A fast Petrov-Galerkin spectral method for the multi-dimensional Boltzmann equation using mapped Chebyshev functions*

Jingwei Hu[†]   Xiaodong Huang[‡]   Jie Shen[§]   Haizhao Yang[¶]

May 17, 2021

Abstract

Numerical approximation of the Boltzmann equation presents a challenging problem due to its high-dimensional, nonlinear, and nonlocal collision operator. Among the deterministic methods, the Fourier-Galerkin spectral method stands out for its relative high accuracy and possibility of being accelerated by the fast Fourier transform. However, this method requires a domain truncation which is unphysical since the collision operator is defined in $\mathbb{R}^d$. In this paper, we introduce a Petrov-Galerkin spectral method for the Boltzmann equation in the unbounded domain. The basis functions (both test and trial functions) are carefully chosen mapped Chebyshev functions to obtain desired convergence and conservation properties. Furthermore, thanks to the close relationship of the Chebyshev functions and the Fourier cosine series, we are able to construct a fast algorithm with the help of the non-uniform fast Fourier transform (NUFFT). We demonstrate the superior accuracy of the proposed method in comparison to the Fourier spectral method through a series of 2D and 3D examples.

Keywords. Boltzmann equation; Petrov-Galerkin spectral method; mapped Chebyshev function; unbounded domain, NUFFT.

AMS subject classifications. 35Q20, 65M70.

1   Introduction

In multi-scale modeling, kinetic theory serves as a basic building block that bridges microscopic particle models and macroscopic continuum models. By tracking the probability density function, kinetic equations describe the non-equilibrium dynamics of the complex particle systems and have been widely used in disparate fields such as rarefied gas dynamics [7], plasma physics [5], nuclear reactor modeling [8], chemistry [16], biology, and socioeconomics [25].

In this paper, we consider the numerical approximation of the Boltzmann equation, one of the fundamental equations in kinetic theory [6, 32]. The complete equation include both particle transport and collisions which are often treated separately by operator splitting. Since the collision part is the

main difficulty when numerically solving the equation, we focus on the following spatially homogeneous Boltzmann equation:

$$\partial_t f = Q(f, f), \quad t > 0, \quad \boldsymbol{v} \in \mathbb{R}^d, \quad d = 2, 3, \tag{1.1}$$

where $f = f(t, \boldsymbol{v})$ is the probability density function of time t and velocity $\boldsymbol{v}$, and $Q(f, f)$ is the collision operator whose bilinear form is given by

$$Q(g, f)(\boldsymbol{v}) = \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) \left[g(\boldsymbol{v}'_*) f(\boldsymbol{v}') - g(\boldsymbol{v}_*) f(\boldsymbol{v}) \right] \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v}_*, \tag{1.2}$$

where the post-collisional velocities $(\boldsymbol{v}', \boldsymbol{v}'_*)$ are defined in terms of the pre-collisional velocities $(\boldsymbol{v}, \boldsymbol{v}_*)$ as

$$\begin{cases} \boldsymbol{v}' = \frac{1}{2}(\boldsymbol{v} + \boldsymbol{v}_*) + \frac{1}{2}|\boldsymbol{v} - \boldsymbol{v}_*|\boldsymbol{\sigma}, \\ \boldsymbol{v}'_* = \frac{1}{2}(\boldsymbol{v} + \boldsymbol{v}_*) - \frac{1}{2}|\boldsymbol{v} - \boldsymbol{v}_*|\boldsymbol{\sigma}, \end{cases} \tag{1.3}$$

with $\boldsymbol{\sigma}$ being a vector over the unit sphere S^{d-1}. The collision kernel $\mathcal{B}$ takes the form

$$\mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) = B(|\boldsymbol{v} - \boldsymbol{v}_*|, \cos\theta), \quad \cos\theta = \left\langle \frac{\boldsymbol{v} - \boldsymbol{v}_*}{|\boldsymbol{v} - \boldsymbol{v}_*|}, \boldsymbol{\sigma} \right\rangle, \tag{1.4}$$

i.e., it is a function depending only on the relative velocity $|\boldsymbol{v} - \boldsymbol{v}_*|$ and cosine of the scattering angle. The collision operator $Q(f, f)$ satisfies many important physical properties, including conservation of mass, momentum, and energy:

$$\int_{\mathbb{R}^d} Q(f, f) \, \mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^d} Q(f, f) \boldsymbol{v} \, \mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^d} Q(f, f)|\boldsymbol{v}|^2 \, \mathrm{d}\boldsymbol{v} = 0, \tag{1.5}$$

and the Boltzmann's H-theorem:

$$\int_{\mathbb{R}^d} Q(f, f) \log f \, \mathrm{d}\boldsymbol{v} \leq 0. \tag{1.6}$$

In the physically relevant case ($d = 3$), the collision operator is a five-fold quadratic integral whose numerical approximation can be extremely challenging. The stochastic methods, such as the direct simulation Monte Carlo (DSMC) methods proposed by Nanbu [26] and Bird [3], have been historically popular due their simplicity and efficiency. However, like any Monte Carlo methods, they suffer from slow convergence and high statistical noise especially for low-speed and unsteady flows. In the past two decades, the deterministic methods have undergone extensive development largely due to the advance in computing powers, see [9] for a recent review.

Among the deterministic methods for the Boltzmann equation, the Fourier-Galerkin spectral method stands out for its relative high accuracy and possibility of being accelerated by the fast Fourier transform (see, for instance, [27, 23, 13, 12] for major algorithmic development and [28, 10, 1, 17] for stability and convergence analysis). Although being a method with reasonable efficiency and accuracy tradeoff, the Fourier spectral method requires a domain truncation which is unphysical since the original collision operator is defined in $\mathbb{R}^d$. This truncation changes the structure of the equation and often comes with an accuracy loss. Inspired by the recent work [18] of the two authors here where a spectral method was introduced for the 1D inelastic Boltzmann equation, we develop in this paper a Petrov-Galerkin spectral method for the Boltzmann equation (1.1)[1] using mapped Chebyshev functions in $\mathbb{R}^d$. Both the test functions and trial functions are carefully chosen to obtain desired approximation properties. Furthermore, thanks to the close relationship of the Chebyshev functions and the Fourier cosine series, we are able

---

[1]Unlike the inelastic Boltzmann equation which has a non-trivial solution in 1D, the classical Boltzmann equation (1.1) must be considered at least for $d \geq 2$.

to construct a fast algorithm with the help of the non-uniform fast Fourier transform (NUFFT). This speedup is critical as the direct implementation of the proposed method would require excessive storage for precomputation and significant online computational cost that soon become a bottleneck for larger N (the number of spectral modes). Extensive numerical tests in 2D and 3D are performed to demonstrate the accuracy and efficiency of the proposed method. In particular, the comparison with the Fourier spectral method in [12] indeed confirms the better approximation properties of the proposed method. Finally, we mention some recent spectral methods [11, 15, 21, 19, 20] for the Boltzmann equation that use other orthogonal polynomial bases in $\mathbb{R}^d$. Up to our knowledge, our method is the first with a fast implementation and a consistency analysis.

The rest of this paper is organized as follows. In Section 2, we introduce the mapped Chebyshev functions in $\mathbb{R}^d$ along with their approximation properties. In Section 3, we construct the Petrov-Galerkin spectral method for the Boltzmann equation using the mapped Chebyshev functions as trial and test functions. The approximation properties for the collision operator and moments are proved as well. The numerical realization including the fast algorithm is described in detail in Section 4. In Section 5, several numerical tests in 2D and 3D are performed to demonstrate the accuracy and efficiency of the proposed method. The paper is concluded in Section 6.

2 Multi-dimensional mapped Chebyshev functions

In this section, we introduce the mapped Chebyshev functions in $\mathbb{R}^d$ and discuss their approximation properties. These functions are extension of the one-dimensional mapped Chebyshev functions introduced in [18] based on tensor product formulation [29, 31]. Later in Section 3, they will serve as the trial functions and test functions in the Petrov-Galerkin spectral method for the Boltzmann equation.

2.1 Mapped Chebyshev functions in $\mathbb{R}^d$

To define the mapped Chebyshev functions in $\mathbb{R}^d$, we start with the one-dimensional Chebyshev polynomials on the interval $I = (-1, 1)$:

$$T_0(\xi) = 1, \quad T_1(\xi) = \xi, \quad T_{k+1}(\xi) = 2\xi T_k(\xi) - T_{k-1}(\xi), \quad k \geq 1. \tag{2.1}$$

Define the inner product $(\cdot, \cdot)_\omega$ as

$$(F, G)_\omega := \int_I F(\xi)G(\xi)\omega(\xi)\,\mathrm{d}\xi, \quad \omega(\xi) = (1 - \xi^2)^{-\frac{1}{2}}, \tag{2.2}$$

then $\{T_k(\xi)\}_{k \geq 0}$ satisfy the orthogonality condition

$$(T_k, T_l)_\omega = c_k \delta_{k,l}, \quad \forall \, k, l \geq 0, \tag{2.3}$$

where $c_0 = \pi$ and $c_k = \pi/2$ for $k \geq 1$.

We then introduce a one-to-one mapping $\xi \to v(\xi)$ (its inverse is denoted as $v \to \xi(v)$) from I to $\mathbb{R}$ such that

$$\frac{\mathrm{d}v}{\mathrm{d}\xi} = \frac{S}{(1 - \xi^2)^{1 + \frac{r}{2}}} := \frac{\omega(\xi)}{[\mu(\xi)]^2}, \quad v(\pm 1) = \pm\infty, \tag{2.4}$$

where $S > 0$ is a scaling parameter, $r \geq 0$ is the tail parameter, and the function μ is given by

$$\mu(\xi) = \frac{(1 - \xi^2)^{\frac{1+r}{4}}}{\sqrt{S}}. \tag{2.5}$$

3

With this mapping we define two sets of mapped Chebyshev functions in $\mathbb{R}$ as

$$\widetilde{T}_k(v) := \frac{[\mu(\xi(v))]^4}{\sqrt{c_k}} T_k(\xi(v)), \quad \widehat{T}_k(v) := \frac{[\mu(\xi(v))]^{-2}}{\sqrt{c_k}} T_k(\xi(v)). \tag{2.6}$$

Define the inner product $(\cdot, \cdot)_{\mathbb{R}}$ as

$$(f, g)_{\mathbb{R}} := \int_{\mathbb{R}} f(v) g(v) \, \mathrm{d}v, \tag{2.7}$$

then it is easy to check that $\{\widetilde{T}_k(v)\}_{k \geq 0}$ and $\{\widehat{T}_k(v)\}_{k \geq 0}$ satisfy the orthonormal condition:

$$(\widetilde{T}_k, \widehat{T}_l)_{\mathbb{R}} = \delta_{k,l}, \quad \forall \, k, l \geq 0. \tag{2.8}$$

Remark 2.1. Two one-to-one mappings between I and $\mathbb{R}$ often used in practice are

- **logarithmic mapping** $(r = 0)$:

$$v = \frac{S}{2} \ln\left(\frac{1 + \xi}{1 - \xi}\right), \quad \xi = \tanh\left(\frac{v}{S}\right), \quad \mu(\xi) = \frac{1}{\sqrt{S}} (1 - \xi^2)^{\frac{1}{4}}, \tag{2.9}$$

- **algebraic mapping** $(r = 1)$:

$$v = \frac{S\xi}{\sqrt{1 - \xi^2}}, \quad \xi = \frac{v}{\sqrt{S^2 + v^2}}, \quad \mu(\xi) = \frac{1}{\sqrt{S}} (1 - \xi^2)^{\frac{1}{2}}. \tag{2.10}$$

In the multi-dimensional case, we denote the multi-vector as $\boldsymbol{v} = (v_1, \ldots, v_d)$ and multi-index as $\boldsymbol{k} = (k_1, \ldots, k_d)$, where k_j is a non-negative integer for each $j = 1, \ldots, d$; $0 \leq \boldsymbol{k} \leq N$ means $0 \leq k_j \leq N$ for each $j = 1, \ldots, d$. We define the mapped Chebyshev functions in $\mathbb{R}^d$ using (2.6) via the tensor product as

$$\widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) := \prod_{j=1}^{d} \widetilde{T}_{k_j}(v_j), \quad \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) := \prod_{j=1}^{d} \widehat{T}_{k_j}(v_j). \tag{2.11}$$

The inner products $(\cdot, \cdot)_{\boldsymbol{\omega}}$ in $I^d = (-1, 1)^d$ and $(\cdot, \cdot)_{\mathbb{R}^d}$ in $\mathbb{R}^d$ are defined, respectively, by

$$(F, G)_{\boldsymbol{\omega}} := \int_{I^d} F(\boldsymbol{\xi}) G(\boldsymbol{\xi}) \boldsymbol{\omega}(\boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\xi}, \quad (f, g)_{\mathbb{R}^d} := \int_{\mathbb{R}^d} f(\boldsymbol{v}) g(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}, \tag{2.12}$$

with the weight function $\boldsymbol{\omega}(\boldsymbol{\xi}) := \prod_{j=1}^{d} \omega(\xi_j)$. Then we have

$$\left(\widetilde{\boldsymbol{T}}_{\boldsymbol{k}}, \widehat{\boldsymbol{T}}_{\boldsymbol{l}}\right)_{\mathbb{R}^d} = \prod_{j=1}^{d} \left(\widetilde{T}_{k_j}, \widehat{T}_{l_j}\right)_{\mathbb{R}} = \prod_{j=1}^{d} \delta_{k_j, l_j} =: \boldsymbol{\delta}_{\boldsymbol{k}, \boldsymbol{l}}. \tag{2.13}$$

Therefore, any d-variate function $f(\boldsymbol{v})$ can be expanded using $\{\widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})\}_{\boldsymbol{k} \geq 0}$ as

$$f(\boldsymbol{v}) = \sum_{\boldsymbol{k} \geq 0} \widetilde{f}_{\boldsymbol{k}} \widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) = \sum_{\boldsymbol{k} \geq 0} \widetilde{f}_{\boldsymbol{k}} \frac{[\boldsymbol{\mu}(\boldsymbol{\xi})]^4}{\sqrt{c_{\boldsymbol{k}}}} \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}), \tag{2.14}$$

and the expansion coefficients $\{\widetilde{f}_{\boldsymbol{k}}\}_{\boldsymbol{k} \geq 0}$ are determined by

$$\widetilde{f}_{\boldsymbol{k}} = \left(f, \widehat{\boldsymbol{T}}_{\boldsymbol{k}}\right)_{\mathbb{R}^d} = \frac{1}{\sqrt{c_{\boldsymbol{k}}}} \left([\boldsymbol{\mu}(\boldsymbol{\xi})]^{-4} f(\boldsymbol{v}(\boldsymbol{\xi})), \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi})\right)_{\boldsymbol{\omega}}, \tag{2.15}$$

where

$$\boldsymbol{\mu}(\boldsymbol{\xi}) := \prod_{j=1}^{d} \mu(\xi_j), \quad c_{\boldsymbol{k}} := \prod_{j=1}^{d} c_{k_j}, \quad \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}) := \prod_{j=1}^{d} T_{k_j}(\xi_j), \tag{2.16}$$

4

and $\boldsymbol{v}(\boldsymbol{\xi})$ is the mapping from I^d to $\mathbb{R}^d$ such that each component ξ_j is mapped to v_j via the 1D mapping (2.4). The inverse mapping $\boldsymbol{\xi}(\boldsymbol{v})$ is understood similarly.

In Section 3, we will introduce the Petrov-Galerkin spectral method for the Boltzmann equation in $\mathbb{R}^d$, where the trial function space and test function space are chosen, respectively, as

$$\widetilde{\mathbb{T}}_N^d := \{\widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})\}_{0 \leq \boldsymbol{k} \leq N}, \quad \widehat{\mathbb{T}}_N^d := \{\widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})\}_{0 \leq \boldsymbol{k} \leq N}. \tag{2.17}$$

The choice of these functions is motivated by their decay/growth properties at large $|\boldsymbol{v}|$. The following result is a straightforward extension of the 1D result in [18].

Lemma 2.2. *For any $\boldsymbol{k} \geq 0$ and $|\boldsymbol{v}| \gg 1$, we have*

$$\left|\widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})\right| \sim \begin{cases} e^{-\frac{2}{S}(\sum_{j=1}^d |v_j|)}, & r = 0, \\ \prod_{j=1}^d |v_j|^{-4}, & r = 1; \end{cases} \quad \left|\widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})\right| \sim \begin{cases} e^{\frac{1}{S}(\sum_{j=1}^d |v_j|)}, & r = 0, \\ \prod_{j=1}^d |v_j|^2, & r = 1, \end{cases} \tag{2.18}$$

where $r = 0$ corresponds to the logarithmic mapping (2.9) and $r = 1$ to the algebraic mapping (2.10).

2.2 Approximation properties

We describe below some approximation properties of the mapped Chebyshev functions in $\mathbb{R}^d$.

For a function $f(\boldsymbol{v})$ defined in $\mathbb{R}^d$, the transform $\boldsymbol{v}(\boldsymbol{\xi})$ maps it to a function in I^d. Hence, we introduce the linked function pair (f, F) such that $f(\boldsymbol{v}) = f(\boldsymbol{v}(\boldsymbol{\xi})) \equiv F(\boldsymbol{\xi})$. In addition, we introduce another function pair $(\widehat{f}^\alpha, \widehat{F}^\alpha)$ as

$$\widehat{f}^\alpha(\boldsymbol{v}) := f(\boldsymbol{v})[\boldsymbol{\mu}(\boldsymbol{\xi}(\boldsymbol{v}))]^{-\alpha} = F(\boldsymbol{\xi})[\boldsymbol{\mu}(\boldsymbol{\xi})]^{-\alpha} =: \widehat{F}^\alpha(\boldsymbol{\xi}). \tag{2.19}$$

We define the approximation space in $\mathbb{R}^d$ with a parameter α as

$$\mathbb{V}_N^{\alpha,d}(\mathbb{R}^d) := \operatorname{span}\left\{\boldsymbol{T}_{\boldsymbol{k}}^\alpha(\boldsymbol{v}) := [\boldsymbol{\mu}(\boldsymbol{\xi}(\boldsymbol{v}))]^\alpha \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}(\boldsymbol{v})), \ 0 \leq \boldsymbol{k} \leq N\right\}. \tag{2.20}$$

Therefore, the trial function space $\widetilde{\mathbb{T}}_N^d$ and test function space $\widehat{\mathbb{T}}_N^d$ introduced in the previous section correspond to $\mathbb{V}_N^{4,d}$ and $\mathbb{V}_N^{-2,d}$, respectively.

In the following, the L^2 space with a given weight $\boldsymbol{w}$ is equipped with norm

$$\|f\|_{L_{\boldsymbol{w}}^2(I^d)} = \left(\int_{I^d} |f(\boldsymbol{\xi})|^2 \boldsymbol{w}(\boldsymbol{\xi}) \, d\boldsymbol{\xi}\right)^{1/2} \quad \text{or} \quad \|f\|_{L_{\boldsymbol{w}}^2(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \boldsymbol{w}(\boldsymbol{v}) \, d\boldsymbol{v}\right)^{1/2}, \tag{2.21}$$

depending on the domain of interest.

Let $\mathbb{P}_N^d(I^d)$ denote the set of d-variate polynomials in I^d with degree $\leq N$ in each direction, and $\Pi_N^d : L_{\boldsymbol{\omega}}^2(I^d) \to \mathbb{P}_N^d(I^d)$ be the Chebyshev orthogonal projection operator such that

$$\left(\Pi_N^d F - F, \phi\right)_{\boldsymbol{\omega}} = 0, \quad \forall \phi \in \mathbb{P}_N^d(I^d). \tag{2.22}$$

Then we define another projection operator $\pi_N^{\alpha,d} : L_{\boldsymbol{\mu}^{2-2\alpha}}^2(\mathbb{R}^d) \to \mathbb{V}_N^{\alpha,d}(\mathbb{R}^d)$ by

$$\pi_N^{\alpha,d} f := \boldsymbol{\mu}^\alpha \Pi_N^d (F\boldsymbol{\mu}^{-\alpha}) = \boldsymbol{\mu}^\alpha \Pi_N^d \widehat{F}^\alpha. \tag{2.23}$$

One can verify using the definition that

$$\begin{aligned}
\left(\pi_N^{\alpha,d} f - f, \boldsymbol{\mu}^{2-2\alpha} \boldsymbol{T}_{\boldsymbol{k}}^\alpha\right)_{\mathbb{R}^d} &= \int_{\mathbb{R}^d} (\pi_N^{\alpha,d} f - f)\boldsymbol{\mu}^{2-\alpha} \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}(\boldsymbol{v})) \, d\boldsymbol{v} \\
&= \int_{I^d} \left[\boldsymbol{\mu}^\alpha \Pi_N^d \widehat{F}^\alpha - \boldsymbol{\mu}^\alpha \widehat{F}^\alpha\right] \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}) \boldsymbol{\mu}^{2-\alpha} \frac{\boldsymbol{\omega}(\boldsymbol{\xi})}{\boldsymbol{\mu}^2} \, d\boldsymbol{\xi} \\
&= \left(\Pi_N^d \widehat{F}^\alpha - \widehat{F}^\alpha, \boldsymbol{T}_{\boldsymbol{k}}\right)_{\boldsymbol{\omega}} = 0, \quad \forall \, 0 \leq \boldsymbol{k} \leq N. \tag{2.24}
\end{aligned}$$

5

Next, we introduce the function space $\boldsymbol{B}_\alpha^m(\mathbb{R}^d)$ equipped with the norm

$$\|f\|_{\boldsymbol{B}_\alpha^m(\mathbb{R}^d)} = \left(\sum_{0 \leq \boldsymbol{k} \leq m} \left\| \boldsymbol{D}_{\alpha,\boldsymbol{v}}^{\boldsymbol{k}} f \right\|_{L^2_{\varpi^{\boldsymbol{k}+\frac{1+r}{2}\mathbf{1}}}(\mathbb{R}^d)}^2 \right)^{1/2}, \tag{2.25}$$

and semi-norm

$$|f|_{\boldsymbol{B}_\alpha^m(\mathbb{R}^d)} = \left(\sum_{j=1}^d \left\| D_{\alpha,v_j}^m f \right\|_{L^2_{\varpi^{m\boldsymbol{e}_j+\frac{1+r}{2}\mathbf{1}}}(\mathbb{R}^d)}^2 \right)^{1/2}, \tag{2.26}$$

where $\mathbf{1}$ is an all-one vector, $\boldsymbol{e}_j = (0,\ldots,1,\ldots,0)$ with 1 in the j-th position and 0 elsewhere, and

$$\boldsymbol{D}_{\alpha,\boldsymbol{v}}^{\boldsymbol{k}} f := D_{\alpha,v_1}^{k_1} \cdots D_{\alpha,v_d}^{k_d} f, \quad \boldsymbol{\varpi}^{\boldsymbol{k}} := \prod_{j=1}^d (1 - \xi(v_j)^2)^{k_j}, \tag{2.27}$$

with

$$D_{\alpha,v_j}^{k_j} f := \underbrace{a(v_j)\frac{\partial}{\partial v_j}\left(a(v_j)\frac{\partial}{\partial v_j}\left(\ldots \left(a(v_j)\frac{\partial \widehat{f}^\alpha}{\partial v_j}\right)\ldots\right)\right)}_{k_j \text{ times derivatives}} = \frac{\partial^{k_j} \widehat{F}^\alpha}{\partial \xi_j}, \tag{2.28}$$

where $a(v_j) := \frac{\mathrm{d}v_j}{\mathrm{d}\xi_j}$ is determined by the mapping.

We have the following approximation result.

Theorem 2.3. *Let $\alpha \in \mathbb{R}$, $r \geq 0$. If $f \in \boldsymbol{B}_\alpha^m(\mathbb{R}^d)$, we have*

$$\left\| \pi_N^{\alpha,d} f - f \right\|_{L^2_{\boldsymbol{\mu}^{2-2\alpha}}(\mathbb{R}^d)} \leq C N^{-m} |f|_{\boldsymbol{B}_\alpha^m(\mathbb{R}^d)}. \tag{2.29}$$

Proof. Note that

$$\begin{aligned}
\left\| \pi_N^{\alpha,d} f - f \right\|_{L^2_{\boldsymbol{\mu}^{2-2\alpha}}(\mathbb{R}^d)}^2 &= \int_{\mathbb{R}^d} (\pi_N^{\alpha,d} f - f)^2 \boldsymbol{\mu}^{2-2\alpha} \, \mathrm{d}\boldsymbol{v} \\
&= \int_{I^d} \left[\boldsymbol{\mu}^\alpha \Pi_N^d \widehat{F}^\alpha - \boldsymbol{\mu}^\alpha \widehat{F}^\alpha \right]^2 \boldsymbol{\mu}^{2-2\alpha} \frac{\boldsymbol{\omega}(\boldsymbol{\xi})}{\boldsymbol{\mu}^2} \, \mathrm{d}\boldsymbol{\xi} \\
&= \left\| \Pi_N^d \widehat{F}^\alpha - \widehat{F}^\alpha \right\|_{L^2_{\boldsymbol{\omega}}(I^d)}^2.
\end{aligned}$$

By the multi-variate (full tensor product) Chebyshev approximation result (Theorem 2.1 in [30]), we know

$$\left\| \Pi_N^d \widehat{F}^\alpha - \widehat{F}^\alpha \right\|_{L^2_{\boldsymbol{\omega}}(I^d)} \leq C N^{-m} \left(\sum_{j=1}^d \left\| \partial_{\xi_j}^m \widehat{F}^\alpha \right\|_{L^2_{\varpi^{m\boldsymbol{e}_j-\frac{1}{2}\mathbf{1}}}(I^d)}^2 \right)^{1/2}.$$

Hence,

$$\begin{aligned}
\left\| \pi_N^{\alpha,d} f - f \right\|_{L^2_{\boldsymbol{\mu}^{2-2\alpha}}(\mathbb{R}^d)} &= \left\| \Pi_N^d \widehat{F}^\alpha - \widehat{F}^\alpha \right\|_{L^2_{\boldsymbol{\omega}}(I^d)} \\
&\leq C N^{-m} \left(\sum_{j=1}^d \left\| \partial_{\xi_j}^m \widehat{F}^\alpha \right\|_{L^2_{\varpi^{m\boldsymbol{e}_j-\frac{1}{2}\mathbf{1}}}(I^d)}^2 \right)^{1/2} \\
&\leq C N^{-m} \left(\sum_{j=1}^d \left\| D_{\alpha,v_j}^m f \right\|_{L^2_{\varpi^{m\boldsymbol{e}_j+\frac{1+r}{2}\mathbf{1}}}(\mathbb{R}^d)}^2 \right)^{1/2} \\
&= C N^{-m} |f|_{\boldsymbol{B}_\alpha^m(\mathbb{R}^d)}.
\end{aligned}$$

$\square$

3 A Petrov-Galerkin spectral method for the Boltzmann equation

We consider the initial value problem

$$\begin{cases} \partial_t f(t, \boldsymbol{v}) = Q(f, f), & t > 0, \quad \boldsymbol{v} \in \mathbb{R}^d, \\ f(0, \boldsymbol{v}) = f^0(\boldsymbol{v}), \end{cases} \tag{3.1}$$

where $Q(f, f)$, in a strong form, is given by (1.2). To construct the Petrov-Galerkin spectral method, the following weak form of the collision operator is more convenient:

$$(Q(f, f), \phi)_{\mathbb{R}^d} = \int_{\mathbb{R}^d} Q(f, f)(\boldsymbol{v}) \phi(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) f(\boldsymbol{v}) f(\boldsymbol{v}_*)[\phi(\boldsymbol{v}') - \phi(\boldsymbol{v})] \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v} \, \mathrm{d}\boldsymbol{v}_*, \tag{3.2}$$

where $\phi(\boldsymbol{v})$ is a test function.

We look for an approximation of f in the trial function space $\widetilde{\mathbb{T}}_N^d$ as

$$f(t, \boldsymbol{v}) \approx f_N(t, \boldsymbol{v}) = \sum_{0 \leq \boldsymbol{k} \leq N} \widetilde{f}_{\boldsymbol{k}}(t) \widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) \in \widetilde{\mathbb{T}}_N^d. \tag{3.3}$$

Substituting f_N into (3.1) and requiring the residue of the equation to be orthogonal to the test function space $\widehat{\mathbb{T}}_N^d$, we obtain

$$\left(\partial_t f_N - Q(f_N, f_N), \widehat{\boldsymbol{T}}_{\boldsymbol{k}} \right)_{\mathbb{R}^d} = 0 \quad \text{for all } \widehat{\boldsymbol{T}}_{\boldsymbol{k}} \in \widehat{\mathbb{T}}_N^d. \tag{3.4}$$

By the orthogonality condition (2.13), we find that the coefficients $\{\widetilde{f}_{\boldsymbol{k}}(t)\}$ satisfy the following ODE system

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t} \widetilde{f}_{\boldsymbol{k}}(t) = \mathcal{Q}_{\boldsymbol{k}}^N, \\ \widetilde{f}_{\boldsymbol{k}}(0) = \widetilde{f}_{\boldsymbol{k}}^0, \end{cases} \quad 0 \leq \boldsymbol{k} \leq N, \tag{3.5}$$

where

$$\begin{aligned} \mathcal{Q}_{\boldsymbol{k}}^N &:= \left(Q(f_N, f_N), \widehat{\boldsymbol{T}}_{\boldsymbol{k}} \right)_{\mathbb{R}^d} \\ &= \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) f_N(\boldsymbol{v}) f_N(\boldsymbol{v}_*) \left[\widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}') - \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) \right] \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v} \, \mathrm{d}\boldsymbol{v}_*, \end{aligned} \tag{3.6}$$

and

$$\widetilde{f}_{\boldsymbol{k}}^0 := \left(f^0, \widehat{\boldsymbol{T}}_{\boldsymbol{k}} \right)_{\mathbb{R}^d} = \frac{1}{\sqrt{c_{\boldsymbol{k}}}} \left([\boldsymbol{\mu}(\boldsymbol{\xi})]^{-4} f^0(\boldsymbol{v}(\boldsymbol{\xi})), \boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}) \right)_{\boldsymbol{\omega}}. \tag{3.7}$$

Note that we used the weak form (3.2) in (3.6).

Remark 3.1. An equivalent way of writing the ODE system (3.5) is

$$\begin{cases} \partial_t f_N(t, \boldsymbol{v}) = \pi_N^{4;d} Q(f_N, f_N), \\ f_N(0, \boldsymbol{v}) = \pi_N^{4,d} f^0(\boldsymbol{v}), \end{cases} \tag{3.8}$$

where $\pi_N^{4,d}$ is the projection operator defined in (2.23) (with $\alpha = 4$). Indeed, for any $f \in L_{\boldsymbol{\mu}^{-6}}^2(\mathbb{R}^d)$,

$$\pi_N^{4,d} f = \sum_{0 \leq \boldsymbol{k} \leq N} \left(f, \widehat{\boldsymbol{T}}_{\boldsymbol{k}} \right)_{\mathbb{R}^d} \widetilde{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) \in \mathbb{V}_N^{4,d}(\mathbb{R}^d) = \widetilde{\mathbb{T}}_N^d. \tag{3.9}$$

3.1 Approximation property for the collision operator

In this subsection, we establish a consistency result of the spectral approximation for the collision operator. We will show that if f and $Q(f, f)$ have certain regularity, the proposed approximation of the collision operator $\pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f)$ enjoys spectral accuracy. We will only prove this result under the algebraic mapping (2.10) with $S = 1$, that is in 1D,

$$v = \frac{\xi}{\sqrt{1 - \xi^2}}, \quad \xi = \frac{v}{\sqrt{1 + v^2}}, \quad \mu = \sqrt{1 - \xi^2} = \frac{1}{\sqrt{1 + v^2}}. \tag{3.10}$$

The reason of this choice is strongly motivated by the existing regularity result of the Boltzmann collision operator under an exponentially weighted Lebesgue norm:

$$\|f\|_{\mathcal{L}_k^p(\mathbb{R}^d)} = \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^p (1 + |\boldsymbol{v}|^2)^{kp/2} \, d\boldsymbol{v} \right)^{1/p}, \quad k \in \mathbb{R}, \ 1 \leq p < \infty. \tag{3.11}$$

Specifically, we write the collision operator (1.2) as $Q(g, f) = Q^+(g, f) - Q^-(g, f)$, where the gain part and loss part are given by

$$
\begin{aligned}
Q^+(g, f)(\boldsymbol{v}) &= \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) g(\boldsymbol{v}_*') f(\boldsymbol{v}') \, d\boldsymbol{\sigma} \, d\boldsymbol{v}_*, \\
Q^-(g, f)(\boldsymbol{v}) &= \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) g(\boldsymbol{v}_*) f(\boldsymbol{v}) \, d\boldsymbol{\sigma} \, d\boldsymbol{v}_*.
\end{aligned}
\tag{3.12}
$$

Then we have the following regularity result for the gain operator $Q^+(g, f)$ established in [24].

Theorem 3.2 (Theorem 2.1, [24]). *Let $k, \eta \in \mathbb{R}$, $1 \leq p < \infty$, and let the collision kernel $\mathcal{B}$ satisfy certain cut-off assumption[2]. Then the following estimate holds*

$$\|Q^+(g, f)\|_{\mathcal{L}_\eta^p(\mathbb{R}^d)} \leq C_{k,\eta,p}(\mathcal{B}) \|g\|_{\mathcal{L}_{|k+\eta|+|\eta|}^1(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{k+\eta}^p(\mathbb{R}^d)}, \tag{3.13}$$

where $C_{k,\eta,p}(\mathcal{B})$ is a constant that depends only on the kernel $\mathcal{B}$ and k, η and p.

To obtain a similar estimate for the loss operator $Q^-(g, f)$, we restrict ourselves to the variable hard sphere (VHS) collision model [4]. Note that this kernel falls into the assumption in Theorem 3.2. We have the following result.

Proposition 3.2. *Let $\eta \in \mathbb{R}$, $1 \leq p < \infty$, and let the collision kernel takes the form $\mathcal{B} = C_\lambda |\boldsymbol{v} - \boldsymbol{v}_*|^\lambda$, where $0 \leq \lambda \leq 1$ and C_λ is a positive constant. Then the following estimate holds*

$$\|Q^-(g, f)\|_{\mathcal{L}_\eta^p(\mathbb{R}^d)} \leq C_\lambda \|g\|_{\mathcal{L}_\lambda^1(\mathbb{R}^d)} \|f\|_{L_{\lambda+\eta}^p(\mathbb{R}^d)}. \tag{3.14}$$

Proof. Note that

$$|\boldsymbol{v} - \boldsymbol{v}_*| \leq |\boldsymbol{v}| + |\boldsymbol{v}_*| = (|\boldsymbol{v}|^2 + |\boldsymbol{v}_*|^2 + 2|\boldsymbol{v}||\boldsymbol{v}_*|)^{1/2} \leq \left(1 + |\boldsymbol{v}|^2\right)^{1/2} \left(1 + |\boldsymbol{v}_*|^2\right)^{1/2}.$$

Then

$$
\begin{aligned}
Q^-(g, f)(\boldsymbol{v}) &= C_\lambda f(\boldsymbol{v}) \int_{\mathbb{R}^d} g(\boldsymbol{v}_*) |\boldsymbol{v} - \boldsymbol{v}_*|^\lambda \, d\boldsymbol{v}_* \\
&\leq C_\lambda f(\boldsymbol{v}) \left(1 + |\boldsymbol{v}|^2\right)^{\lambda/2} \left[\int_{\mathbb{R}^d} g(\boldsymbol{v}_*) \left(1 + |\boldsymbol{v}_*|^2\right)^{\lambda/2} \, d\boldsymbol{v}_* \right] \\
&= C_\lambda f(\boldsymbol{v}) \left(1 + |\boldsymbol{v}|^2\right)^{\lambda/2} \|g\|_{L_\lambda^1(\mathbb{R}^d)}.
\end{aligned}
$$

---

[2]To avoid technicality, we do not spell out the condition here and only mention that most of the collision kernels used in numerical simulations satisfy this assumption.

Therefore, for any $\eta \in \mathbb{R}$, $1 \le p < \infty$,

$$\|Q^-(g,f)\|_{\mathcal{L}^p_\eta(\mathbb{R}^d)} \le C_\lambda \|g\|_{\mathcal{L}^1_\lambda(\mathbb{R}^d)} \|f\|_{L^p_{\lambda+\eta}(\mathbb{R}^d)}.$$

$\square$

Combining the previous two results, we can obtain the following theorem.

Theorem 3.3. *Let the collision kernel takes the form $\mathcal{B} = C_\lambda |\boldsymbol{v} - \boldsymbol{v}_*|^\lambda$, where $0 \le \lambda \le 1$ and C_λ is a positive constant. Then the collision operator $Q(g,f)$ satisfies*

$$\|Q(g,f)\|_{\mathcal{L}^2_{3d}(\mathbb{R}^d)} \le C_d(\mathcal{B}) \|g\|_{\mathcal{L}^1_{\lambda+6d}(\mathbb{R}^d)} \|f\|_{\mathcal{L}^2_{\lambda+3d}(\mathbb{R}^d)}, \tag{3.15}$$

where $C_d(\mathcal{B})$ is a constant that depends only on the kernel $\mathcal{B}$ and the dimension d.

Proof. Choosing $k = \lambda$, $\eta = 3d$, $p = 2$ in (3.13), we have

$$\|Q^+(g,f)\|_{\mathcal{L}^2_{3d}(\mathbb{R}^d)} \le C_{\lambda,3d,2}(\mathcal{B}) \|g\|_{\mathcal{L}^1_{\lambda+6d}(\mathbb{R}^d)} \|f\|_{\mathcal{L}^2_{\lambda+3d}(\mathbb{R}^d)}.$$

Choosing $\eta = 3d$, $p = 2$ in (3.14), we have

$$\|Q^-(g,f)\|_{\mathcal{L}^2_{3d}(\mathbb{R}^d)} \le C_\lambda \|g\|_{\mathcal{L}^1_\lambda(\mathbb{R}^d)} \|f\|_{\mathcal{L}^2_{\lambda+3d}(\mathbb{R}^d)}.$$

Combining both, we obtain

$$\|Q(g,f)\|_{\mathcal{L}^2_{3d}(\mathbb{R}^d)} \le C_d(\mathcal{B}) \|g\|_{\mathcal{L}^1_{\lambda+6d}(\mathbb{R}^d)} \|f\|_{\mathcal{L}^2_{\lambda+3d}(\mathbb{R}^d)}.$$

$\square$

Before we proceed to the consistency proof, we need the following lemmas.

Lemma 3.4. *Under the algebraic mapping (2.10) with $S = 1$, we have*

$$\|f\|_{\mathcal{L}^2_\eta(\mathbb{R}^d)} \le \|f\|_{L^2_{\boldsymbol{\mu}^{-2\eta}}(\mathbb{R}^d)} \le \|f\|_{\mathcal{L}^2_{d\eta}(\mathbb{R}^d)} \quad \text{for any } \eta \ge 0. \tag{3.16}$$

Proof. Note that

$$1 + \sum_{j=1}^d |v_j|^2 \le \prod_{j=1}^d (1 + |v_j|^2) \le (1 + \sum_{j=1}^d |v_j|^2)^d.$$

Then we have

$$
\begin{aligned}
\|f\|_{L^2_{\boldsymbol{\mu}^{-2\eta}}(\mathbb{R}^d)} &= \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \boldsymbol{\mu}^{-2\eta} \, d\boldsymbol{v} \right)^{1/2} \\
&= \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \prod_{j=1}^d (1 + |v_j|^2)^\eta \, d\boldsymbol{v} \right)^{1/2} \\
&\ge \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \left(1 + \sum_{j=1}^d |v_j|^2 \right)^\eta d\boldsymbol{v} \right)^{1/2} = \|f\|_{\mathcal{L}^2_\eta(\mathbb{R}^d)}.
\end{aligned}
$$

Also,

$$
\begin{aligned}
\|f\|_{L^2_{\boldsymbol{\mu}^{-2\eta}}(\mathbb{R}^d)} &= \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \prod_{j=1}^d (1 + |v_j|^2)^\eta \, d\boldsymbol{v} \right)^{1/2} \\
&\le \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 \left(1 + \sum_{j=1}^d |v_j|^2 \right)^{d\eta} d\boldsymbol{v} \right)^{1/2} = \|f\|_{\mathcal{L}^2_{d\eta}(\mathbb{R}^d)}.
\end{aligned}
$$

$\square$

9

Lemma 3.5. *For any $\eta \geq 0$, there exist $\epsilon > 0$ and $C_\epsilon > 0$ such that*

$$\|f\|_{\mathcal{L}_\eta^1(\mathbb{R}^d)} \leq C_\epsilon \|f\|_{\mathcal{L}_{\eta+\frac{1+\epsilon}{2}}^2(\mathbb{R}^d)}. \tag{3.17}$$

Proof. Note that

$$
\begin{aligned}
\|f\|_{\mathcal{L}_\eta^1(\mathbb{R}^d)}^2 &= \left(\int_{\mathbb{R}^d} |f(\boldsymbol{v})|(1+|\boldsymbol{v}|^2)^{\frac{\eta}{2}} \, d\boldsymbol{v} \right)^2 \\
&\leq \int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 (1+|\boldsymbol{v}|^2)^{\eta+\frac{1+\epsilon}{2}} \, d\boldsymbol{v} \int_{\mathbb{R}^d} (1+|\boldsymbol{v}|^2)^{-\frac{1+\epsilon}{2}} \, d\boldsymbol{v} \\
&\leq C_\epsilon \int_{\mathbb{R}^d} |f(\boldsymbol{v})|^2 (1+|\boldsymbol{v}|^2)^{\eta+\frac{1+\epsilon}{2}} \, d\boldsymbol{v} \\
&= C_\epsilon \|f\|_{\mathcal{L}_{\eta+\frac{1+\epsilon}{2}}^2(\mathbb{R}^d)}^2,
\end{aligned}
$$

where we used the Cauchy-Schwarz inequality. $\qquad\square$

We are ready to present a consistency result.

Theorem 3.6. *Let the collision kernel takes the form $\mathcal{B} = C_\lambda |\boldsymbol{v} - \boldsymbol{v}_*|^\lambda$, where $0 \leq \lambda \leq 1$ and C_λ is a positive constant. Then under the algebraic mapping (2.10) with $S = 1$, we have*

$$\left\| Q(f,f) - \pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f) \right\|_{\mathcal{L}_3^2(\mathbb{R}^d)}$$

$$\leq C_{d,\epsilon}(\mathcal{B}) N^{-m} \left(|f|_{\boldsymbol{B}_{\lambda+6d+\frac{3+\epsilon}{2}}^m(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + |f|_{\boldsymbol{B}_{\lambda+3d+1}^m(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} + |Q(f,f)|_{\boldsymbol{B}_4^m(\mathbb{R}^d)} \right), \tag{3.18}$$

where m is a positive integer, d is the dimension, $\epsilon > 0$ is a constant, and $C_{d,\epsilon}(\mathcal{B})$ is a constant depending only on the kernel $\mathcal{B}$, d and ϵ.

Proof. By the triangle inequality

$$
\begin{aligned}
&\|Q(f,f) - \pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{\mathcal{L}_3^2(\mathbb{R}^d)} \\
&\leq \|Q(f,f) - \pi_N^{4,d} Q(f,f)\|_{\mathcal{L}_3^2(\mathbb{R}^d)} + \|\pi_N^{4,d} Q(f,f) - \pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{\mathcal{L}_3^2(\mathbb{R}^d)}.
\end{aligned}
$$

For the first term, by Lemma 3.4 and Theorem 2.3, we have

$$\|Q(f,f) - \pi_N^{4,d} Q(f,f)\|_{\mathcal{L}_3^2(\mathbb{R}^d)} \leq \left\| Q(f,f) - \pi_N^{4,d} Q(f,f) \right\|_{L_{\boldsymbol{\mu}^{-6}}^2(\mathbb{R}^d)} \leq C N^{-m} |Q(f,f)|_{\boldsymbol{B}_4^m(\mathbb{R}^d)}.$$

For the second term, using again Lemma 3.4, Theorem 2.3, and Lemma 3.5, Theorem 3.3, we have

$$\|\pi_N^{4,d} Q(f,f) - \pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{\mathcal{L}_3^2(\mathbb{R}^d)} \leq \|\pi_N^{4,d} Q(f,f) - \pi_N^{4,d} Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{L^2_{\mu^{-6}}(\mathbb{R}^d)}$$

$$\leq \|Q(f,f) - Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{L^2_{\mu^{-6}}(\mathbb{R}^d)} \leq \|Q(f,f) - Q(\pi_N^{4,d} f, \pi_N^{4,d} f)\|_{\mathcal{L}_{3d}^2(\mathbb{R}^d)}$$

$$\leq \|Q(f - \pi_N^{4,d} f, f)\|_{\mathcal{L}_{3d}^2(\mathbb{R}^d)} + \|Q(\pi_N^{4,d} f, f - \pi_N^{4,d} f)\|_{\mathcal{L}_{3d}^2(\mathbb{R}^d)}$$

$$\leq C_d(\mathcal{B}) \left(\|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + \|\pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} \right)$$

$$\leq C_d(\mathcal{B}) \left(\|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} (\|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)}) + \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} \right)$$

$$\leq C_{d,\epsilon}(\mathcal{B}) \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+6d+\frac{1+\epsilon}{2}}^2(\mathbb{R}^d)} \left(\|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} \right)$$

$$+ C_d(\mathcal{B}) \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \|f - \pi_N^{4,d} f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)}$$

$$\leq C_{d,\epsilon}(\mathcal{B}) \|f - \pi_N^{4,d} f\|_{L^2_{\mu^{-2(\lambda+6d+\frac{1+\epsilon}{2})}}(\mathbb{R}^d)} (\|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + \|f - \pi_N^{4,d} f\|_{L^2_{\mu^{-2(\lambda+3d)}}(\mathbb{R}^d)})$$

$$+ C_d(\mathcal{B}) \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \|f - \pi_N^{4,d} f\|_{L^2_{\mu^{-2(\lambda+3d)}}(\mathbb{R}^d)}$$

$$\leq C_{d,\epsilon}(\mathcal{B}) N^{-m} |f|_{\mathbf{B}_{\lambda+6d+\frac{3+\epsilon}{2}}^m(\mathbb{R}^d)} \left(\|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + C N^{-m} |f|_{\mathbf{B}_{\lambda+3d+1}^m(\mathbb{R}^d)} \right)$$

$$+ C_d(\mathcal{B}) N^{-m} \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} |f|_{\mathbf{B}_{\lambda+3d+1}^m(\mathbb{R}^d)}$$

$$\leq C_{d,\epsilon}(\mathcal{B}) N^{-m} \left(|f|_{\mathbf{B}_{\lambda+6d+\frac{3+\epsilon}{2}}^m(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{\lambda+3d}^2(\mathbb{R}^d)} + |f|_{\mathbf{B}_{\lambda+3d+1}^m(\mathbb{R}^d)} \|f\|_{\mathcal{L}_{\lambda+6d}^1(\mathbb{R}^d)} \right).$$

Combining the above inequalities, we arrive at the desired result. $\qquad\square$

3.2 Approximation property for the moments

For the Boltzmann equation, the moments or macroscopic observables are important physical quantities. Still, under the algebraic mapping (2.10), we can show that the spectral method (3.5) preserves mass and energy.

Theorem 3.7. *If using the algebraic mapping (2.10) with $N \geq 2$, the spectral method (3.5) preserves mass and energy, i.e., $\rho(t)$ and $E(t)$ defined by*

$$\rho(t) := \int_{\mathbb{R}^d} f_N(t, \boldsymbol{v}) \, d\boldsymbol{v}, \quad E(t) = \int_{\mathbb{R}^d} f_N(t, \boldsymbol{v}) |\boldsymbol{v}|^2 \, d\boldsymbol{v} \tag{3.19}$$

remain constant in time. Furthermore,

$$\rho(t) \equiv \int_{\mathbb{R}^d} f^0(\boldsymbol{v}) \, d\boldsymbol{v}, \quad E(t) \equiv \int_{\mathbb{R}^d} f^0(\boldsymbol{v}) |\boldsymbol{v}|^2 \, d\boldsymbol{v}. \tag{3.20}$$

Proof. In 1D, the first few Chebyshev polynomials read

$$T_0(\xi) = 1, \quad T_1(\xi) = \xi, \quad T_2(\xi) = 2\xi^2 - 1.$$

With the algebraic mapping (2.10), we have

$$T_0(\xi(v)) = 1, \quad T_1(\xi(v)) = \frac{v}{\sqrt{v^2 + S^2}}, \quad T_2(\xi(v)) = \frac{v^2 - S^2}{v^2 + S^2}, \quad \mu(\xi(v)) = \frac{\sqrt{S}}{\sqrt{v^2 + S^2}}.$$

Then

$$\widehat{T}_k(v) = \frac{[\mu(\xi(v))]^{-2}}{\sqrt{c_k}} T_k(\xi(v)) = \frac{v^2 + S^2}{\sqrt{c_k} S} T_k(\xi(v)).$$

Specifically,

$$\widehat{T}_0(v) = \frac{v^2 + S^2}{\sqrt{c_0}S}, \quad \widehat{T}_1(v) = \frac{v\sqrt{v^2 + S^2}}{\sqrt{c_1}S}, \quad \widehat{T}_2(v) = \frac{v^2 - S^2}{\sqrt{c_2}S}.$$

Therefore,

$$1 = \frac{\sqrt{c_0}}{2S}\widehat{T}_0(v) - \frac{\sqrt{c_2}}{2S}\widehat{T}_2(v), \quad v^2 = \frac{\sqrt{c_0}S}{2}\widehat{T}_0(v) + \frac{\sqrt{c_2}S}{2}\widehat{T}_2(v).$$

Hence we can replace $(\widehat{T}_0(v), \widehat{T}_2(v))$ by $(1, v^2)$ as basis functions, namely,

$$\widehat{\mathbb{T}}_N^1 = \text{span}\{1, \widehat{T}_1, v^2, \widehat{T}_3, \widehat{T}_4, \cdots, \widehat{T}_N\}.$$

In d dimensions, it is easy to see

$$1, \ v_1^2, \ v_2^2, \ \ldots, \ v_d^2 \in \widehat{\mathbb{T}}_N^d \quad \text{for } N \geq 2.$$

In other words, we have shown that $1, |\boldsymbol{v}|^2 \in \widehat{\mathbb{T}}_N^d$ for $N \geq 2$.

On the other hand, by (3.19) and (3.4), we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho(t) = (\partial_t f_N(t, \boldsymbol{v}), 1)_{\mathbb{R}^d} = (Q(f_N, f_N), 1)_{\mathbb{R}^d} = 0;$$

$$\frac{\mathrm{d}}{\mathrm{d}t}E(t) = (\partial_t f_N(t, \boldsymbol{v}), |\boldsymbol{v}|^2)_{\mathbb{R}^d} = (Q(f_N, f_N), |\boldsymbol{v}|^2)_{\mathbb{R}^d} = 0,$$

where in the last equality we used the conservation property (1.5) of the collision operator.

It remains to show

$$\int_{\mathbb{R}^d} f_N(0, \boldsymbol{v})\,\mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^d} f^0(\boldsymbol{v})\,\mathrm{d}\boldsymbol{v}, \quad \int_{\mathbb{R}^d} f_N(0, \boldsymbol{v})|\boldsymbol{v}|^2\,\mathrm{d}\boldsymbol{v} = \int_{\mathbb{R}^d} f^0(\boldsymbol{v})|\boldsymbol{v}|^2\,\mathrm{d}\boldsymbol{v}.$$

Note that $f_N(0, \boldsymbol{v}) = \pi^{4,d}f^0$, it suffices to show

$$(\pi^{4,d}f^0 - f^0, 1)_{\mathbb{R}^d} = (\pi^{4,d}f^0 - f^0, |\boldsymbol{v}|^2)_{\mathbb{R}^d} = 0,$$

which is true by (2.24) (with $\alpha = 4$). $\qquad\square$

4   Numerical realization

To implement the proposed spectral method, one needs to solve the ODE system (3.5). For time discretization, one can just use the explicit Runge-Kutta methods. Hence, the key is the efficient evaluation of $\mathcal{Q}_{\boldsymbol{k}}^N$ as defined in (3.6).

In this section, we introduce two algorithms to compute $\mathcal{Q}_{\boldsymbol{k}}^N$. The first one is a direct algorithm that treats $\mathcal{Q}_{\boldsymbol{k}}^N$ as a matrix/tensor-vector multiplication. Since the weight matrix/tensor does not depend on the numerical solution f_N, it can be precomputed and stored for repeated use. This approach is simple but will soon meet a bottleneck when N increases since the memory requirement as well as the online computational cost can get extremely high. To alleviate this, we propose a fast algorithm, where the key idea is to recognize the gain term of the collision operator as a non-uniform discrete Fourier cosine transform to be accelerated by the non-uniform FFT (NUFFT). Note that this is possible because we are using the mapped Chebyshev functions as a basis, which is related to the Fourier cosine series.

4.1 A direct algorithm

To derive the direct algorithm, we substitute (3.3) into (3.6) to obtain

$$
\begin{aligned}
\mathcal{Q}_{\boldsymbol{k}}^N & = \sum_{0 \leq \boldsymbol{i},\boldsymbol{j} \leq N} \widetilde{f}_{\boldsymbol{i}} \widetilde{f}_{\boldsymbol{j}} \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v},\boldsymbol{v}_*,\boldsymbol{\sigma}) \, \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}(\boldsymbol{v}) \widetilde{\boldsymbol{T}}_{\boldsymbol{j}}(\boldsymbol{v}_*) [\widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}') - \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v})] \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v} \, \mathrm{d}\boldsymbol{v}_* \\
& = \sum_{0 \leq \boldsymbol{i},\boldsymbol{j} \leq N} \widetilde{f}_{\boldsymbol{i}} \widetilde{f}_{\boldsymbol{j}} \left[\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) - \widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) \right], \quad 0 \leq \boldsymbol{k} \leq N,
\end{aligned} \tag{4.1}
$$

where

$$
\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v},\boldsymbol{v}_*,\boldsymbol{\sigma}) \, \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}(\boldsymbol{v}) \widetilde{\boldsymbol{T}}_{\boldsymbol{j}}(\boldsymbol{v}_*) \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}') \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v} \, \mathrm{d}\boldsymbol{v}_*, \tag{4.2}
$$

$$
\widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) := \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v},\boldsymbol{v}_*,\boldsymbol{\sigma}) \, \widetilde{\boldsymbol{T}}_{\boldsymbol{i}}(\boldsymbol{v}) \widetilde{\boldsymbol{T}}_{\boldsymbol{j}}(\boldsymbol{v}_*) \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v} \, \mathrm{d}\boldsymbol{v}_*. \tag{4.3}
$$

Since the tensors $\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$ and $\widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$ do not depend on coefficients $\{\widetilde{f}_{\boldsymbol{k}}\}_{0 \leq \boldsymbol{k} \leq N}$, a straightforward way to evaluate $\mathcal{Q}_{\boldsymbol{k}}^N$ is to precompute $\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$ and $\widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$, and then evaluate the sum in (4.1) directly in the online computation. This is what we refer to as the direct algorithm. We observe that this algorithm requires $\mathcal{O}(N^{3d})$ memory to store the tensors $\widetilde{I}_1$ and $\widetilde{I}_2$; and to evaluate (4.1), it requires $\mathcal{O}(N^{3d})$ operations. Both the memory requirement and online computational cost can be quite demanding especially for $d = 3$ and large N.

We give some details on how to approximate $\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$ and $\widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k})$, though this step can be completed in advance and does not take the actual computational time. We first perform a change of variables $(\boldsymbol{v},\boldsymbol{v}_*) \to (\boldsymbol{v}(\boldsymbol{\xi}),\boldsymbol{v}_*(\boldsymbol{\eta}))$ to transform the integrals of $(\boldsymbol{v},\boldsymbol{v}_*) \in \mathbb{R}^d \times \mathbb{R}^d$ into integrals of $(\boldsymbol{\xi},\boldsymbol{\eta}) \in I^d \times I^d$, using the mapping introduced in Section 2.1:

$$
\widetilde{I}_1(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) = \int_{I^d} \int_{I^d} G_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta}) \frac{\boldsymbol{T}_{\boldsymbol{i}}(\boldsymbol{\xi}) \boldsymbol{T}_{\boldsymbol{j}}(\boldsymbol{\eta})}{\sqrt{c_{\boldsymbol{i}} c_{\boldsymbol{j}}}} \boldsymbol{\omega}(\boldsymbol{\xi}) \boldsymbol{\omega}(\boldsymbol{\eta}) \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{\eta}, \tag{4.4}
$$

$$
\widetilde{I}_2(\boldsymbol{i},\boldsymbol{j},\boldsymbol{k}) = \int_{I^d} \int_{I^d} L_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta}) \frac{\boldsymbol{T}_{\boldsymbol{i}}(\boldsymbol{\xi}) \boldsymbol{T}_{\boldsymbol{j}}(\boldsymbol{\eta})}{\sqrt{c_{\boldsymbol{i}} c_{\boldsymbol{j}}}} \boldsymbol{\omega}(\boldsymbol{\xi}) \boldsymbol{\omega}(\boldsymbol{\eta}) \, \mathrm{d}\boldsymbol{\xi} \, \mathrm{d}\boldsymbol{\eta}, \tag{4.5}
$$

where

$$
G_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta}) := [\boldsymbol{\mu}(\boldsymbol{\xi})\boldsymbol{\mu}(\boldsymbol{\eta})]^2 \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}),\boldsymbol{v}_*(\boldsymbol{\eta}),\boldsymbol{\sigma}) \frac{\boldsymbol{T}_{\boldsymbol{k}}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)}{\sqrt{c_{\boldsymbol{k}}} [\boldsymbol{\mu}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)]^2} \, \mathrm{d}\boldsymbol{\sigma}, \tag{4.6}
$$

$$
L_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta}) := \frac{\boldsymbol{T}_{\boldsymbol{k}}(\boldsymbol{\xi}) [\boldsymbol{\mu}(\boldsymbol{\eta})]^2}{\sqrt{c_{\boldsymbol{k}}}} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}),\boldsymbol{v}_*(\boldsymbol{\eta}),\boldsymbol{\sigma}) \, \mathrm{d}\boldsymbol{\sigma}. \tag{4.7}
$$

Notice that in (4.6), $\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})} \in I^d$ is the value transformed from

$$
\boldsymbol{v}' = \frac{1}{2}(\boldsymbol{v}(\boldsymbol{\xi}) + \boldsymbol{v}_*(\boldsymbol{\eta})) + \frac{1}{2} |\boldsymbol{v}(\boldsymbol{\xi}) - \boldsymbol{v}_*(\boldsymbol{\eta})| \boldsymbol{\sigma} \in \mathbb{R}^d
$$

under the same mapping. To approximate the above integrals in $\boldsymbol{\xi}$, $\boldsymbol{\eta}$, and $\boldsymbol{\sigma}$, we choose $M_{\boldsymbol{v}}$ Chebyshev-Gauss-Lobatto quadrature points in each dimension of I^d for both $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$; and $M_{\boldsymbol{\sigma}}$ quadrature points on the unit sphere S^{d-1} (for $d = 2$, this can be the uniform points in polar angle; for $d = 3$, this can be the Lebedev quadrature [22]). Therefore, for each fixed index $\boldsymbol{k}$, (4.4) and (4.5) are forward Chebyshev transforms of the functions $G_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta})$ and $L_{\boldsymbol{k}}(\boldsymbol{\xi},\boldsymbol{\eta})$, respectively. Thus, they can be evaluated efficiently using the fast Chebyshev transform.

4.2 A fast algorithm

To introduce the fast algorithm, we take the original form (3.6) and split $\mathcal{Q}_{\boldsymbol{k}}^{N}$ into a gain term and a loss term as $\mathcal{Q}_{\boldsymbol{k}}^{N} = \mathcal{Q}_{\boldsymbol{k}}^{N,+} - \mathcal{Q}_{\boldsymbol{k}}^{N,-}$, where

$$\mathcal{Q}_{\boldsymbol{k}}^{N,+} = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*) \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}') \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v}_* \right) f_N(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}, \tag{4.8}$$

$$\mathcal{Q}_{\boldsymbol{k}}^{N,-} = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*) \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{v}_* \right) f_N(\boldsymbol{v}) \widehat{\boldsymbol{T}}_{\boldsymbol{k}}(\boldsymbol{v}) \, \mathrm{d}\boldsymbol{v}. \tag{4.9}$$

We propose to evaluate $\mathcal{Q}_{\boldsymbol{k}}^{N,+}$ and $\mathcal{Q}_{\boldsymbol{k}}^{N,-}$ following the above expressions. To this end, given the coefficients $\{\widetilde{f}_{\boldsymbol{k}}\}_{0\leq \boldsymbol{k}\leq N}$ at each time step, we first reconstruct f_N as in (3.3) at $M_{\boldsymbol{v}}$ Chebyshev-Gauss-Lobatto quadrature points in each dimension of $\boldsymbol{v}$ (for an accurate approximation we choose $M_{\boldsymbol{v}} = N+2$). This can be achieved by the fast Chebyshev transform in $\mathcal{O}(M_{\boldsymbol{v}}^d \log M_{\boldsymbol{v}})$ operations.

To evaluate the gain term $\mathcal{Q}_{\boldsymbol{k}}^{N,+}$, we change the integrals of $(\boldsymbol{v}, \boldsymbol{v}_*) \in \mathbb{R}^d \times \mathbb{R}^d$ into integrals of $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in I^d \times I^d$ in (4.8) (similarly as in the previous subsection for $\widetilde{I}_1(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$):

$$\begin{aligned}
\mathcal{Q}_{\boldsymbol{k}}^{N,+} &= \int_{I^d} \left(\int_{I^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta})) \frac{T_{\boldsymbol{k}}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)}{\sqrt{c_{\boldsymbol{k}}} \left[\boldsymbol{\mu}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)\right]^2} \frac{\omega(\boldsymbol{\eta})}{[\boldsymbol{\mu}(\boldsymbol{\eta})]^2} \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{\eta} \right) f_N(\boldsymbol{v}(\boldsymbol{\xi})) \frac{\omega(\boldsymbol{\xi})}{[\boldsymbol{\mu}(\boldsymbol{\xi})]^2} \, \mathrm{d}\boldsymbol{\xi} \\
&= \int_{I^d} \left(\int_{S^{d-1}} F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi}) \, \mathrm{d}\boldsymbol{\sigma} \right) f_N(\boldsymbol{v}(\boldsymbol{\xi})) \frac{\omega(\boldsymbol{\xi})}{[\boldsymbol{\mu}(\boldsymbol{\xi})]^2} \, \mathrm{d}\boldsymbol{\xi},
\end{aligned} \tag{4.10}$$

where

$$F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi}) := \int_{I^d} \mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta})) \frac{T_{\boldsymbol{k}}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)}{\sqrt{c_{\boldsymbol{k}}} \left[\boldsymbol{\mu}\left(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta},\boldsymbol{\sigma})}\right)\right]^2} \frac{\omega(\boldsymbol{\eta})}{[\boldsymbol{\mu}(\boldsymbol{\eta})]^2} \, \mathrm{d}\boldsymbol{\eta}. \tag{4.11}$$

Suppose $M_{\boldsymbol{v}}$ quadrature points are used in each dimension of $\boldsymbol{v}$ and $\boldsymbol{v}_*$ and $M_{\boldsymbol{\sigma}}$ points are used on the sphere S^{d-1}, a direct evaluation of (4.11) would require $\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^{2d} N^d)$ operations. Given $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$, a direct evaluation of (4.10) would take $\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^d N^d)$ operations. Therefore, the major bottleneck is to compute $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$, which is prohibitively expensive without a fast algorithm. Our main idea is to recognize (4.11) as a non-uniform discrete Fourier cosine transform so it can be evaluated by the non-uniform fast Fourier transform (NUFFT). We will see that the total complexity to evaluate $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$ can be brought down to $\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^{2d} |\log \epsilon| + M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^d N^d \log N)$, where ϵ is the requested precision in the NUFFT algorithm.

Applying the Chebyshev-Gauss-Lobatto quadrature $(\boldsymbol{\eta}_j, w_j)_{1\leq j \leq M_{\boldsymbol{v}}}$, (4.11) becomes

$$\begin{aligned}
F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi}) &\approx \sum_{1\leq j \leq M_{\boldsymbol{v}}} w_j \frac{\mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}_j), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta}_j))}{\sqrt{c_{\boldsymbol{k}}} [\boldsymbol{\mu}(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma})})]^2 [\boldsymbol{\mu}(\boldsymbol{\eta}_j)]^2} T_{\boldsymbol{k}}(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma})}) \\
&= \sum_{1\leq j \leq M_{\boldsymbol{v}}} w_j \frac{\mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}_j), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta}_j))}{\sqrt{c_{\boldsymbol{k}}} [\boldsymbol{\mu}(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma})})]^2 [\boldsymbol{\mu}(\boldsymbol{\eta}_j)]^2} \prod_{l=1}^{d} \cos\left(k_l \arccos(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma}),l})\right) \\
&= \frac{1}{\sqrt{c_{\boldsymbol{k}}}} \sum_{1\leq j \leq M_{\boldsymbol{v}}} q_j \prod_{l=1}^{d} \cos\left(k_l \boldsymbol{z}_{j,l}\right),
\end{aligned} \tag{4.12}$$

where $\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma}),l}$ is the l-th component of $\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma})}$, and

$$q_j := w_j \frac{\mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}_j), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta}_j))}{[\boldsymbol{\mu}(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma})})]^2 [\boldsymbol{\mu}(\boldsymbol{\eta}_j)]^2}, \quad \boldsymbol{z}_{j,l} := \arccos(\boldsymbol{\zeta}_{(\boldsymbol{\xi},\boldsymbol{\eta}_j,\boldsymbol{\sigma}),l}). \tag{4.13}$$

Note that q_j and $\boldsymbol{z}_j$ depend on $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$. (4.12) is almost like a non-uniform discrete Fourier cosine transform. Indeed, we propose to evaluate $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$ as follows.

14

For each fixed $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$, we compute

$$\widetilde{F}_{\boldsymbol{K}} := \sum_{1 \le \boldsymbol{j} \le M_{\boldsymbol{v}}} q_{\boldsymbol{j}} \exp(\mathbf{i}\boldsymbol{K} \cdot \boldsymbol{z_j}), \quad -N \le \boldsymbol{K} \le N, \tag{4.14}$$

which is a non-uniform discrete Fourier transform mapping non-uniform samples $\boldsymbol{z_j} \in [0, \pi]^d$ into frequencies $\boldsymbol{K} \in [-N, N]^d$. This can be done efficiently using the NUFFT. In recent years, various NUFFT algorithms have been developed. In our numerical realization, we employ an efficient algorithm called FINUFFT [2]. The general idea is to apply an interpolation between non-uniform samples and an equispaced grid, and then perform the uniform FFT on the new grid. This algorithm only costs $\mathcal{O}(M_{\boldsymbol{v}}^d |\log \epsilon| + N^d \log N)$ operations in computing (4.14) with the requested precision ϵ. Once we obtain $\widetilde{F}_{\boldsymbol{K}}$, $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$ can be retrieved as

- in 2D

$$F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi}) = \frac{1}{\sqrt{c_{\boldsymbol{k}}}} \frac{1}{2} \mathbf{Re}\left(\widetilde{F}_{(k_1, k_2)} + \widetilde{F}_{(-k_1, k_2)} \right); \tag{4.15}$$

- in 3D

$$F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi}) = \frac{1}{\sqrt{c_{\boldsymbol{k}}}} \frac{1}{4} \mathbf{Re}\left(\widetilde{F}_{(k_1, k_2, k_3)} + \widetilde{F}_{(-k_1, k_2, k_3)} + \widetilde{F}_{(k_1, -k_2, k_3)} + \widetilde{F}_{(k_1, k_2, -k_3)} \right). \tag{4.16}$$

This procedure needs to be repeated for every $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$, hence the overall computational cost for getting $F_{\boldsymbol{k}}(\boldsymbol{\sigma}, \boldsymbol{\xi})$ is $\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^{2d} |\log \epsilon| + M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^d N^d \log N)$.

To evaluate the loss term $\mathcal{Q}_{\boldsymbol{k}}^{N,-}$, we change the integrals of $(\boldsymbol{v}, \boldsymbol{v}_*) \in \mathbb{R}^d \times \mathbb{R}^d$ into integrals of $(\boldsymbol{\xi}, \boldsymbol{\eta}) \in I^d \times I^d$ in (4.9) (similarly as in the previous subsection for $\widetilde{I}_2(\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k})$):

$$\mathcal{Q}_{\boldsymbol{k}}^{N,-} = \int_{I^d} \left(\int_{I^d} \int_{S^{d-1}} \mathcal{B}(\boldsymbol{v}(\boldsymbol{\xi}), \boldsymbol{v}_*(\boldsymbol{\eta}), \boldsymbol{\sigma}) f_N(\boldsymbol{v}_*(\boldsymbol{\eta})) \frac{\omega(\boldsymbol{\eta})}{[\boldsymbol{\mu}(\boldsymbol{\eta})]^2} \, \mathrm{d}\boldsymbol{\sigma} \, \mathrm{d}\boldsymbol{\eta} \right) f_N(\boldsymbol{v}(\boldsymbol{\xi})) \frac{\boldsymbol{T_k}(\boldsymbol{\xi}) \omega(\boldsymbol{\xi})}{\sqrt{c_{\boldsymbol{k}}} [\boldsymbol{\mu}(\boldsymbol{\xi})]^4} \, \mathrm{d}\boldsymbol{\xi}. \tag{4.17}$$

Then one can just evaluate the terms in the parentheses directly with complexity $\mathcal{O}(M_{\boldsymbol{v}}^{2d} M_{\boldsymbol{\sigma}})$. The outer integral in $\boldsymbol{\xi}$ can be viewed as the Chebyshev transform of some function thus can be evaluated efficiently by the fast Chebyshev transform in $\mathcal{O}(M_{\boldsymbol{v}}^d \log M_{\boldsymbol{v}})$. In particular, if we consider the Maxwell kernel, i.e., $\mathcal{B}(\boldsymbol{v}, \boldsymbol{v}_*, \boldsymbol{\sigma}) \equiv$ constant, terms in the parentheses only requires $\mathcal{O}(M_{\boldsymbol{v}}^d)$ complexity.

4.3 Comparison of direct and fast algorithms

To summarize, we list the storage requirement and (online) computational complexity for both the direct and fast algorithms in Table 1. Note that we only list the dominant complexity for each term. It is clear that the main cost of the fast algorithm comes from evaluating the gain term. Compared with the direct algorithm, the fast algorithm is generally faster as $M_{\boldsymbol{\sigma}}$ can be chosen much smaller than N^d in practice (see Section 5). Most importantly, the fast algorithm does not require any precomputation with excessive storage requirement and everything can be computed on the fly.

	direct algorithm		fast algorithm
	storage	(online) operation	(online) operation
gain term	$\mathcal{O}(N^{3d})$	$\mathcal{O}(N^{3d})$	$\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^{2d} |\log \epsilon| + M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^d N^d \log N)$
loss term	$\mathcal{O}(N^{3d})$	$\mathcal{O}(N^{3d})$	$\mathcal{O}(M_{\boldsymbol{\sigma}} M_{\boldsymbol{v}}^{2d})$

Table 1: Storage requirement and (online) computational cost of the direct and fast algorithms. N is the number of spectral modes in each dimension of $\boldsymbol{v}$; $M_{\boldsymbol{v}} = \mathcal{O}(N)$ is the number of quadrature points in each dimension; $M_{\boldsymbol{\sigma}} \ll N^d$ is the number of quadrature points on the sphere S^{d-1}; and ϵ is the requested precision in the NUFFT algorithm. The proposed fast algorithm does not require extra storage other than that storing the computational target, e.g., the gain and loss terms.

5 Numerical examples

In this section, we perform extensive numerical tests to demonstrate the accuracy and efficiency of the proposed Petrov-Galerkin spectral method in both 2D and 3D.

Recall that the main motivation of the current work is to obtain better accuracy by considering approximations in an unbounded domain. To illustrate this point, we will compare three methods to solve the Boltzmann equation:

(1) **Fast Fourier-Galerkin spectral method proposed in [14]:** this method can achieve a good accuracy-efficiency tradeoff among the current deterministic methods for the Boltzmann equation. However, it requires the truncation of the domain to $[-L, L]^d$, where L is often chosen empirically such that the solution is close to zero at the boundary.

(2) **Fast Chebyshev-0 method:** the method proposed in the current paper using the logarithmic mapping (2.9), where $r = 0$ and the scaling parameter S needs to be properly chosen.

(3) **Fast Chebyshev-1 method:** the method proposed in the current paper using the algebraic mapping (2.10), where $r = 1$ and the scaling parameter S needs to be properly chosen.

In all three methods, the choice of truncation or mapping/scaling parameters has a great impact on the numerical accuracy. In the following tests, we first determine L in the Fourier spectral method. Then for the two Chebyshev methods, we propose an adaptive strategy to determine the scaling parameter S: for example, in 1D, the Chebyshev-Gauss-Lobatto quadrature points on the interval $[-1, 1]$ are given by

$$-1 = \xi_1 < \xi_2 < \ldots < \xi_N = 1.$$

We choose S such that the two quadrature points ξ_2 and ξ_{N-1} are mapped to the boundary of $[-L, L]$, i.e.,

$$v(\xi_1) = -\infty, \quad v(\xi_2) = -L, \quad v(\xi_{N-1}) = L, \quad v(\xi_N) = -\infty.$$

Note that this S is adaptive in the sense that different N will correspond to different S.

5.1 2D examples

5.1.1 2D BKW solution

We consider first the 2D BKW solution. This is one of the few known analytical solutions to the Boltzmann equation and a perfect example to verify the accuracy of a numerical method.

When $d = 2$ and the collision kernel $\mathcal{B} \equiv 1/(2\pi)$, the following is a solution to the initial value problem (3.1):

$$f_{\text{BKW}}(t, \boldsymbol{v}) = \frac{1}{2\pi K^2} \exp\left(-\frac{\boldsymbol{v}^2}{2K}\right)\left(2K - 1 + \frac{1-K}{2K}\boldsymbol{v}^2\right), \tag{5.1}$$

where $K = 1 - \exp(-t/8)/2$. By taking the time derivative of $f_{\text{BKW}}(t, \boldsymbol{v})$, we can obtain the exact collision operator as

$$Q_{\text{BKW}}(f) = \left\{\left(-\frac{2}{K} + \frac{\boldsymbol{v}^2}{2K^2}\right)f_{\text{BKW}} + \frac{1}{2\pi K^2}\exp\left(-\frac{\boldsymbol{v}^2}{2K}\right)\left(2 - \frac{1}{2K^2}\boldsymbol{v}^2\right)\right\}K', \tag{5.2}$$

where $K' = \exp(-t/8)/16$. This way we can apply the numerical method to compute $Q_{\text{BKW}}(f)$ directly and check its error without worrying about the time discretization.

In the fast Fourier spectral method, we take $N_\rho = N$ quadrature points in the radial direction and $N_\theta = 8$ quadrature points on the unit circle (see [14] for more details). In the fast Chebyshev methods, we take $M_{\boldsymbol{v}} = N + 2$ quadrature points for each dimension of $(\boldsymbol{v}, \boldsymbol{v}_*)$ and $M_{\boldsymbol{\sigma}} = N$ quadrature points on the unit circle. The precision in NUFFT is selected as $\epsilon = 1e - 14$. The numerical error of $Q_{\text{BKW}}(f)$ is estimated on a 200×200 uniform grid in the rectangular domain $[-6.3, 6.3]^2$ at time $t = 2$.

Test 01: In this test, we examine thoroughly the numerical errors concerning different truncation parameters L in the Fourier method, and scaling parameters S in the Chebyshev methods. The L^2 errors of $Q_{\text{BKW}}(f)$ for three methods are presented in Figure 1. It is obvious that the accuracy is not good when L and S are too small or too large. When L and S are chosen appropriately, the accuracy is close to the machine precision. In Table 2, we record the best accuracy for a given N of each method. One can see that the fast Chebyshev-0 method can always achieve the best accuracy.

	fast Fourier	fast Chebyshev-0	fast Chebyshev-1
$N = 8$	8.9223e-03	2.2289e-03	4.1388e-03
$N = 16$	1.0989e-04	2.2033e-05	2.9713e-04
$N = 24$	3.2104e-06	1.9843e-07	1.5004e-05
$N = 32$	9.4720e-08	3.0082e-09	5.3946e-07
$N = 40$	1.9836e-09	2.5434e-11	1.6120e-08
$N = 48$	6.1797e-11	1.6255e-13	6.0320e-10
$N = 56$	1.4315e-12	2.1482e-14	2.5213e-11

Table 2: (2D BKW: Test 01) The L^2 error of $Q_{\text{BKW}}(f)$ at time $t = 2$. The best accuracy for a given N of each method.

Test 02: In this test, we fix the computational domain and examine the numerical errors concerning different N. In the Fourier method, we test $L = 8.83$ and $L = 13.24$. In the Chebyshev methods, we use the same L to select the scaling parameter S accordingly. The L^∞ errors of $Q_{\text{BKW}}(f)$ for three methods are presented in Figure 2. Among these three methods, the fast Chebyshev-0 method can achieve the best accuracy when N is small. The fast Chebyshev-1 method doesn't provide a good approximation. This is because the quadrature points in Chebyshev-1 method are much more clustered near the origin compared to Chebyshev-0 method and apparently points located far away from the origin play an important role in the unbounded domain problem.

Test 03: In this test, we examine the numerical errors of the Chebyshev methods with a fixed scaling parameter: $S = 4$ in the fast Chebyshev-0 method; $S = 5$ in the fast Chebyshev-1 method. These two values are selected based on the results in Figure 1. The L^∞ errors of $Q_{\text{BKW}}(f)$ for both methods are
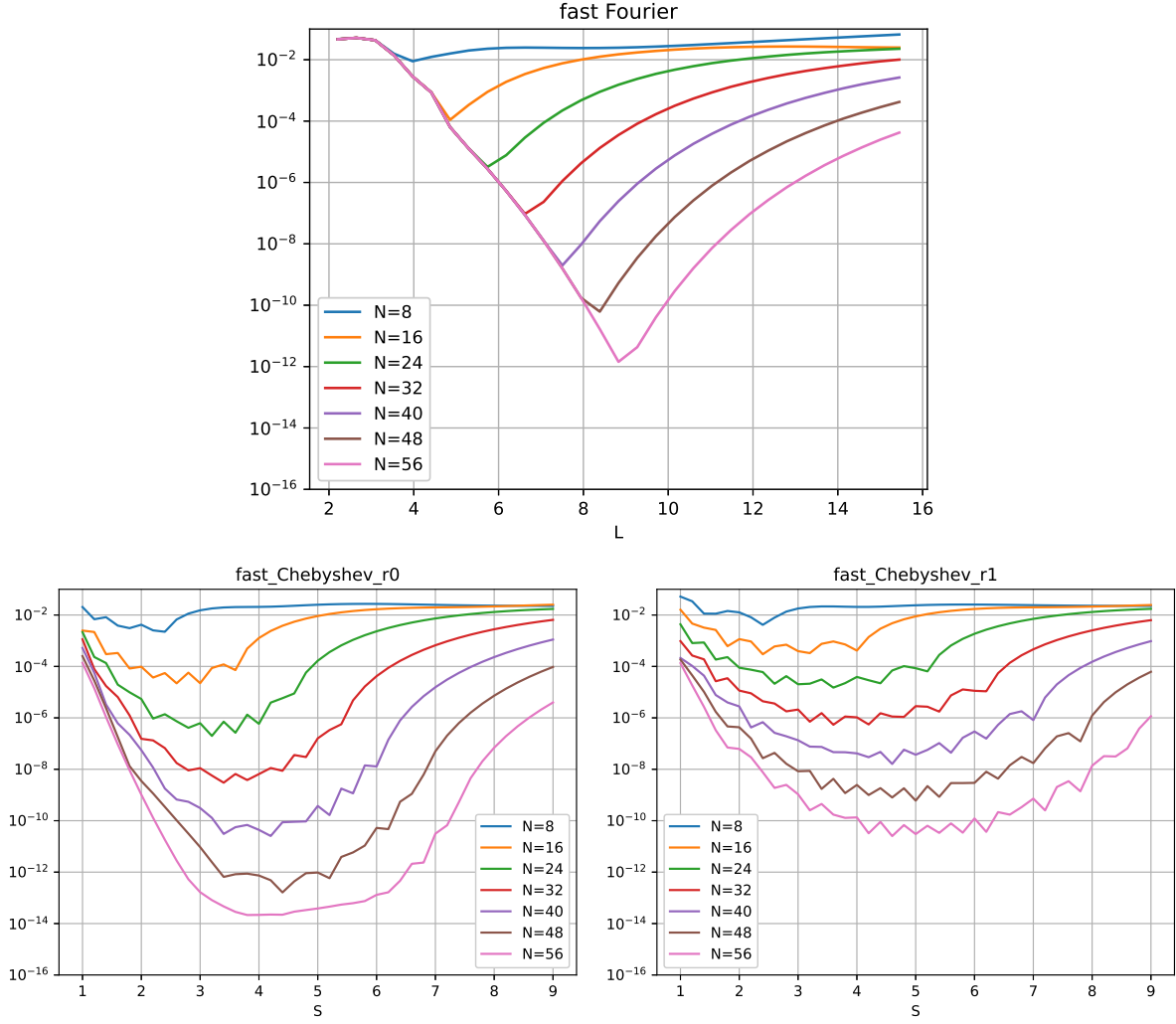
Figure 1: (2D BKW: Test 01) The L^2 error of $Q_{\mathrm{BKW}}(f)$ at time $t = 2$. Top: fast Fourier method. Bottom: fast Chebyshev methods.
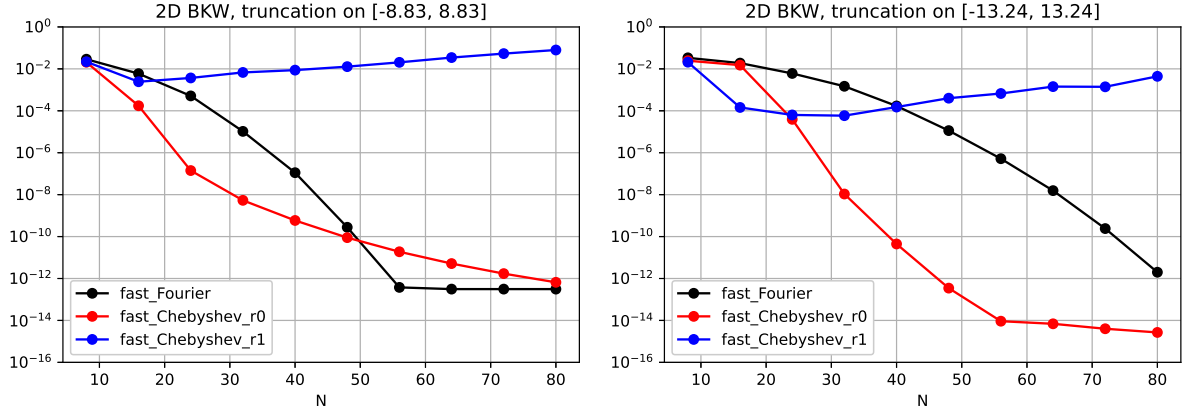

Figure 2: (2D BKW: Test 02) The L^∞ error of $Q_{\mathrm{BKW}}(f)$ at time $t = 2$. Left: $L = 8.83$; Right: $L = 13.24$.

presented in Figure 3. As a comparison, results of the fast Fourier method are also plotted. Again for small N, the fast Chebyshev-0 method provides the best accuracy among three methods.
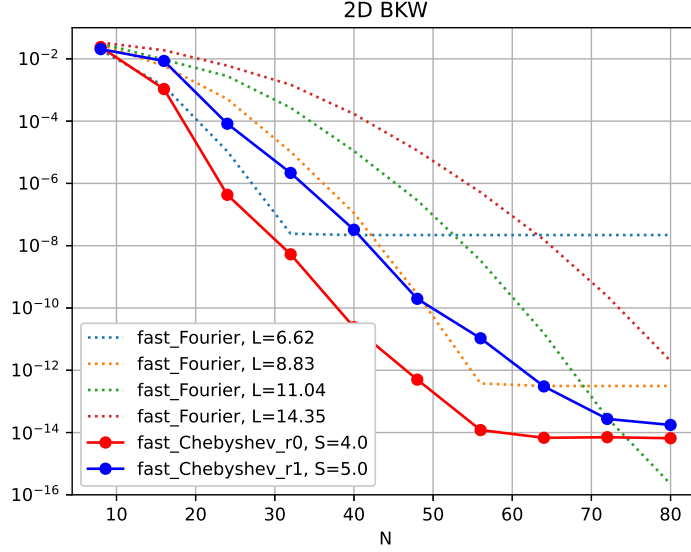

Figure 3: (2D BKW: Test 03) The L^∞ error of $Q_{\mathrm{BKW}}(f)$ at time $t = 2$.

Test 04: In this test, we report the computational time of the direct (Chebyshev) algorithm and the fast (Chebyshev) algorithm. The computations were done on Intel(R) Core(TM) i7-6700 CPU in a single thread. Table 3 shows the running time of the direct and fast algorithms concerning different N. Note that the direct algorithm is left out when $N \geq 32$ due to the memory constraint.

	direct algorithm		fast algorithm
	online (sec)	precomputation (sec)	online (sec)
$N = 8$	0.0047	1.5956	0.194955
$N = 16$	0.1207	62.7991	2.036821
$N = 32$	-	-	24.779492
$N = 64$	-	-	4.937722e+02
$N = 128$	-	-	1.331576e+04

Table 3: (2D BKW: Test 04) Running time in second for a single evaluation of the gain term.

5.1.2 Computing the moments

We next consider the time evolution problem and check the accuracy for moments approximation. Since the fast Chebyshev-0 method performs generally better than the fast Chebyshev-1 method, we will restrict to the former in the following tests. The comparison with the fast Fourier method will still be considered.

In (3.1), we choose the collision kernel $\mathcal{B} \equiv 1/(2\pi)$ and the initial condition as

$$f^0(\boldsymbol{v}) = \frac{\rho_1}{2\pi T_1} \exp\left(-\frac{(\boldsymbol{v} - V_1)^2}{2T_1}\right) + \frac{\rho_2}{2\pi T_2} \exp\left(-\frac{(\boldsymbol{v} - V_2)^2}{2T_2}\right), \tag{5.3}$$

where $\rho_1 = \rho_2 = 1/2$, $T_1 = T_2 = 1$ and $V_1 = (x_1, y_1) = (-1, 2)$, $V_2 = (x_2, y_2) = (3, -3)$.

19

Then for the momentum flow and energy flow defined as

$$P_{ij} = \int_{\mathbb{R}^2} f \boldsymbol{v}_i \boldsymbol{v}_j \, d\boldsymbol{v}, \quad (i, j = 1, 2), \quad q_i = \int_{\mathbb{R}^2} f \boldsymbol{v}_i |\boldsymbol{v}|^2 \, d\boldsymbol{v}, \quad (i = 1, 2), \tag{5.4}$$

we have their exact formulas

$$P_{11} = -\frac{9}{8} e^{-t/2} + \frac{57}{8}, \quad P_{12} = P_{21} = -5 e^{-t/2} - \frac{1}{2}, \quad P_{22} = \frac{9}{8} e^{-t/2} + \frac{51}{8}, \tag{5.5}$$

and

$$q_1 = \frac{1}{4} \left(11 e^{-t/2} + 103 \right), \quad q_2 = -\frac{1}{8} \left(89 e^{-t/2} + 103 \right). \tag{5.6}$$

In the fast Fourier method, we take $N_\rho = N$ quadrature points in the radial direction and $N_\theta = N$ quadrature points on the unit circle. The truncation domain $[-L, L]^2$ is selected as $L = 14.35$. In the fast Chebyshev-0 method, we take $M_{\boldsymbol{v}} = N + 2$ quadrature points for each dimension of $(\boldsymbol{v}, \boldsymbol{v}_*)$ and $M_{\boldsymbol{\sigma}} = N$ quadrature points on the unit circle. The precision in NUFFT is selected as $\epsilon = 1e - 14$. The scaling parameter S is adaptively chosen based on L. For both methods, we use the 4th-order Runge-Kutta method with $\Delta t = 0.02$ for time integration.

The absolute errors of the moments are presented in Figure 4–8. The fast Chebyshev-0 method clearly provides a better approximation in comparison to the Fourier method for fixed N.
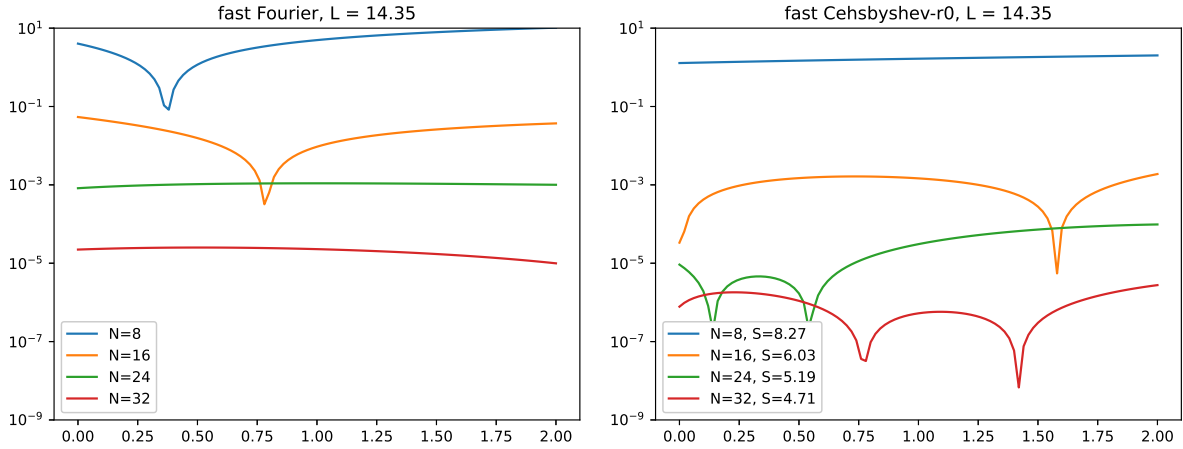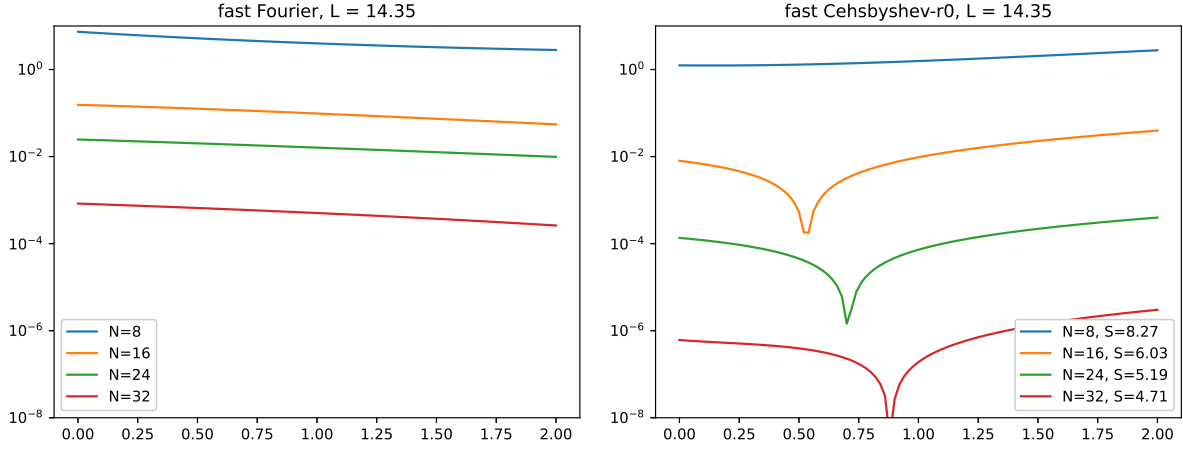

Figure 4: (2D moments) The time evolution for the absolute error of the momentum flow P_{11}. Left: the fast Fourier method. Right: the fast Chebyshev-0 method.

5.2 3D BKW solution

We finally consider the 3D BKW solution. When $d = 3$ and the collision kernel $\mathcal{B} \equiv 1/(4\pi)$, the following is a solution to the initial value problem (3.1):

$$f_{\text{BKW}}(t, \boldsymbol{v}) = \frac{1}{2(2\pi K)^{3/2}} \exp\left(-\frac{\boldsymbol{v}^2}{2K} \right) \left(\frac{5K - 3}{K} + \frac{1 - K}{K^2} \boldsymbol{v}^2 \right), \tag{5.7}$$

where $K = 1 - \exp(-t/6)$. As in 2D, we can obtain the exact collision operator as

$$Q_{\text{BKW}}(f) = \left\{ \left(-\frac{3}{2K} + \frac{\boldsymbol{v}^2}{2K^2} \right) f_{\text{BKW}} + \frac{1}{2(2\pi K)^{3/2}} \exp\left(-\frac{\boldsymbol{v}^2}{2K} \right) \left(\frac{3}{K^2} + \frac{K - 2}{K^3} \boldsymbol{v}^2 \right) \right\} K', \tag{5.8}$$

with $K' = \exp(-t/6)/6$.

Figure 5: (2D moments) The time evolution for the absolute error of the momentum flow P_{12}. Left: the fast Fourier method. Right: the fast Chebyshev-0 method.
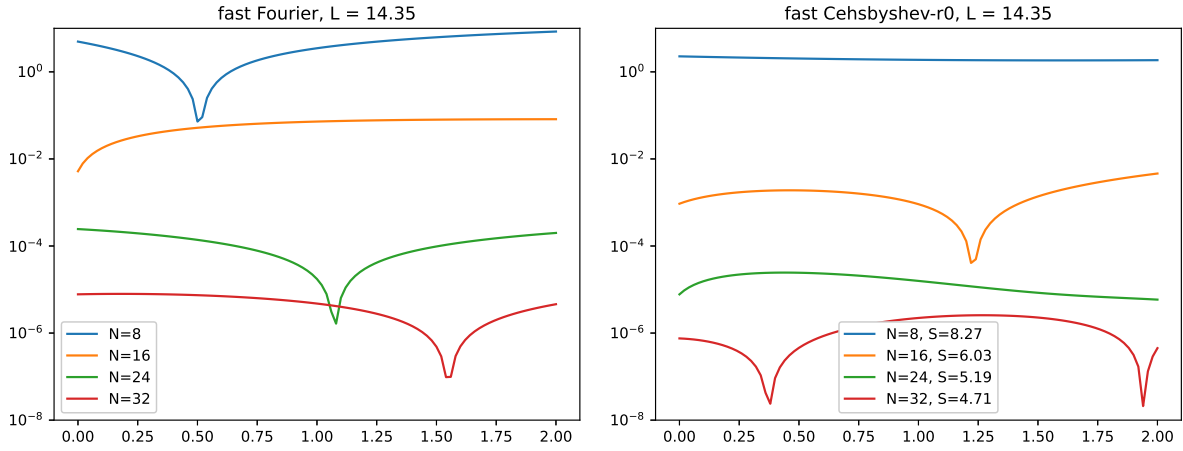
Figure 6: (2D moments) The time evolution for the absolute error of the momentum flow P_{22}. Left: the fast Fourier method. Right: the fast Chebyshev-0 method.
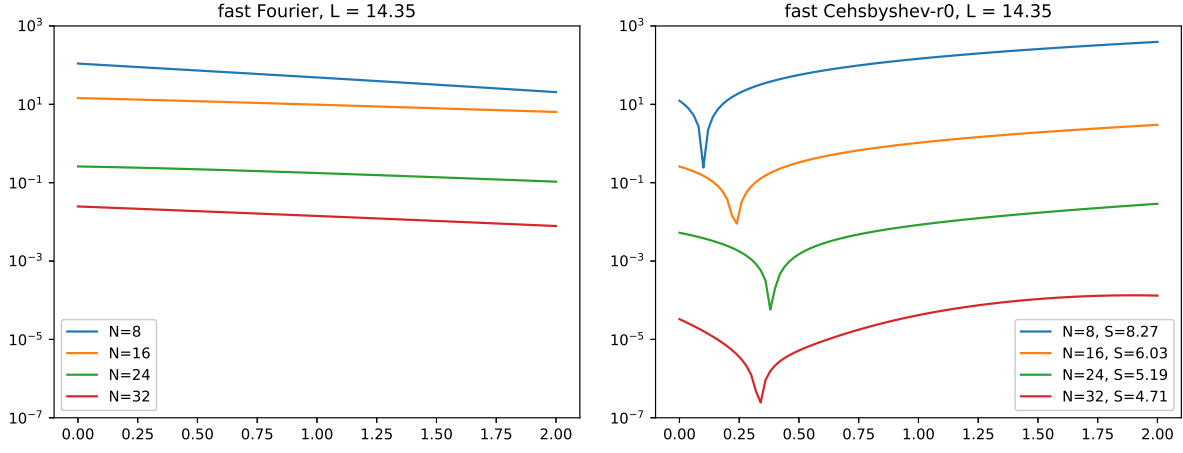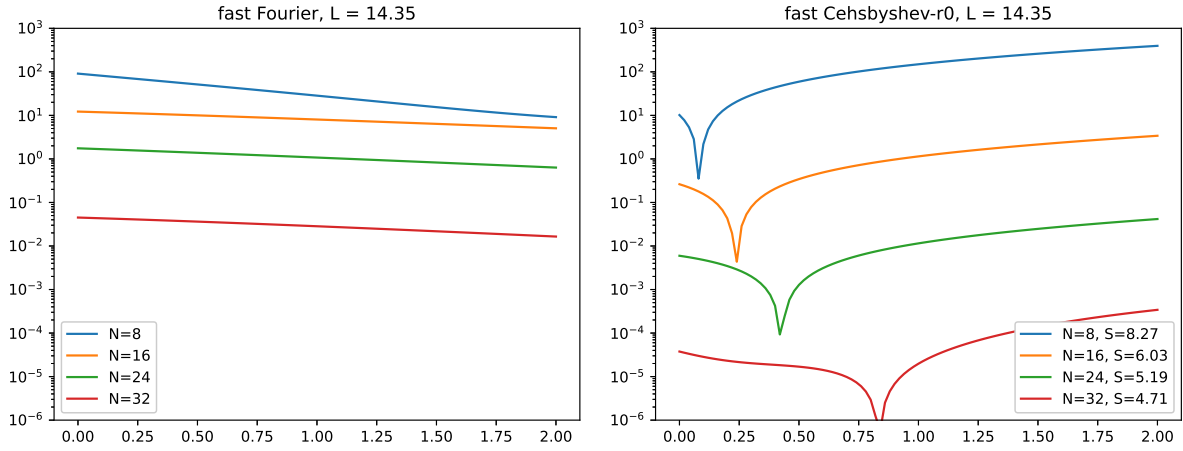
Figure 7: (2D moments) The time evolution for the absolute error of the momentum flow q_1. Left: the fast Fourier method. Right: the fast Chebyshev-0 method.

Figure 8: (2D moments) The time evolution for the absolute error of the momentum flow q_2. Left: the fast Fourier method. Right: the fast Chebyshev-0 method.

Here we again compare the fast Fourier spectral method with the fast Chebyshev-0 method. In the former, we take domain $L = 6.62$, $N_\rho = N$ quadrature points in the radial direction and $M_{\boldsymbol{\sigma}} = 38$ Lebedev quadrature points on the unit sphere. In the latter, we choose S adaptively based on L, $M_{\boldsymbol{v}} = N + 2$ quadrature points for each dimension of $(\boldsymbol{v}, \boldsymbol{v}_*)$ and $M_{\boldsymbol{\sigma}}$ Lebedev quadrature points on the unit sphere. The precision in NUFFT is selected as $\epsilon = 1e-14$. The L^∞ error of $Q_{\mathrm{BKW}}(f)$ is estimated on a $30 \times 30 \times 30$ uniform grid in the rectangular domain $[-6.3, 6.3]^3$ at time $t = 6.5$.

	fast Fourier ($M_{\boldsymbol{\sigma}} = 38$)	fast Chebyshev-0
$N = 12$	2.36e-03	1.61e-02 ($M_{\boldsymbol{\sigma}} = 14$)
$N = 16$	4.37e-04	2.72e-03 ($M_{\boldsymbol{\sigma}} = 38$)
$N = 20$	3.62e-05	3.08e-06 ($M_{\boldsymbol{\sigma}} = 86$)
$N = 24$	3.61e-06	3.10e-08 ($M_{\boldsymbol{\sigma}} = 146$)
$N = 28$	1.64e-07	1.58e-08 ($M_{\boldsymbol{\sigma}} = 170$)
$N = 32$	3.82e-08	7.16e-10 ($M_{\boldsymbol{\sigma}} = 230$)

Table 4: (3D BKW) The L^∞ error of $Q_{\mathrm{BKW}}(f)$ at time $t = 6.5$.

The results are reported in Table 4. Unlike the Fourier method for which $M_{\boldsymbol{\sigma}} = 38$ is enough (we have tested that larger value of $M_{\boldsymbol{\sigma}}$ would not further increase the accuracy), we observe that more quadrature points on the sphere are needed to get the best accuracy in the Chebyshev method. As soon as $N \geq 20$, the Chebyshev method can always obtain better accuracy than the Fourier method.

6 Conclusion

We introduced a Petrov-Galerkin spectral method for the spatially homogeneous Boltzmann equation in multi-dimensions. The mapped Chebyshev functions in $\mathbb{R}^d$ were carefully chosen to serve as the trial functions and test functions in the approximation. In the case of the algebraic mapping, we established a consistency result for approximation of the collision operator as well as the conservation property for the moments. Thanks to the close relation between the Chebyshev functions and the Fourier cosine series, we proposed a fast algorithm to alleviate the memory constraint in the precomputation and accelerate the online computation in the direct implementation. Through a series of numerical examples in 2D and 3D, we demonstrated that the proposed method can provide better accuracy (at least one or two digits for small N) in comparison to the popular Fourier spectral method.

References

[1] ALONSO, R., GAMBA, I., AND THARKABHUSHANAM, S. Convergence and error estimates for the Lagrangian-based conservative spectral method for Boltzmann equations. *SIAM J. Numer. Anal. 56* (2018), 3534–3579.

[2] BARNETT, A. H., MAGLAND, J., AND AF KLINTEBERG, L. A parallel nonuniform fast Fourier transform library based on an "exponential of semicircle" kernel. *SIAM Journal on Scientific Computing 41*, 5 (2019), C479–C504.

[3] BIRD, G. A. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows.* Clarendon Press, Oxford, 1994.

[4] BIRD, G. A., AND BRADY, J. *Molecular gas dynamics and the direct simulation of gas flows*, vol. 42. Clarendon press Oxford, 1994.

[5] BIRDSALL, C. K., AND LANGDON, A. B. *Plama Physics via Computer Simulation.* CRC Press, 2018.

[6] CERCIGNANI, C. *The Boltzmann Equation and Its Applications.* Springer-Verlag, New York, 1988.

[7] CERCIGNANI, C. *Rarefied Gas Dynamics: From Basic Concepts to Actual Calculations.* Cambridge University Press, Cambridge, 2000.

[8] CHANDRASEKHAR, S. *Radiative Transfer.* Dover Publications, 1960.

[9] DIMARCO, G., AND PARESCHI, L. Numerical methods for kinetic equations. *Acta Numer. 23* (2014), 369–520.

[10] FILBET, F., AND MOUHOT, C. Analysis of spectral methods for the homogeneous Boltzmann equation. *Trans. Amer. Math. Soc. 363* (2011), 1947–1980.

[11] FONN, E., GROHS, P., AND HIPTMAIR, R. Polar spectral scheme for the spatially homogeneous Boltzmann equation. *Research Report 2014* (2014).

[12] GAMBA, I., HAACK, J., HAUCK, C., AND HU, J. A fast spectral method for the Boltzmann collision operator with general collision kernels. *SIAM J. Sci. Comput. 39* (2017), B658–B674.

[13] GAMBA, I., AND THARKABHUSHANAM, S. Spectral-Lagrangian methods for collisional models of non-equilibrium statistical states. *J. Comput. Phys. 228* (2009), 2012–2036.

[14] GAMBA, I. M., HAACK, J. R., HAUCK, C. D., AND HU, J. A fast spectral method for the Boltzmann collision operator with general collision kernels. *SIAM Journal on Scientific Computing 39*, 4 (2017), B658–B674.

[15] GAMBA, I. M., AND RJASANOW, S. Galerkin–Petrov approach for the Boltzmann equation. *Journal of Computational Physics 366* (2018), 341–365.

[16] GIOVANGIGLI, V. *Multicomponent Flow Modeling.* Springer Science & Business Media, 1999.

[17] HU, J., QI, K., AND YANG, T. A new stability and convergence proof of the Fourier-Galerkin spectral method for the spatially homogeneous Boltzmann equation. *SIAM J. Numer. Anal. 59*, 2 (2021), 613–633.

[18] HU, J., SHEN, J., AND WANG, Y. A Petrov-Galerkin spectral method for the inelastic Boltzmann equation using mapped Chebyshev functions. *Kinetic & Related Models 13*, 4 (2020).

[19] HU, Z., AND CAI, Z. Burnett spectral method for high-speed rarefied gas flows. *SIAM J. Sci. Comput. 42* (2020), B1193–B1226.

[20] HU, Z., CAI, Z., AND WANG, Y. Numerical simulation of microflows using Hermite spectral methods. *SIAM J. Sci. Comput. 42* (2020), B105–B134.

[21] KITZLER, G., AND SCHÖBERL, J. A polynomial spectral method for the spatially homogeneous Boltzmann equation. *SIAM Journal on Scientific Computing 41*, 1 (2019), B27–B49.

[22] LEBEDEV, V. Quadratures on a sphere. *USSR Computational Mathematics and Mathematical Physics 16*, 2 (1976), 10–24.

[23] MOUHOT, C., AND PARESCHI, L. Fast algorithms for computing the Boltzmann collision operator. *Math. Comp. 75* (2006), 1833–1852.

[24] MOUHOT, C., AND VILLANI, C. Regularity theory for the spatially homogeneous Boltzmann equation with cut-off. *Arch. Rational Mech. Anal. 173* (2004), 169–212.

[25] NALDI, G., PARESCHI, L., AND TOSCANI, G., Eds. *Mathematical Modeling of Collective Behavior in Socio-Economic and Life Sciences.* Birkhäuser Basel, 2010.

[26] NANBU, K. Direct simulation scheme derived from the Boltzmann equation. I. Monocomponent gases. *J. Phys. Soc. Jpn. 49* (1980), 2042–2049.

[27] PARESCHI, L., AND RUSSO, G. Numerical solution of the Boltzmann equation I: spectrally accurate approximation of the collision operator. *SIAM J. Numer. Anal. 37* (2000), 1217–1245.

[28] PARESCHI, L., AND RUSSO, G. On the stability of spectral methods for the homogeneous Boltzmann equation. *Transport Theory Statist. Phys. 29* (2000), 431–447.

[29] SHEN, J., TANG, T., AND WANG, L.-L. *Spectral methods: algorithms, analysis and applications*, vol. 41. Springer Science & Business Media, 2011.

[30] SHEN, J., AND WANG, L.-L. Sparse spectral approximations of high-dimensional problems based on hyperbolic cross. *SIAM Journal on Numerical Analysis 48*, 3 (2010), 1087–1109.

[31] SHEN, J., WANG, L.-L., AND YU, H. Approximations by orthonormal mapped Chebyshev functions for higher-dimensional problems in unbounded domains. *Journal of Computational and Applied Mathematics 265* (2014), 264–275.

[32] VILLANI, C. A review of mathematical topics in collisional kinetic theory. In *Handbook of Mathematical Fluid Mechanics*, S. Friedlander and D. Serre, Eds., vol. I. North-Holland, 2002, pp. 71–305.