# Model Reduction of a Two-Dimensional Kinetic Swarming Model by Operator Projections

Junming Duan[1], Yangyu Kuang[1]  and Huazhong Tang[2,*]

[1] *LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China.*

[2] *HEDPS, CAPT & LMAM, School of Mathematical Sciences, Peking University, Beijing 100871, P.R. China; School of Mathematics and Computational Science, Xiangtan University, Xiangtan 411105, P.R. China.*

**Abstract.** Hyperbolic, rotation invariant moment systems are derived for a non-linear kinetic description of two-dimensional Vicsek swarming model. The systems also preserve mass conservation, and numerical experiments show that this approach captures the key features of the model such as shock waves, contact discontinuities, rarefaction waves, vortex formations. If the system order increases, the solutions of the moment systems converge to the solution of the corresponding kinetic equation.

## 1. Introduction

Swarming is a collective behaviour exhibited by entities, particularly animals of similar size which aggregate together, milling about the same spot, or moving en masse or migrating in some direction — cf. Ref. [1]. Various swarming models have been discussed in literature [8, 11, 17], but numerical methods for such models are not well developed. The first numerical investigation, published only recently by Gamba *et al.* [14], deals with the kinetic description of Vicsek swarming model. On the other hand, kinetic theory, which plays an important role in many applications, has been widely studied — cf. Refs. [9, 10]. The kinetic equation determines the distribution function and, consequently, the transport coefficients. The moment method [5, 16, 24] is a reduction of the kinetic equation based on the expansion of the distribution function into the series of tensorial Hermite polynomials and introduction of balance equations for higher order moments of this function. A major disadvantage of the Grad moment method is the loss of hyperbolicity, which causes the solution blow-up for distributions distant from the equilibrium state — cf. Refs. [7, 25], and

---

*Corresponding author. Email addresses:* `duanjm@pku.edu.cn` (J. Duan), `kyy@pku.edu.cn` (Y. Kuang), `hztang@math.pku.edu.cn` (H. Tang)

even the increase of the moment numbers does not help to prevent this effect [6]. Nevertheless, recently various approaches to the Grad moment method for the kinetic equation have been proposed — e.g. numerically regularized moment model of arbitrary order for Boltzmann-BGK equation [5] and high Mach number flow [6], the regularization of one- and multi-dimensional Grad moment systems leading to global hyperbolicity [2–4], quadrature based projection methods for Boltzmann equation [13, 18, 21], similar to 1-D regularization in Ref. [2]. These works contributed to a better understanding of the hyperbolicity of the Grad moment systems. On the other hand, Fan *et al.* [13] developed a general model reduction technique based on operator projections, where time and space derivatives are synchronously projected onto a finite-dimensional weighted polynomial space. Such a method was successfully applied to 1- and 3-D special relativistic Boltzmann equations and the globally hyperbolic moment models of arbitrary order were derived [22, 23].

In this paper we apply the projection operator model reduction method to the two-dimensional kinetic description of the Vicsek swarming model determining a globally hyperbolic moment system of arbitrary order. The moment system is obtained by the Grad type expansion near the equilibrium, and if the ratio $\varepsilon$ of micro to macro variables is small, the system with small moment numbers delivers a good approximation for the original kinetic model. The main difficulties here are to find weight functions which would determine appropriate function spaces, suitable bases and projection operators in these spaces.

The paper is organized as follows. Section 2 introduces the kinetic and macroscopic equations for Vicsek model. Section 3 deals with orthogonal functions and presents globally hyperbolic moment systems of arbitrary order for the kinetic description of Vicsek model. The properties of the moment systems such as hyperbolicity, rotational invariance and mass-conservation are considered in Section 4. Numerical experiments presented in Section 5 confirm the convergence of the hyperbolic moment systems, and Section 6 contains our concluding remarks.

## 2. Kinetic and Macroscopic Equations for Vicsek Model

### 2.1. Kinetic equation

Let $t$ denote time, $\boldsymbol{x} \in \mathbb{R}^d$ spatial variable, $\boldsymbol{\omega}$ unit velocity vector, $\sigma$ scaled diffusion constant describing the intensity of noise in Brownian motion, $f = f(t, \boldsymbol{x}, \boldsymbol{\omega})$ the particle distribution function, and let $J(t, \boldsymbol{x})$ refer to the mean flux at $\boldsymbol{x}$ — i.e.

$$J(t, \boldsymbol{x}) := \int_{\mathbb{R}^d} \int_{S^{d-1}} K(\boldsymbol{y} - \boldsymbol{x}) \boldsymbol{\omega} f(t, \boldsymbol{y}, \boldsymbol{\omega}) \mathrm{d}\boldsymbol{y} \, \mathrm{d}\boldsymbol{\omega}, \tag{2.1}$$

where $S^{d-1}$ is the unit sphere in $\mathbb{R}^d$, $K(\boldsymbol{x}) = \mathbf{1}_{|\boldsymbol{x}| < R}$ the characteristic function of the ball $B(0, R) = \{\boldsymbol{x} : |\boldsymbol{x}| \leq R\}$, with $R$ being the radius of alignment interaction between particles. By $\boldsymbol{F}[f](t, \boldsymbol{x}, \boldsymbol{\omega})$ we denote the mean-field interaction force between the particles,

$$\boldsymbol{F}[f](t, \boldsymbol{x}, \boldsymbol{\omega}) := (\mathrm{Id} - \boldsymbol{\omega} \otimes \boldsymbol{\omega}) \Omega(t, \boldsymbol{x}),$$

where Id is the identity operator and $\Omega(t, x) := J(t, x)/|J(t, x)|$ the mean velocity. The vector field $F[f](t, x, \omega)$ tends to align particles in the direction $\Omega$ and becomes spatially local in the large scale space-time region. Therefore, in computations the mean flux $J$ is often approximated as

$$J(t, x) = \int_{S^{d-1}} \omega f(t, x, \omega) \mathrm{d}\omega, \tag{2.2}$$

when the step size of spatial mesh is smaller than radius $R$ — cf. Ref. [14].

The kinetic equation for Vicsek model [14] has the form

$$\partial_t f + \omega \cdot \nabla_x f + \nabla_\omega \cdot (F[f]f) = \sigma \Delta_\omega f. \tag{2.3}$$

It is known [15] that if $J(t, x)$ is defined by (2.1) or (2.2) and does not always equal 0 for $t \in [0, T]$, then for any non-negative initial value $f(0, x, \omega)$ in $L^1(D) \cap L^\infty(D)$, $D = \mathbb{R}^d \times S^{d-1}$ and for any $T > 0$, the kinetic equation (2.3) has a non-negative global weak solution in $C\left(0, T; L^1(D) \cap L^\infty((0, T) \times D)\right)$.

Here we only consider the case where $J(t, x)$ has the form (2.2), so that the direction of the mean velocity of $f$ is

$$\Omega(t, x) = \frac{\int_{S^{d-1}} \omega f \, \mathrm{d}\omega}{\left|\int_{S^{d-1}} \omega f \, \mathrm{d}\omega\right|}.$$

Besides, we rewrite the equation (2.3) as

$$\partial_t f + \omega \cdot \nabla_x f = Q(f), \tag{2.4}$$

with the collision term

$$Q(f) = -\nabla_\omega \cdot \left(F[f]f\right) + \sigma \Delta_\omega f.$$

**Proposition 2.1** (cf. Gamba *et al.* [14]). *The collision term $Q(f)$ can be represented in the form*

$$Q(f) = \sigma \nabla_\omega \cdot \left(M_\Omega \nabla_\omega \left(\frac{f}{M_\Omega}\right)\right), \tag{2.5}$$

*and satisfies the inequality*

$$\int_{S^{d-1}} Q(f)f \frac{\mathrm{d}\omega}{M_\Omega} \leq 0,$$

*where*

$$M_\Omega(\omega) = C_0 \exp\left(\frac{\omega \cdot \Omega}{\sigma}\right),$$

*is the equilibrium function, also called the von Mises distribution, and $C_0$ is a normalization constant.*

**Remark 2.1.** Once a direction $\boldsymbol{\Omega}$ is fixed, $Q$ becomes a linear operator. Note that although the equilibria of the operator $Q$ is the set $\{\rho M_{\boldsymbol{\Omega}} : \rho \in \mathbb{R}, \boldsymbol{\Omega} \in S^{d-1}\}$ of dimension $d$, the collision invariants of $Q$ have dimension one only. Thus, in general,

$$\int_{S^{d-1}} Q(f) \mathrm{d}\boldsymbol{\omega} = 0, \quad \int_{S^{d-1}} Q(f) \boldsymbol{\omega} \mathrm{d}\boldsymbol{\omega} \neq 0,$$

so that the operator $Q$ preserves mass but not flux.

**Remark 2.2.** Setting $\boldsymbol{\Omega} = (\cos\bar{\theta}, \sin\bar{\theta})^T$ for $d = 2$, one represents the equilibrium as

$$M(\theta - \bar{\theta}) = C_0 \exp\left(\frac{\cos(\theta - \bar{\theta})}{\sigma}\right),$$

so that the kinetic equation (2.4) takes the form

$$f_t + \cos\theta f_x + \sin\theta f_y = Q(f),$$

$$Q(f) = \sigma \partial_\theta \left(M_{\bar{\theta}} \partial_\theta \left(\frac{f}{M_{\bar{\theta}}}\right)\right) = \partial_\theta \left(\sin(\theta - \bar{\theta})f\right) + \sigma \partial_\theta^2 f,$$

where $\theta$ and $\bar{\theta}$ are, respectively, microscopic and macroscopic velocity angles.

## 2.2. Macroscopic equations

The kinetic equation (2.4) is written at the level of individual particles. However, in order to study the system dynamics for large time and at large length scale, one can introduce new variables $\tilde{t} = \varepsilon t$ and $\tilde{x} = \varepsilon x$, with the ratio of micro to macro variable $\varepsilon \ll 1$ — cf. Ref. [12]. In new setting, the kinetic equation (2.4) takes the form

$$\partial_t f^\varepsilon + \boldsymbol{\omega} \cdot \nabla_{\boldsymbol{x}} f^\varepsilon = \frac{1}{\varepsilon} Q(f^\varepsilon), \tag{2.6}$$

where tilde is dropped to simplify notation. If $\varepsilon$ tends to 0, then $f^\varepsilon$ tends to the equilibrium state in the local space — i.e.

$$\lim_{\varepsilon \to 0} f^\varepsilon = f(t, \boldsymbol{x}, \boldsymbol{\omega}) = \rho(t, \boldsymbol{x}) M_{\boldsymbol{\Omega}}(\boldsymbol{\omega}).$$

The evolution of the macroscopic density $\rho = \int_{S^{d-1}} f \mathrm{d}\boldsymbol{\omega}$ and the velocity $\boldsymbol{\Omega}$ satisfies the following equations

$$\partial_t \rho + \nabla_{\boldsymbol{x}} \cdot (c_1 \rho \boldsymbol{\Omega}) = 0, \tag{2.7}$$

$$\rho\left(\partial_t \boldsymbol{\Omega} + c_2(\boldsymbol{\Omega} \cdot \nabla_{\boldsymbol{x}})\boldsymbol{\Omega}\right) + \lambda(\mathrm{Id} - \boldsymbol{\Omega} \otimes \boldsymbol{\Omega})\nabla_{\boldsymbol{x}}\rho = 0,$$

$$|\boldsymbol{\Omega}| = 1, \tag{2.8}$$

where $c_1$, $c_2$ and $\lambda$ are constants depending on $\sigma$. Such a macroscopic system is hyperbolic but non-conservative, and the operator $\mathrm{Id} - \boldsymbol{\Omega} \otimes \boldsymbol{\Omega}$ ensures the constraint (2.8).

**Remark 2.3.** In 2-D case, the macroscopic density $\rho$ and the velocity angle $\bar{\theta}$ are connected to the particle distribution function $f$ as follows

$$\rho(t,\boldsymbol{x}) = \int_0^{2\pi} f(t,\boldsymbol{x},\theta)\mathrm{d}\theta, \tag{2.9a}$$

$$\boldsymbol{J}(t,\boldsymbol{x}) = \int_0^{2\pi} \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} f(t,\boldsymbol{x},\theta)\mathrm{d}\theta, \tag{2.9b}$$

$$\boldsymbol{\Omega} = (\cos\bar{\theta}, \sin\bar{\theta})^T = \frac{\boldsymbol{J}(t,\boldsymbol{x})}{|\boldsymbol{J}(t,\boldsymbol{x})|}. \tag{2.9c}$$

## 3. Derivation of the Moment System

### 3.1. Orthogonal functions

If the distribution function is the local Maxwellian, Hermite polynomials are used in order to obtain Grad's moment system [16]. In this work, we use two systems of trigonometric polynomials $\{H_n^c(\theta-\bar{\theta})\}_{n=0}^{\infty}$ and $\{H_n^s(\theta-\bar{\theta})\}_{n=1}^{\infty}$ orthogonal with weight $M(\theta-\bar{\theta})$, $\theta \in [0,2\pi]$. The elements of these systems have the form

$$H_n^c(\theta-\bar{\theta}) \triangleq \sum_{k=0}^n a_{k,n} \cos\left(k(\theta-\bar{\theta})\right), \quad H_n^s(\theta-\bar{\theta}) \triangleq \sum_{k=1}^n b_{k,n} \sin\left(k(\theta-\bar{\theta})\right), \tag{3.1}$$

and satisfy the relations

$$\int_0^{2\pi} H_m^l(\theta-\bar{\theta})H_n^k(\theta-\bar{\theta})M(\theta-\bar{\theta})\mathrm{d}\theta = \delta_{m,n}\delta_{l,k}, \quad l,k = c \text{ or } s,$$

where $m$ is a non-negative integer if $l = c$ and a positive integer if $l = s$. The range of $n$ is defined similarly.

**Remark 3.1.** The coefficients $a_{k,n}$ and $b_{k,n}$ in (3.1) can be derived by the Gram-Schmidt process using usual trigonometric identities and the modified regular cylindrical Bessel functions

$$I_n(x) = \frac{1}{\pi} \int_0^{\pi} \exp(x\cos\theta)\cos(n\theta)\mathrm{d}\theta$$

of order $n$.

### 3.2. Hilbert space and orthogonal basis

Let $\mathscr{H}$ be the closed linear span of the set

$$\mathscr{P} = \left\{ \cos\left(n(\theta-\bar{\theta})\right)M(\theta-\bar{\theta}), \sin\left(n(\theta-\bar{\theta})\right)M(\theta-\bar{\theta}) : \theta \in [0,2\pi], \, n \in \mathbb{N} \right\},$$

with respect to the norm generated by the inner product

$$\langle f(\theta-\bar\theta), g(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} \triangleq \int_0^{2\pi} f(\theta-\bar\theta)g(\theta-\bar\theta)\frac{\mathrm{d}\theta}{M(\theta-\bar\theta)}.$$

This inner product is symmetric — i.e. $\langle f,g\rangle_M = \langle g,f\rangle_M$, and also satisfies the relations

$$\langle Q(f),g\rangle_M = -\sigma\int_0^{2\pi} M\partial_\theta\left(\frac{f}{M}\right)\partial_\theta\left(\frac{g}{M}\right)\mathrm{d}\theta = \langle f,Q(g)\rangle_M, \tag{3.2}$$

$$\langle Q(f),f\rangle_M = -\sigma\int_0^{2\pi} M\partial_\theta^2\left(\frac{f}{M}\right)\mathrm{d}\theta \leqslant 0, \tag{3.3}$$

as follows from Eq. (2.5). Similarly, for any positive integer $N$, we define the weighted trigonometric polynomial space $\mathscr{H}_N$ as

$$\mathscr{H}_N := \mathrm{span}\big\{\cos\big(n(\theta-\bar\theta)\big)M(\theta-\bar\theta), \sin\big(n(\theta-\bar\theta)\big)M(\theta-\bar\theta) : n\in\mathbb{N}, n\leq N\big\}.$$

Setting $P_n^c(\theta-\bar\theta) := H_n^c(\theta-\bar\theta)M(\theta-\bar\theta)$ and $P_n^s(\theta-\bar\theta) := H_n^s(\theta-\bar\theta)M(\theta-\bar\theta)$, we easily obtain the following result.

**Proposition 3.1.** *The systems* $\{P_n^c(\theta-\bar\theta), P_n^s(\theta-\bar\theta)\}_{n=1}^N$ *and* $\{P_n^c(\theta-\bar\theta), \{P_n^s(\theta-\bar\theta)\}_{n=1}^\infty\}$ *constitute standard orthogonal bases in* $\mathscr{H}_N$ *and* $\mathscr{H}$, *respectively. Moreover, the following relations*

$$\langle\cos(k(\theta-\bar\theta))M(\theta-\bar\theta), P_n^c(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} = 0, \quad k\leqslant n-1,$$
$$\langle\sin(k(\theta-\bar\theta))M(\theta-\bar\theta), P_n^s(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} = 0, \quad k\leqslant n-1,$$
$$\langle\cos(k(\theta-\bar\theta))M(\theta-\bar\theta), P_n^s(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} = 0, \quad \forall k,n,$$
$$\langle\sin(k(\theta-\bar\theta))M(\theta-\bar\theta), P_n^c(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} = 0, \quad \forall k,n.$$

*hold.*

**Remark 3.2.** Proposition 3.1 can be used to derive various recurrence relations for $P_n^c(\theta)$ and $P_n^s(\theta)$. For example, the series representation of $\cos\theta P_n^c(\theta)$ in $\mathscr{H}$ is

$$\cos\theta P_n^c(\theta) = \sum_{k=0}^\infty c_k^c P_k^c(\theta) + \sum_{k=1}^\infty c_k^s P_k^s(\theta),$$

where $c_k^c = \langle\cos\theta P_n^c(\theta), P_k^c(\theta)\rangle$, $c_k^s = 0$. Using the trigonometric formula $\cos\theta\cos(n\theta) = \cos((n+1)\theta) + \cos((n-1)\theta)$ we obtain

$$\left(\cos\theta - c_n^c\right)P_n^c = c_{n-1}^c P_{n-1}^c + c_{n+1}^c P_{n+1}^c.$$

In what follows, we will also use three-term recurrence relations for the terms $\sin\theta P_n^c(\theta)$, $\cos\theta P_n^s(\theta)$ and $\sin\theta P_n^s(\theta)$.

Considering the expansion

$$f(t,\boldsymbol{x},\theta) = \sum_{k=0}^{\infty} f_k^c(t,\boldsymbol{x},\bar{\theta}) P_k^c(\theta-\bar{\theta}) + \sum_{k=1}^{\infty} f_k^s(t,\boldsymbol{x},\bar{\theta}) P_k^s(\theta-\bar{\theta}),$$

of a function $f \in \mathscr{H}$ into the series with the coefficients $f_k^c = \langle f, P_k^c(\theta-\bar{\theta})\rangle_{M(\theta-\bar{\theta})}$ and $f_k^s = \langle f, P_k^s(\theta-\bar{\theta})\rangle_{M(\theta-\bar{\theta})}$, we define the projection operator $\Pi_N[\bar{\theta}] : \mathscr{H} \mapsto \mathscr{H}_N$ as

$$\Pi_N[\bar{\theta}]f \triangleq \sum_{k=0}^{N} f_k^c P_k^c(\theta-\bar{\theta}) + \sum_{k=1}^{N} f_k^s P_k^s(\theta-\bar{\theta}). \tag{3.4}$$

Introducing notation $\tilde{\boldsymbol{f}}_N$ for $(f_0^c, \cdots, f_N^c, f_1^s, \cdots, f_N^s)^{\mathrm{T}}$, we rewrite Eq. (3.4) as

$$\Pi_N[\bar{\theta}]f = \left[\tilde{\boldsymbol{f}}_N, \boldsymbol{P}_N(\theta-\bar{\theta})\right]_N,$$

where $[\cdot,\cdot]_N$ is the common inner product of two vectors,

$$\boldsymbol{P}_N(\theta-\bar{\theta}) \triangleq \begin{pmatrix} \boldsymbol{P}_N^c(\theta-\bar{\theta}) \\ \boldsymbol{P}_N^s(\theta-\bar{\theta}) \end{pmatrix},$$

and

$$\boldsymbol{P}_N^c(\theta-\bar{\theta}) \triangleq \left(P_0^c(\theta-\bar{\theta}), P_1^c(\theta-\bar{\theta}), \cdots, P_N^c(\theta-\bar{\theta})\right)^{\mathrm{T}},$$
$$\boldsymbol{P}_N^s(\theta-\bar{\theta}) \triangleq \left(P_1^s(\theta-\bar{\theta}), P_2^s(\theta-\bar{\theta}), \cdots, P_N^s(\theta-\bar{\theta})\right)^{\mathrm{T}}.$$

### 3.3. The moment system

In this section we apply the operator projection method [13] to the kinetic equation (2.4) in the space $\mathscr{H}$ with the projection operators $\Pi_N[\bar{\theta}]$ defined in (3.4). The variables $\rho$ and $\bar{\theta}$ are two physical quantities, required in the calculation of the equilibrium distribution $M(\theta-\bar{\theta})$. Therefore, the $(2N+1)$-dimensional vector $\boldsymbol{F}_N \triangleq [\rho, \cdots, f_N^c, \bar{\theta}, f_2^s, \cdots, f_N^s]^{\mathrm{T}}$ is considered instead of $\tilde{\boldsymbol{f}}_N$. According to Proposition 3.5 below,

$$\tilde{\boldsymbol{f}}_N = (I - \boldsymbol{e}_{N+2}\boldsymbol{e}_{N+2}^{\mathrm{T}})\boldsymbol{F}_N, \tag{3.5}$$

where $\boldsymbol{e}_{N+2}$ is $(N+2)$-th column of the unit matrix of order $2N+1$.

The moment system for the kinetic equation (2.4) is derived in the following way.

(i) **Distribution projection:** Find the projection $\Pi_N[\bar{\theta}]f$ of the distribution function $f$ in $\mathscr{H}_N$.

(ii) **Derivative projections:** Compute the partial derivatives of $\Pi_N[\bar{\theta}]f$ and project them onto the subspace $\mathscr{H}_N$, thus obtaining

$$\Pi_N[\bar{\theta}]\partial_s \Pi_N[\bar{\theta}]f = \left[\frac{\partial \tilde{\boldsymbol{f}}}{\partial s} - \frac{\partial \bar{\theta}}{\partial s}\tilde{D}_N^T \tilde{\boldsymbol{f}}_N, \boldsymbol{P}_N(\theta-\bar{\theta})\right]_N, \quad s = t, x, y. \tag{3.6}$$

(iii) **Product projections:** For $s = x$ and $s = y$, respectively, multiply the corresponding relations (3.6) by $\cos\theta$ and $\sin\theta$ and project the resulting functions onto subspace $\mathcal{H}_N$, so that

$$\Pi_N[\bar{\theta}]\cos\theta\,\Pi_N[\bar{\theta}]\partial_x\Pi_N[\bar{\theta}]f = \left[J_N^c(\bar{\theta})\left(\frac{\partial\tilde{f}_N}{\partial x} - \frac{\partial\bar{\theta}}{\partial x}\tilde{D}_N^T\tilde{f}_N\right), \boldsymbol{P}_N(\theta - \bar{\theta})\right]_N,$$

$$\Pi_N[\bar{\theta}]\sin\theta\,\Pi_N[\bar{\theta}]\partial_y\Pi_N[\bar{\theta}]f = \left[J_N^s(\bar{\theta})\left(\frac{\partial\tilde{f}_N}{\partial y} - \frac{\partial\bar{\theta}}{\partial y}\tilde{D}_N^T\tilde{f}_N\right), \boldsymbol{P}_N(\theta - \bar{\theta})\right]_N.$$

(iv) **Collision term projection:** Substitute $\Pi_N f$ into the collision term and project it onto $\mathcal{H}_N$, so that

$$\Pi_N[\bar{\theta}]Q(\Pi_N[\bar{\theta}]f) = \left[Q_N\tilde{f}_N, \boldsymbol{P}_N(\theta - \bar{\theta})\right]_N.$$

(v) **Coefficient comparison:** Substitute the found expressions into equation (2.4), use (3.5) and equate the coefficients at each basis function, thus obtaining the moment system

$$\Pi_N[\bar{\theta}]\partial_t(\Pi_N[\bar{\theta}]f) + \Pi_N[\bar{\theta}]\big(\cos\theta\,\Pi_N[\bar{\theta}](\partial_x(\Pi_N[\bar{\theta}]f))\big)$$
$$+ \Pi_N[\bar{\theta}]\big(\sin\theta\,\Pi_N[\bar{\theta}](\partial_y(\Pi_N[\bar{\theta}]f))\big) = \Pi_N[\bar{\theta}]Q\big(\Pi_N[\bar{\theta}]f\big), \qquad (3.7)$$

or

$$D_N(\boldsymbol{F}_N)_t + J_N^c D_N(\boldsymbol{F}_N)_x + J_N^s D_N(\boldsymbol{F}_N)_y = \tilde{Q}_N^{\mathrm{T}}\boldsymbol{F}_N. \qquad (3.8)$$

where

$$D_N = (I - \boldsymbol{e}_{N+2}\boldsymbol{e}_{N+2}^{\mathrm{T}} - \tilde{\tilde{D}}_N^{\mathrm{T}}\boldsymbol{F}_N\boldsymbol{e}_{N+2}^{\mathrm{T}}), \qquad (3.9)$$

and $\tilde{\tilde{D}}_N$ and $\tilde{Q}_N$ are, respectively, obtained from the matrices $\tilde{D}_N$ and $Q_N$ in (3.11)-(3.12), by replacing $(N+2)$-th row with the zero vector.

Now we describe the terms $\tilde{D}_N, J_N^c, J_N^s$ and $Q_N$.

**Proposition 3.2** (Product projections). *If $\Pi_N$ is the above defined projection operator, then*

$$\Pi_N[\bar{\theta}]\cos\theta\,\boldsymbol{P}_N(\theta - \bar{\theta}) = J_N^c(\bar{\theta})\boldsymbol{P}_N(\theta - \bar{\theta}),$$
$$\Pi_N[\bar{\theta}]\sin\theta\,\boldsymbol{P}_N(\theta - \bar{\theta}) = J_N^s(\bar{\theta})\boldsymbol{P}_N(\theta - \bar{\theta}),$$

*where*

$$J_N^c(\bar{\theta}) = \cos\bar{\theta}J_N^1 - \sin\bar{\theta}J_N^2, \quad J_N^s(\bar{\theta}) = \sin\bar{\theta}J_N^1 + \cos\bar{\theta}J_N^2, \qquad (3.10)$$

*with the matrices $J_N^1$ and $J_N^2$ defined by*

$$\Pi_N[0]\cos\theta\,\boldsymbol{P}_N(\theta) = J_N^1\boldsymbol{P}_N(\theta), \quad \Pi_N[0]\sin\theta\,\boldsymbol{P}_N(\theta) = J_N^2\boldsymbol{P}_N(\theta).$$

*Moreover, the matrices $J_N^c(\bar{\theta}) \in \mathbb{R}^{(2N+1)\times(2N+1)}$, $J_N^s(\bar{\theta}) \in \mathbb{R}^{(2N+1)\times(2N+1)}$ are symmetric, hence diagonalisable with real eigenvalues $\lambda$, all of which satisfy the inequality $|\lambda| \leq 1$.*

*Proof.* It follows from Refs. [4,13] that the matrix $J_N^c$ is diagonalisable with real eigenvalues. The entries of the vector $\boldsymbol{P}_N$ constitute an orthonormal basis, hence $\int_0^{2\pi}(\boldsymbol{P}_N\boldsymbol{P}_N^{\mathrm{T}}/M)\mathrm{d}\theta = I$ and

$$\boldsymbol{x}^{\mathrm{T}}\big(\lambda I - J_N^c(\bar\theta)\big)\boldsymbol{x} = \boldsymbol{x}^{\mathrm{T}}\left(\int_0^{2\pi}(\lambda-\cos\theta)\boldsymbol{P}_N\boldsymbol{P}_N^{\mathrm{T}}\frac{\mathrm{d}\theta}{M}\right)\boldsymbol{x} = \int_0^{2\pi}(\lambda-\cos\theta)(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{P}_N)^2\frac{\mathrm{d}\theta}{M},$$

for any $\boldsymbol{x}\in\mathbb{R}^{2N+1}$ and $\lambda\in\mathbb{R}$. Consequently, if $\lambda\notin[-1,1]$, the matrix $\lambda I - J_N^c(\bar\theta)$ is positive or negative definite, so that any eigenvalue $\lambda$ of the matrix $J_N^c(\bar\theta)$ satisfies the inequality $|\lambda|\le 1$. The eigenvalues of $J_N^s(\bar\theta)$ have the same property.

Let us now compute the entries of the matrix $J_N^c(\bar\theta)$. Taking into account the definitions of $J_N^1$ and $J_N^2$ and the trigonometric identities $\cos\theta = \cos\bar\theta\cos(\theta-\bar\theta) - \sin\bar\theta\sin(\theta-\bar\theta)$, $\sin\theta = \sin\bar\theta\cos(\theta-\bar\theta) + \cos\bar\theta\sin(\theta-\bar\theta)$, we obtain

$$\begin{aligned}
(J_N^c)_{ij}(\bar\theta) &= \langle\cos\theta(\boldsymbol{P}_N)_i(\theta-\bar\theta),(\boldsymbol{P}_N)_j(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} \\
&= \cos\bar\theta\big(\langle\cos\theta(\boldsymbol{P}_N)_i(\theta),(\boldsymbol{P}_N)_j(\theta)\rangle_{M(\theta)}\big) - \sin\bar\theta\big(\langle\sin\theta(\boldsymbol{P}_N)_i(\theta),(\boldsymbol{P}_N)_j(\theta)\rangle_{M(\theta)}\big) \\
&= \cos\bar\theta\left(J_N^1\right)_{ij} - \sin\bar\theta\left(J_N^2\right)_{ij},
\end{aligned}$$

where $(\boldsymbol{P}_N)_i$ is the $i$-th component of $\boldsymbol{P}_N$. Similar computations for the matrix $(J_N^s)_{ij}(\bar\theta)$ produce the second formula in (3.10). □

Using the recurrence relations in Remark 3.2 and the orthogonality of the basis, we can represent $J_N^1$ and $J_N^2$ in the block form

$$J_N^1 = \begin{pmatrix} J_{11} & O \\ O & J_{12} \end{pmatrix}, \quad J_N^2 = \begin{pmatrix} O & J_{21} \\ J_{22} & O \end{pmatrix},$$

with the tridiagonal matrices $J_{11}\in\mathbb{R}^{(N+1)\times(N+1)}, J_{12}\in\mathbb{R}^{N\times N}, J_{21}\in\mathbb{R}^{(N+1)\times N}, J_{22}\in\mathbb{R}^{N\times(N+1)}$, so that $J_N^c$ and $J_N^s$ consist of four tridiagonal matrices.

**Proposition 3.3** (Derivative projections)**.** *If $\Pi_N$ is the above defined projection operator, then*

$$\Pi_N[\bar\theta]\mathrm{d}\big(\boldsymbol{P}_N(\theta-\bar\theta)\big) = \tilde{D}_N\boldsymbol{P}_N(\theta-\bar\theta)\mathrm{d}(\theta-\bar\theta),$$

*where $\tilde{D}_N\in\mathbb{R}^{(2N+1)\times(2N+1)}$ is a constant matrix, the $(i,j)$-entry of which is*

$$-\int_0^{2\pi}(\boldsymbol{P}_N)_i(\theta)\mathrm{d}\frac{(\boldsymbol{P}_N)_j(\theta)}{M(\theta)}. \tag{3.11}$$

*Proof.* The $(i,j)$-entry of the matrix $\tilde{D}_N$ has the form

$$\begin{aligned}
(\tilde{D}_N)_{ij} &= \langle\mathrm{d}\big((\boldsymbol{P}_N)_i(\theta-\bar\theta)\big),(\boldsymbol{P}_N)_j(\theta-\bar\theta)\rangle_{M(\theta-\bar\theta)} \\
&= \int_0^{2\pi}(\boldsymbol{P}_N)_j(\theta)\frac{\mathrm{d}((\boldsymbol{P}_N)_i(\theta))}{M(\theta)} = -\int_0^{2\pi}(\boldsymbol{P}_N)_i(\theta)\mathrm{d}\frac{(\boldsymbol{P}_N)_j(\theta)}{M(\theta)}.
\end{aligned}$$

Note that above we used the periodicity of the integrand and the integration by parts. □

A notable observation here is that non-zero entries of $\tilde{D}_N$ and $J_N^2$ are located at the same positions.

**Proposition 3.4** (Collision term projection). *If*

$$Q(\boldsymbol{P}_N) \triangleq \left[ Q((\boldsymbol{P}_N)_0(\theta - \bar{\theta})), \cdots, Q((\boldsymbol{P}_N)_{2N}(\theta - \bar{\theta})) \right]^{\mathrm{T}},$$

*then*

$$\Pi_N[\bar{\theta}]Q(\boldsymbol{P}_N) = Q_N \boldsymbol{P}_N(\theta - \bar{\theta}), \tag{3.12}$$

*where $Q_N \in \mathbb{R}^{(2N+1) \times (2N+1)}$ is a symmetric negative semi-definite matrix with the zero first column and the zero first row.*

*Proof.* Recalling the property (3.2), we obtain

$$\begin{aligned}
(Q_N)_{ij} &= \langle Q((\boldsymbol{P}_N)_i(\theta - \bar{\theta})), (\boldsymbol{P}_N)_j(\theta - \bar{\theta}) \rangle_{M(\theta - \bar{\theta})} \\
&= \langle (\boldsymbol{P}_N)_i(\theta - \bar{\theta}), Q((\boldsymbol{P}_N)_j(\theta - \bar{\theta})) \rangle_{M(\theta - \bar{\theta})} \\
&= \langle Q((\boldsymbol{P}_N)_j(\theta - \bar{\theta})), (\boldsymbol{P}_N)_i(\theta - \bar{\theta}) \rangle_{M(\theta - \bar{\theta})} = (Q_N)_{ji}.
\end{aligned}$$

It is also worth noting that the first column and the first row in the matrix $Q_N$ have only zero entries since $\boldsymbol{P}_0^c = M(\theta - \bar{\theta})$ and $Q(\boldsymbol{P}_0^c) = 0$. On the other hand, the property (3.3) yields

$$\begin{aligned}
\boldsymbol{x}^{\mathrm{T}} Q_N \boldsymbol{x} &= \boldsymbol{x}^{\mathrm{T}} \left( \int_0^{2\pi} Q(\boldsymbol{P}_N) \boldsymbol{P}_N^{\mathrm{T}} \frac{\mathrm{d}\theta}{M} \right) \boldsymbol{x} = \int_0^{2\pi} Q(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_N)(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_N) \frac{\mathrm{d}\theta}{M} \\
&= \langle Q(\boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_N), \boldsymbol{x}^{\mathrm{T}} \boldsymbol{P}_N \rangle \leqslant 0,
\end{aligned}$$

for any vector $\boldsymbol{x} \in \mathbb{R}^{2N+1}$. Hence, the matrix $Q_N$ is negative semi-definite. $\qquad\square$

**Remark 3.3.** The three-term recurrence relation and orthogonality of the basis elements imply that $Q_N$ is a tridiagonal matrix. In applications, it can be calculated by numerical integration methods — e.g. the adaptive integration function `gsl_integration_qag` in GNU Scientific Library.

**Proposition 3.5** (Constraints). *If $\rho$ is defined by (2.9a), then $\rho = f_0^c$. Moreover, $f_1^s = 0$ and $a_0 f_0^c + a_1 f_1^c > 0$, where $a_0 = \int_0^{2\pi} \cos(\theta) P_0^c(\theta) \mathrm{d}\theta$, $a_1 = \int_0^{2\pi} \cos(\theta) P_1^c(\theta) \mathrm{d}\theta$.*

*Proof.* The direct use of the definition (2.9a) implies

$$\begin{aligned}
\rho &= \int_0^{2\pi} 1 \cdot \Pi_N[\bar{\theta}] f \, \mathrm{d}\theta = \int_0^{2\pi} P_0^c(\theta - \bar{\theta}) \cdot \left[ \tilde{\boldsymbol{f}}_N, \boldsymbol{P}_N(\theta - \bar{\theta}) \right]_N \frac{\mathrm{d}\theta}{M(\theta - \bar{\theta})} \\
&= \sum_{k=0}^N f_k^c \langle P_0^c(\theta), P_k^c(\theta) \rangle_{M(\theta)} + \sum_{k=1}^N f_k^s \langle P_0^c(\theta), P_k^s(\theta) \rangle_{M(\theta)} = f_0^c.
\end{aligned}$$

Similarly, Eqs. (2.9b), (2.9c) lead to the relations

$$
\int_0^{2\pi} \cos\theta\, \Pi_N[\bar\theta] f\, d\theta = \cos\bar\theta \int_0^{2\pi} \cos(\theta-\bar\theta)\cdot\left[\tilde{\boldsymbol{f}}_N, \boldsymbol{P}_N(\theta-\bar\theta)\right]_N \frac{d\theta}{M(\theta-\bar\theta)}
$$

$$
-\sin\bar\theta \int_0^{2\pi} \sin(\theta-\bar\theta)\cdot\left[\tilde{\boldsymbol{f}}_N, \boldsymbol{P}_N(\theta-\bar\theta)\right]_N \frac{d\theta}{M(\theta-\bar\theta)}
$$

$$
= \cos\bar\theta(a_0 f_0^c + a_1 f_1^c) - \sin\bar\theta\, a_2 f_1^s = \cos\bar\theta\, |J(t,\boldsymbol{x})|, \qquad (3.13)
$$

$$
\int_0^{2\pi} \sin\theta\, \Pi_N[\bar\theta] f\, d\theta = \sin\bar\theta(a_0 f_0^c + a_1 f_1^c) + \cos\bar\theta\, a_2 f_1^s = \sin\bar\theta\, |J(t,\boldsymbol{x})|, \qquad (3.14)
$$

where $a_2 = \int_0^{2\pi} \sin(\theta) P_1^s(\theta) d\theta$. Multiplying equations (3.13) and (3.14) by $\sin\bar\theta$ and $\cos\bar\theta$, respectively, and subtracting one from another gives $a_2 f_1^s = 0$, consequently, $|J(t,\boldsymbol{x})| = a_0 f_0^c + a_1 f_1^c$. Since $a_2 \neq 0$, we also have $f_1^s = 0$, and the condition $|J(t,\boldsymbol{x})| \neq 0$ implies $a_0 f_0^c + a_1 f_1^c > 0$. $\qquad\square$

**Remark 3.4.** The moment system can now be established for various values of $N$. For example, if $N = 0$, the moment system consists of the equation $\partial f_0^c/\partial t = 0$ only. For $N = 1$, it takes the form

$$
\frac{\partial f_0^c}{\partial t} + \nabla_{\boldsymbol{x}} \cdot (a_6 \Omega) = 0,
$$

$$
\frac{\partial f_1^c}{\partial t} + \left( a_1 \cos\bar\theta \frac{\partial f_0^c}{\partial x} + a_3 \cos\bar\theta \frac{\partial f_1^c}{\partial x} - a_4 a_6 \sin\bar\theta \frac{\partial\bar\theta}{\partial x} \right)
$$

$$
+ \left( a_1 \sin\bar\theta \frac{\partial f_0^c}{\partial y} + a_3 \sin\bar\theta \frac{\partial f_1^c}{\partial y} + a_4 a_6 \cos\bar\theta \frac{\partial\bar\theta}{\partial y} \right) = (Q_N)_{11} f_1^c, \qquad (3.15)
$$

$$
\frac{\partial(a_6\bar\theta)}{\partial t} + \left( -a_2 \sin\bar\theta \frac{\partial f_0^c}{\partial x} - a_4 \sin\bar\theta \frac{\partial f_1^c}{\partial x} + a_5 a_6 \cos\bar\theta \frac{\partial\bar\theta}{\partial x} \right)
$$

$$
+ \left( a_2 \cos\bar\theta \frac{\partial f_0^c}{\partial y} + a_4 \cos\bar\theta \frac{\partial f_1^c}{\partial y} + a_5 a_6 \sin\bar\theta \frac{\partial\bar\theta}{\partial y} \right) = (Q_N)_{12} f_1^c,
$$

where $a_0, a_1$ and $a_2$ are derived in Proposition 3.5, and

$$
a_3 = \int_0^{2\pi} \cos(\theta)(P_1^c)^2(\theta) \frac{d\theta}{M(\theta)}, \qquad a_4 = \int_0^{2\pi} \sin(\theta) P_1^s(\theta) P_1^c(\theta) \frac{d\theta}{M(\theta)},
$$

$$
a_5 = \int_0^{2\pi} \cos(\theta)(P_1^s)^2(\theta) \frac{d\theta}{M(\theta)}, \qquad a_6 = a_0 f_0^c + a_1 f_1^c.
$$

Let us compare it with the macroscopic equation (2.7). The first equation in (2.7) can be rewritten as $(\partial\rho/\partial t) + \nabla_{\boldsymbol{x}} \cdot (a_0\rho\Omega) = 0$. This differs from the first equation in (3.15) but if $\varepsilon$ tends to 0, the equations are the same. Indeed, the system (2.7) has been obtained by integrating Eq. (2.6) with respect to $\theta$ and passing to the limit when $\varepsilon$ tends to 0. Hence,
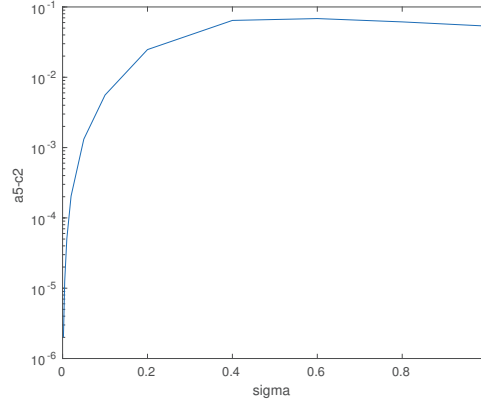
Figure 1: The difference between $a_5$ and $c_2$ for various values of $\sigma \to 0$.

Eq. (2.7) corresponds the situation where $f = \rho M$ or $f_1^c = 0$. On the other hand, if $\varepsilon$ tends to 0 but $\sigma$ does not, the coefficient $a_5$ differs from the related coefficient $c_2$ in (2.15) as they are

$$a_5 = \frac{\int_0^{2\pi} \cos\theta \sin\theta M(\theta) \sin\theta \, d\theta}{\int_0^{2\pi} \sin\theta M(\theta) \sin\theta \, d\theta}, \qquad c_2 = \frac{\int_0^{2\pi} \cos\theta \sin\theta M(\theta)\psi(\theta) d\theta}{\int_0^{2\pi} \sin\theta M(\theta)\psi(\theta) d\theta},$$

where $\psi$ is the generalized collisional invariant — cf. Ref. [14]. Fig. 1 shows the dependence of $a_5 - c_2$ on $\sigma$. One can see that if $\sigma \to 0$ (i.e. no noise), then $a_5 - c_2 \to 0$. The equation (2.15) can be recovered from the moment model if one pass to the limit as $\varepsilon \to 0$ and $\sigma \to 0$. The constraint $|\Omega| = 1$ in (2.8) is also included the moment system.

## 4. Properties of Moment Systems

### 4.1. Hyperbolicity

The hyperbolicity of the moment system (3.8) will be established if we show that the matrix $D_N$ is invertible and for any $\alpha, \beta \in \mathbb{R}$, the matrix $\alpha J_N^c + \beta J_N^s$ can be diagonalised with real eigenvalues. Let us start with the invertibility of $D_N$.

**Proposition 4.1.** *The matrix $D_N$ defined by Eq. (3.9) is invertible.*

*Proof.* The determinant of the matrix $D_N$ in (3.9) is $\det(D_N) = -(\tilde{\tilde{D}}_N)_{N+2}^{\mathrm{T}} F_N$, where $(\tilde{\tilde{D}}_N)_{N+2}^{\mathrm{T}}$ denotes the $(N+2)$-th row of the matrix $\tilde{\tilde{D}}_N^{\mathrm{T}}$. According to Proposition 3.3, we

obtain

$$
\begin{aligned}
\left(\tilde{D}_N\right)_{i,N+2} &= -\int_0^{2\pi} (\boldsymbol{P}_N)_i(\theta)\mathrm{d}\frac{(\boldsymbol{P}_N)_{N+2}(\theta)}{M(\theta)} = -\int_0^{2\pi} (\boldsymbol{P}_N)_i(\theta)\mathrm{d}(B_{1,1}\sin\theta) \\
&= -b\int_0^{2\pi} (\boldsymbol{P}_N)_i(\theta)\cos\theta\,\mathrm{d}\theta = -b\langle \cos\theta M(\theta), (\boldsymbol{P}_N)_i(\theta)\rangle_{M(\theta)} \\
&= \begin{cases} -ba_0, & \text{if } i=0, \\ -ba_1, & \text{if } i=1, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
$$

Note that $b$ is the $(1,1)$-entry of the matrix $B_N$, and $a_0, a_1$, evaluated in Proposition 3.5, satisfy the inequality $a_0 f_0^c + a_1 f_1^c > 0$. Thus, $\det(D_N) = b(a_0 f_0^c + a_1 f_1^c) \neq 0$. $\qquad\square$

Now we can formulate the main result of this section.

**Proposition 4.2.** *The moment system* (3.8) *is hyperbolic in time.*

*Proof.* Since $J_N^c$ and $J_N^s$ are real symmetric matrices — cf. Proposition 3.2, so is the linear combination $\alpha J_N^c + \beta J_N^s$, $\alpha, \beta \in \mathbb{R}$. Therefore, the matrix $\alpha J_N^c + \beta J_N^s$ can be diagonalised with real eigenvalues. Combing this with Proposition 4.1, we obtain the result. $\qquad\square$

**Remark 4.1.** Hyperbolicity plays an important role in theoretical analysis. The absence of it could lead to ill-posed equations and unphysical solutions — cf. Ref. [3].

## 4.2. Rotational invariance

**Proposition 4.3.** *The moment system* (3.8) *is invariant under the rotation transformation*

$$
\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \\ -\sin\alpha & \cos\alpha \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix},
$$
$$
\theta' = \theta - \alpha, \quad \bar{\theta}' = \bar{\theta} - \alpha.
$$

*Proof.* Let $\tilde{\boldsymbol{f}}_N = (f_0^c, \cdots, f_N^c, f_1^s, \cdots, f_N^s)^{\mathrm{T}}$ and $\tilde{\boldsymbol{f}}'_N = ((f_0^c)', \cdots, (f_N^c)', (f_1^s)', \cdots, (f_N^s)')^{\mathrm{T}}$, respectively, be the expansion coefficients of the functions $f(t, x, y, \theta)$ and $f'(t, x', y', \theta')$. The identity $f(t, x, y, \theta) = f'(t, x', y', \theta')$ yields

$$
\frac{\partial \tilde{\boldsymbol{f}}_N^{\mathrm{T}}}{\partial t} = \frac{\partial \tilde{\boldsymbol{f}}'^{\mathrm{T}}_N}{\partial t}, \qquad \frac{\partial \bar{\theta}}{\partial t} = \frac{\partial \bar{\theta}'}{\partial t}, \qquad \frac{\partial \tilde{\boldsymbol{f}}_N^{\mathrm{T}}}{\partial x} = \cos\alpha\frac{\partial \tilde{\boldsymbol{f}}'^{\mathrm{T}}_N}{\partial x'} - \sin\alpha\frac{\partial \tilde{\boldsymbol{f}}'^{\mathrm{T}}_N}{\partial y'},
$$

$$
\frac{\partial \bar{\theta}}{\partial x} = \cos\alpha\frac{\partial \bar{\theta}'}{\partial x'} - \sin\alpha\frac{\partial \bar{\theta}'}{\partial y'}, \qquad \frac{\partial \tilde{\boldsymbol{f}}_N^{\mathrm{T}}}{\partial y} = \sin\alpha\frac{\partial \tilde{\boldsymbol{f}}'^{\mathrm{T}}_N}{\partial x'} + \cos\alpha\frac{\partial \tilde{\boldsymbol{f}}'^{\mathrm{T}}_N}{\partial y'},
$$

$$
\frac{\partial \bar{\theta}}{\partial y} = \sin\alpha\frac{\partial \bar{\theta}'}{\partial x'} + \cos\alpha\frac{\partial \bar{\theta}'}{\partial y'}.
$$

Using the equations

$$J_N^c(\bar{\theta}) = \cos\bar{\theta} J_N^1 - \sin\bar{\theta} J_N^2, \qquad J_N^s(\bar{\theta}) = \sin\bar{\theta} J_N^1 + \cos\bar{\theta} J_N^2,$$
$$J_N^c(\bar{\theta}') = \cos\bar{\theta}' J_N^1 - \sin\bar{\theta}' J_N^2, \quad J_N^s(\bar{\theta}') = \sin\bar{\theta}' J_N^1 + \cos\bar{\theta}' J_N^2,$$

we obtain

$$\frac{\partial \tilde{\boldsymbol{f}}_N^{\mathrm{T}}}{\partial x} J_N^c(\bar{\theta}) + \frac{\partial \tilde{\boldsymbol{f}}_N^{\mathrm{T}}}{\partial y} J_N^s(\bar{\theta})$$
$$= \cos\bar{\theta}' \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial x} J_N^1(\bar{\theta}) - \sin\bar{\theta}' \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial x} J_N^2(\bar{\theta}) + \sin\bar{\theta}' \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial y} J_N^1(\bar{\theta}) + \cos\bar{\theta}' \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial y} J_N^2(\bar{\theta})$$
$$= \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial x'} J_N^c(\bar{\theta}') + \frac{\partial \tilde{\boldsymbol{f}}'_N^{\mathrm{T}}}{\partial y'} J_N^s(\bar{\theta}'),$$
$$\frac{\partial \bar{\theta}}{\partial x} J_N^c(\bar{\theta}) + \frac{\partial \bar{\theta}}{\partial y} J_N^s(\bar{\theta}) = \frac{\partial \bar{\theta}'}{\partial x'} J_N^c(\bar{\theta}') + \frac{\partial \bar{\theta}'}{\partial y'} J_N^s(\bar{\theta}'),$$
$$\tilde{\boldsymbol{f}}_N^{\mathrm{T}} Q_N = \tilde{\boldsymbol{f}}'_N^{\mathrm{T}} Q_N,$$

so that $\boldsymbol{F}'_N = (I - \boldsymbol{e}_{N+2}\boldsymbol{e}_{N+2}^T)\tilde{\boldsymbol{f}}'_N + \bar{\theta}'\boldsymbol{e}_{N+2}$. We can substitute it into Eq. (3.8) and the proof of Proposition 4.3 follows. $\qquad\square$

## 4.3. Mass conservation

Let us note another property of the moment system.

**Proposition 4.4.** *The moment system* (3.8) *preserves the mass-conservation.*

*Proof.* Since $\rho = f_0^c$, one has to check whether $f_0^c$ stays unchanged. Simple calculations show that the first equation in the moment system is

$$\frac{\partial f_0^c}{\partial t} + \nabla_{\boldsymbol{x}} \cdot \left( (a_0 f_0^c + a_1 f_1^c)\Omega \right) = 0,$$

where $a_0$ and $a_1$ are the same as in Proposition 3.5 and $\Omega = (\cos\bar{\theta}, \sin\bar{\theta})^{\mathrm{T}}$. Since the equation is in the divergence form, the mass conservation is kept by the convective term. According to Proposition 3.4, the matrix $Q_N$ has zero first row and zero first column. Thus the first entry in $\tilde{Q}_N^{\mathrm{T}} \boldsymbol{F}_N$ is equal to 0, which does not change $f_0^c$, hence the moment system (3.8) is mass-conservative. $\qquad\square$

## 5. Numerical Experiments

In this section we carry out numerical experiments to examine the main features of the hyperbolic moment systems (3.7)-(3.8).

## 5.1. Grad type expansions for various $\bar{\theta}$

Let us start from the relationship between the Grad type expansions of density distribution for different $\bar{\theta}$.

**Proposition 5.1.** *If*

$$\Pi_N[\bar{\theta}_1]f = \sum_{k=0}^{2N} \tilde{f}_k^1 (\boldsymbol{P}_N)_k(\theta - \bar{\theta}_1), \quad \Pi_N[\bar{\theta}_2]f = \sum_{k=0}^{2N} \tilde{f}_k^2 (\boldsymbol{P}_N)_k(\theta - \bar{\theta}_2),$$

*then*

$$[\tilde{f}_0^2, \cdots, \tilde{f}_{2N}^2]^{\mathrm{T}} = T_N(\bar{\theta}_1 - \bar{\theta}_2)[\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]^{\mathrm{T}}, \tag{5.1}$$

*where*

$$T_N(\bar{\theta}) = \begin{pmatrix} A_N & O \\ O & B_N \end{pmatrix} X_N(\bar{\theta}) \begin{pmatrix} A_N^{-1} & O \\ O & B_N^{-1} \end{pmatrix},$$

*the entries $a_{i,j}$, $b_{i,j}$ of the matrices $A_N = (a_{i,j})_{(N+1)\times(N+1)}$, $B_N = (b_{i,j})_{N\times N}$, are defined by relations* (3.1) *and*

$$X_N(\theta) = \begin{pmatrix} 1 & 0 & \cdots & \cdots & & 0 \\ 0 & \cos(\theta) & O & -\sin(\theta) & & O \\ & & \ddots & & \ddots & \\ \vdots & O & \cos(N\theta) & O & & -\sin(N\theta) \\ \vdots & \sin(\theta) & O & \cos(\theta) & & O \\ & & \ddots & & \ddots & \\ 0 & O & \sin(N\theta) & O & & \cos(N\theta) \end{pmatrix}.$$

*Proof.* We first consider the connections between basis functions. Setting

$$\boldsymbol{E}_N^c(\theta) \triangleq \big(1, \cos\theta, \cdots, \cos(N\theta)\big)^{\mathrm{T}}, \qquad \boldsymbol{E}_N^s(\theta) \triangleq \big(\sin\theta, \cdots, \sin(N\theta)\big)^{\mathrm{T}},$$

$$\boldsymbol{H}_N^c(\theta) \triangleq \big(H_0^c(\theta), H_1^c(\theta), \cdots, H_N^c(\theta)\big)^{\mathrm{T}}, \quad \boldsymbol{H}_N^s(\theta) \triangleq \big(H_1^s(\theta), H_2^s(\theta), \cdots, H_N^s(\theta)\big)^{\mathrm{T}},$$

and taking into account the representations

$$\boldsymbol{E}_N^c(\theta - \bar{\theta}_2)$$
$$= \begin{pmatrix} 1 & 0 & \cdots & \cdots & & 0 \\ 0 & \cos(\bar{\theta}_1 - \bar{\theta}_2) & O & -\sin(\bar{\theta}_1 - \bar{\theta}_2) & & O \\ \vdots & & \ddots & & \ddots & \\ 0 & O & \cos(N(\bar{\theta}_1 - \bar{\theta}_2)) & O & & -\sin(N(\bar{\theta}_1 - \bar{\theta}_2)) \end{pmatrix} \times \begin{pmatrix} \boldsymbol{E}_N^c(\theta - \bar{\theta}_1) \\ \boldsymbol{E}_N^s(\theta - \bar{\theta}_1) \end{pmatrix},$$

$$\boldsymbol{E}_N^s(\theta - \bar{\theta}_2)$$
$$= \begin{pmatrix} \sin(\bar{\theta}_1 - \bar{\theta}_2) & O & \cos(\bar{\theta}_1 - \bar{\theta}_2) & O \\ & \ddots & & \ddots \\ O & \sin(N(\bar{\theta}_1 - \bar{\theta}_2)) & O & \cos(N(\bar{\theta}_1 - \bar{\theta}_2)) \end{pmatrix} \times \begin{pmatrix} \boldsymbol{E}_N^c(\theta - \bar{\theta}_1) \\ \boldsymbol{E}_N^s(\theta - \bar{\theta}_1) \end{pmatrix},$$

we obtain

$$
\begin{pmatrix} \boldsymbol{H}_N^c(\theta - \bar\theta_2) \\ \boldsymbol{H}_N^s(\theta - \bar\theta_2) \end{pmatrix} = \begin{pmatrix} A_N & O \\ O & B_N \end{pmatrix} X_N(\bar\theta_1 - \bar\theta_2) \begin{pmatrix} A_N^{-1} & O \\ O & B_N^{-1} \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_N^c(\theta - \bar\theta_1) \\ \boldsymbol{H}_N^s(\theta - \bar\theta_1) \end{pmatrix}
$$

$$
= T_N(\bar\theta_1 - \bar\theta_2) \begin{pmatrix} \boldsymbol{H}_N^c(\theta - \bar\theta_1) \\ \boldsymbol{H}_N^s(\theta - \bar\theta_1) \end{pmatrix}.
$$

Let $\tilde{T}_N(\bar\theta_1 - \bar\theta_2)$ be the matrix connecting the vectors $\boldsymbol{P}_N(\theta - \bar\theta_1)$ and $\boldsymbol{P}_N(\theta - \bar\theta_2)$ — i.e. $\boldsymbol{P}_N(\theta - \bar\theta_1) = \tilde{T}_N(\bar\theta_1 - \bar\theta_2)\boldsymbol{P}_N(\theta - \bar\theta_2)$. Then

$$
(\tilde{T}_N)_{ij}(\bar\theta_1 - \bar\theta_2) = \int_0^{2\pi} (\boldsymbol{P}_N)_i(\theta - \bar\theta_1)(\boldsymbol{P}_N)_j(\theta - \bar\theta_2)/M(\theta - \bar\theta_2)\mathrm{d}\theta
$$

$$
= \int_0^{2\pi} (\boldsymbol{H}_N)_i(\theta - \bar\theta_1)(\boldsymbol{H}_N)_j(\theta - \bar\theta_2)M(\theta - \bar\theta_1)\mathrm{d}\theta = (T_N)_{ji}(\bar\theta_1 - \bar\theta_2),
$$

where $(\boldsymbol{H}_N)_i$ the $i$-th component of the vector $\left(\boldsymbol{H}_N^c, \boldsymbol{H}_N^s\right)^T$. Thus $\tilde{T}_N(\bar\theta_1 - \bar\theta_2) = T_N'(\bar\theta_1 - \bar\theta_2)$ and since

$$
\Pi_N f = [\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]\boldsymbol{P}_N(\theta - \bar\theta_1) = [\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]\tilde{T}_N(\bar\theta_1 - \bar\theta_2)\boldsymbol{P}_N(\theta - \bar\theta_2)
$$

$$
= [\tilde{f}_0^2, \cdots, \tilde{f}_{2N}^2]\boldsymbol{P}_N(\theta - \bar\theta_2),
$$

we obtain

$$
[\tilde{f}_0^2, \cdots, \tilde{f}_{2N}^2]^{\mathrm{T}} = T_N(\bar\theta_1 - \bar\theta_2)[\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]^{\mathrm{T}},
$$

which completes the proof.                                                                                                 $\square$

## 5.2. Numerical scheme

We consider a uniform spatial grid $\{(x_i, y_j), i, j \in \mathbb{Z}\}$ with the stepsizes $\Delta x = x_{i+1} - x_i$ and $\Delta y = y_{j+1} - y_j$. The grid $\{t_{n+1} = t_n + \Delta t, n \in \mathbb{N}\}$ in $t$-direction has stepsize $\Delta t = C_{\mathrm{CFL}}\Delta x$, where $C_{\mathrm{CFL}}$ is the Courant-Friedrichs-Lewy number. The approximations of $f$ and $\bar\theta$ at the point $(t_n, x_i, y_j)$ are, respectively, denoted by $f_{i,j}^n$ and $\bar\theta_{i,j}^n$. Consequently, $(\Pi f)_{i,j}^n := \Pi_N[\bar\theta_{i,j}^n]f_{i,j}^n$. Similarly to Ref. [5], here we only consider the first-order semi-implicit operator-splitting numerical scheme for the systems (3.7) and (3.8) divided into convection and collision steps

$$
\Pi_N[\bar\theta_{i,j}^n](\Pi f)_{i,j}^{n*} = (\Pi f)_{i,j}^n - \frac{\Delta t}{\Delta x}\left[\left(\Pi F^-\right)_{i+\frac{1}{2},j}^n - \left(\Pi F^+\right)_{i-\frac{1}{2},j}^n\right],
$$

$$
\Pi_N[\bar\theta_{i,j}^{n*}](\Pi f)_{i,j}^{n**} = (\Pi f)_{i,j}^{n*} - \frac{\Delta t}{\Delta y}\left[\left(\Pi F^-\right)_{i,j+\frac{1}{2}}^{n*} - \left(\Pi F^+\right)_{i,j-\frac{1}{2}}^{n*}\right],
$$

$$
\Pi_N[\bar\theta_{i,j}^{n**}]\left(\frac{(\Pi f)_{i,j}^{n+1} - (\Pi f)_{i,j}^{n**}}{\Delta t}\right) = \Pi_N[\bar\theta_{i,j}^{n**}]Q\left((\Pi f)_{i,j}^{n+1}\right), \tag{5.2}
$$

with the numerical non-conservative HLL fluxes — cf. Refs. [5, 26]. For example, the flux in $x$-direction is

$$
\left(\Pi F^-\right)^n_{i+\frac{1}{2},j} =
\begin{cases}
\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta(\Pi f)^n_{i,j}\right), & 0 \leqslant \lambda^L_{i+\frac{1}{2},j}, \\[2ex]
\dfrac{\lambda^R_{i+\frac{1}{2},j}\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta(\Pi f)^n_{i,j}\right) - \lambda^L_{i+\frac{1}{2},j}\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta\Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i+1,j}\right)}{\lambda^R_{i+\frac{1}{2},j} - \lambda^L_{i+\frac{1}{2},j}} \\[2ex]
\quad + \dfrac{\lambda^L_{i+\frac{1}{2},j}\lambda^R_{i+\frac{1}{2},j}\left(\Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i+1,j} - (\Pi f)^n_{i,j}\right)}{\lambda^R_{i+\frac{1}{2},j} - \lambda^L_{i+\frac{1}{2},j}}, & \lambda^L_{i+\frac{1}{2},j} < 0 < \lambda^R_{i+\frac{1}{2},j}, \\[2ex]
\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta\Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i+1,j}\right), & 0 \geqslant \lambda^R_{i+\frac{1}{2},j},
\end{cases}
$$

$$
\left(\Pi F^+\right)^n_{i-\frac{1}{2},j} =
\begin{cases}
\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta\Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i-1,j}\right), & 0 \leqslant \lambda^L_{i-\frac{1}{2},j}, \\[2ex]
\dfrac{\lambda^R_{i-\frac{1}{2},j}\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta\Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i-1,j}\right) - \lambda^L_{i-\frac{1}{2},j}\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta(\Pi f)^n_{i,j}\right)}{\lambda^R_{i-\frac{1}{2},j} - \lambda^L_{i-\frac{1}{2},j}} \\[2ex]
\quad + \dfrac{\lambda^L_{i-\frac{1}{2},j}\lambda^R_{i-\frac{1}{2},j}\left((\Pi f)^n_{i,j} - \Pi_N[\bar\theta^n_{i,j}](\Pi f)^n_{i-1,j}\right)}{\lambda^R_{i-\frac{1}{2},j} - \lambda^L_{i-\frac{1}{2},j}}, & \lambda^L_{i-\frac{1}{2},j} < 0 < \lambda^R_{i-\frac{1}{2},j}, \\[2ex]
\Pi_N[\bar\theta^n_{i,j}]\left(\cos\theta(\Pi f)^n_{i,j}\right), & 0 \geqslant \lambda^R_{i-\frac{1}{2},j},
\end{cases}
$$

where $\lambda^L_{i\pm\frac{1}{2},j} = \min\{\lambda^{\min}_{i,j}, \lambda^{\min}_{i\pm 1,j}\}$, $\lambda^R_{i\pm\frac{1}{2},j} = \max\{\lambda^{\max}_{i,j}, \lambda^{\max}_{i\pm 1,j}\}$, $\lambda^{\min}_{i,j}$ and $\lambda^{\max}_{i,j}$ refer to minimal and maximal eigenvalues of the matrix $J^c_N(\bar\theta)$ evaluated at the point $(x_i, y_j)$.

**Remark 5.1.** As was shown before, the eigenvalues of the moment system are located in the interval $[-1, 1]$. However, their exact values are not known and numerical calculation is a time-consuming operation. Therefore, in practice, we can take $\lambda^{\max} = 1$ and $\lambda^{\min} = -1$. The HLL scheme then reduces to the Lax-Friedrichs scheme, and numerical experiments show no substantial difference in these approaches.

**Remark 5.2.** The numerical scheme preserves the constraints, since at the moment $t^n$, we first update the value $\bar\theta^{n*}$ and then transform the expansion related to the basis at $\bar\theta^n$ into expansion related to $\bar\theta^{n*}$. The relations (3.13) and (3.14) are also valid for the basis connected with the point $\bar\theta^{n*}$, so that $f^s_1 = 0$ and $a_0 f^c_0 + a_1 f^c_1 > 0$. The same is true for the time level $t^{n*}$ and $t^{n**}$.

**Proposition 5.2.** *For any $\bar\theta_1, \bar\theta_2$, the equation*

$$
\Pi_N[\bar\theta_1]f = \Pi_N[\bar\theta_1]\Pi_N[\bar\theta_2]f \tag{5.3}
$$

*holds.*

*Proof.* Considering the expressions

$$
\Pi_N[\bar\theta_1]f = \sum_{k=0}^{2N} \tilde f^1_k (\boldsymbol{P}_N)_k(\theta - \bar\theta_1), \quad \Pi_N[\bar\theta_2]f = \sum_{k=0}^{2N} \tilde f^2_k (\boldsymbol{P}_N)_k(\theta - \bar\theta_2),
$$

and applying the identity (5.1), we obtain

$$[\tilde{f}_0^2, \cdots, \tilde{f}_{2N}^2]^{\mathrm{T}} = T_N(\bar{\theta}_1 - \bar{\theta}_2)[\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]^{\mathrm{T}}.$$

The mapping $\Pi_N[\bar{\theta}_1]\Pi_N[\bar{\theta}_2]f$ transfers $\Pi_N[\bar{\theta}_2]f$, represented in the $\bar{\theta}_2$-basis, into the corresponding representation in $\bar{\theta}_1$-basis, so that

$$T_N(\bar{\theta}_2 - \bar{\theta}_1)[\tilde{f}_0^2, \cdots, \tilde{f}_{2N}^2]^{\mathrm{T}} = T_N(\bar{\theta}_2 - \bar{\theta}_1)T_N(\bar{\theta}_1 - \bar{\theta}_2)[\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]^{\mathrm{T}} = [\tilde{f}_0^1, \cdots, \tilde{f}_{2N}^1]^{\mathrm{T}},$$

and the proof is completed. $\qquad\square$

Thus in order to determine the coefficients of $f_{i,j}$ in $\Pi_N[\bar{\theta}_{i,j}^{n*}]$, we first calculate the coefficients of $f_{i,j}$ in $\Pi_N[\bar{\theta}_{i,j}^n]$, the value $\bar{\theta}^{n*}$, and use the identity (5.3) with the transition matrix (5.1).

In summary, the numerical scheme is organised as follows:

(a) Update the $x$-convective step and obtain $\Pi_N[\bar{\theta}_{i,j}^n](\Pi f)_{i,j}^{n*}$.

(b) Compute $\bar{\theta}^{n*}$ and establish $(\Pi f)_{i,j}^{n*}$.

(c) Update the $y$-convective step and obtain $\Pi_N[\bar{\theta}_{i,j}^{n*}](\Pi f)_{i,j}^{n**}$.

(d) Compute $\bar{\theta}^{n**}$ and establish $(\Pi f)_{i,j}^{n**}$.

(e) Update the collision step and derive $\Pi_N[\bar{\theta}_{i,j}^{n**}](\Pi f)_{i,j}^{n+1}$.

(f) Compute $\bar{\theta}^{n+1}$ and $(\Pi f)_{i,j}^{n+1}$. Set $n := n+1$ and goto (a).

Note that the collision step is updated via an implicit scheme. The matrix $Q_N$ is substituted into (5.2), so that

$$\Pi_N[\bar{\theta}_{i,j}^{n**}](\Pi f)_{i,j}^{n+1} = \Pi_N[\bar{\theta}_{i,j}^{n**}](I - \Delta t Q_N)^{-1}(\Pi f)_{i,j}^{n**}. \tag{5.4}$$

**Proposition 5.3.** *The implicit discretisation* (5.4) *is unconditionally stable.*

*Proof.* Since $Q_N$ is semi-negative definite matrix, all its eigenvalues $\lambda$ are non-positive. It follows that the eigenvalues of $(I - \Delta t Q_N)^{-1}$ satisfy the inequality $0 \leqslant (1 - \Delta t \lambda)^{-1} \leqslant 1$, implying the unconditional stability of the procedure (5.4). $\qquad\square$

**Remark 5.3.** One of drawbacks of the model reduction by operator projection is the non-conservativity of the moment system derived — cf. Ref. [13]. It is interesting to know how this factor influences the corresponding numerical solution. However, the present work does not deal with numerical methods for non-conservative moment systems, in which case approximate solutions still converge to the solution obtained by spectral methods in our numerical experiments. Let us note Refs. [19, 20], where numerical methods for partially conservative moment systems for the Boltzmann equation are considered.

## 5.3. Reflection boundaries

The reflection boundaries can be treated analogously to upwind schemes. As an example, we consider the left boundary in $x$ direction. For any $j$, the macroscopic numerical flux $\left(\Pi F^+\right)_{-\frac{1}{2},j}$ at the left boundary is defined as $(\Pi f)_{0,j}$ if $\cos\theta \leqslant 0$ and as the reflection of $(\Pi f)_{0,j}$ if $\cos\theta > 0$ — i.e.

$$
\left(\Pi F^+\right)_{-\frac{1}{2},j} = \begin{cases} \cos\theta(\Pi f)_{0,j}(\theta), & \cos\theta \leqslant 0, \\ \cos\theta(\Pi f)_{0,j}(\pi-\theta), & \cos\theta > 0. \end{cases}
$$

Expanding $\left(\Pi F^+\right)^n_{-\frac{1}{2},j}$ in $\boldsymbol{P}_N(\theta - \bar{\theta}_{0,j})$, we obtain

$$
(\Pi F^+)_{-\frac{1}{2},j} = \sum_{k=0}^{2N} \tilde{F}_k(\boldsymbol{P}_N)_k(\theta - \bar{\theta}_{0,j}), \tag{5.5}
$$

where

$$
\begin{aligned}
\tilde{F}_k = &\int_{\cos\theta\leqslant 0} \cos\theta(\Pi f)_{0,j}(\theta)(\boldsymbol{P}_N)_k(\theta - \bar{\theta}_{0,j})\frac{\mathrm{d}\theta}{M(\theta - \bar{\theta}_{0,j})} \\
&+ \int_{\cos\theta>0} \cos\theta(\Pi f)_{0,j}(\pi-\theta)(\boldsymbol{P}_N)_k(\theta - \bar{\theta}_{0,j})\frac{\mathrm{d}\theta}{M(\theta - \bar{\theta}_{0,j})}.
\end{aligned}
$$

In particular, the coefficient $\tilde{F}_0$ in (5.5) is

$$
\begin{aligned}
\tilde{F}_0 &= \int_{\pi/2}^{3\pi/2} \cos\theta(\Pi f)_{0,j}(\theta)\mathrm{d}\theta + \int_{-\pi/2}^{\pi/2} \cos\theta(\Pi f)_{0,j}(\pi-\theta)\mathrm{d}\theta \\
&= \int_{\pi/2}^{3\pi/2} \cos\theta(\Pi f)_{0,j}(\theta)\mathrm{d}\theta + \int_{3\pi/2}^{\pi/2} \cos t(\Pi f)_{0,j}(t)\mathrm{d}t = 0,
\end{aligned}
$$

so that our approach to the reflection boundary does not destroy the mass-conservation.

## 5.4. Numerical results

We apply the 2-D moment system scheme to three Riemann problems with solutions homogeneous in the $y$-direction and also to a vortex formation problem, comparing our results to the spectral method solutions in Ref. [14].

For Riemann problems, the computational domain in the $x$-direction is $[-5, 5]$, the CFL number 0.5, and $\sigma = 0.2$.

**Example 5.1** (Rarefaction wave)**.** The initial density $\rho^\varepsilon$ and velocity angle $\bar{\theta}^\varepsilon$ for the first Riemann problem having a rarefaction wave solution, are

$$
(\rho^\varepsilon, \bar{\theta}^\varepsilon) = \begin{cases} (2, 1.7), & x < 0, \\ (0.218, 0.5), & x > 0, \end{cases}
$$

Table 1: Computational parameters.

| $\varepsilon$ | The interaction time scale |
|---|---|
| $\sigma$ | Noise, 0.2 |
| $C_{CFL}$ | CFL number, 0.5 |
| $\Delta x$ | Mesh step for $x$-direction, $0.00125 \sim 0.01$ |
| $\Delta y$ | Mesh step for $y$-direction, $0.00125 \sim 0.01$ |
| $\Delta t$ | Time step $\Delta t = C_{\text{CFL}}\Delta x$ |
| $N$ | Number of the moments |
| $N_v$ | Number of discrete velocities in the spectral method |
| $n$ | Number of grid points in $x$- or $y$- direction |

Table 2: Example 5.1: Relative $\ell^1$ error of $\rho^\varepsilon, \bar{\theta}^\varepsilon$ and estimated CPU times (seconds) with respect to $N$ or $N_v$ with 2000 cells.

| | | | | $\varepsilon = 1$ | | | | |
|---|---|---|---|---|---|---|---|---|
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 16.3 | 30.7 | 48.2 | 61.8 | 85.5 | 111.9 | 140.5 | 160.7 |
| $\ell^1$ error ($\rho^\varepsilon$) | 6.93e-01 | 8.94e-02 | 2.72e-02 | 2.73e-02 | 2.67e-02 | 2.73e-02 | 2.75e-02 | 2.76e-02 |
| $\ell^1$ error ($\bar{\theta}^\varepsilon$) | 3.15e-01 | 4.07e-02 | 2.50e-02 | 2.32e-02 | 2.37e-02 | 2.40e-02 | 2.42e-02 | 2.43e-02 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 66.2 | 95.6 | 142.2 | 196.5 | 255.2 | 334.6 | 427.1 | 500.0 |
| $\ell^1$ error ($\rho^\varepsilon$) | 5.59e-01 | 1.13e-01 | 2.57e-02 | 1.05e-02 | 1.02e-02 | 1.02e-02 | 9.78e-03 | 1.01e-02 |
| $\ell^1$ error ($\bar{\theta}^\varepsilon$) | 2.61e-01 | 6.15e-02 | 1.36e-02 | 8.55e-03 | 8.17e-03 | 8.72e-03 | 8.28e-03 | 8.65e-03 |
| | | | | $\varepsilon = 0.01$ | | | | |
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 16.9 | 31.3 | 44.5 | 57.1 | 79.9 | 99.8 | 130.4 | 156.6 |
| $\ell^1$ error ($\rho^\varepsilon$) | 8.87e-02 | 7.39e-02 | 7.93e-02 | 8.21e-02 | 8.35e-02 | 8.44e-02 | 8.49e-02 | 8.53e-02 |
| $\ell^1$ error ($\bar{\theta}^\varepsilon$) | 8.64e-02 | 6.22e-02 | 6.55e-02 | 6.75e-02 | 6.86e-02 | 6.93e-02 | 6.98e-02 | 7.00e-02 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 71.8 | 100.1 | 146.3 | 195.1 | 258.7 | 316.2 | 404.2 | 486.5 |
| $\ell^1$ error ($\rho^\varepsilon$) | 3.31e+00 | 3.69e-01 | 1.84e-02 | 2.15e-02 | 1.81e-02 | 2.10e-02 | 1.89e-02 | 2.07e-02 |
| $\ell^1$ error ($\bar{\theta}^\varepsilon$) | 3.63e-01 | 2.30e-02 | 1.10e-02 | 1.52e-02 | 1.25e-02 | 1.47e-02 | 1.31e-02 | 1.45e-02 |

and the von Mises distribution serves as an initial particle distribution function.

Figs. 2 and 3 show the densities $\rho^\varepsilon$ and macroscopic velocity angles $\bar{\theta}^\varepsilon$ at $t = 4$ for $\varepsilon = 1$ and $\varepsilon = 0.01$, respectively. The solid line represents the reference solution obtained by a spectral method with 4000 cells and $N_v = 32$.

Note that for $N > 1$, our results agree well with the reference solution. Moreover, for small $\varepsilon$, the reference solutions are better approximated by our method and even a small number of moments provides better approximation for original kinetic model than spectral methods. Table 2 lists the relative $\ell^1$ errors of $\rho^\varepsilon, \bar{\theta}^\varepsilon$ and CPU time for the moment and spectral methods. The error dynamic for $\rho^\varepsilon$ and $\bar{\theta}^\varepsilon$ is similar. Therefore, in what follows we only display it for $\rho^\varepsilon$. If $\varepsilon = 1$, the error of the moment method with $N = 3$ is lower
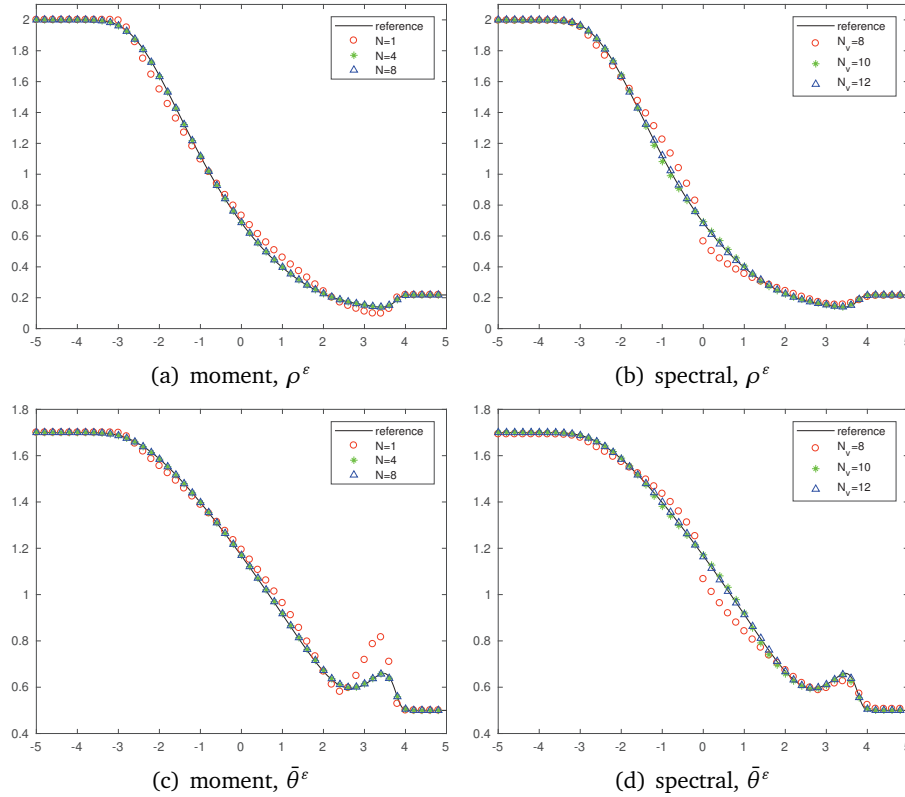
Figure 2: Example 5.1: The densities and macroscopic velocity angles for moment and spectral methods with 2000 space cells; $\varepsilon = 1$, $t = 4$. The moment numbers $N$ are $1, 4, 8$; the grid numbers in velocity for spectral method are $N_v = 8, 10, 12$. The solid line represents the reference solution obtained by the spectral method with $N_v = 32$ and 4000 space cells.

than the spectral method with $N_v = 10$ and CPU time is 48.2 and 95.6, respectively. If $\varepsilon = 0.01$, the error of the moment method with $N = 2$ is lower than the spectral method with $N_v = 10$. The CPU time is 31.3 and 100.1, respectively. Thus the moment method requires smaller number $N$ to achieve similar results.

**Example 5.2** (Shock wave). The initial data for the second Riemann problem are

$$(\rho^\varepsilon, \bar{\theta}^\varepsilon) = \begin{cases} (1, 1.5), & x < 0, \\ (2, 1.83), & x > 0. \end{cases}$$

Figs. 4 and 5 show the densities $\rho^\varepsilon$ and macroscopic velocity angles $\bar{\theta}^\varepsilon$ at $t = 4$ for $\varepsilon = 1$ and $\varepsilon = 0.01$. The solid line represents the reference solution obtained by the spectral method with 4000 cells, $N_v = 32$.

Note for $N > 1$, the results obtained by the moment system method are consistent with the reference solution. Moreover, for small $\varepsilon$, the moment method provides better approximation for the shock wave solution than the spectral method. Table 3 shows the relative $\ell^1$-error for $\rho^\varepsilon$ and CPU time. If $\varepsilon = 1$, the results for $N = 2$ are better than for

(a) moment, $\rho^\varepsilon$     (b) spectral, $\rho^\varepsilon$

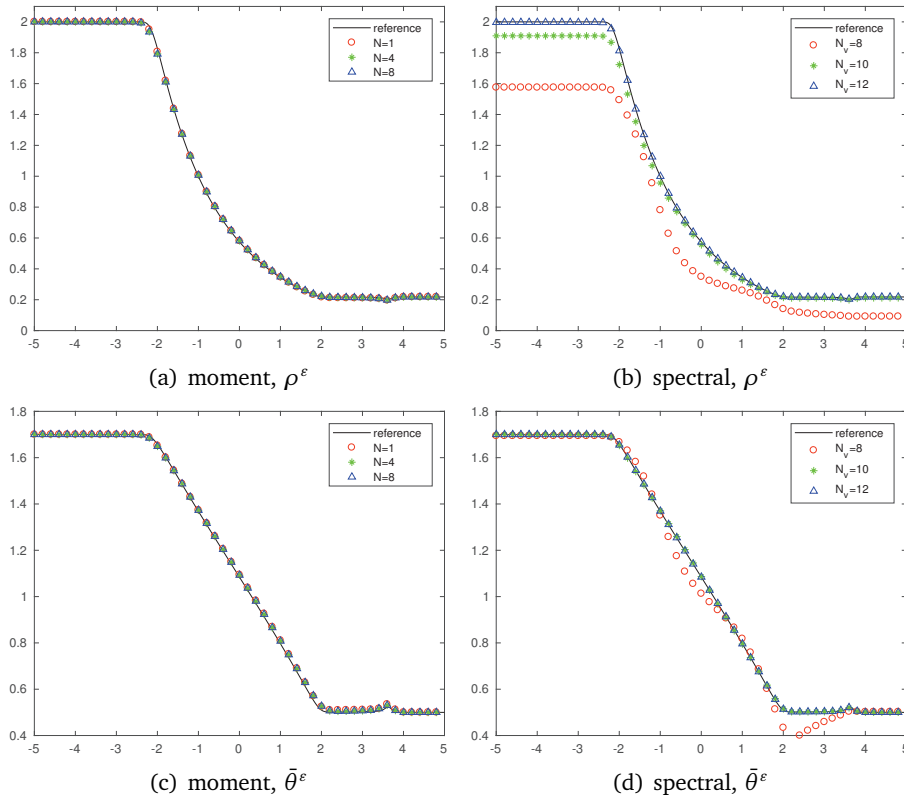(c) moment, $\bar{\theta}^\varepsilon$     (d) spectral, $\bar{\theta}^\varepsilon$

Figure 3: Same as in Fig. 2 but for $\varepsilon = 0.01$.

Table 3: Example 5.2: Same as in Tab. 2 but for Example 5.2.

| | | | | $\varepsilon = 1$ | | | | |
|---|---|---|---|---|---|---|---|---|
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 19.0 | 30.4 | 44.1 | 57.5 | 79.1 | 99.1 | 130.5 | 155.4 |
| $\ell^1$ error | 9.17e-02 | 4.06e-03 | 4.79e-03 | 5.09e-03 | 5.21e-03 | 5.27e-03 | 5.32e-03 | 5.35e-03 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 65.0 | 87.9 | 124.6 | 162.6 | 217.5 | 268.2 | 342.9 | 420.5 |
| $\ell^1$ error | 1.68e-01 | 1.13e-02 | 2.38e-03 | 1.72e-03 | 1.69e-03 | 1.72e-03 | 1.69e-03 | 1.72e-03 |
| | | | | $\varepsilon = 0.01$ | | | | |
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 19.8 | 30.9 | 43.4 | 56.8 | 79.1 | 98.6 | 129.6 | 154.9 |
| $\ell^1$ error | 1.96e-02 | 1.47e-02 | 1.54e-02 | 1.59e-02 | 1.62e-02 | 1.64e-02 | 1.66e-02 | 1.67e-02 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 55.4 | 79.7 | 116.1 | 152.1 | 207.5 | 267.9 | 333.6 | 408.8 |
| $\ell^1$ error | 3.57e+00 | 2.57e-01 | 2.82e-02 | 5.41e-03 | 4.88e-03 | 5.08e-03 | 4.92e-03 | 5.07e-03 |

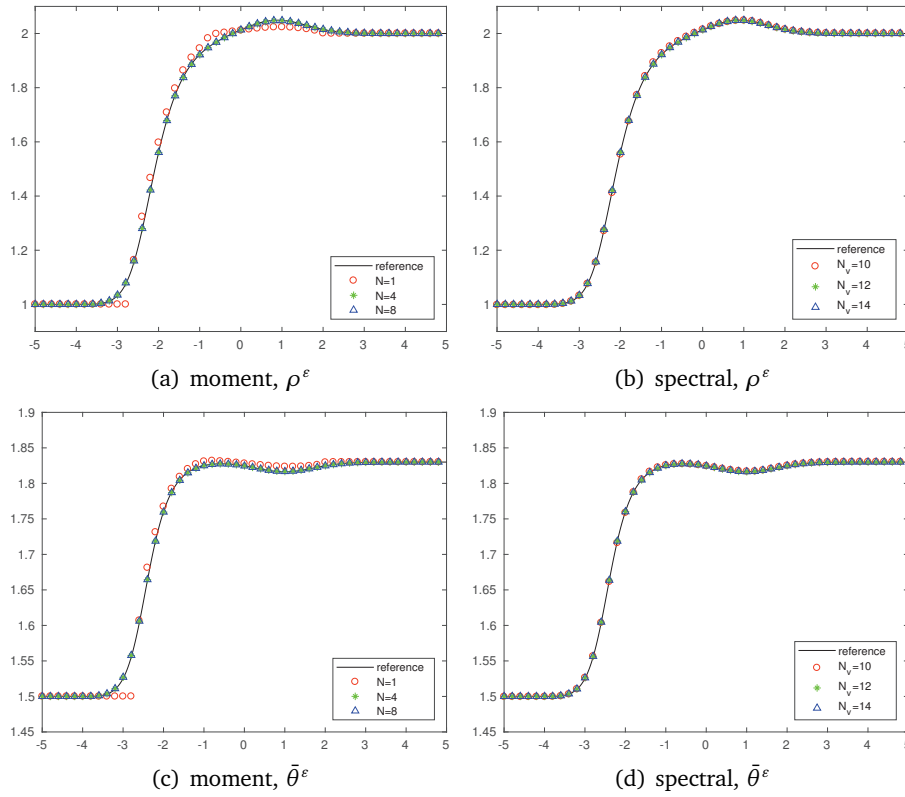$N_v = 10$ and CPU time is, respectively, 30.4 and 87.9. If $\varepsilon = 0.01$, CPU time for $N = 2$

Figure 4: Example 5.2: The densities and macroscopic velocity angles for 2000 space cells; $t = 4$, $\varepsilon = 1$. The moment numbers $N$ are $1, 4, 8$; the grid numbers in velocity for the spectral method $N_v$ are $10, 12, 14$. The solid line represents the reference solution obtained by the spectral method with $N_v = 32$ and 4000 cells.

and $N_v = 12$ is, respectively, 30.9 and 116.1. Thus the moment method overperforms the spectral method.

For $N = 1$, there is a sub-shock approximately at the point $x = -2.5$. The increase of $\varepsilon$ and the density ratio produces more solution sub-shocks in both moment and spectral methods — cf. Refs. [2, 19].

**Example 5.3** (Contact discontinuity). The initial data for the third Riemann problem are

$$(\rho^\varepsilon, \bar{\theta}^\varepsilon) = \begin{cases} (1, 1), & x < 0, \\ (1, -1), & x > 0. \end{cases}$$

Figs. 6 and 7 shows the densities $\rho^\varepsilon$ and macroscopic velocity angles $\bar{\theta}^\varepsilon$ at $t = 4$ for $\varepsilon = 1$ and $\varepsilon = 0.01$. The solid line represents the reference solution obtained by the spectral method with 8000 cells, $N_v = 32$.

Note for $\varepsilon = 1$, $N > 5$ and for $\varepsilon = 0.01$, $N > 4$ the results for the moment system method are consistent with the reference solution. The convergence is faster for smaller $\varepsilon$. Table 4 shows the relative $\ell^1$-error for $\rho^\varepsilon$ and CPU time. Comparing CPU time for $N = 7$,

(a) moment, $\rho^\varepsilon$

(b) spectral, $\rho^\varepsilon$

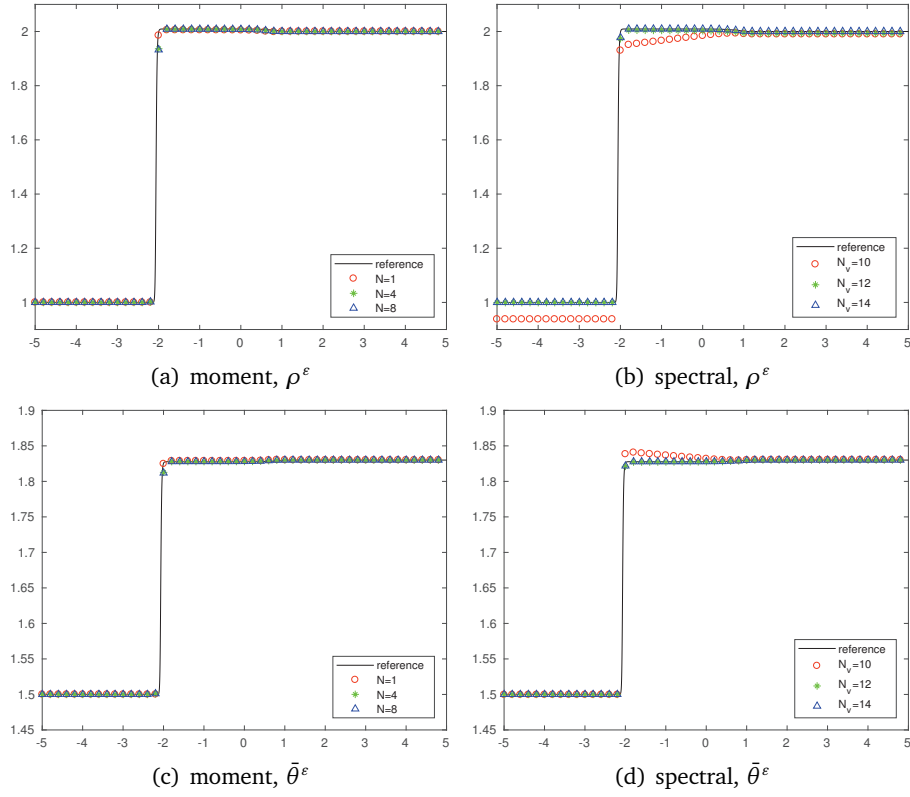(c) moment, $\bar{\theta}^\varepsilon$

(d) spectral, $\bar{\theta}^\varepsilon$

Figure 5: Same as in Fig. 4 but for $\varepsilon = 0.01$.

Table 4: Example 5.3: Same as in Tab. 2 but for Example 5.3 with 4000 cells.

| | | | | $\varepsilon = 1$ | | | | |
|---|---|---|---|---|---|---|---|---|
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 78.4 | 121.7 | 170.9 | 228.4 | 308.5 | 392.4 | 506.5 | 608.9 |
| $\ell^1$ error | 1.41e+00 | 5.92e-01 | 5.11e-01 | 8.86e-02 | 2.92e-02 | 1.27e-02 | 7.72e-03 | 9.95e-03 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 232.8 | 342.6 | 511.7 | 701.8 | 1000.3 | 1170.9 | 1506.4 | 1783.1 |
| $\ell^1$ error | 2.69e-01 | 1.22e-01 | 1.05e-02 | 5.82e-03 | 2.68e-03 | 2.76e-03 | 2.71e-03 | 2.66e-03 |
| | | | | $\varepsilon = 0.01$ | | | | |
| $N$ | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| CPU time | 76.0 | 119.7 | 167.5 | 224.0 | 300.8 | 384.4 | 499.2 | 606.5 |
| $\ell^1$ error | 4.81e-01 | 7.21e-02 | 4.51e-02 | 4.17e-02 | 4.24e-02 | 4.30e-02 | 4.34e-02 | 4.36e-02 |
| $N_v$ | 8 | 10 | 12 | 14 | 16 | 18 | 20 | 22 |
| CPU time | 237.2 | 335.6 | 516.4 | 636.5 | 875.2 | 1099.9 | 1389.7 | 1687.1 |
| $\ell^1$ error | 4.99e+00 | 8.91e-01 | 1.57e-02 | 1.19e-02 | 1.04e-02 | 1.12e-02 | 1.06e-02 | 1.11e-02 |

$N_v = 12$, $\varepsilon = 1$ and for $N = 4$, $N_v = 10$, $\varepsilon = 0.01$, we observe that for smaller $\varepsilon$ the moment method is more efficient.
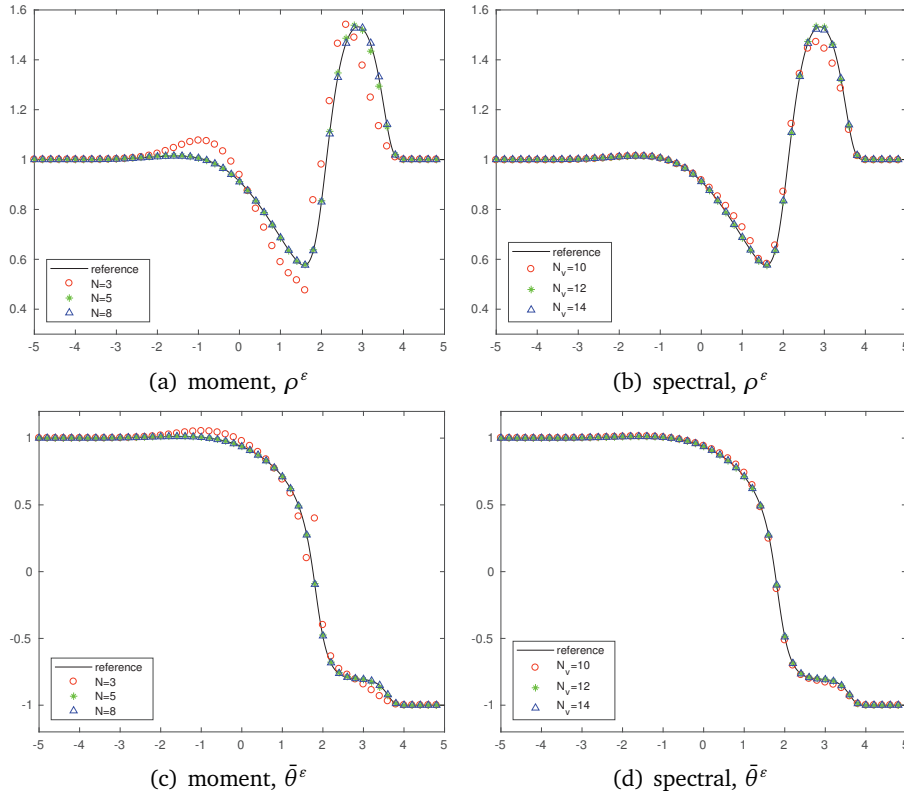
Figure 6: Example 5.3: The densities and macroscopic velocity angles for $\varepsilon = 1$, $t = 4$. Both the moment method and spectral method use 4000 cells in space. The moment number $N$ is $3, 5, 8$ and the grid number in velocity of the spectral method is $N_v = 10, 12, 14$. The solid line is the reference solution obtained by using the spectral method with $N_v = 32$ and 8000 cells.
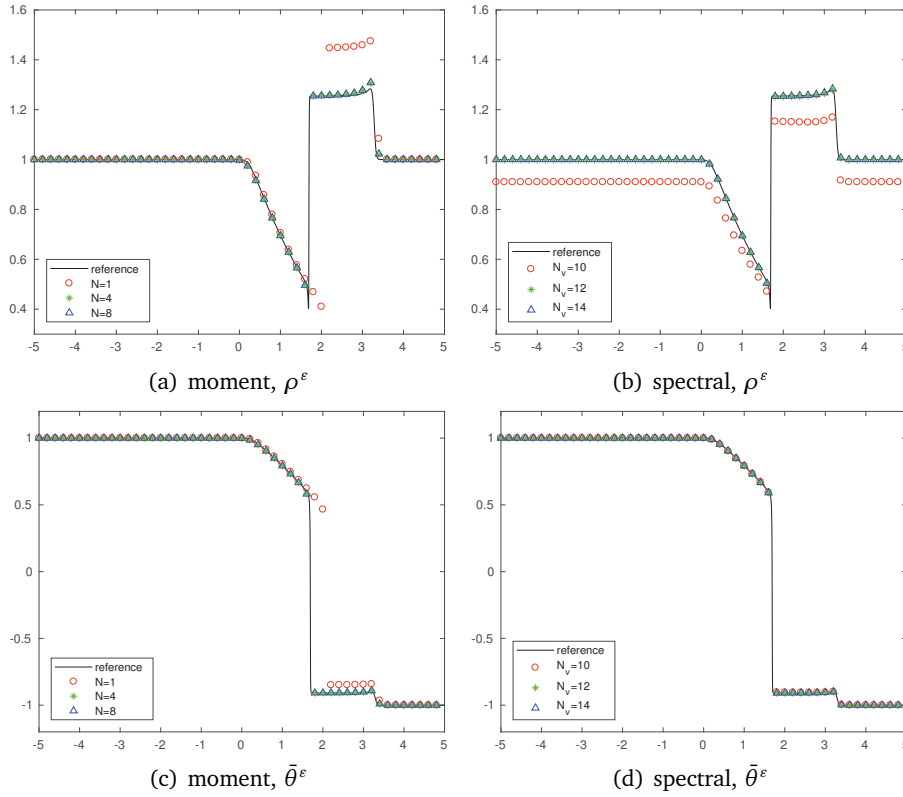
**Example 5.4** (Vortex formation). The computational domain is the square $[-5, 5] \times [-5, 5]$ with reflection boundary conditions, divided by the uniform square mesh $\{(x_i, y_j) | x_i = -5 + i\Delta x, y_j = -5 + j\Delta y, i, j = 0, 1, \cdots, n-1\}$. The initial data are

$$f(0, x_i, y_j, \theta) = \frac{1}{2\pi}, \quad \bar{\theta}(0, x_i, y_j) = \begin{cases} 0, & x < 4.5, \\ \dfrac{\pi}{2}, & \text{otherwise.} \end{cases}$$

After a transient period, the solution converges to a steady state vortex-type formation.

In numerical simulation, we add a perturbation to the initial velocity direction on the right boundary to ensure that the final steady state is a counterclockwise rotation. Fig. 13(b) shows the density $\rho$ at time $t$. The computations stop if the relative $\ell^2$-error for the density for two successive iterations — evaluated at $t = 200, 400, 800, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2800, 3000, 3200, 3400, 3600, 4000$ — is smaller than $1.5 \times 10^{-3}$.

Figs. 8-9 show the density and velocity for the moment methods with $\varepsilon = 1$, $n = 20$, where thirteen equally spaced contours are chosen from 0 to 2.4 with the step 0.2. Figs. 10-

(a) moment, $\rho^\varepsilon$     (b) spectral, $\rho^\varepsilon$

(c) moment, $\bar{\theta}^\varepsilon$     (d) spectral, $\bar{\theta}^\varepsilon$

Figure 7: Same as in Fig. 6 but for $\varepsilon = 0.01$ and $N = 1, 4, 8$.

11 show the corresponding results for $n = 50$. In Fig. 12 we also present the results obtained by the spectral method with $n = 50$. Fig. 13 shows the average density $M(l)$ over the square $\Omega(l) = \{(x, y) | \max\{|x|, |y|\} = l\}$ and the relative $\ell^2$-errors for the density, where

$$M(l) = \frac{1}{4l} \int_{\Omega(l)} \rho \, \mathrm{d}s,$$

N2n20 is the abbreviation for $N = 2, n = 20$ and specn50 for the spectral method with $n = 50$. The profiles of $M(l)$ are close to each other with an aberration at the boundary of the domain ($l \approx 5$), which got smaller as the number of moments and mesh cells increases. The numerical error plots also show that the increase of the time steps diminishes the error, but the increase of the mesh number $n$ or the moment number $N$ slows the convergence to the steady state solution.

## 6. Conclusion

We apply the model reduction method by operator projections to a non-linear kinetic description of the Vicsek swarming model, deriving the moment systems of arbitrary order.
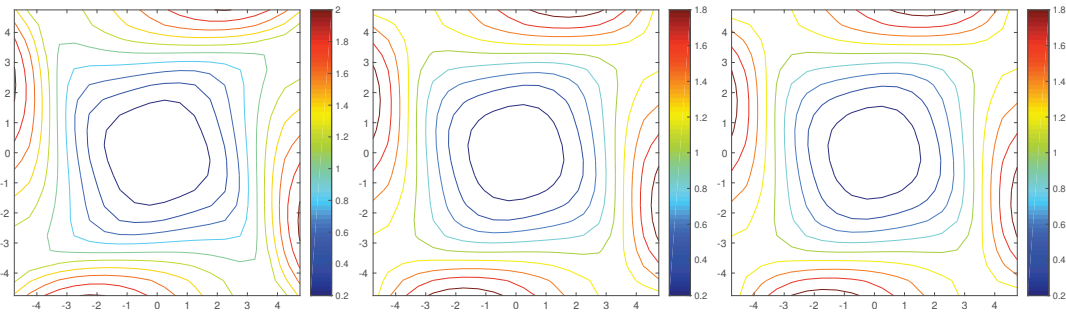
Figure 8: Example 5.4: The contours of density obtained by the moment method with $20 \times 20$ cells. From left to right: $N = 2, 3, 4$.
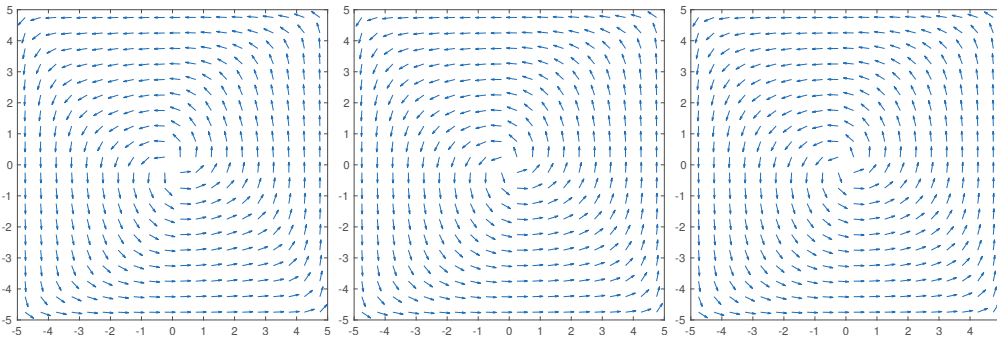


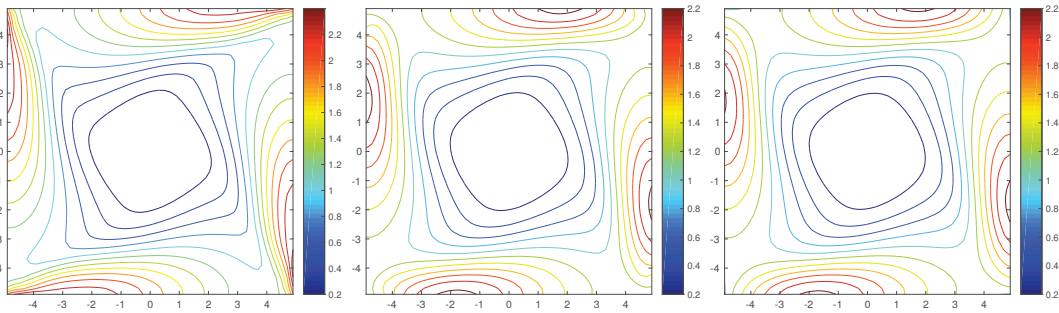Figure 9: Same as in Fig. 8 but for the velocity arrow diagrams.



Figure 10: Same as in Fig. 8 but for $n = 50$.

We study properties of these systems such as hyperbolicity, rotational invariance, mass-conservation and also the connections between Grad type expansions in different bases. In order to find solution of the Cauchy problem for the hyperbolic moment systems, we introduce a semi-implicit scheme and compare it with the existing spectral method for the kinetic equation. Numerical experiments show that the solutions of the hyperbolic moment systems can converge to the solutions of the kinetic equation for Vicsek swarming model when the order of the moment systems increases. Moreover, the moment method cap-
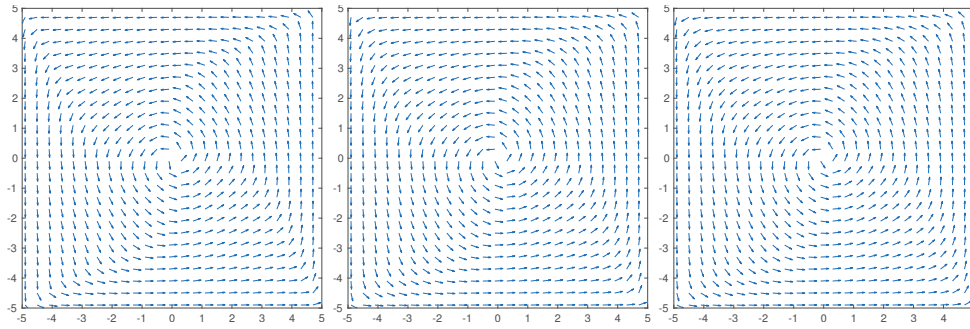
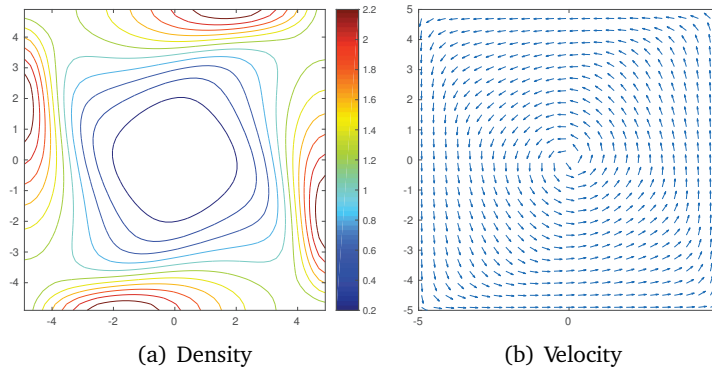Figure 11: Same as in Fig. 10 but for the velocity arrow diagrams.



(a) Density                                    (b) Velocity

Figure 12: Same as in Figs. 8-11 but for spectral method with $N_v = 32$, $n = 50$.



(a) Average density $M(l)$                    (b) Relative $\ell^2$ errors of $\rho$ with respect to $t$
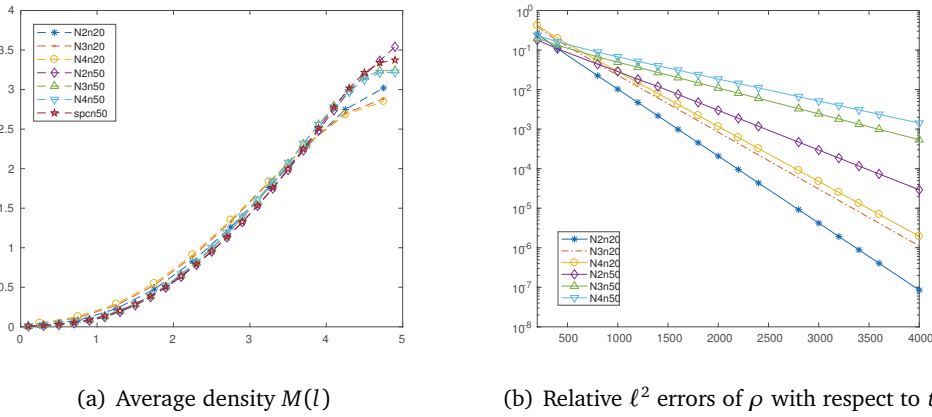
Figure 13: Example 5.4: The average density $M(l)$ in the square $\Omega(l)$ and the relative $\ell^2$ errors of density to time $t$.

tures key features of the model, including shock waves, contact discontinuities, rarefaction waves, and vortex formations. It is also noted that for small $\varepsilon$, the moment method is more efficient than the spectral method in the CPU time.

## Acknowledgments

This work was partially supported by the Special Project on High-performance Computing under the National Key R&D Program (No. 2016YFB0200603), Science Challenge Project (No. JCKY2016212A502), and the National Natural Science Foundation of China (Nos. 91330205, 91630310, & 11421101).

## References

[1] R. Bouffanais, *Design and Control of Swarm Dynamics*, Springer (2016).

[2] Z. Cai, Y. Fan and R. Li, *Globally hyperbolic regularization of Grad's moment system in one dimensional space*, Commun. Math. Sci. **11**, 547–571 (2013).

[3] Z. Cai, Y. Fan and R. Li, *Globally hyperbolic regularization of Grad's moment system*, Comm. Pure Appl. Math. **67**, 464–518 (2014).

[4] Z. Cai, Y. Fan and R. Li, *A framework on moment model reduction for kinetic equation*, SIAM J. Appl. Math. **75**, 2001–2023 (2014).

[5] Z. Cai and R. Li, *Numerical regularized moment method of arbitrary order for Boltzmann-BGK equation*, SIAM J. Sci. Comput. **32**, 2875–2907 (2010).

[6] Z. Cai, R. Li and Y. Wang, *Numerical regularized moment method for high Mach number flow*, Commun. Comput. Phys. **11**, 1415–1438 (2012).

[7] Z. Cai, Y. Fan and R. Li, *On hyperbolicity of 13-moment system*, Kinet. Relat. Mod. **7**, 415–432 (2014).

[8] J.A. Cañizo, J. Carrillo and J. Rosado, *A well-posedness theory in measures for some kinetic models of collective motion*, Math. Models Meth. Appl. Sci. **21**, 515–539 (2011).

[9] C. Cercignani, *The Boltzmann Equation and Its Applications*, Springer (1988).

[10] S. Chapman and T.G. Cowling, *The Mathematical Theory of Non-uniform Gases*, Cambridge University Press (1991).

[11] P. Degond, J.G. Liu, S. Motsch and V. Panferov, *Hydrodynamic models of self-organized dynamics: Derivation and existence theory*, Meth. and Appl. Anal. **20**, 89–114 (2013).

[12] P. Degond and S. Motsch, *Continuum limit of self-driven particles with orientation interaction*, Math. Models Meth. Appl. Sci. **18**, 1193–1215 (2008).

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

[14] I.M. Gamba, J.R. Haack and S. Motsch, *Spectral method for a kinetic swarming model*, J. Comput. Phys. **297**, 32–46 (2015).

[15] I.M. Gamba and M.J. Kang, *Global weak solutions for Kolmogorov-Vicsek type equations with orientational interactions*, Arch. Rational Mech. Anal. **222**, 1–26 (2015).

[16] H. Grad, *On the kinetic theory of rarefied gases*, Commun. Pure Appl. Math. **2**, 331–407 (1949).

[17] S.Y. Ha and E. Tadmor, *From particle to kinetic and hydrodynamic descriptions of flocking*, Kinet. Relat. Mod. **1**, 415–435 (2008).

[18] J. Koellermeier and M. Torrilhon, *Hyperbolic moment equations using quadrature based projection methods*, AIP Conf. Proc. **1628**, 626–633 (2014).

[19] J. Koellermeier and M. Torrilhon, *Numerical study of partially conservative moment equations in kinetic theory*, Commun. Comput. Phys. **21**, 981–1011 (2017).

[20] J. Koellermeier and M. Torrilhon, *Numerical solution of hyperbolic moment models for the Boltzmann equation*, Euro. J. Mech. - B/Fluids **64**, 41–46 (2017).

[21] J. Koellermeier, R. Schaerer and M. Torrilhon, *A framework for hyperbolic approximation of kinetic equations using quadrature-based projection methods*, Kinet. Relat. Mod. **7**, 531–549 (2014).

[22] Y.Y. Kuang and H.Z.Tang, *Globally hyperbolic moment model of arbitrary order for one-dimensional special relativistic Boltzmann equation*, J. Stat. Phys. **167**, 1303–1353 (2017).

[23] Y.Y. Kuang and H.Z.Tang, *Globally hyperbolic moment model of arbitrary order for three-dimensional special relativistic Boltzmann equation*, arXiv **1705.03990**, 2017.

[24] D. Levermore, *Moment closure hierarchies for kinetic theories*, J. Stat. Phys. **83**, 1021–1065 (1996).

[25] I. Müller and T. Ruggeri, *Rational Extended Thermodynamics*, Springer-Verlag (1998).

[26] S. Rhebergen, O. Bokhove and J. J. W. Van Der Vegt, *Discontinuous Galerkin finite element methods for hyperbolic nonconservative partial differential equations*, J. Comput. Phys. **227**, 1887–1922 (2008).