θ_0 and length scale l. This introduces the Knudsen number Kn, the inverse of which scales the collision operator $Q(f)$, and reads $Kn := \tau_0/\left(\sqrt{\theta_0}l\right)$—we refer to [32] for the details of non-dimensionalization. In the definition of the collision frequency $\tau(x, t)^{-1}$ given in (2.14), we choose $C = 1$ and $\omega = 1$. Our choice of C and

ω does not necessarily corresponds to a physical system and is made for demonstration purposes.

We consider the following test cases.

1. **Test case-1** We consider the pdf

$$f(\xi)=\frac{1}{\sqrt{2\pi\theta_0}}\exp\left(-\frac{(\xi-u_0)^2}{2\theta_0}\right)+\frac{1}{\sqrt{2\pi\theta_1}}\exp\left(-\frac{(\xi-u_1)^2}{2\theta_1}\right). \tag{5.1}$$

   Given the first M moments of f and using the pos-L2-MOM, we approximate the $M+1$-st moment of f. We study the error of this approximation with respect to the number of moments M. We choose $\theta_0=3$, $u_0=-4$, $\theta_1=4$ and $u_1=5$, which ensures that f is far away from a Maxwell-Boltzmann distribution function in the Kullback–Leibler divergence sense.
2. **Test case-2** For a one-dimensional space-velocity domain, we consider the Sod's shock tube problem from [35]. We set $\Omega=[-2,2]$ and $D=[0,0.3]$. Recall that D is the time domain. As the initial data, we consider a gas at rest and at equilibrium. We initialize the temperature θ with a constant value of one and we initialize density as

$$\rho(x,t=0)=\begin{cases}7, & x\leq 0\\ 1, & x>0\end{cases}. \tag{5.2}$$

   As the boundary data f_{in}, we consider a Maxwell-Boltzmann pdf. At $x=x_{\min}$ and for all $t\in D$, we set density to 7, velocity to 0 and temperature to 1. The velocity and the temperature at the right boundary remains the same but the density changes to 1. We consider two different values of the Knudsen number—$Kn=0.1$ and $Kn=0.01$.
3. **Test case-3** For a one-dimensional space-velocity domain, we consider the two-beam interaction experiment from [30]. The space-time domain $\Omega\times D$ remains the same as the previous test case. As the initial data, we consider a gas at equilibrium with a constant density and temperature of one. As the initial velocity, we consider

$$v(x,t=0)=\begin{cases}1, & x\leq 0\\ -1, & x>0\end{cases}. \tag{5.3}$$

   As the boundary data f_{in}, we consider a Maxwell-Boltzmann pdf. At $x=x_{\min}$ and for all $t\in D$, we set density to 1, velocity to 1 and temperature to 1. The density and the temperature at the right boundary remains the same but the velocity changes to -1. We consider two different values of the Knudsen number—$Kn=0.1$ and $Kn=0.01$.
4. **Test case-4** We consider a two-dimensional spatial domain and a planar flow regime. We choose $\Omega=[0,2]\times[0,2]$ and $D=[0,0.2]$. We consider a micro-bubble dispersion problem where we start with a fluid at equilibrium and at rest. We consider a constant temperature of one and consider a density given as

$$\rho(x,t=0)=\rho_0+\exp(-|x-1|^2\times 10^2),\quad \forall x\in\Omega. \tag{5.4}$$

   As the ground state density, we set $\rho_0=1$. As the boundary data f_{in}, we consider a Maxwell-Boltzmann pdf with a density ρ_0, velocity zero and temperature one. We consider a Knudsen number of 0.1.

We emphasis that for this test case, it is crucial that the moment-closure problem has a positive solution. Otherwise, the density can get negative resulting in a breakdown of the solution algorithm. We refer to [27] for a similar experiment involving the linearized BE and the Grad's Hermite expansion, which is not necessarily positive. There, the deviation in density gets negative for small values of M. However, since the BE is linearized, negative densities do not crash the solution algorithm.

5.1 Test case-1 We truncate the velocity domain to $\Omega_\xi = [-20, 20]$. This ensures that the support of f (upto machine precision) is contained inside Ω_ξ. We compute $W(f_M)$ using the optimization problem given in (2.7).

5.1.1 Error in the higher order moment Recall that we used f_M to close the moment system in (2.18) by approximating the M-th order moment of f. The relative error of this approximation is given as

$$\mathcal{E}(M) := \left| \frac{\langle \xi^M (f_M - f) \rangle_N}{\langle \xi^M f \rangle_N} \right|. \tag{5.5}$$

We study $\mathcal{E}(M)$ for different values of M. We vary M from 3 to 22 in steps of one, and we fix N at a sufficiently large value of 40.

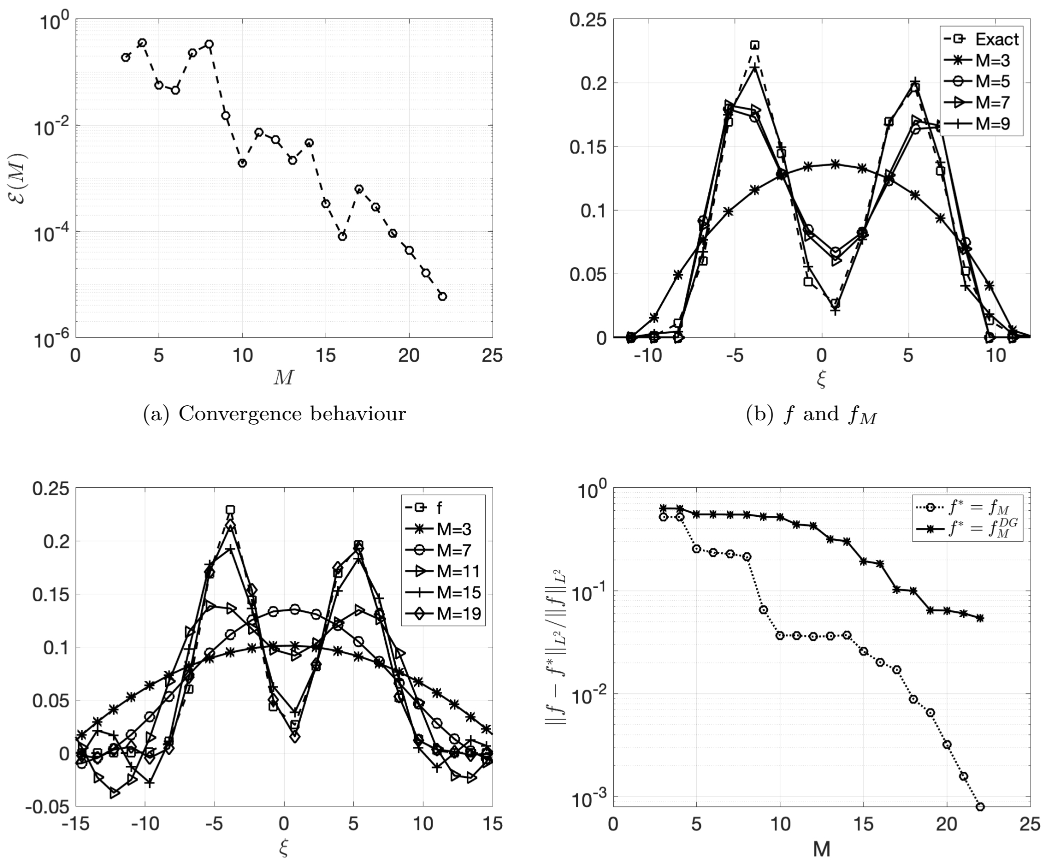
As M increases, $\mathcal{E}(M)$ appears to converge to zero, although not monotonically—see Figure 1a. Note that this non-monotonic convergence is typical also for a Grad's moment approximation [11, 27, 36]. However, unlike the Grad's moment approximation where the error convergences monotonically for either the even or the odd values of M, the convergence behaviour of the pos-L2-MOM is rather random. For instance, the error (slightly) increases from $M = 5$ to $M = 7$. Similarly, the error (slightly) increases from $M = 15$ to $M = 17$. Nevertheless, for $M \geq 16$, the error appears to converge monotonically.

5.1.2 Error in approximating the pdf For different values of M, Figure 1b compares f to f_M. To extend the discrete values of f_M to Ω_ξ, we perform a piecewise linear interpolation between the quadrature points. For $M = 3$, pos-L2-MOM is unable to capture the general shape of the function. Nevertheless, increasing the value of M improves the results. Already for $M = 5$, we observe that f_M has two distinct peaks and starts to capture the shape of the function. Increasing M from 5 to 7 does not show much of an improvement. However, increasing M from 7 to 9 improves the results significantly. The result for $M = 9$ almost overlaps the exact solution with little deviations. Let us mention that for all values of M, f_M remains positive.

For a comparison, we compute a DG approximation of f. We represent the DG approximation by f_M^{DG} and compute it by projecting f (under the $L^2(\Omega_\xi)$ inner-product) onto the first M Legendre polynomials in ξ. For the different values of M, Figure 1c compares f to f_M^{DG}. Since a DG approximation does not penalize negativity (see Remark 1), for all values of M, f_M^{DG} is negative for some part of the velocity domain. Furthermore, only for $M \geq 11$, the DG approximation starts to capture the general shape of the function. Compare this to f_M, which, already for $M = 5$, accurately represents the shape of the function.

The superior accuracy of f_M—as compared to f_M^{DG}—in approximating f is clearly visible in Figure 1d, which compares the relative L2-error in approximating f. The difference between the error values becomes larger as the value of M increases. For the largest value of M equals 22, the relative L2-error resulting from the approximation f_M

is 8×10^{-4}, which is $\approx 10^{-2}$ times smaller than that resulting from the approximation f_M^{DG}.



(a) Convergence behaviour

(b) f and f_M

(c) f and f_M^{DG}

(d) Comparison of the relative L2-error

Fig. 1: Results for test case-1. (a) and (d) y-axis is on a log-scale.

5.2 Test case-2

5.2.1 Reference solution We compute the reference solution using a DVM proposed in [22]. We consider an explicit Euler time-stepping scheme and a first-order FV spatial discretization. We truncate the velocity domain to $[-7, 7]$, and place $N = 350$ velocity grid points inside the truncated velocity domain. As the velocity grid points, we consider Gauss-Legendre quadrature nodes. We discretize the space domain with $N_x = 10^3$ uniform cells and consider a constant time-step of $\Delta t = 0.5 \times \Delta x/7$. To arrive at these discretization parameters, we performed a convergence study that consisted of the following steps. (i) With the velocity and the temperature field taken from the initial data, estimate $\xi_{\max/\min}$ using the relation in (2.20). For the present test case, this provides $\xi_{\max} = 3.5$ and $\xi_{\min} = -3.5$. (ii) Fix N_x at 10^3 and Δt to $0.5 \times \Delta x/\xi_{\max}$. (iii) Choose $N = 50$ and increase it to 350 in steps of 50. (iv) Terminate the refinement as soon as the relative change in mass, momentum and energy between two subsequent refinement cycles drops below a tolerance of 10^{-5}. (v) If the tolerance is not reached, increase $\xi_{\max}$ by 0.5, decrease $\xi_{\min}$ by 0.5 and repeat the process from step-(ii). Note that if the refinement cycle does not terminate then one should increase the value of N_x and repeat the entire process. For the all the test cases mentioned earlier, the value of $N_x = 10^3$ was sufficiently large to terminate the refinement cycle.

5.2.2 Convergence study We are interested in the relative L2 error in the different macroscopic quantities that we define as

$$\mathcal{E}_{cons}(M, N_x) := \frac{\|\langle P_{\text{cons}}(f_{M,N_x}(\cdot, t = T, \cdot) - f_{DVM}(\cdot, t = T, \cdot))\rangle_N\|_{L^2(\Omega;\mathbb{R}^3)}}{\|\langle P_{\text{cons}} f_{DVM}(\cdot, t = T, \cdot)\rangle_N\|_{L^2(\Omega;\mathbb{R}^3)}}. \tag{5.6}$$

Above, P_{cons} and f_{M,N_x} are as defined in (2.13) and (3.4), respectively. We keep the value of N fixed at 30.

We first consider $\text{Kn} = 0.1$. We increase M from 3 to 10 in steps of 1 and N_x from 200 to 10^3 in steps of 200. We choose $\Delta t = 0.5 \times \Delta x/7$. Figure 2a shows the error $\mathcal{E}_{cons}(M, N_x)$ for the different values of M and N_x. Fixing N_x at a small value—200 for instance—and increasing M does not reduce the error. This is because for small values of N_x, the error is dominated by the error in our space-time discretization. Furthermore, for a small value of M, increasing N_x beyond a certain limit does not decrease the error. On the other hand, choosing a large value of N_x—10^3 for instance—and increasing M, or increasing both M and N_x simultaneously, reduces the error. Note that similar to the previous test case, the error decay is not monotonic. Our results suggest that to balance the accuracy with the computational cost, an adaptive choice of M and an adaptive spatial grid is desirable. We plan to develop such an adaptive framework in the future—see [2] for an adaptive moment method. Let us also mention that at $N_x = 10^3$ and $M = 10$, we attain a minimum relative error of 2.4×10^{-2}. We find this error value acceptable, given that $M = 10$ is less than 10% of the velocity grid points used in our reference DVM.

We now consider $\text{Kn} = 0.01$. We choose M and N_x as before. Figure 2a shows the error $\mathcal{E}_{cons}(M, N_x)$ for the different values of M and N_x. As compared to $\text{Kn} = 0.1$, the smaller values of M perform much better, which is in accordance with similar studies conducted in the previous works [36]. For instance, consider the results for $M = 4$ and $N_x = 10^3$. For $\text{Kn} = 0.1$, we find $\mathcal{E}_{cons}(4, 10^3) = 1.3 \times 10^{-1}$, whereas for $\text{Kn} = 0.01$ we find $\mathcal{E}_{cons}(4, 10^3) = 2.5 \times 10^{-2}$, which is almost an order-of-magnitude better than the result for $\text{Kn} = 0.1$.

Although the lower values of M perform better for $\text{Kn} = 0.01$ than for $\text{Kn} = 0.1$, the minimum error attained is almost the same for both the Knudsen number—for $\text{Kn} = 0.01$ the minimum error is 2.3×10^{-3}, which is 0.95 times that of the minimum error for $\text{Kn} = 0.1$. This is because for $\text{Kn} = 0.01$, the error at $N_x = 10^3$ is already dominated by the error in our spatial discretization and we see almost no error reduction upon increasing M from 7 to 10. By increasing N_x from 10^3 to 1.5×10^3, we could remove this error stagnation and for $M = 10$, achieve an error of 1.2×10^{-3}.

5.2.3 Sub-shocks Shock speeds that are faster than the characteristic speeds in a moment system result in sub-shocks—we refer to [34] for an exhaustive study of sub-shocks for the Grad's MOM. Similar to the Grad's MOM, the pos-L2-MOM shows sub-shocks-type structures—see the density profile shown in Figure 3. These structures have a staircase-type shape, and increasing M from 3 to 5 has a smoothing

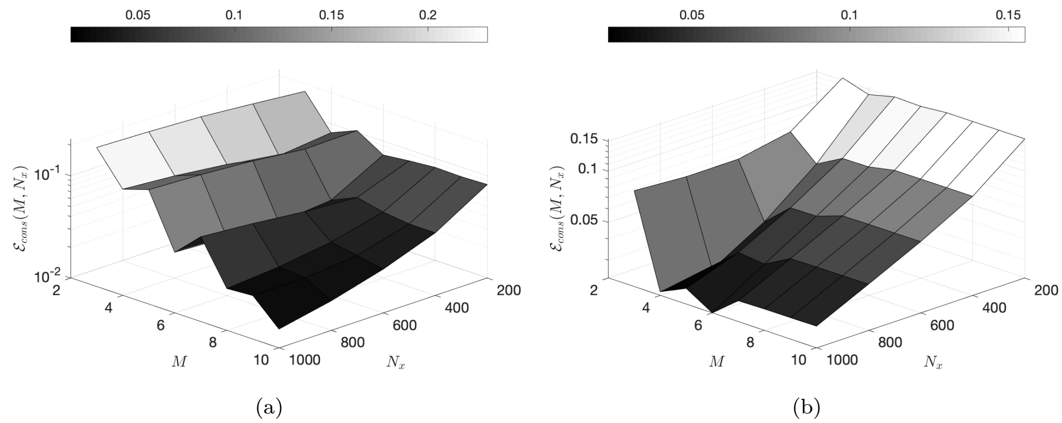


Fig. 2: Results for test case-2. Convergence of the relative error with N_x and M. Computations performed using the pos-L2-MOM. (a) Kn = 0.1 and (b) Kn = 0.01. The z-axis on both the plots is on a log-scale.

effect that reduces the staircase effect. To conclude that these structures are indeed sub-shocks, one needs to study the characteristic speeds of the moment system given in (2.18). Note that these sub-shocks can be removed by introducing second-order spatial derivatives in the moment equations via regularization—see the discussion on the regularized-13 moment equations [39].

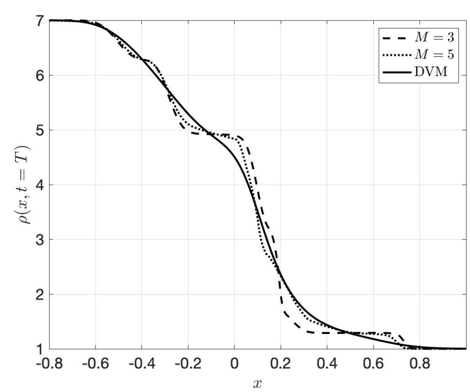
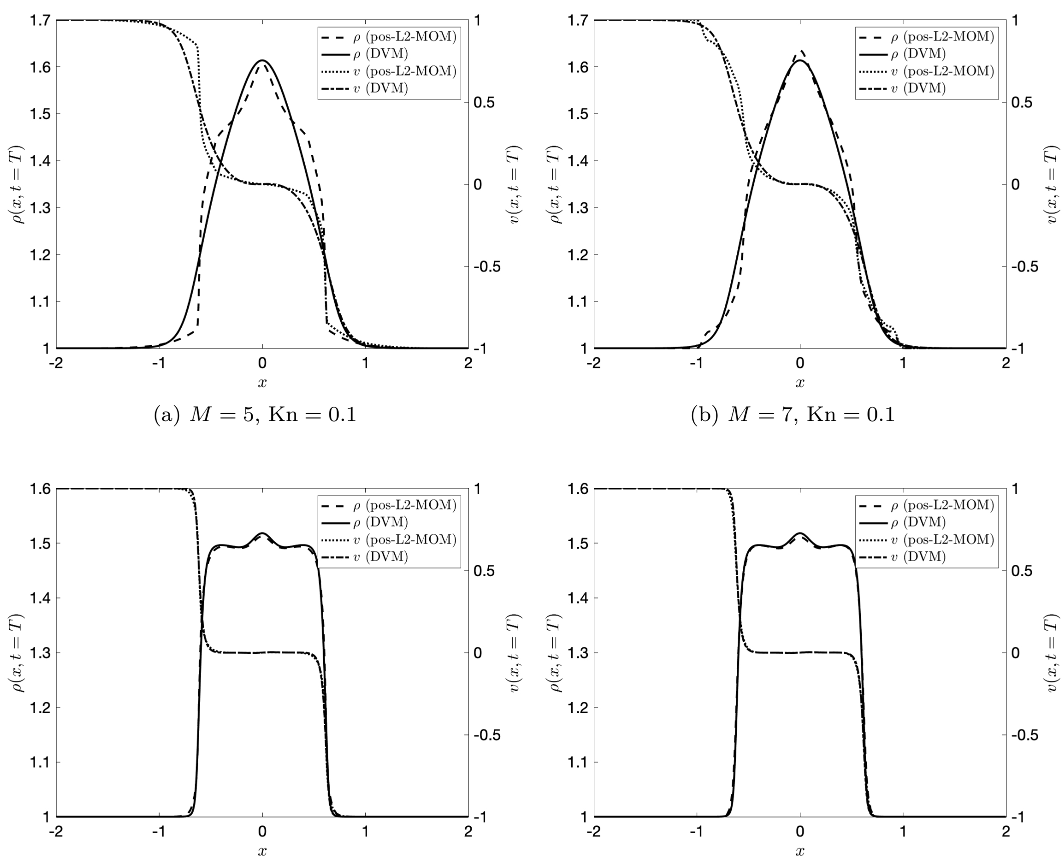


Fig. 3: Results for test case-2. Density profile for $Kn = 0.1$ and at $t = T$. Computations performed with $N_x = 10^3$ grid-cells.

5.3 Test case-3 As before, we construct a reference solution using the DVM. The convergence study discussed in Subsection 5.2.1 lead to $N_x = 10^3$, $\xi_{\max} = 5$, $\xi_{\min} = -5$ and $N = 350$. For the pos-L2-MOM, we fix $N = 30$ and $N_x = 10^3$, and study the results for two different values of M, $M = 5$ and $M = 7$. We choose $\Delta t = 0.5\Delta x/\xi_{\max}$. The convergence behaviour is similar to the previous test case and not discussed for brevity.

For Kn = 0.1 and $M = 5$, Figure 4a compares the density and the velocity computed using the DVM and the pos-L2-MOM. The results for temperature are similar and are not shown for brevity. The pos-L2-MOM performs well and results in an error of $\mathcal{E}_{cons}(5, 10^3) = 6.8 \times 10^{-2}$. Furthermore, increasing the value of M from 5 to 7 improves the results and the error reduces to $\mathcal{E}_{cons}(7, 10^3) = 2.5 \times 10^{-2}$—Figure 4b shows the result for $M = 7$. Reducing the Knudsen number to 0.01, improves the results for both $M = 5$ and $M = 7$—see Figure 4c and Figure 4d. For both the values of M, we obtained an error of $\mathcal{E}_{cons}(5/7, 10^3) = 9 \times 10^{-3}$, which is approximately 1/3 of the error for Kn = 0.1. Note that similar to the previous test case, the error for Kn = 0.01 is dominated by the error in the space-time discretization. Therefore, increasing M from 5 to 7 does not offer any improvement.



(a) $M = 5$, Kn = 0.1 (b) $M = 7$, Kn = 0.1

(c) $M = 5$, Kn = 0.01 (d) $M = 7$, Kn = 0.01

Fig. 4: Results for test case-3. Density and velocity profiles for different values of M and different Knudsen numbers. The left and the right y-axis is for density and velocity, respectively.

5.4 Test case-4 Under the limited computational resources, we were unable to compute a highly-refined reference solution in multi-dimensions. For this reason, we refrain from performing a convergence study for the present test case. Rather, we compare our moment method to a sufficiently refined DVM and showcase an improvement in the moment solution by increasing M. For both the DVM and the moment method, we consider tensorized Gauss-Legendre quadrature points with $N = 40$ quadrature points in each direction. We place these quadrature points inside $\Omega_\xi = [\xi_{\min}, \xi_{\max}] \times [\xi_{\min}, \xi_{\max}]$ with $\xi_{\max} = 7$ and $\xi_{\min} = -7$. We discretize the spatial domain with 150×150 uniform elements with grid-size $\Delta x = 1.3 \times 10^{-2}$. We consider a constant time-step of $\Delta = \Delta x/(4 \times \xi_{\max})$.

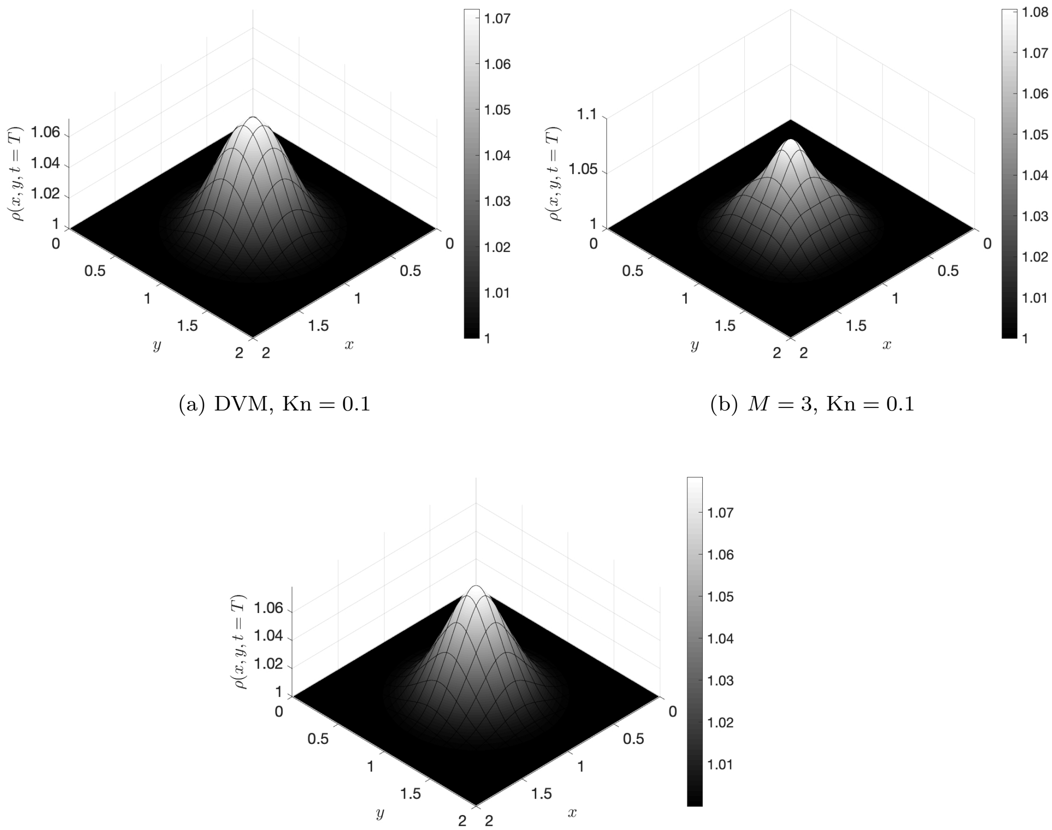
As time progresses, the density disperses into the spatial domain. This is made clear by Figure 5a that shows the density profile at $t = T$ computed using the DVM. At the same time-instance, Figure 5b and Figure 5c show the density profile at $t = T$ computed using the pos-L2-MOM with $M = 3$ and $M = 5$, respectively. As expected, both the density profiles are positive. Furthermore, the moment solution appears to improve upon increasing the value of M. The improvement is quantified by the decrease in the relative L2-error in density shown in Table 1.

The dispersion of the micro-bubble triggers a flow velocity and a temperature gradient. Figure 6 compares the x_1 velocity component and the temperature along a cross-section of the spatial domain computed using the pos-L2-MOM and the DVM. The results for the x_2 velocity component are similar and are not shown for brevity. As expected, similar to density, the results for both the velocity and the temperature appear to improve as M is increased from 3 to 5, the relative L2-error shown in Table 1 indicates the same. We note that, as compared to the previous test cases, the moment method performs better for the present test case. A possible reason for this could be that our DVM solution is not as refined as for the previous test cases—the previous test cases consider a 1D velocity grid of 350 points whereas the present test case considers a tensorized grid of 40×40 points.

M	ρ	v_1	v_2	θ
3	1.6×10^{-3}	2×10^{-1}	2.1×10^{-1}	2.8×10^{-3}
5	5.3×10^{-4}	5×10^{-2}	4.8×10^{-2}	5.4×10^{-4}

Table 1: Results for test case-4. Relative $L^2(\Omega)$-error in different macroscopic quantities at $t = T$ and $Kn = 0.1$.

6 Conclusions We proposed a positive moment method for the Boltzmann-BGK equation based upon L2-minimization. We showed that on a space-time discrete level both the feasibility of the minimization problem and the stability of the moment approximation can be ensured via a CFL-type condition. Our proof of booth these properties relied on relating our moment method to a discrete-velocity-method. Through a entropy-minimization based discretization of the collision operator, we ensured that our moment approximation conserves mass, momentum and energy. We also extended our method to a multi-dimensional space-velocity domain. With the help of numerical experiments, we studied the accuracy of our method for both single and multi-dimensional space-velocity domains. Our method performed well for a broad range of problems involving strong shocks, beam interaction and micro-bubble dispersion. Furthermore, it retained accuracy for a broad range of Knudsen numbers.



(a) DVM, $\mathrm{Kn} = 0.1$

(b) $M = 3$, $\mathrm{Kn} = 0.1$

(c) $M = 5$, $\mathrm{Kn} = 0.1$

Fig. 5: Results for test case-4. Density profiles at $t = T$.

References

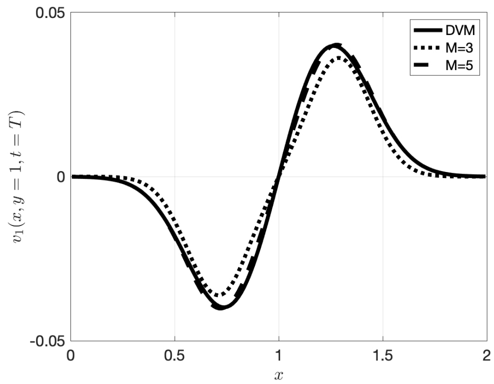
[1] M. R. A. Abdelmalik and E. H. van Brummelen. Moment closure approximations of the Boltzmann equation based on ϕ-divergences. *Journal of Statistical Physics*, 164(1):77–104, Jul 2016.

[2] M. R. A. Abdelmalik and E. H. van Brummelen. Error estimation and adaptive moment hierarchies for goal-oriented approximations of the Boltzmann equation. *Computer Methods in Applied Mechanics and Engineering*, 325(Supplement C):219 − 239, 2017.

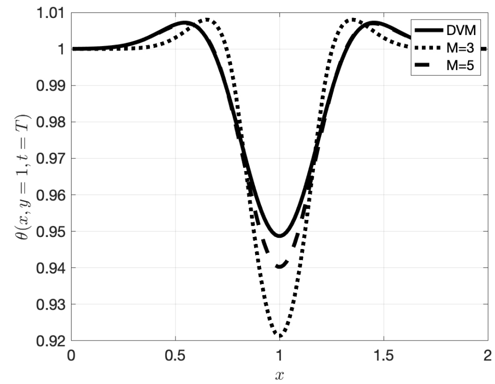
[3] A. Alekseenko and E. Josyula. Deterministic solution of the spatially homogeneous Boltzmann equation using discontinuous Galerkin discretizations in the velocity space. *Journal of Computational Physics*, 272:170 − 188, 2014.

[4] G. W. Alldredge, M. Frank, and C. D. Hauck. A regularized entropy-based moment method for kinetic equations. *SIAM Journal on Applied Mathematics*, 79(5):1627–1653, 2019.

[5] G. W. Alldredge, C. D. Hauck, D. P. O. Leary, and A. L. Tits. Adaptive change of basis in entropy-based moment closures for linear kinetic equations. *Journal*



(a) $v_1(x, y = 1, t = T)$, Kn = 0.1



(b) $\theta(x, y = 1, t = T)$, Kn = 0.1

Fig. 6: Results for test case-4. v_1 and θ profiles.

of Computational Physics, 258:489 – 508, 2014.

[6] G. W. Alldredge, C. D. Hauck, and A. L. Tits. High-order entropy-based closures for linear transport in slab geometry ii: A computational study of the optimization problem. *SIAM Journal on Scientific Computing*, 34(4):B361–B391, 2012.

[7] C. Baranger, J. Claudel, N. Hérouard, and L. Mieussens. Locally refined discrete velocity grids for stationary rarefied flow simulations. *Journal of Computational Physics*, 257:572 – 593, 2014.

[8] N. Bohmer and M. Torrilhon. Entropic quadrature for moment approximations of the Boltzmann-BGK equation. *Journal of Computational Physics*, 401:108992, 2020.

[9] S. Brull and L. Mieussens. Local discrete velocity grids for deterministic rarefied flow simulations. *Journal of Computational Physics*, 266:22 – 46, 2014.

[10] Z. Cai, Y. Fan, and L. Ying. An entropic Fourier method for the Boltzmann equation. *SIAM Journal on Scientific Computing*, 40(5):A2858–A2882, 2018.

[11] Z. Cai and M. Torrilhon. Numerical simulation of microflows using moment methods with linearized collision operator. *Journal of Scientific Computing*, 2017.

[12] C. Cercignani. *The Boltzmann Equation and Its Applications.* Springer, 67 edition, 1988.

[13] S. Chapman, T. G. Cowling, and D. Burnett. *The mathematical theory of non-uniform gases: an account of the kinetic theory of viscosity, thermal conduction and diffusion in gases.* Cambridge university press, 1990.

[14] F. Filbet and S. Jin. A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources. *Journal of Computational Physics*, 229(20):7625 – 7648, 2010.

[15] H. Grad. On the kinetic theory of rarefied gases. *Communications on Pure and Applied Mathematics*, 2(4):331–407, 1949.

[16] C. Groth and J. McDonald. Towards physically realizable and hyperbolic moment closures for kinetic theory. *Continuum Mech. Thermodyn.*, 21(467), 2009.

[17] J. Han, C. Ma, Z. Ma, and W. E. Uniformly accurate machine learning-based hydrodynamic models for kinetic equations. *Proceedings of the National Academy of Sciences*, 116(44):21983–21991, 2019.

[18] C. Hauck and R. McClarren. Positive PN closures. *SIAM Journal on Scientific Computing*, 32(5):2603–2626, 2010.

[19] M. Junk. Maximum entropy for reduced moment problems. *Mathematical Models and Methods in Applied Sciences*, 10(07):1001–1025, 2000.

[20] C. D. Levermore. Moment closure hierarchies for kinetic theories. *Journal of Statistical Physics*, 83(5):1021–1065, Jun 1996.

[21] L. R. Mead and N. Papanicolaou. Maximum entropy in the problem of moments. *Journal of Mathematical Physics*, 25(8):2404–2417, 1984.

[22] L. Mieussens. Discrete velocity model and implicit scheme for the BGK equation of rarefied gas dynamics. *Mathematical Models and Methods in Applied Sciences*, 10(08):1121–1149, 2000.

[23] L. Mieussens. Convergence of a discrete-velocity model for the Boltzmann-BGK equation. *Computers & Mathematics with Applications*, 41(1):83 – 96, 2001.

[24] C. Ringhofer, C. Schmeiser, and A. Zwirchmayr. Moment methods for the semiconductor Boltzmann equation on bounded position domains. *SIAM Journal on Numerical Analysis*, 39(3):1078–1095, 2001.

[25] M. Sadr, M. Torrilhon, and M. H. Gorji. Gaussian process regression for maximum entropy distribution. *Journal of Computational Physics*, 418:109644, 2020.

[26] N. Sarna, J. Giesselmann, and M. Torrilhon. Convergence analysis of Grad's Hermite expansion for linear kinetic equations. *SIAM Journal on Numerical Analysis*, 58(2):1164–1194, 2020.

[27] N. Sarna, H. Kapadia, and M. Torrilhon. Simultaneous-approximation-term based boundary discretization for moment equations of rarefied gas dynamics. *Journal of Computational Physics*, 407:109243, 2020.

[28] N. Sarna and M. Torrilhon. On stable wall boundary conditions for the Hermite discretization of the linearised Boltzmann equation. *Journal of Statistical Physics*, 170(1):101–126, 2018.

[29] R. P. Schaerer, P. Bansal, and M. Torrilhon. Efficient algorithms and implementations of entropy-based moment closures for rarefied gases. *Journal of Computational Physics*, 340:138 – 159, 2017.

[30] R. P. Schaerer and M. Torrilhon. The 35-moment system with the maximum-entropy closure for rarefied gas flows. *European Journal of Mechanics - B/Fluids*, 64:30 – 40, 2017.

[31] J. Schneider. Entropic approximation in kinetic theory. *ESAIM: M2AN*, 38(3):541–561, 2004.

[32] H. Struchtrup. *Macroscopic Transport Equations for Rarefied Gas Flows.* Springer Ltd, 2010.

[33] E. Tadmor. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numerica*, 12:451–512, 2003.

[34] M. Torrilhon. Characteristic waves and dissipation in the 13-moment case. *Continuum Mech. Thermodyn.*, 12:289–301, 2000.

[35] M. Torrilhon. Two-dimensional bulk microflow simulations based on regularized Grad's 13-moment equations. *Multiscale Modeling & Simulation*, 5(3):695–728, 2006.

[36] M. Torrilhon. Convergence study of moment approximations for boundary value problems of the Boltzmann-BGK equation. *Communications in Computational Physics*, 18(03):529–557, 2015.

[37] M. Torrilhon. Modeling nonequilibrium gas flow based on moment equations. *Annual Review of Fluid Mechanics*, 48(1):429–458, 2016.

[38] M. Torrilhon and N. Sarna. Hierarchical Boltzmann simulations and model error estimation. *Journal of Computational Physics*, 342:66 – 84, 2017.
[39] M. Torrilhon and H. Struchtrup. Regularized 13-moment equations: shock structure calculations and comparison to Burnett models. *Journal of Fluid Mechanics*, 513:171–198, 2004.
[40] S. Ukai. Solutions of the Boltzmann equation. In *Patterns and Waves*, volume 18 of *Studies in Mathematics and Its Applications*, pages 37 – 96. Elsevier, 1986.
[41] J. Yang and J. Huang. Rarefied flow computations using nonlinear model Boltzmann equations. *Journal of Computational Physics*, 120(2):323 – 339, 1995.
[42] Y. Ye and E. Tse. An extension of Karmarkar's projective algorithm for convex quadratic programming. *Mathematical Programming*, 44:157–179, 1989.
[43] C. Zhenning, Y. Fan, and R. Li. Globally Hyperbolic Regularization of Grad's Moment System. *Communications on Pure and Applied Mathematics*, 67(3):464–518, 2014.