The Dissertation Committee for Chenglong Zhang
certifies that this is the approved version of the following dissertation:

On Study of Deterministic Conservative Solvers for The Nonlinear Boltzmann and Landau Transport Equations

Committee:

Irene M. Gamba, Supervisor

Clint Dawson

Omar Ghattas

Philip J. Morrison

Kui Ren

Philip L. Varghese

On Study of Deterministic Conservative Solvers for The Nonlinear Boltzmann and Landau Transport Equations

by

Chenglong Zhang, B.S., M.S. C.S.E.M.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

August 2014

Dedicated to my beloved family, my parents and my wife, Ying Gabrielle Li.

Acknowledgments

I wish to thank the multitudes of people who helped me. First of all, the greatest and innermost gratitude must be dedicated to my graduate advisor, Professor Irene M. Gamba, without whom it would have been impossible to complete this thesis. Her generous support, professional and personal respect, and constant encouraging smiles bring endless power. I would also wish to thank Professor Philip Morrison and Professor Philip L. Varghese on the discussion of the physical background and application aspects, which provide me a better picture on my projects. Thanks should be also given to Professor Clint Dawson and Professor Omar Ghattas, from whom I learned a lot regarding finite element methods and super-computing techniques, which helped greatly on the projects. I wish to thank Professor Kui Ren for his patient discussions on some mathematical aspects. I owe special thanks to Jeff Haack, who has provided plenty of patient instructions to me and helped me understand very well on spectral methods.

Finally, my constant and forever gratitude always goes to my beloved family, my parents and my wife Gaby for their supportive words every day.

On Study of Deterministic Conservative Solvers for The Nonlinear Boltzmann and Landau Transport Equations

Chenglong Zhang, Ph.D.
The University of Texas at Austin, 2014

Supervisor: Irene M. Gamba

The Boltzmann Transport Equation (BTE) has been the keystone of the kinetic theory, which is at the center of Statistical Mechanics bridging the gap between the atomic structures and the continuum-like behaviors. The existence of solutions has been a great mathematical challenge and still remains elusive. As a grazing limit of the Boltzmann operator, the Fokker-Planck-Landau (FPL) operator is of primary importance for collisional plasmas. We have worked on the following three different projects regarding the most important kinetic models, the BTE and the FPL Equations.

(1). A Discontinuous Galerkin Solver for Nonlinear BTE We propose a deterministic numerical solver based on Discontinuous Galerkin (DG) methods, which has been rarely studied. As the key part, the weak form of the collision operator is approximated within subspaces of piecewise polynomials. To save the tremendous computational cost with increasing order of polynomials and number of mesh nodes, as well as to resolve loss of conservations due to

domain truncation and DG approximation, the following combined procedures are applied. First, the collision operator is projected onto a subspace of basis polynomials up to first order. Then, at every time step, a conservation routine is employed to enforce the preservation of desired moments (mass, momentum and/or energy), with only linear complexity. The asymptotic error analysis shows the validity and guarantees the accuracy of these two procedures. We applied the property of "shifting symmetries" in the weight matrix, which consists in finding a minimal set of basis matrices that can exactly reconstruct the complete family of collision weight matrix. This procedure, together with showing the sparsity of the weight matrix, reduces the computation and storage of the collision matrix from $O(N^3)$ down to $O(N^2)$.

(2). Spectral Gap for Linearized Boltzmann Operator Spectral gaps provide information on the relaxation to equilibrium. This is a pioneer field currently unexplored form the computational viewpoint. This work, for the first time, provides numerical evidence on the existence of spectral gaps and corresponding approximate values. The linearized Boltzmann operator is projected onto a Discontinuous Galerkin mesh, resulting in a "collision matrix". The original spectral gap problem is then approximated by a constrained minimization problem, with objective function the Rayleigh quotient of the "collision matrix" and with constraints the conservation laws. A conservation correction then applies. We also study the convergence of the approximate Rayleigh quotient to the real spectral gap.

(3). A Conservative Scheme for Approximating Collisional Plasmas

We have developed a deterministic conservative solver for the inhomogeneous Fokker-Planck- Landau equations coupled with Poisson equations. The original problem is splitted into two subproblems: collisonless Vlasov problem and collisonal homogeneous Fokker-Planck-Landau problem. They are handled with different numerical schemes. The former is approximated using Runge-Kutta Discontinuous Galerkin (RKDG) scheme with a piecewise polynomial basis subspace covering all collision invariants; while the latter is solved by a conservative spectral method. To link the two different computing grids, a special conservation routine is also developed.

All the projects are implemented with hybrid MPI and OpenMP. Numerical results and applications are provided.

Table of Contents

List of Tables

List of Figures

Chapter 1

Background Introduction

After centuries of work, people eventually found that the study of fluid mechanics contributed in an essential way, with the work of Boltzmann and Maxwell, to the understanding of the motion of atoms. All the equations involved are undoubtedly valid, since they are just consequence of the Newton laws of mechanics either applied directly to the molecules of the fluid, or, at a more macroscopic level to elementary volumes of fluid (even if it requires some non-obvious work to go from the atomic description to the continuous one). There exists one "chain" of equations [36],

I Hamiltonian system of particles

$\Downarrow$

II Boltzmann equations

$\Downarrow$

III Euler/Navier-Stokes equations

$\Downarrow$

IV Models of turbulence

where each step is deduced from the previous one with introduction of hierarchy of equations and a process of closure which in some cases leads to the appearance of irreversibility.

The above models have their own regimes of validity. The continuum description is valid as long as the smallest significant volume in the flow contains a sufficient number of molecules to establish meaningful averages. Thus, the gas can be modeled in its macroscopic level, for near statistical equilibrium states. In such case, many flow and heat transfer problems can be described by a rather low number of partial differential equations, namely the well known Navier-Stokes equations, revealing a formal link between the macroscopic and microscopic descriptions. However, the conservation equations are not closed systems unless the shear stresses and heat flux can be expressed in terms of the other macroscopic quantities. In many cases, it fails to meet this requirement. And what's more, the Navier-Stokes equations will fail for rapidly changing processes, when gradients of the macroscopic variables become so steep that their scale length is of the same oder of *mean free path* l_0. The regimes of validity can be categorized by the the dimensionless measure *Knudsen number*, $Kn = \frac{l_0}{L_{flow}}$ (L_{flow} is the characteristic dimension of the flow), as follows [69].

- $Kn \ll 1$, i.e. $Kn \preceq 0.01$. Hydrodynamic regime; well described by the Navier-Stokes equations.

- $0.01 \preceq Kn \preceq 0.1$. The slip flow regime, where the Navier-Stokes equations can describe the flow well, but must be supplied with boundary

conditions that describe velocity slip and temperature jumps at gas-wall interfaces (rarefaction effects).

- $0.1 \preceq Kn \preceq 10$. The transition regime, where the Navier-Stokes equations fail, and the gas must be described in greater detail, e.g., by the Boltzmann equation, or by extended macroscopic models.

- $Kn \succeq 10$. Free molecular flow, where collisions between particles do not play an important role and the flow is dominated by particle-wall interactions.

The *kinetic theories of gases* arise from study of *rarefied gases*, which are outside the hydrodynamic regime, i.e. $Kn \succeq 0.01$. Rarefied gas flows play an important role in applications like aerospace design (space shuttle reentry; Figure 1.1), vacuum engineering (material processing, pumps), or, more recently, nanotechnology. For example, planetary vehicles such as the space shuttle typically operate in rarefied gas environments at the outer limits of the atmosphere. During re-entry such craft are subject to extremes of velocity and altitude, so it is important that the aerodynamic and thermal loads on the vehicle are properly characterized if the feasibility of the vehicle design is to be accurately assessed.

Mathematically such flows are described by the *Boltzmann Transport Equations* (BTE). The BTE can be used to determine how physical quantities change, such as heat energy and momentum, when a fluid is in transport, and other properties characteristic to fluids such as viscosity, thermal conductivity

also electrical conductivity (by treating the charge carriers in a material as a gas) can be derived. Its descriptive power makes it indispensable for predicting non-continuum phenomena in gases when experimental data is limited or not available. Its applications range from external aerodynamics and thruster plume flows to vacuum facilities and microscale devices.

If we have a "gas" of charged particle or charge carriers, the binary collisions are then replaced by Coulombic interactions and the collisions become grazing. This is the well-known *Fokker-Planck-Landau equation* (FPL), which is of primary importance for plasma applications. Plasma is an ionized gas that can occur and apply in many cases (99% of the visible universe!), such as semiconductors, controlled fusion (long-term sustainable energy sources, Figure 1.2; plasma sheath problem, Figure 1.3), space shuttle reentry (plasma sheath problems; Figure 1.1), etc.

Figure 1.1: Space Shuttle reentry and its glowing plasma trail (Space Shuttle Atlantis in the sky on July 21, 2011, to its final landing) [52]

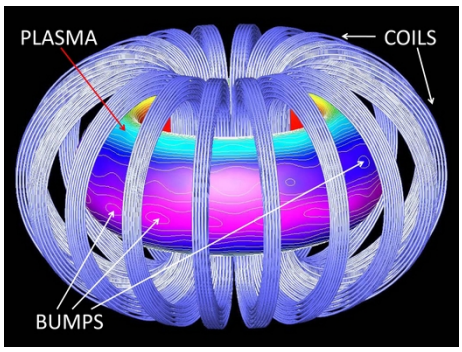


Figure 1.2: The fusion plasma [118]

Similar to molecular gas dynamics, plasmas can be also described on fluid level or kinetic level. Fluid theory on plasma is relatively simple and the approximations are accurate for majority of observed phenomena, whose distributions are assumed to be Maxwellians. However, there are still many cases when fluid description is not adequate. In such cases, the kinetic theory on plasmas will play an important role. For plasmas of sufficiently high temperatures, collisions are negligible and thus a collisionless model, e.g. Vlasov equation (with an electromagnetic force), becomes fundamental. When collisions take effects, things can be quite different. There are collisions with neutral atoms (modeled by *Krook collision terms* [40]) or collisions due to Coulomb potentials, which is the Fokker-Planck-Landau operator. Collisions can cause many different phenomena which cannot be tracked through Vlasov models, for instance wave-particle interactions (e.g. "electron trapping" and nonlinear damping), wave-wave interactions and some other nonlinear effects [40].

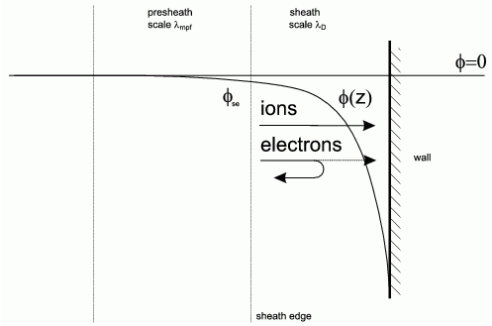


Figure 1.3: The plasma sheath [85]

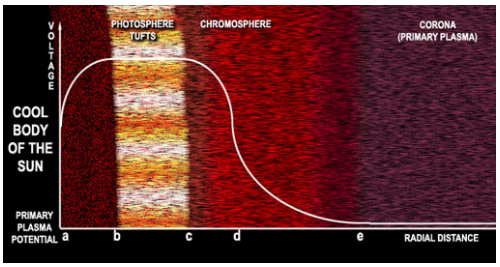


Figure 1.4: The Solar Plasma Sheath [1]

One of our interests is the "plasma sheath" problem (or, Debye sheath, electrostatic sheath, Langmuir sheath; Figure 1.3). It occurs to the plasma at the "wall" – contact surface or interface between two plasmas of different charge densities and energies, and is the transition from a plasma to a solid surface (any cold wall) or to a different plasma. When ions and electrons hit the wall, they recombine and vanish. However, electrons have much lighter mass and faster thermal speed, so more electrons will fly out of the plasma, charging the wall negative relative to the bulk plasma, and leave the plasma with a potential positive with respect to the wall. As the potential increases, more and more electrons are reflected by the sheath potential. An equilibrium is finally reached, and a layer is formed. Due to Debye shielding, the thickness of the layer will be several Debye length. Study on such phenomena has a wide range of importance in, for example, seeking for materials which can cope with the very demanding requirements of a fusion device or understanding how the Sun's plasma sheath (Figure 1.4) modulate the solar current (The solar wind disappearance event on May 10-12, 1999 [101]).

Still, many subjects about such nonlinear effects in collisional plasmas remain to be well understood.

The outline for the rest of the dissertation is organized as follows.

- Chapter 2 provides some basics on the Boltzmann equations;
- Chapter 3 describes our conservative DG solver for the Boltzmann equation;

- Chapter 4 goes to the computation of spectral gaps for the linearized Boltzmann operator;
- Chapter 5 contributes to our conservative solver for the collisional plasma model, i.e inhomogeneous FPL equations coupled with Poisson equation.
- Chapter 6 summarizes the significance and originality of the thesis and also plans for future work.

Chapter 2

Preliminaries On Boltzmann Equations

The BTE is an integro-differential transport equation, with the solution a phase probability density distribution $f(x,v,t) \in \Omega_x \times \mathbb{R}^{d_v} \times \mathbb{R}^+$ (where $\Omega_x \subseteq \mathbb{R}^{d_x}$) measuring the likelihood to find molecules at a location x with molecular velocities v at a given time t. The classical BTE models interactions or collisions through a bilinear collision operator, where the collisional kernel models the intermolecular potentials and angular scattering mechanisms known as the angular cross-section. These intramolecular potentials model from hard spheres to soft potentials up to Coulombic interactions (important for plasma collisional modeling). The scattering angular function models the anisotropic nature of the interactions. The angular cross sections could be integrable (Grad cutoff kernels) or non-integrable (Grad non-cutoff kernels).

The BTE with initial boundary values reads

$$\begin{aligned} \frac{\partial f(x,v,t)}{\partial t} + v \cdot \nabla_x f(x,v,t) + F(x,t) \cdot \nabla_v f(x,v,t) &= Q(f,f)(x,v,t) \quad (2.1) \\ f(x,v,0) &= f_0(x,v)\,, \\ f(x,v,t) &= f_B(x,v,t) \quad x \in \partial\Omega_x\,, \end{aligned}$$

where the right hand side, the bilinear integral collisional operator, can

be defined weakly or strongly. The strong form goes

$$Q(f,f) = \int_{v_*\in\mathbb{R}^d,\sigma\in S^{d-1}} [\frac{1}{'e\,'J}\,'f'f_* - ff_*]\,B(|v-v_*|,\sigma)\,d\sigma dv_*\,, \qquad (2.2)$$

where, for simplicity, here and in the following, denote $'f = f('v)$, $f_* = f(v_*)$ and $'f_* = f('v_*)$ (the right prime means dependency on post-collisional velocity v', that is, $f' = f(v')$, $f'_* = f(v'_*)$) and drop the dependencies on x, t. $'v$, $'v_*$ are pre-collisional velocities corresponding to v, v_*. The integration is parametrized in terms of the center of mass and relative velocity. And on the $d-1$ dimensional sphere, integration is done with respect to the unit direction given by the elastic post collisional relative velocity. The parameter $'e$ (depending on pre relative velocity $'u$) is the restitution coefficient covering the range from sticky ($e=0$) to elastic ($e=1$) interactions. $'J$ is the Jacobian of pre-collisional velocities w.r.t post-collisional ones.

The pre- and post-collisional velocities obey

$$u = v - v_*, \quad v' = v + \frac{\beta}{2}(|u|\sigma - u), \quad v'_* = v_* - \frac{\beta}{2}(|u|\sigma - u), \quad \beta = \frac{1+e}{2} \quad (2.3)$$

Here is the key for the model, the *collision kernel*

$$B(|u|,\sigma) = |u|^\gamma b(\cos(\theta)), \quad \gamma \in (-d, +\infty)\,, \qquad (2.4)$$

with *angular cross-sections*

$$\cos(\theta) = \frac{u\cdot\sigma}{|u|}\,, \quad b(\cos(\theta)) \sim \sin^{-(d-1)-\alpha}(\frac{\theta}{2}) \text{ as } \theta \sim 0\,, \quad \alpha \in (-\infty, 2) \quad (2.5)$$

Without loss of generality, we can assume

$$b(\cos(\theta)) = \frac{1}{2^{d-1}\pi}\sin^{-(d-1)-\alpha}(\frac{\theta}{2})\,. \qquad (2.6)$$

The regularity parameters γ and α actually corresponds to different types of interactions and different power-law molecular potentials. For interaction potentials obeying spherical repulsive laws

$$\phi(r) = r^{-(s-1)}, \quad s \in [2, +\infty)\,,$$

the collision kernel and angular cross-section are explicit for $d = 3$, that is, $\gamma = (s-5)/(s-1)$ and $\alpha = 2/(s-1)$ (see [36]). As a convention, $-d < \gamma < 0$ defines Soft Potentials, $\gamma = 0$ is the Maxwell Molecules type interaction, $0 < \gamma < 1$ describes Variable Hard Potentials and $\gamma = 1$ is the classical Hard Sphere model. Also, the angular cross-sections can be of short range or long range, that is, $b(\cos(\theta))$ can be integrable for $\alpha < 0$ and non-integrable when $\alpha \geq 0$. When $\alpha = 2$, together with $\gamma = -3$, models the grazing collisions under Coulombian potentials, which deduces to the Fokker-Planck-Landau equations (shown later).

The weak form for (2.2), or called Maxwell form, after a change of variable $u = v - v_*$ is given by

$$\int_{\mathbb{R}^d} Q(f,f)(v)\phi(v)dv = \int_{v,u\in\mathbb{R}^d} f(v)f(v-u)\int_{\sigma\in S^{d-1}} [\phi(v')-\phi(v)]B(|u|,\sigma)d\sigma du dv\,, \tag{2.7}$$

which is a *double mixing convolution*. Such a structure will be the base for future design of solvers.

Remark. As is proposed recently by I.M. Gamba, such double mixing convolution structures (with various quantified state transition probabilities B) are pretty universal for kinetic evolutions of non-equilibrium systems of

birth-death dynamics, including Boltzmann equations, Landau equations (introduced in Chapter 5), etc. Such evolutions can be all derived from the *Kac Master equation*, which represents a *Markov process*.

For the time being, in the whole report, we only consider elastic collisions as examples, i.e. $\beta = 1$ in (2.3).

In spite of its complicated form, $Q(f, f)$ enjoys many interesting and remarkable properties. Among them, the followings are most fundamental and important [36].

2.1 Collision Invariants and Conservation Laws

By symmetry of (2.7), one can find that

$$\int_{\mathbb{R}^d} Q(f,f)(v)\phi(v)dv = \frac{1}{2}\int_{\mathbb{R}^{2d}}\int_{S^{d-1}} ff_*[\phi+\phi_*-\phi'-\phi'_*]B(|v-v_*|,\sigma)d\sigma dv_* dv\,. \tag{2.8}$$

So, one can see (2.8) is identical to zero if

$$\phi + \phi_* = \phi' + \phi'_*\,. \tag{2.9}$$

Functions ϕ satisfying (2.9) are called "*collision invariants*". The right-hand side of (2.8) is the average change of $\phi(v)$ in unit time due to collisions. The Boltzmann theorem [38] tells us that, (2.9) holds with v and v_* satisfying (2.3) if and only if ϕ is given by

$$\phi(v) = a + b\cdot v + c|v|^2\,, \tag{2.10}$$

of which, $\phi(v) = 1, v, |v|^2$ are the $d+2$ elementary collision invariants, which correspond to the conservation of mass, momentum and kinetic energy.

2.2 Entropy Dissipation

For any $f > 0$, take $\phi = \log f$, then from the weak form (2.8)

$$\int_{\mathbb{R}^d} \log f Q(f,f) dv = \frac{1}{4} \int_{\mathbb{R}^{2d}} \int_{S^{d-1}} (f f_* - f' f'_*) \log(f' f'_* / f f_*) B(|v - v_*|, \sigma) d\sigma dv_* dv \,. \tag{2.11}$$

We know $(x-y)\log(x/y) \geq 0$ and equality holds if and only if $x = y$; and notice that $B(|v - v_*|, \sigma) \geq 0$ ('=' holds only at $\theta = 0$), so,

$$\int_{\mathbb{R}^d} \log f Q(f,f) dv \leq 0 \,, \tag{2.12}$$

and equality holds iff

$$f' f'_* = f f_* \tag{2.13}$$

holds almost everywhere. Taking the logarithm of both sides gives $\log f' + \log f'_* = \log f + \log f_*$. Similar with what we did for (2.9), it's provable that such f's exist and are given by

$$f(v) = \exp(a + b \cdot v + c|v|^2) \,. \tag{2.14}$$

In order to make f integrable over the whole velocity space, c must be negative, which gives

$$f(v) = A \exp(-\alpha(v - \xi)^2) \,, \tag{2.15}$$

where $\alpha > 0$ and ξ is a constant vector. This is called *Maxwellian distribution*, usually written by

$$M(v) = \frac{\rho}{(2\pi T)^{\frac{d}{2}}} \exp(-\frac{|v - \bar{v}|^2}{2T}) \,, \tag{2.16}$$

where ρ is the macroscopic density, $\bar{v}$ the macroscopic velocity and T the macroscopic temperature ($= R\vartheta$ where ϑ is the absolute temperature, R is a gas constant).

Since the entropy dissipation rate $\int_{\mathbb{R}^d} \log f Q(f, f) dv$ is nonpositive, then the Boltzmann inequality (2.12) holds for all positive f, and equality holds if and only if f is a Maxwellian distribution, which is a solution to a vanishing collision integral $Q(f, f) = 0$.

2.3 The H-theorem

If we define

$$\mathscr{H} = \int_{\mathbb{R}^d} f \log f dv \,, \tag{2.17}$$

and for the i-th velocity component,

$$\mathscr{H}_i = \int_{\mathbb{R}^d} v_i f \log f dv \,, \tag{2.18}$$

where f is any function satisfying the Boltzmann equation

$$\partial_t f + v \cdot \nabla_x f + F \cdot \nabla_v f = Q(f, f) \,, \tag{2.19}$$

where, for the sake of generality, the (velocity-independent) body force term $\mathbf{F}$ which is usually left out is introduced here. Then, multiplying by $1 + \log f$

on both sides of the Boltzmann equation, integrating over the whole velocity domain, noticing $\nabla(f\log f) = (1+\log f)\nabla f$ and the special collision invariant "1", we obtain

$$\frac{\partial \mathscr{H}}{\partial t} + \nabla_x \cdot \mathbf{H} = \int_{\mathbb{R}^d} \log f Q(f,f) dv \leq 0\,, \tag{2.20}$$

where $\mathbf{H} = (\mathscr{H}_1, \mathscr{H}_2, \mathscr{H}_3)$. Now, we introduce a quantity

$$H = \int_{\mathbb{R}^d} \mathscr{H} dx\,. \tag{2.21}$$

In the case of Maxwellian distribution, or $Q(f,f) = 0$, H is conservative, since we can treat $\mathscr{H}$ as "density" and then $\mathbf{H}$ will be the corresponding flow of H (now H is like the "mass"; $\mathbf{H} = \mathscr{H}\mathbf{u}$, $\frac{\partial \mathscr{H}}{\partial t} + \nabla\cdot(\mathscr{H}\mathbf{u}) = 0$)(here "macroscopic velocity" $\mathbf{u}$ is just the corresponding interpretation, not the true one), or we just take the total time derivative of H and notice the vanishing collision operator. In general case (2.20), however, we can say molecular collisions act as a negative source for the quantity H.

We now split $\mathbf{H}$ into a macroscopic (convective) flow of H, $\mathscr{H}\mathbf{u}$ and a microscopic flow of H, $\mathbf{H} - \mathscr{H}\mathbf{u}$.

We integrate both sides of (2.20) w.r.t x, if the boundary $\partial\Omega_x$ of the whole integration region Ω_x moves with velocity $\mathbf{u}_0$, then we get

$$\frac{dH}{dt} - \int_{\partial\Omega_x} (\mathbf{H} - \mathscr{H}\mathbf{u}_0)\cdot \mathbf{n} dS = \int_{\Omega_x}\int_{\mathbb{R}^d} Q(f,f)\log f dv dx \leq 0\,, \tag{2.22}$$

where $\mathbf{n}$ is the inward normal.

In conclusion, we get two classical forms of the celebrated H-theorem of Boltzmann equation:

1. If the gas is homogeneous ($\partial f/\partial x = 0$ and hence $\nabla_x \cdot \mathbf{H}$ in (2.20) is zero), the quantity $\mathscr{H}$ never increases with time and is steady iff the distribution is Maxwellian.

2. If the gas is enclosed in a region such that (e.g. molecules are specularly reflected at the boundary)

$$\int_{\partial\Omega} (\mathbf{H} - \mathscr{H}\mathbf{u}_0) \cdot \mathbf{n} dS \leq 0\,, \tag{2.23}$$

   then the quantity H never increases with time and is steady iff the distribution is Maxwellian.

The reason why Boltzmann equation is of basic importance is that it shows the Boltzmann equation has a basic feature of *irreversibility*: the quantity H always decreases even when it's not released to the surroundings when no energy exchange takes place between gas and surroundings.

At the end, we want to mention that the decreasing of H in the absence of energy change with the surroundings shows that the Boltzmann equation describes an evolution towards a state of minimum H, provided no additional H flows in from the exterior. The final state ($t \to \infty$) will probably be a steady state provided such a state is compatible with boundary conditions and is stable. More particular than steady state is the equilibrium state defined as the steady state with energy exchange with the surroundings. We mention that *the distribution function must be Maxwellian in an equilibrium*; and no steady solutions of the Boltzmann equation exist when the gas is bounded by

specularly reflecting walls except for Maxwellians. Such a propety is quite basic but of great importance for justifying any numerical solutions. This will be one of the benchmark tests for our numerical sovers.

2.4 Macroscopic Fluid Dynamic Limits

Now we consider the problem of evaluating the macroscopic quantities with the distribution function given so that we hope to get a picture of how the microscopic description and macroscopic description of gas dynamics are connected.

The macroscopic limits are obtained when the fluid becomes dense enough that particles undergo many collisions over the scales of interest. This situation is described by *Knudsen number* mentioned above.

It's easy to understand the definitions of mass density and mean velocity or flow velocity

$$\rho(x,t) = \int_{v\in\mathbb{R}^d} f dv\, \rho(x,t)\bar{V}(x,t) = \int_{v\in\mathbb{R}^d} v f dv\,, \tag{2.24}$$

which can be directly observed.

In fact, each molecule has its own velocity v that can be decomposed into the sum of macroscopic velocity u and a random term ξ (or called peculiar velocity) which describes the random deviation of the molecular velocity from the ordered motion with a velocity u:

$$v = u + \xi\,. \tag{2.25}$$

Obviously, the random velocity coincides with molecular velocity when the gas is macroscopically at rest. More, the random velocity satisfies

$$\int_{\mathbb{R}^d} \xi f dv = 0\,. \tag{2.26}$$

The "momentum density" $\rho(x,t)\bar{V}(x,t)$ actually can be interpreted as a mass flow. Simiarly, we can define "momentum flow". Since momentum is a vector, we have to consider the flow of the j-th component of momentum in the i-th direction:

$$\int_{\mathbb{R}^d} v_i(v_j f)dv = \int_{\mathbb{R}^d} v_i v_j f dv\,. \tag{2.27}$$

It shows that the momentum flow is described by a symmetric tensor of second order. In order to find out how it will appear in a macroscopic description, following the idea in (2.25)

$$\int_{\mathbb{R}^d} v_i v_j f dv = \rho u_i u_j + \int_{\mathbb{R}^d} \xi_i \xi_j f dv\,. \tag{2.28}$$

Now, we get two parts: the first is the macroscopic momentum flow (momentum density times velocity); while the second is a hidden momentum flow due to the random motion of the molecules. To understand the latter term, we assume no external body force and take a fixed region of the gas and observe the change of momentum inside it. We find that, the change can be only partially attributed to the matter that enters and leaves the region, leaving the other part of the change which has no macroscopic explanation, unless we attribute it to the action of a force exerted on the boundary of the region of interest by the neighboring regions of the gas. That means, the integral of $\int_{\mathbb{R}^d} \xi_i \xi_j f dv$

now contributes to the stress tensor. We write

$$p_{ij} = \int_{\mathbb{R}^d} \xi_i \xi_j f dv \,, \tag{2.29}$$

which, in the macroscopic equations derived from the Boltzmann equations, plays the same role as the stress tensor in the conservation equations derived from macroscopic considerations.

Similarly, we can also define "energy density" and "energy flow", which gives rise to the macroscopic heat flow.

We conclude and list macroscopic quantities or informations of interest, which are contained in the single distribution function f, for $d = 3$,

$$\begin{aligned}
&\text{Density} \quad \rho(x,t) = \int_{v\in\mathbb{R}^d} f(x,v,t)dv \\
&\text{Flow velocity vector} \quad \bar{V}(x,t) = \frac{1}{\rho(x,t)} \int_{v\in\mathbb{R}^d} v f(x,v,t)dv \\
&\text{Temperature} \quad T(x,t) = \frac{1}{3\rho(x,t)} \int_{v\in\mathbb{R}^d} |v-\bar{V}|^2 f(x,v,t)dv \\
&\text{Pressure} \quad p(x,t) = \frac{1}{3} \int_{v\in\mathbb{R}^d} |v-\bar{V}|^2 f(x,v,t)dv = \rho(x,t)T \\
&\text{Specific internal energy} \quad e(x,t) = \frac{1}{2\rho(x,t)} \int_{v\in\mathbb{R}^d} |v-\bar{V}|^2 f(x,v,t)dv = \frac{3}{2}T \\
&\text{Stress tensor, } \mathbf{p}(\text{x, t}) = \{p_{ij}\}(x,t) \quad p_{ij} = \int_{v\in\mathbb{R}^d} (v_i - V_i)(v_j - V_j) f(x,v,t)dv \\
&\text{Heat-flow vector} \quad \mathbf{q}(x,t) = \frac{1}{2} \int_{v\in\mathbb{R}^d} (v-\bar{V})|v-\bar{V}|^2 f(x,v,t)dv
\end{aligned} \tag{2.30}$$

Then, we multiply the 5 elementary collision invariants on both sides of the BTE (2.1), and noticing the above definitions, we get

$$\begin{aligned} \frac{\partial \rho}{\partial t} + \frac{\partial}{\partial x_j}(\rho u_j) &= 0 \\ \frac{\partial}{\partial t}(\rho u_i) + \frac{\partial}{\partial x_j}(\rho u_i u_j + p_{ij}) &= \rho F_i \\ \frac{\partial}{\partial t}[\rho(\frac{1}{2}u^2 + e)] + \frac{\partial}{\partial x_j}[\rho u_j(\frac{1}{2}u^2 + e) + p_{ji}u_i + q_j] &= \rho F_j u_j \end{aligned} \tag{2.31}$$

which are the basic equations of continuum mechanics, particularly of macroscopic gas dynamics and physically interpreted as conservation of mass, momentum and energy.

However, they are not closed unless the so-called "constitutive equations" are introduced. In the case of gas, or more generally, a fluid, two models are well known:

1. Euler (or ideal) fluid:

$$p_{ij} = p\delta_{ij}; \qquad q_i = 0\,. \tag{2.32}$$

2. Navier-Stokes-Fourier (or viscous and thermally conducting) fluid:

$$p_{ij} = p\delta_{ij} - \mu(\frac{\partial u_i}{\partial u_j} + \frac{\partial u_j}{\partial u_i}) - \lambda\frac{\partial u_k}{\partial x_k}\delta_{ij}; \qquad q_i = -\kappa\frac{\partial T}{\partial x_i}\,, \tag{2.33}$$

where μ and λ are the viscosity coefficients, and κ is the heat conducting coefficient.

Chapter 3

A Conservative Solver for Boltzmann Equation Based On Discontinuous Galerkin Scheme

In this chapter, we describe our deterministic numerical solver based on Discontinuous Galerkin (DG) methods, which has been rarely studied.

3.1 Introduction

The BTE is of primary importance in rarefied gas dynamics. For a gas flow, when the Knudsen number is far less than unity, the bulk quantities can be deduced from the microscopic level and it's enough to work under the hydrodynamic regime. However, when the Knudsen number is of order unity, classical macroscopic models, the Navier-Stokes equations for example, fail to correctly capture the macroscopic quantities. In such cases, a kinematics approach based on the Boltzmann equation modeling rarefied gases rapidly dominates.

The existence of solutions has been a great mathematical challenge and still remains elusive. That makes the numerical approximation to solutions a very challenging problem. Albeit these, solving the BTE and studying the

evolution properties are among the most fundamental problems in fluid dynamics. Extensive efforts have been put onto the numerical treatment of BTE and other kinetic equations. The main challenges include, but not limited to, the high dimensionality in the collision operator and revealing the collision mechanism through suitable formulating.

Basically, there are a few classes of computational methods for solving the BTE. One of them is the well-known *Direct Simulation Monte Carlo* (DSMC) method, which was developed initially by Bird [11] and Nanbu [100] and more recently by [109, 110]. Currently, there is extensive work from Rjasanow and Wagner [110] and references therein, to determine accurately the high-velocity tail behavior of the distribution functions from DSMC data. DSMC developed to calculate statistical moments under near stationary regimes, but are not efficient to capture transients as well as details of the solution $f(x, v, t)$. In addition these methods inherit statistical fluctuations that become a bottleneck in the presence of non-stationary flows or close to continuum regimes.

During the last decade, deterministic methods, such as Discrete Velocity or Spectral Methods, have been attracting more attention. Discrete velocity models were developed by Broadwell [28] and mathematically studied by Cabannes, Illner and Kawashima among many authors [30, 84, 88]. More recently these models have been studied for many other applications on kinetic elastic theory in [17, 37, 80, 93–96, 123]. Spectral methods, which have been originally developed by Pareschi, Gabetta and Toscani [64], and later

by Bobylev and Rjasanow[25] and Pareschi and Russo[106], are supported by the ground breaking work of Bobylev[14] using the Fourier Transformed Boltzmann Equation to analyze its solutions in the case of Maxwell type of interactions. More recent implementations of spectral methods for the non-linear Boltzmann equation are due to Bobylev and Rjasanow [27], who developed a method using the Fast Fourier Transform (FFT) for Maxwell type interactions, and then for Hard-Sphere interactions [26] using generalized Radon and X-ray transforms via FFT. Simultaneously, L. Pareschi and B. Perthame [105] developed a similar scheme using FFT for Maxwell type interactions. Later, I. Ibragimov and S. Rjasanow [83] developed a numerical method to solve the space homogeneous Boltzmann Equation on a uniform grid for variable hard potential (VHP) interactions with elastic collisions. We mention that, most recently, Filbet and Russo[59, 61] implemented a method to solve the space inhomogeneous Boltzmann equation using the previously developed spectral methods in [105, 106]. Conservative Lagrangian-Spectral Method, which uses Fourier Transform as the main tool, was introduced by I.Gamba and Sri H.Tharkabhushanam [68, 69], and more recently was extended to anisotropic interactions by I. Gamba and J. Haack [86, 87]. It has the capability of approximating solutions to elastic and inelastic collisional models, for both isotropic and anisotropic non-cutoff angular cross-sections. For other deterministic schemes, we suggest refer to [7].

While the behavior of the spectral methods may rely on the smoothness of the underlying solutions, in order to capture more irregular features,

the discontinuous Galerkin (DG) [51] method may be more appropriate, due to its locality and flexibility. It is a finite element method using discontinuous piecewise polynomials as basis functions and numerical fluxes based on upwinding for stability. Please refer to [51] for more details. For problems of charge transport in semiconductor devices, DG methods are very promising and have provided accurate results at a comparable computational cost [45–48, 50]. It seems, DG could be a potential method for kinetic equations. However, there are barely any previous work on full nonlinear BTE. To our best knowledge, one try might be [82], which is only dealing with 1D prototype of BTE. Most recently, A. Majorana [91] published a work on a DG-based BTE solver. He derived a set of partial differential equations on t, x by a partial application of the DG method, which is only on variable v. The collision invariants are used as basis to guarantee the conservations laws. However, it's unclear how the collisional integrals are evaluated and this evaluation actually requires $O(N^3)$ operations. Also very few and limited numerical results were provided. Another most recent work comes from A. Alekseenko et al [2]. Our scheme was developed independently and is different than the one in [2], in the way of constructing basis functions, evaluating angular cross-section integrals and the enforcing of conservation routines. In addition, we are able to provide asymptotic error analysis.

3.2 The Discontinuous Galerkin Projections and Evaluations of the Collision Integrals

We are working in the velocity domain $v \in \mathbb{R}^d$. The general theory of homogeneous Boltzmann equation tells us [22, 67] that if Ω_v is a sufficiently large velocity domain such that the initial state f_0 enjoys most of its mass and energy inside it, then Ω_v (or one of comparable size) will also contain most of the mass and energy of the solution f for any given time $t > 0$. For example, if the initial state $f_0 \in L^1_{e^{a|v|^2}}(\mathbb{R}^d)$, then $f(v,t) \in L^1_{e^{b|v|^2}}(\mathbb{R}^d)$ for some positive constant $b \leq a$ [4]. Thus, it's reasonable to assume a compact support for the solution and truncate the whole velocity domain to finite $\Omega_v = [-L, L)^d$.

A regular mesh is applied, that is, we divide each direction into n disjoint elements uniformly, such that $[-L, L] = \bigcup_k I_k$, where interval $I_k = [w_{k-\frac{1}{2}}, w_{k+\frac{1}{2}})$, $w_k = -L + (k + \frac{1}{2})\Delta v$, $\Delta v = \frac{2L}{n}$, $k = 0 \ldots n-1$ and thus there is a Cartesian partitioning $\mathcal{T}_h = \bigcup_k E_k$, with uniform cubic element $E_k = I_{k_1} \otimes I_{k_2} ... \otimes I_{k_d}$, $k = (k_1, k_2, ..., k_d)$.

Discontinuous Galerkin methods assume piecewisely defined basis functions, that is

$$f(v,t) = \sum_k \mathbf{u}_k(t) \cdot \Phi(v)\chi_k(v), \tag{3.1}$$

where multi-index $k = (k_1, k_2, ..., k_d)$, $0 \leq |k| < (n-1)^3$; $\chi_k(v)$ is the characteristic function over element E_k; coefficient vector $\mathbf{u}_k = (\mathbf{u}_k^0, ..., \mathbf{u}_k^p)$, where p is the total number of basis functions locally defined on E_k; basis vector $\Phi(v) = (\phi_0(v), ..., \phi_p(v))$. Usually, we choose element of basis vector $\Phi(v)$

as local polynomial in $P^p(E_k)$, which is the set of polynomials of total degree at most p on E_k. For sake of convenience, we select the basis such that $\{\phi_i(v) : i = 0, ..., p\}$ are orthogonal. For example, when $d = 3$, $p = 1$, local linear basis over element E_k can be set as

$$\{1, \frac{v_1 - w_{k_1}}{\Delta v}, \frac{v_2 - w_{k_2}}{\Delta v}, \frac{v_3 - w_{k_3}}{\Delta v}\}. \tag{3.2}$$

One fact should be addressed that, no matter what types of numerical schemes for Boltzmann-type equations, the treatment of the collision operator always remains the most important and challenging part. The remaining left hand side, i.e. advection part of (2.1) would just follow the standard way of DG-FEM. So, first of all, let's consider the homogeneous problem

$$\frac{\partial f(v,t)}{\partial t} = Q(f,f)(v,t), \tag{3.3}$$

and focus on the weak form (2.7).

Apply the i-th basis function on element E_m, $\phi_i(v)\chi_k(v)$, to (2.7) and operate a change of variables $(v, u) \leftarrow (v, v_*)$, where $u = v - v_*$ is the relative velocity,

$$\begin{aligned}
&\int_{v\in E_m} Q(f,f)\phi_i(v)dv \\
&= \int_{v\in E_m, v_*\in\mathbb{R}^d} f(v)f(v-u)\int_{\sigma\in\mathbb{S}^{d-1}} [\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)]|u|^\gamma b(\frac{u\cdot\sigma}{|u|})d\sigma du dv \\
&= \sum_k \sum_{\bar{k}} \mathbf{u}_k^T \mathbf{G}_{m,i}(k,\bar{k})\mathbf{u}_{\bar{k}}\,,
\end{aligned} \tag{3.4}$$

where we recall that the post-collisional velocity $v' = v + \frac{1}{2}(|u|\sigma - u)$. Here, for fixed $k, \bar{k}, m, i$, the entry $\mathbf{G}_{m,i}(k,\bar{k})$ is actually a $(p+1)\times(p+1)$ matrix,

defined as

$$\mathbf{G}_{m,i}(k,\bar{k}) = \int_{v\in E_k}\int_{v-u\in E_{\bar{k}}} \Phi(v)\otimes\Phi(v-u)\chi_k(v)\chi_{\bar{k}}(v-u)|u|^\gamma \int_{\mathbb{S}^{d-1}}[\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)]b(\tfrac{u\cdot\sigma}{|u|})d\sigma du dv\,. \tag{3.5}$$

The key is to evaluate the block entry $\mathbf{G}_{m,i}(k,\bar{k})$ in (3.5). Due to the convolution formulation, the integrals w.r.t v, u can be approximated through *Triangular quadratures*. Indeed, along each dimension, if $v_i \in I_{k_i}$, $v_* \in I_{\bar{k}_i}$, then (v_i, u_i) will form a parallelegram which can be divided into two triangles. See Figure 3.1.

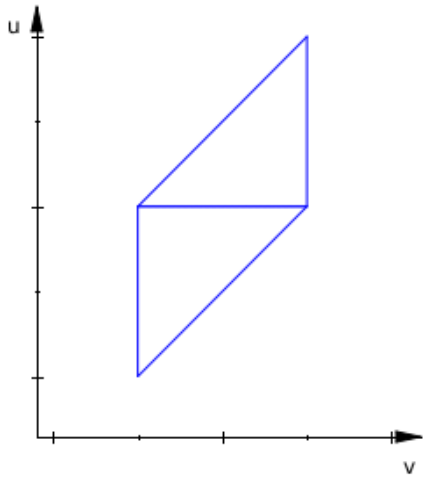


Figure 3.1: Along each dimension, (v_i, u_i) forms two right triangles

The integrals on the sphere take the most effort, because one has to figure out how the Cartesian cubes intersect with the sphere. Let's extract the angular integrals in (3.5), denoted by $g_{m,i}(v,u)$, and study it separately

$$g_{m,i}(v,u) = \int_{\mathbb{S}^{d-1}}[\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)]b(\frac{u\cdot\sigma}{|u|})d\sigma. \tag{3.6}$$

For any fixed v, u, the post-collisional velocity v' will be on the surface of a ball centered at $v - \frac{u}{2}$ with radius $\frac{|u|}{2}$. The angular cross-section $b(\cos\theta)$ itself may contain non-integrable singularity at $\theta = 0$. However, the "gain-loss" terms in the above square bracket will absorb the singularity in $b(\cos\theta)$ and make it integrable. Our scheme has to take this issue into account and design a careful way of computing.

1. Integrable $b(\cos\theta)$.

   This case allows to split the "gain" and "loss" terms. Only "gain" terms involve post-collisional velocity v' and can be studied separately.

   For $d = 2$, the angular integrals (4.70) can be evaluated analytically. Indeed, for fixed v, u, the regions over the cycle $\sigma = (\sin\theta, \cos\theta)$ such that $v' = v + 0.5(|u|\sigma - u) \in E_m$ can be exactly figured out, by solving a system of trigonometric inequalities

   $$\begin{cases} v_1 + 0.5(|u|\sin\theta - u_1) \in I_{m_1} \\ v_2 + 0.5(|u|\cos\theta - u_2) \in I_{m_2} \end{cases} \tag{3.7}$$

   We have built a programmable routine of deriving all possible overlapped intervals of θ.

   The case $d = 3$ performs similarly. We solve the following nonlinear trigonometric inequalities

   $$\begin{cases} v_1 + 0.5(|u|\sin\theta\cos\varphi - u_1) \in I_{m_1} \\ v_2 + 0.5(|u|\sin\theta\sin\varphi - u_2) \in I_{m_2} \\ v_3 + 0.5(|u|\cos\theta - u_3) \in I_{m_3} \end{cases} \tag{3.8}$$

The third inequality will give a range for the polar angle θ, and all integrals w.r.t θ will be performed by adaptive quadratures, say, CQUAD in GSL [58]; for any fixed θ, the first two inequalities will decide the range of azimuthal angle φ exactly (by invoking the routine mentioned above).

Note: The angle θ above is NOT the scattering angle defined in (2.5).

2. Non-integrable $b(\cos\theta)$.

   Consider a local spherical coordinate system with u being the polar direction. Then, consider a transformation which rotates the polar direction back onto z-axis of the Cartesian coordinate system. The orthogonal rotation matrix A can be constructed explicitly

   $d = 2$:

$$A = \frac{1}{|u|}\begin{pmatrix} -u_2 & u_1 \\ u_1 & u_2 \end{pmatrix} \tag{3.9}$$

   $d = 3$:

$$A = \frac{1}{|u|}\begin{pmatrix} \frac{u_1 u_3}{\sqrt{u_1^2+u_2^2}} & \frac{u_2 u_3}{\sqrt{u_1^2+u_2^2}} & -\sqrt{u_1^2+u_2^2} \\ -\frac{u_2|u|}{\sqrt{u_1^2+u_2^2}} & \frac{u_1|u|}{\sqrt{u_1^2+u_2^2}} & 0 \\ u_1 & u_2 & u_3 \end{pmatrix} \tag{3.10}$$

   where we assume $u_1^2 + u_2^2 \neq 0$, otherwise, the rotation matrix is reduced to the identity matrix.

   Then, consider a change of variable $\sigma \leftarrow A^{-1}\sigma = A^T\sigma$, for which the Jocobian is 1. If denote by θ the angle between u and σ , as exactly defined in (2.5), recalling post-collisional velocity $v' = v + \frac{1}{2}(|u|\sigma - u)$,

we have

$$g_{m,i}(v,u) = \int_{\mathbb{S}^{d-1}} \left[\phi_i \circ \chi_m(v+z) - \phi_i \circ \chi_m(v)\right] b(\cos\theta) d\sigma$$
$$= \int_{\mathbb{S}^{d-1}} \left[\phi_i \circ \chi_m(v - \frac{u}{2} + \frac{|u|}{2}\sigma) - \phi_i \circ \chi_m(v)\right] b(\cos\theta) d\sigma,$$

where, if $d = 2$: $z = \frac{|u|}{2} A^T \left(\sin\theta, \cos\theta - 1\right)^T$, $\sigma = A^T \left(\sin\theta, \cos\theta\right)^T$; if $d = 3$: $z = \frac{|u|}{2} A^T \left(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta - 1\right)^T$, $\sigma = A^T \left(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta\right)^T$.

We take $d = 3$ for example. The whole domain of (θ, φ), i.e. the sphere, can be divided into the following four subdomains: (1) $S_1 = [0, \theta_0] \times [0, 2\pi]$; (2) $S_2 = [\theta_0, \theta_1] \times I_\varphi(\theta)$; (3) $S_3 = [\theta_0, \theta_1] \times ([0, 2\pi] \setminus I_\varphi(\theta))$; and (4) $S_4 = [\theta_1, \pi] \times [0, 2\pi]$. Here θ_0 is determined according to the following strategy: when $v \in E_m$, $\sin\frac{\theta_0}{2} = \min(1, \frac{1}{|u|}\text{dist}(v, \partial E_m))$ by noticing that $|z| = |u| \sin\frac{\theta}{2}$; when $v \notin E_m$, θ_0 is the smallest possible θ such that v' lies in E_m. θ_1 is the largest possible θ such that v' lies in E_m. $I_\varphi(\theta)$ are effective intervals for φ, depending on θ, such that v' lies in E_m.

Due to the characteristic functions in the integrands of $g_{m,i}(v,u)$ (3.11), we have the following four cases

(a) '0-0': when $v' \notin E_m$ and $v \notin E_m$. It's trivial because it contributes nothing to the final weight matrix.

(b) '1-0': when $v' \in E_m$ but $v \notin E_m$. In this case, the effective domain (where $g_{m,i}(v,u) \neq 0$) is $(\theta, \varphi) \in S_2$

$$g_{m,i}(v,u) = \int_{S_2} \phi_i(v') b(\cos\theta) \sin\theta d\theta d\varphi$$

(c) '0-1': when $v' \notin E_m$ but $v \in E_m$. In this case, the effective domain is $(\theta, \varphi) \in S_3 \cup S_4$.

$$g_{m,i}(v,u) = -\int_{S_3 \cup S_4} \phi_i(v) b(\cos\theta) \sin\theta d\theta d\varphi$$

(d) '1-1': when $v' \in E_m$ and $v \in E_m$. In this case, the effective domain is $(\theta, \varphi) \in S_1 \cup S_2$.

$$g_{m,i}(v,u) = \int_{S_1 \cup S_2} \left[\phi_i(v') - \phi_i(v)\right] b(\cos\theta) \sin\theta d\varphi d\theta$$

We have to pay special attention to integrals over S_1, where the singularity is absorbed. Recall $\phi_i(v)$ are polynomial basis locally defined on each element E_m and $v' = v + z$. Since $z \sim 0$, we take the Taylor expansion of $\phi_i(v')$ around v,

$$\phi_i(v') - \phi_i(v) = \nabla\phi_i(v) \cdot z + \frac{1}{2} z^T \nabla^2 \phi_i(v) z + O(|z|^3)$$

So, it's not hard to observe that, for terms with lowest power of $\sin\theta$, the azimuthal angle φ will be integrated out and leaves only powers of $1 - \cos\theta$, which will help cancel the singularity in $b(\cos\theta)$.

That is,

$$
\begin{aligned}
&\int_0^{\theta_0}\int_0^{2\pi} [\phi_i(v') - \phi_i(v)]\, b(\cos\theta)\sin\theta d\varphi d\theta \\
&= \int_0^{\theta_0}\int_0^{2\pi} [\phi_i(v') - \phi_i(v)] \sin^{-2-\alpha}\frac{\theta}{2}\sin\theta d\varphi d\theta \\
&= 4\int_0^{t_0}\int_0^{2\pi} [\phi_i(v') - \phi_i(v)]\, t^{-1-\alpha} d\varphi dt \quad (\text{change } t = \sin\frac{\theta}{2}) \\
&\le C\int_0^{\theta_0} (1-\cos\theta)\sin^{-2-\alpha}\frac{\theta}{2}\sin\theta d\varphi d\theta \\
&\le C\int_0^{t_0} t^{1-\alpha} dt \quad (\text{ change } t = \sin\frac{\theta}{2}, \quad t_0 = \sin\frac{\theta_0}{2}) \\
&= \frac{C}{2-\alpha} t_0^{2-\alpha} \quad (\text{ notice } \alpha < 2).
\end{aligned}
$$

The sets S_1 and S_2 can be combined. The outer integration w.r.t the polar angle θ is performed using adaptive quadratures , say CQUAD in GSL [58], and the inner integration w.r.t φ is done analytically by calling a similar routine that derives all possible intervals of φ.

Remark. In practice, the above routine can be only applied to the case when v, v' fall onto the same mesh element (when collision is almost grazing); for other cases, the angular cross-sections can be regarded as integrable (far away from grazing collisions) and thus can call routines in "Integrable $b(\cos\theta)$".

Once $g_{m,i}(v,u)$ is done, plugging it back into (3.5), we get the block matrix $\mathbf{G}_{m,i}(k,\bar{k})$.

If denote the whole coefficient vector $\mathbf{U} = (\mathbf{u}_0, \ldots, \mathbf{u}_{N-1})^T$, where $\mathbf{u}_k = (\mathbf{u}_k^0, ..., \mathbf{u}_k^p)^T$, then the semi-discrete DG form of the homogeneous BTE (3.3)

goes

$$\frac{d\mathbf{U}}{dt} = \mathbf{Q}(\mathbf{U}), \tag{3.11}$$

with initial data the L^2 projection of $f_0 = f(v, 0)$; where the collision vector $\mathbf{Q} = (\mathbf{Q}_0, \ldots, \mathbf{Q}_{M-1})^T$, $M = ((p+1)n)^d$ is total degrees of freedom and each block $\mathbf{Q}_m$ is of size $(p+1)^d \times 1$, with its components $\mathbf{Q}_m^i$, for $i = 0 \ldots l$,

$$\mathbf{Q}_m^i = \sum_k \sum_{\bar{k}} \mathbf{u}_k^T \mathbf{G}_{m,i}(k, \bar{k}) \mathbf{u}_{\bar{k}}. \tag{3.12}$$

Or, for each component of the coefficient vector,

$$\frac{d\mathbf{u}_m^i}{dt} = \sum_k \sum_{\bar{k}} \mathbf{u}_k^T \mathbf{G}_{m,i}(k, \bar{k}) \mathbf{u}_{\bar{k}}, \tag{3.13}$$

where we call the matrix $\mathbf{G}_{m,i}$ *"Boltzmann collision matrix"* or simply *"weight matrix"*.

3.3 Reductions on the Computing and Storage Complexity of Collision Matrix

For every test function $\phi_i(v)$ defined over element E_m, we have to compute a collision matrix $\mathbf{G}_{m,i}$ of size $N \times N$, with $N = (p+1)n$. So, theoretically, the computing and storage complexity for the weights in total would be $O(N^3)$. This could be hugely expensive. However, the following features of the collision matrices are crucial for reducing the complexity, i.e

- Temporally independent and precomputed;
- "Shifting symmetries";

- Sparse;
- Parallelizable.

3.3.1 Shifting Symmetry Property for Uniform Meshes

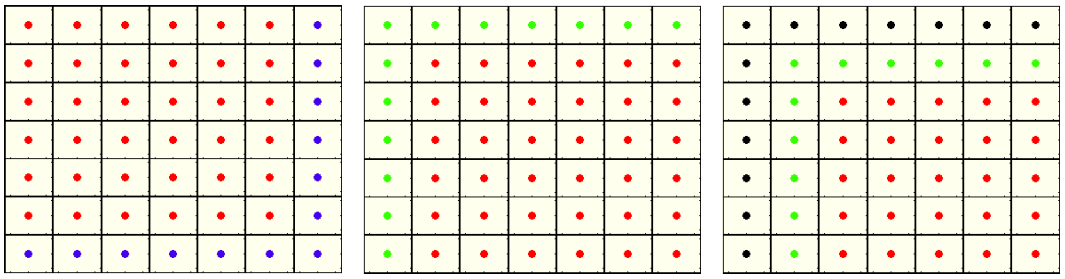
Here we assume a uniform mesh. Recall the fact that the post-collisional velocity $v' = \frac{v+v_*}{2} + \frac{|v-v_*|}{2}\sigma$. Thus, after doing a same shift on both $v \in E_k$ and $v_* \in E_{\bar{k}}$, v' will end up with shifting the same as well. So, as long as the relative positions between E_k ($E_{\bar{k}}$) and test element E_m keep unchanged, and at the same time, the piecewise basis functions $\phi(v)$ on E_m are only valued locally upon the relative position of v inside E_m, then the evaluation of (3.5) will be unchanged. This is summarized as the following property.

Shifting Symmetry Property If the basis piecewise polynomials $\phi(v)$, defined over element E_m, are functions of $\frac{v-w_m}{\Delta v}$ (where w_m is the center of cube E_m), then, the family of collision matrix $\{\mathbf{G}_{m,i}\}$ satisfies the "shifting symmetry" property

$$\mathbf{G}_{m,i}(k, \bar{k}) = \mathbf{G}_{\tilde{m},i}(k - (m - \tilde{m}), \bar{k} - (m - \tilde{m})), \tag{3.14}$$

where $m, \tilde{m}, k, \bar{k}$ are d-dimensional multi-indices; $i = 0, \ldots, p$.

This shifting procedure can be illustratively shown in Figure 3.2, for a 1D problem with piecewise constant basis functions. Figure 3.2 shows that the lower-right $(n-1) \times (n-1)$ submatrix of Matrix G_1 is equivalent to the upper-left submatrix of Maxtrix G_0, while only leaving the first row and first column of G_1 to be determined. This rule applies again to Matrix G_2.



Matrix G_0 Matrix G_1 Matrix G_2

Figure 3.2: Dots (entries) of the same color are shifted to the neighboring matrices, showing illustratively for 1-D

This property inspires us to seek for possible ways to reduce the actual computing complexity of all the collision matrices.

Theorem 3.3.1 (Minimal Basis Set)**.** *There exists a minimal basis set of matrices*

$$\mathfrak{B} = \{G_{m,i}(k,\bar{k}) : \textit{For } j = 1..d,\ \textit{if } m_j \neq 0, k_j \times \bar{k}_j = 0;\ \textit{if } m_j = 0, k_j, \bar{k}_j = 0, 1, \ldots, n-1\},$$

which can exactly reconstruct the complete family $\{\mathbf{G}_{m,i}\}$*, through shifting.*

Proof. Indeed, without loss of generality, let's only consider piecewise constant basis functions, i.e $i = 0$. And we start from only one "layer" on one dimension or let's imagine a 1-D prototype problem, i.e $G_m(k,\bar{k})$ where $m, k, \bar{k}$ are 1-d indices. This is corresponding to one velocity component. The complete family $\{\mathbf{G}_m\}$ will be the tensor product of all "layers".

For any $m \neq 0$ ($m = 1, \ldots, n-1$), the entries $G_m(k,\bar{k})$ are obtained through shifting according to the following policy

$$G_m(k,\bar{k}) = \begin{cases} G_0(k-m,\bar{k}-m), \text{ if } k,\bar{k} \geq m \\ G_l(0,\bar{k}-k) \text{ with } l=1...m, \text{ if } \bar{k} \geq k,\ k < m \\ G_l(k-\bar{k},0) \text{ with } l=1...m, \text{ if } k \geq \bar{k},\ \bar{k} < m \end{cases} \tag{3.15}$$

which recover the complete set of entries $G_m(k,\bar{k})$. None of the entries in the basis set are shifting-equivalent. And it's not hard to observe that if one drops any entry in the basis set, it will be impossible to recover the original complete family. Thus, we conclude that the set $\mathfrak{B}$ is one minimal basis set. □

As seen from Theorem 3.3.1, along each dimension, we only need to compute and store the full matrix for $m = 0$, and the first rows and columns for all other m's. This requires a computing and stroage complexity bounded by $3n^2$. For d dimensions, the total complexity will be bounded by $3^d N^2$ with $N = n^d$. Hence, in the actual algorithm, we only need to compute the minimal set $\mathfrak{B}$, which requires a computing complexity of only $O(N^2)$.

3.3.2 Sparsity

The matrices in the set $\mathfrak{B}$ are actually highly sparse. The sparsity of $\mathfrak{B}$, again, comes from $v' = v - \frac{u}{2} + \frac{|u|}{2}\sigma$. The post-collisional velocity v' is on the sphere parametrized by $\sigma \in S^{d-1}$, centered at $\frac{v+v_*}{2}$ and radius given by $|u|/2$. Thus, not all binary particle collisions between velocities $v \in E_k$ and $v_* \in E_{\bar{k}}$ could collide ending up with a post-collisional velocity v' lying in a given fixed element E_m. Since, for each v and v_* fixed, the sphere that contains v' and v'_* in a binary collision is a $(d-1)$-manifold embedded in d-dimensions, the

counting of such interactions are non-zero when such sphere intersects with element E_m. This results in only an $O(n^{2d-1})$ of nonzeros in the set $\mathfrak{B}$.

Therefore, while by Theorem 3.3.1, the calculations of the weights $G_m(k, \bar{k})$ can be made in an algorithm with computational complexity of $O(n^{2d})$, we conjecture that the corresponding storage complexity is of $O(n^{2d-1})$. Indeed, we verify this order complexity with test run for $d = 3$ in Table 3.1, done on a single core of Xeon E5-2680 2.7GHz processor (on cluster Stampede-TACC [103]).

n	wall clock time (s)	order	♯ of nonzeros	order
8	3.14899	\	812884	\
12	39.3773	6.2301	6826904	5.2484
16	228.197	6.1075	30225476	5.1717
20	893.646	6.1176	94978535	5.1311
24	2686.72	6.0375	241054134	5.1054

Table 3.1: The computing and storage complexity of "basis" $\mathfrak{B}$.

3.3.3 Parallelization

The whole weight matrices are only computed once and stored for further use. Due to the locality of DG schemes, the whole process of computing $\mathfrak{B}$ can be well performed using hybrid MPI [65] and OpenMP [13]. The collision weight matrices quantify the contributions of the binary collisions to the evolution of the distribution functions. For each grid point on the distribution fuction, the time evolution is attributed to all possible binary collisions. Furthermore, different grid points do not need to communicate with each other. Thus, one can distribute all grid points across the computing node community

while keeping the grid information accessible to each computing node within the community. This is done using MPI. To further parallelize the computing, on each node, the working load, for example, computing of matrix entries and matrix-vector computations are shared among threads, using OpenMP.

Figure 3.3 shows the parallel efficiency of strong scaling for computing some sets of "basis matrix".

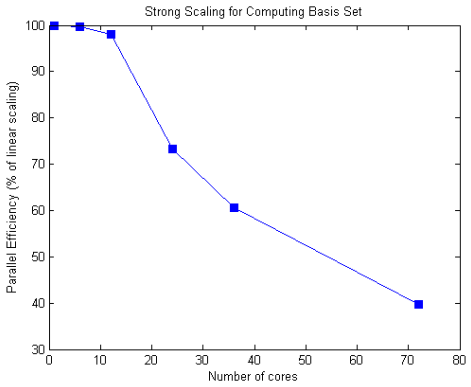


Figure 3.3: The strong scalability of computing collision matrix (n=18)

3.4 Conservation Routines

The above approximate collision operator Q doesn't preserve the moments as needed, due to the truncation of velocity domain. To achieve the conservation properties, following the ideas in [68], we design an intermediate routine to enforce the conservations. This routine will be implemented as a L^2-*distance minimization problem* with the constraints the preservation of desired moments. The optimization problem can be solved through *Lagrange*

multiplier method.

The conservation of moments for the approximate solution $f_h(t,v)$ goes, for any time t,

$$\int_{\Omega_v} f_h(t,v)\varphi(v)dv = \int_{\Omega_v} f_{h,0}(v)\varphi(v)dv, \tag{3.16}$$

where $\varphi(v)$ is one of the $d+2$ collision invariants $1, v_1, \ldots, v_d, |v|^2$.

So, our objective is to solve

Conservation Routine [Functional Level]: Minimize in the Banach space

$$\mathcal{B}^e = \left\{ X \in L^2(\Omega_v) : \int_{\Omega_v} X = \int_{\Omega_v} Xv = \int_{\Omega_v} X|v|^2 = 0 \right\},$$

the functional

$$\mathcal{A}^e(X) := \int_{\Omega_v} \left(Q_{uc}(f)(v) - X\right)^2 dv. \tag{3.17}$$

Recall the DG approximation for $f_h(t,v)$ in (4.67) and the time evolution for $f_h(t,v)$ (3.11), one can get the conservation requirements on the approximation collision voctor $\mathbf{Q}$, defined in (3.11),

$$\mathbf{CQ} = \mathbf{0} \tag{3.18}$$

where the $(d+2) \times N$ dimensional constraint matrix writes

$$\mathbf{C}_{:,j} = \begin{pmatrix} \int_{E_k} \phi_l(v)dv \\ \int_{E_k} \phi_l(v)vdv \\ \int_{E_k} \phi_l(v)|v|^2 dv \end{pmatrix}, \tag{3.19}$$

with ϕ_l the $l-th$ basis function on element E_k and the column index $j = (p+1)k + l = 0, \ldots, N-1$.

To enforce the conservation, we seek for the L^2-distance closest $\mathbf{Q}_c$, which is the minimizer of the following constrained optimization problem:

Conservation Routine [Discrete Level]: Find $\mathbf{Q}_c$, the minimizer of the problem

$$\begin{aligned} &\min \frac{1}{2}(\mathbf{Q}_c - \mathbf{Q})^T \mathbf{D}(\mathbf{Q}_c - \mathbf{Q}) \\ &\text{s.t.} \quad \mathbf{C}\mathbf{Q}_c = \mathbf{0}. \end{aligned}$$

Due to the orthogonality of the local basis, $\mathbf{D}$ is a positive definite diagonal matrix with its j-th entry $\frac{1}{|E_k|}\int_{E_k}(\phi_l(v))^2 dv$, $j = (p+1)k + l$. For example, in 3D, when $p = 0$, $\mathbf{D}$ is reduced to an identity matrix; while $p = 1$, with the orthogonal basis chosen in (3.2),

$$\mathbf{D} = \text{Diag } (1, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, 1, \frac{1}{12}, \frac{1}{12}, \frac{1}{12}, 1, ...).$$

Remark. Note that with spectral method in [68], the corresponding discrete optimization problem actually takes D to be an identity matrix. This is because the L^2 norm is aymptotically preserved by the l^2 norm of its Fourier coefficients.

To solve the minimization problem, we employ the Lagrange multiplier method. Denote by $\lambda \in \mathbb{R}^{d+2}$ the multiplier vector. Then the objective function writes

$$\mathcal{L}(\mathbf{Q}_c, \lambda) = \frac{1}{2}(\mathbf{Q}_c - \mathbf{Q})^T \mathbf{D}(\mathbf{Q}_c - \mathbf{Q}) - \lambda^T \mathbf{C}\mathbf{Q}_c. \tag{3.20}$$

Solving by finding the critical value of $\mathcal{L}$ gives

$$\begin{cases} \dfrac{\partial \mathcal{L}}{\partial \mathbf{Q_c}} = \mathbf{0} \\ \dfrac{\partial \mathcal{L}}{\partial \lambda} = \mathbf{0} \end{cases} \Longrightarrow \begin{cases} \mathbf{Q}_c = \mathbf{Q} + \mathbf{D}^{-1}\mathbf{C}^T\lambda \\ \mathbf{C}\mathbf{Q}_c = \mathbf{0} \end{cases} \Longrightarrow \lambda = -(\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T)^{-1}\mathbf{C}\mathbf{Q}.$$

(Here, notice that $\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T$ is symmetric and positive definite and hence exists the inverse.)

Thus, we get the minimizer $\mathbf{Q}_c$

$$\mathbf{Q}_c = [\mathbf{I} - \mathbf{D}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T)^{-1}\mathbf{C}]\mathbf{Q}, \tag{3.21}$$

where $\mathbf{I}$ is an identity matrix of size $N \times N$. So, $\mathbf{Q}_c$ is a perturbation of $\mathbf{Q}$.

So, the final conservative semi-discrete DG formulation for the homogeneous equation writes

$$\frac{d\mathbf{U}}{dt} = \mathbf{Q}_c, \tag{3.22}$$

which preserves the desired moments. Furthermore, the approximate solution approaches a stationary state. This is guaranteed by analyzing the convergence behavior.

3.5 Temporal Evolution

The approximate solution will be solved at the level of discrete time. That is, $t_{n+1} = t_n + \triangle t$, where $\triangle t$ is the time step size. Since there is no high order derivatives or diffusive natures in the homogeneous BTE, no CFL condition is imposed. The only restriction on time step size maybe that Δt should be less than the dimensionalized mean free time. So, we can choose the

simplest *Euler scheme*. At each time step, the conservation routine, denoted *CONSERVE*, designed in the last section will be called to force conservations.

So, suppose $\mathbf{U}_n$ is the coefficient vector (thus the solution) computed at the current time t_n, then the solution for the next time step is obtained through the following routines

$$\mathbf{Q}_n = \text{COMPUTE}(\mathbf{U}_n)\,,$$

$$\mathbf{Q}_{c,n} = \text{CONSERVE}(\mathbf{Q}_n)\,,$$

$$\mathbf{U}_{n+1} = \mathbf{U}_n + \triangle t \mathbf{Q}_{c,n}\,.$$

The Euler scheme is formally first order in time. For higher order accuracy, a higher order Runge Kutta scheme can be used whenever necessary. The conservation routine has to be invoked at every intermediate step of the Runge Kutta scheme.

At each time step, for the evolution of each mesh element, we have to compute a quadratic form (3.13) which inevitably involves $O(n^9)$ (in $d = 3$) operations in total. However, due to the sparsity, the actual order of number of operations for each time step is $O(n^8)$, which is indeed a large number. Fortunately, the reconstructions of collision matrices and computing of quadratic form (3.13) are well parallelizable for each Euler step. See Table 3.2 for the results on time for one single temporal step of evolution, and Table 3.3 for results on the parallelization for one step of time evolution. Both tests run on Xeon Intel 3.33GHz Westmere processors (on cluster Lonestar-TACC [103]).

From Table 3.2 we can see, the time consumed for one single step grows with an order slightly less than 8. This is normal, because during

n	wall clock time (s)	order
16	18.3362	\
18	47.6001	8.0993
20	105.155	7.5227
22	216.818	7.5923
24	419.533	7.5862
26	781.282	7.7683

Table 3.2: The wall clock time for one temporal evolution step

number of cores	wall clock time (s)
1	459.967
2	341.771
6	181.561
12	144.485
24	129.691
36	107.907
48	90.2676
72	74.2794

Table 3.3: The parallelization for one temporal evolution step, for $n = 24$

time evolution and reconstruction of the whole collision matrix, we only need to retrieve those "effective" (non-zero) matrix entries through shifting of the basis set $\mathfrak{B}$ (see Theorem 3.3.1). The grid points (index m) and associated "effective" entries (indices $k, \bar{k}$) are shifted together. Thus, in practice, not every grid point requires a full ergodic of the weight matrix, or in other words, many grid points only need a partial access to the weight matrix. Table 3.3 actually shows a low strong scaling efficiency (speedup is far from linear). This is not surprising because we need to call a parallelized reconstruction process for each time step, then gather information together and re-distribute them to the computing community. And furthermore, we have to call the

conservation routine, which is essentially serial, at each time step. In addition, when computing the basis set $\mathfrak{B}$, we choose to distribute grid points across the computing nodes while the basis information associated with each grid point m (see Theorem 3.3.1) is not of equivalent size, for example for $m = 0$ the full matrix is computed, while for other m's, only first row and column are computed. Hence, some processing elements, for example the one containing $m = 0$, have to be accessed much more frequently than others, also causing an unbalanced distribution of computing resources.

3.6 Asymptotic Behaviors and Error Analysis

The asymptotic error analysis is based on the work [5]. Readers can find more details of the proofs for many theorems and estimates invoked here.

Since we are working under a DG framework, it might be necessary to summerize some of the notations and properties regarding DG approximations, see Appendix for details.

Next we analyze the asymptotic error. We assume $f \in C([0,T]; L^2(\mathbb{R}^d))$ be the solution to the homogeneous BTE (3.3) with initial $f_0 = f(v,0)$. The Galerkin method allows us to take the L^2 projection, $P_h : L^2(\Omega_v) \to L^2(\Omega_v)$, on both sides of the BTE (3.3)

$$\frac{\partial}{\partial t} P_h f(v,t) = P_h Q(f)(v,t) \text{ in } [0,T] \times \Omega_v. \tag{3.23}$$

We introduce the concept of *extension operator* $E : H^\alpha(\Omega_v) \to L^2(\mathbb{R}^d)$, which will be used in future derivations. See appendix or [5] for properties of

extension operators and definition of moments.

The collision operator $Q(f)$ is global in velocity. It's reasonable to expect $P_hQ(f) \sim P_hQ(EP_hf)$ for "accurate" enough projectors (or small enough mesh size h). Thus the discrete solution $g(v,t)$ to the problem

$$\frac{\partial}{\partial t}g(v,t) = P_hQ(Eg)(v,t) \tag{3.24}$$

is expected to be a good approximation to P_hf, the solution to projected equation (3.23).

This is not enough, because we are limited to the conservation properties. Hence, the following initial value problem is studied in our asymptotic analysis, whose solution is expected to approximate the solution f of the original homogeneous BTE (3.3).

$$\begin{aligned} \frac{\partial}{\partial t}g(v,t) &= Q_c(g)(v,t) \\ g_0(v) &= P_hf(v,0), \end{aligned} \tag{3.25}$$

where $Q_c(g)$ is the conservation correction to the following unconserved operator $Q_{uc}(g)$

$$Q_{uc}(g)(v,t) = P_h(Q(Eg)\chi_{\Omega_v})(v,t), \tag{3.26}$$

where χ_{Ω_v} is the characteristic function on the truncated domain Ω_v. It follows that

$$\|Q_{uc}(f_h)\|_{L^2(\Omega_v)} \lesssim \|Q(Ef_h)\|_{L^2(\Omega_v)} \lesssim \|Q(f)\|_{L^2(\Omega_v)}, \tag{3.27}$$

As is shown in the last section, the conservation correction is the minimizer of the L^2-distance to the projected collision operator subject to mass,

momentum and energy conservation. It can be shown that, the conserved projection operator $Q_c(f_h)$ is a perturbation of $Q_{uc}(f_h)$ by a second order polynomial. See Theorem 3.3 in [5] for the Conservation Correction Estimate.

Let's summerize our major estimate result first.

Theorem 3.6.1 (H^{p+1}-error estimate)**.** *Fix $k', k \geq 0$ and assume nonnegative initial density function $f_0 \in L^1_2 \cap H^{p+1}_q(\mathbb{R}^d)$ with $q = \max\{k + k', 1 + \frac{d}{2\gamma}\}$, $0 < \gamma \leq 1$ defined in collision kernel (2.4). g is the DG solution of the equation (3.25), where the piecewise basis polynomials are of order at most p. For a given simulation time T and index $\alpha \leq p + 1$, there exists an extension E_{p+1}, a lateral size $L_0(T, f_0)$ for domain Ω_v and a small grid diameter $h_0(T, L, f_0, \alpha)$ for triangulation $\mathcal{T}_h$ of Ω_v, such that for any $L \geq L_0$, $h \leq h_0$,*

$$\sup_{t\in[0,T]} \|f - g\|_{H^\alpha_k(\mathcal{T}_h)} \leq C_{k'} e^{C_k T} \left(O(L^{\gamma k+\alpha} h^{p+1-\alpha}) + O(L^{-\gamma k'}) \right)$$

where $h = max_{E_v\in\mathcal{T}_h} diam(E_v)$ is the maximal grid diameter for the regular triangulation $\mathcal{T}_h$ of Ω_v; the constants C_k and $C_{k'}$ depends on H^{p+1}_q-norms and moments of f_0.

Proof. The proofs can be easily extended from the one in [5]. But here to make the work complete, we would like to briefly explain how the proofs go. The readers can refer to [5] for more details.

We will first prove the case $\alpha = 0$, i.e the L^2_k estimate and then follow an induction on the index α.

One can easily observe that, in domain Ω_v,

$$\frac{\partial}{\partial t}(f-g) = Q(f,f) - Q_c(g) = (Q(f,f) - Q(Eg,Eg)) + (Q(Eg,Eg) - Q_c(g)). \tag{3.28}$$

Denote $e_h = \|f-g\|_{L^2_k(\mathcal{T}_h)}$. Multiply on both sides of the above equation (3.28), piecewisely, by $(f-g)\langle v\rangle^{2\gamma k}$ restricted over each element of the triangulation $\mathcal{T}_h$, and sum over all the elements, we get

$$\frac{1}{2}\frac{\partial e_h^2}{\partial t} = I_1 + I_2. \tag{3.29}$$

We estimate I_1 and I_2 separately.

$$\begin{aligned} I_1 &= \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k}(f-g)(Q^+(f+Eg,f-Eg) + Q^+(f-Eg,f+Eg)) \\ &\quad - \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k}(f-g)(Q^-(f+Eg,f-Eg) - \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k}(f-g)Q^-(f-Eg,f+Eg)) \\ &\lesssim \|f-g\|^2_{L^2_k(\mathcal{T}_h)} + \|f-g\|_{L^2_k(\mathcal{T}_h)} \left(\|f\|_{L^2_{k+1/2}(\mathbb{R}^d\setminus\Omega_v)} + \|g\|_{L^2_{k+1/2}(\Omega_v\setminus\delta^{-1}\Omega_v)},\right) \end{aligned}$$

where δ is the dilation parameter of the extension operator; $\lesssim$ means the estimate constants are independent of parameters T, L, h but only information (norms, moments, etc.) of f, g itself. Here, the uniform propagation of higher order moments of f, g are applied. See Lemma 4.2 in [5].

By Holder's inequality and Conservation Correction Estimate,

$$\begin{aligned} I_2 &= \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k}(f-g)(Q(Eg,Eg) - Q_c(g)) \\ &\lesssim L^{\gamma k}\|f-g\|_{L^2_k(\mathcal{T}_h)} \left(\|Q(Eg,Eg) - Q_{uc}(g)\|_{L^2(\mathcal{T}_h)} + \delta^{2k'}\mathrm{O}_{d/2+\gamma(k'-1)}\|g\|_{L^1_{k'}(\mathcal{T}_h)}\right), \end{aligned}$$

where and in the following we apply notation $\mathrm{O}_r := O(L^{-r})$.

So, combining the above estimates for I_1 and I_2 gives us

$$\frac{de_h(t)}{dt} \leq Ce_h(t) + \varepsilon(t) + \varpi(t),$$

where, by the standard approximation theory in the broken Sobolev spaces,

$$\begin{aligned}\varepsilon(t) &:= CL^{\gamma k}\|Q(Eg, Eg) - Q_{uc}(g)\|_{L^2(\mathcal{T}_h)} \\ &\lesssim L^{\gamma k}h^{p+1}\|Q(Eg, Eg)\|_{H^{p+1}(\Omega_v)} \\ &\lesssim L^{\gamma k}h^{p+1}\|g\|^2_{H^{p+1}_{\mu}(\Omega_v)} \quad (\mu > 1 + \frac{d}{2\gamma}),\end{aligned}$$

and

$$\begin{aligned}\varpi(t) &:= C\left(\|f\|_{L^2_{k+1/2}(\mathbb{R}^d\setminus\Omega_v)} + \|g\|_{L^2_{k+1/2}(\Omega_v\setminus\delta^{-1}\Omega_v)}\right) + \mathrm{O}_{d/2+k'-k-s}\|Ef_h\|_{L^1_{k'}(\Omega_v)} \\ &= \delta^{2k'}\mathrm{O}_{\gamma(k'-k-1/2)}\left(\|f\|_{L^2_{k'}(\mathbb{R}^d)} + \|g\|_{L^2_{k'}(\Omega_v)} + \|g\|_{L^1_{k'}(\Omega_v)}\right) \leq \mathrm{O}_{\gamma k'}.\end{aligned}$$

The Gronwall's inequality implies

$$\sup_{t\in[0,T]}\|f - g\|_{L^2_k(\mathcal{T}_h)} \leq \left(\|f_0 - f_{h,0}\|_{L^2_k(\mathcal{T}_h)} + \int_0^T \varepsilon(s)ds + \sup_{t\in[0,T]}\varpi(t)\right)e^{CT}, \tag{3.30}$$

for any $T > 0$. The lateral size $L(T, f_0)$, $h \leq h_0(T, L, f_0)$ are decided following a same argument in Theorem 5.1 in [5].

Additionally, by the standard approximation theory,

$$\|f_0 - f_{h,0}\|_{L^2_k(\mathcal{T}_h)} \lesssim L^{\gamma k}h^{p+1}\|f_0\|_{H^{p+1}(\Omega_v)}.$$

Thus, the case $\alpha = 0$ is proved. Assume the result is true for any multi-index $\beta < \alpha \leq p + 1$. The similarly following the above procedures,

$$\frac{\partial}{\partial t}\|\partial^{\alpha}(f - g)\|^2_{L^2_k(\mathcal{T}_h)} \leq I_1 + I_2 + I_3.$$

Using Leibniz formula and the smoothing effect of the positive collision operator,

$$\begin{aligned} I_1 &:= \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k} \partial^\alpha (f-g)\partial^\alpha (Q(f,f)-Q(Eg,Eg)) \\ &\lesssim \|\partial^\alpha (f-g)\|^2_{L^2_k(\mathcal{T}_h)} + \text{ lower order terms }. \end{aligned}$$

A typical lower order term is given by

$$\|\partial^\alpha (f-g)\|_{L^2_k(\mathcal{T}_h)} \|\partial^{\alpha-\beta}(f+Eg)\|_{L^2_{k+\mu}(\mathbb{R}^d)} \|\partial^\beta (f-Eg)\|_{L^2_{k+1}(\mathbb{R}^d)}.$$

By induction hypothesis,

$$\begin{aligned} \|\partial^\beta (f-Eg)\|_{L^2_{k+1}(\mathbb{R}^d)} &:\leq \|\partial^\beta (f-g)\|_{L^2_{k+1}(\mathcal{T}_h)} + \|\partial^\beta f\|_{L^2_{k+1}(\mathbb{R}^d\setminus\Omega_v)} + \|\partial^\beta Eg\|_{L^2_{k+1}(\mathbb{R}^d\setminus\Omega_v)} \\ &\leq C_{k'} e^{C_k T}\left(O(L^{\gamma(k+1)+\beta} h^{p+1-\beta}) + \delta^{2(k+k')} O(L^{-\gamma k'})\right) \\ &\leq C_{k'} e^{C_k T}\left(O(L^{\gamma k+\alpha} h^{p+1-\alpha}) + \delta^{2(k+k')} O(L^{-\gamma k'})\right), \end{aligned}$$

where the last inequality holds as long las $h \leq L^{1-\gamma}$.

For I_2, by Holder's inequality and the conservation correction estimate,

$$\begin{aligned} I_2 &:= \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k} \partial^\alpha (f-g)\partial^\alpha (Q_c(g)-Q_{uc}(g)) \\ &\leq \|\partial^\alpha (f-g)\|_{L^2_k(\mathcal{T}_h)} \|\partial^\alpha (Q_c(g)-Q_{uc}(g))\|_{L^2_k(\mathcal{T}_h)} \\ &\leq \|\partial^\alpha (f-g)\|_{L^2_k(\mathcal{T}_h)} \left(L^{\gamma k} \|Q(Eg,Eg)-Q_{uc}(g)\|_{L^2(\mathcal{T}_h)} + \delta^{2k''} \mathrm{O}_{d/2+\gamma(k''-k)}\right). \end{aligned}$$

For I_3, by holder's inequality and approximation theory,

$$\begin{aligned} I_3 &:= \int_{\mathcal{T}_h} \langle v\rangle^{2\gamma k} \partial^\alpha (f-g)\partial^\alpha (Q_{uc}(g)-Q(Eg,Eg)) \\ &\leq L^{\gamma k} \|\partial^\alpha (f-g)\|_{L^2_k(\mathcal{T}_h)} \|\partial^\alpha (Q_{uc}(g)-Q(Eg,Eg))\|_{L^2(\mathcal{T}_h)} \\ &\lesssim L^{\gamma k} \|\partial^\alpha (f-g)\|_{L^2_k(\mathcal{T}_h)} h^{p+1-\alpha} \|g\|^2_{H^{p+1}_{d/2+\gamma}}. \end{aligned}$$

Finally, we get

$$\frac{\partial}{\partial t}\|\partial^\alpha(f-g)\|_{L^2_k(\mathcal{T}_h)} \leq C\|\partial^\alpha(f-g)\|_{L^2_k(\mathcal{T}_h)} + C_{k'}e^{C_k T}\left(O(L^{\gamma k+\alpha}h^{p+1-\alpha}) + \delta^{2(k+k')}O(L^{-\gamma k'})\right),$$

therefore, the Gronwall's inequality will give us the final estimate.

□

3.7 Numerical Results

Test 1 is a 2-d Maxwell type of elastic collisions, benchmarked by *Bobylev-Krook-Wu (BKW)* exact solutions. The initial density distribution is

$$f(v,0) = \frac{v^2}{\pi\sigma^2}\exp(-v^2/\sigma^2)\,. \tag{3.31}$$

This problem has an exact solution [57]

$$f(v,t) = \frac{1}{2\pi s^2}\left(2s - 1 + \frac{1-s}{2s}\frac{v^2}{\sigma^2}\right)\exp\left(-\frac{v^2}{2s\sigma^2}\right), \tag{3.32}$$

where $s = 1 - \frac{1}{2}\exp(-\sigma^2 t/8)$. In the test, we choose the scaling parameter $\sigma = \pi/6$ such that the truncation domain is well chosen by $\Omega_v = [-\pi, \pi]$. We let it run for 600 time steps with $\Delta t = 0.1$. This example is used to test the accuracy by calculating the relative L^2 errors compared to its exact solution and relative entropy verifying that the numerical solution will converge to the true equilibrium. See Figure 3.4 for the evolution of the marginal density distributions; Figure 3.5 and Figure 3.6 shows the relative L^2 errors and relative entropy, respectively. The marginal density distribution is defined as

$$f_x(v_x \in I_k) = \frac{1}{(\Delta v)^2}\int_{I_k}\int_{I_{n/2}} f(v,t)dv_x dv_y\,.$$

The relative L^2 error is defined as

$$\frac{\left(\int_{\Omega_v} |f_h(v,t) - f(v,t)|^2 dv\right)^{1/2}}{\left(\int_{\Omega_v} |f(v,t)|^2 dv\right)^{1/2}},$$

and the relative entropy given by

$$\mathcal{H}_{rel}(t) = \int_{\Omega_v} f(v,t) \log f(v,t) - f_M(v) \log f_M(v) dv = \int_{\Omega_v} f(v,t) \log \frac{f(v,t)}{f_M(v)} dv\,, \tag{3.33}$$

where $f_M(v)$ is the true equilibrium density distribution, is expected to converge to zero which implies the solution converges to the true equilibrium in the sense of L^1.

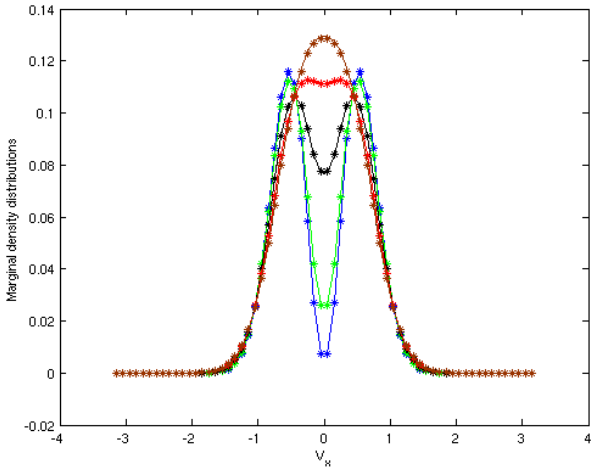


Figure 3.4: **Test 1**: Comparison of solutions at time $t = 0, 1, 5, 10, 15s$. n=44 per direction; solid line: exact solution, stars: p.w. constant approximation

Remark. Through **Test 1**, we would like to mention the positivity issue of numerical solutions. The true density distributions are expected to be positive for any given positive time, if initially so. Our numerical tests show that,

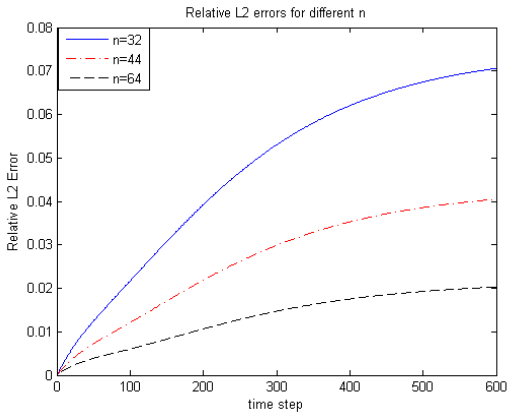


Figure 3.5: **Test 1**: Relative L^2 errors, compared with true solution, for different number of mesh elements

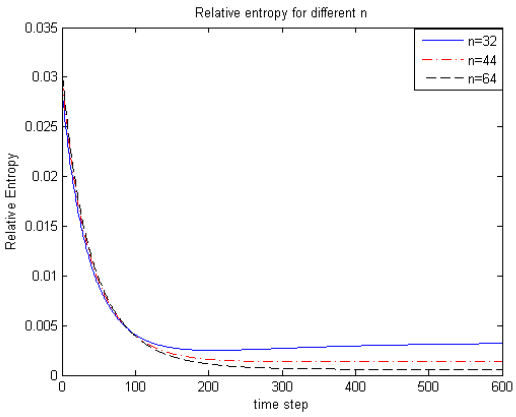


Figure 3.6: **Test 1**: Relative Entropy for different number of mesh elements

positivity can be achieved if we apply piecewise constant basis functions. The conservation laws (here, only mass due to the zero-th order of basis polynomials) are expected to hold but only for a short time, and will be seriously broken in the long run (see more details from the results of the next test problem). With invoke of our conservation routine, the conservations are guaranteed but the positivity is inevitably broken. This seems a common issue for almost all numerical solvers known so far. But fortunately, the negativity only occurs at the tails of the distribution functions, and as long as the "negative energy" (second order moment of the negative part of the density function) stays under controlled by a small ratio to the "positive energy", the accuracy of the numerical approximations is guaranteed.

Test 2 is also 2-d Maxwell type of elastic collisions. This example is

used to show the conservation routines. The initial states we take are convex combinations of two shifted Maxwellian distributions.

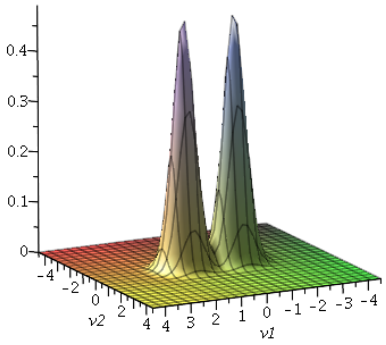


Figure 3.7: **Test 2**: Initial Probability Distribution: two shifted Maxwellians

Truncate the velocity domain $\Omega = [-4.5, 4.5]^2$ and set number of nodes in each velocity direction $n = 32, 40$. The initial density function is a convex combination of two Maxwellians

$$f_0(v) = \lambda M_1(v) + (1-\lambda) M_2(v)\,, \tag{3.34}$$

with $M_i(v) = (2\pi T_i)^{-d/2} e^{-\frac{|v-V_i|^2}{2T_i}}$, $T_1 = T_2 = 0.16$, $V_1 = [-1, 0]$, $V_2 = [1, 0]$ and $\lambda = 0.5$.

We test for $n = 32$ and $n = 40$, for 1000 time steps to compare the results and see the long time behavior as well. The probability density distribution functions are reconstructed with splines.

From Figure 3.10 and Figure 3.11, we can see, the scheme with piecewise constant test functions, as expected conserves moments for short time; in the

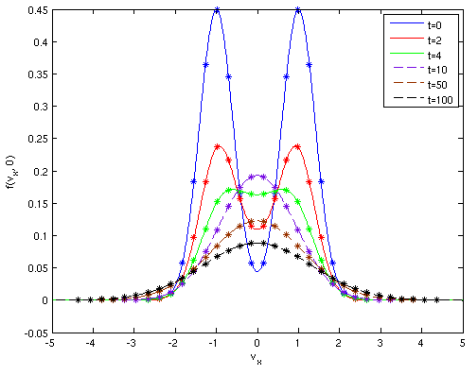


Figure 3.8: **Test 2**: Evolution of pdf without conservation routines.

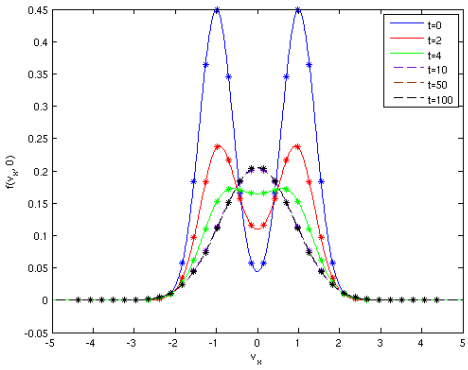


Figure 3.9: **Test 2**: Evolution of pdf with conservation routines

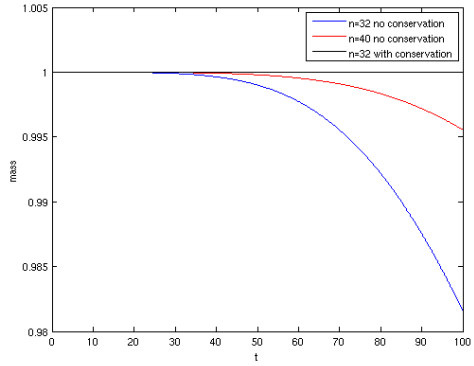


Figure 3.10: **Test 2**: Evolution of mass

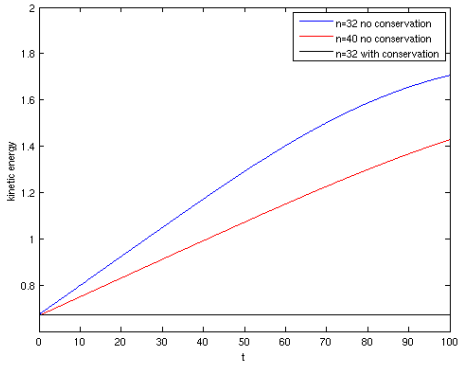


Figure 3.11: **Test 2**: Evolution of kinetic energy

long run, due to the truncation, the tails of the density functions are lifted up and thus moments are expected to lose. At the same time, finer grids indeed give more accuracy. Since the basis polynomials are only zero order, it's expected that mass is much better conserved than higher order moments.

Through the comparison of Figure 3.8 and Figure 3.9 we see, after long time, with no conservation routine, the density distribution collapses due to the

truncation of the domain. While with the invoke of conservation routines, the density function stays stable when reaching equilibrium. So, the conservation routine works and is necessary for stability. However, the cost we pay is the loss of positivity.

Test 3 is initialized by a sudden jump on temperatures, i.e. a jump discontinuity in its initial and far from equilibrium, as shown in Figure 3.12. The initial state is given by

$$f_0(v) = \begin{cases} \dfrac{1}{2\pi T_1} \exp(-\dfrac{|v|^2}{2T_1}), & v_1 \leq 0 \\ \dfrac{1}{2\pi T_2} \exp(-\dfrac{|v|^2}{2T_2}), & v_1 > 0 \end{cases}$$

with $T_1 = 0.3$ and $T_2 = 0.6$. The collision is of type 2d hard spheres.

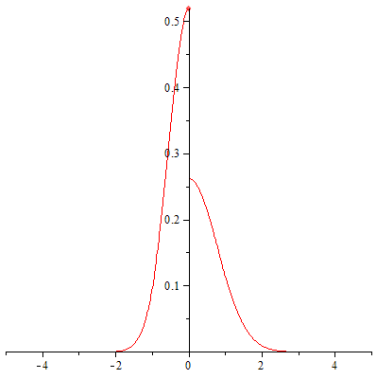


Figure 3.12: **Test 3**: Initial density function

With truncated domain $\Omega_v = [-5, 5]$, $n = 44$ in each direction, the DG solution well captures the discontinuity and converges to equilibrium. See Figure 3.13 and Figure 3.14.

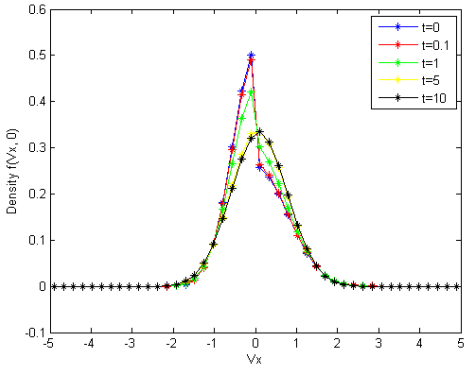


Figure 3.13: **Test 3**: DG solutions

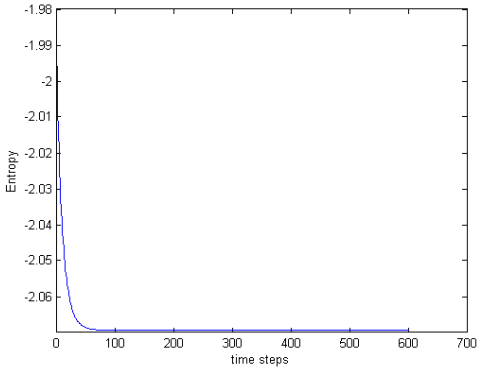


Figure 3.14: **Test 3**: The entropy decay of DG solutions

Test 4 is testing on the 3D homogeneous Boltzmann equation with Maxwell molecular potential, with initial

$$f_0(v) = \frac{1}{2(2\pi\sigma^2)^{3/2}} \left[\exp\left(-\frac{|v - 2\sigma e|^2}{2\sigma^2} \right) + \exp\left(-\frac{|v + 2\sigma e|^2}{2\sigma^2} \right) \right] ,$$

where parameters $\sigma = \pi/10$ and $e = (1, 0, 0)$. $\Omega_v = [-3.4, 3.4]^3$, $n = 30$.

Figure 3.15 shows the evolution of the marginal density distributions, which is defined as

$$f_x(v_x \in I_k) = \frac{1}{(\Delta v)^3} \int_{I_k} \int_{I_{n/2}} \int_{I_{n/2}} f(v, t) dv_x dv_y dv_z \, .$$

Figure 3.16 shows the decay of entropy to its equilibrium state.

Figure 3.17 shows the relaxations of directional temperature, which as expected converge to the averaged temperature.

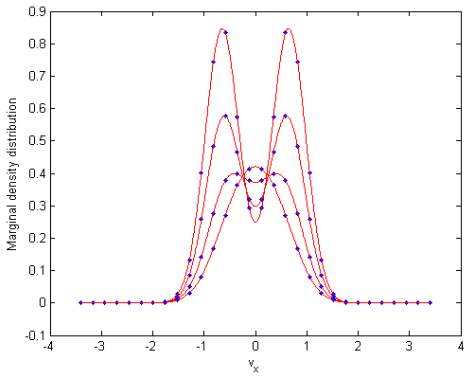


Figure 3.15: **Test 4**: Evolution of marginal distributions at $t = 0, 1, 2.5, 5s$; dots are the piecewise constant value on each element; solid lines are spline reconstructions

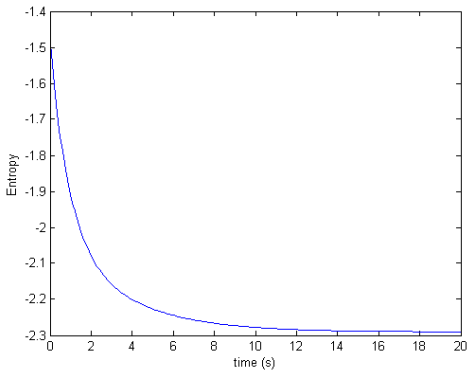


Figure 3.16: **Test 4**: Entropy decay

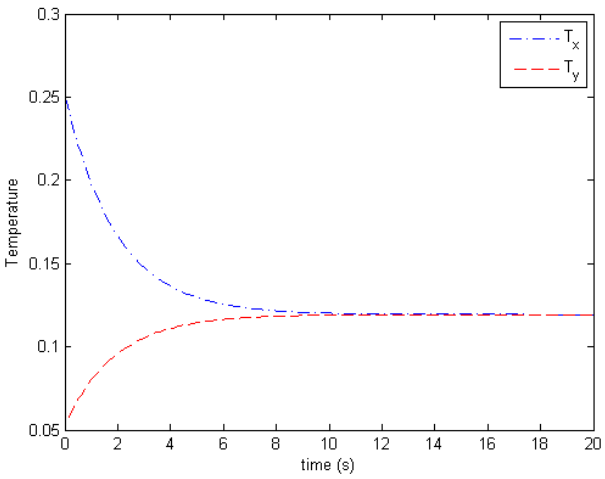


Figure 3.17: **Test 4**: Temperature relaxations along x and y directions

Chapter 4

Computations of Spectral Gaps for Linearized Boltzmann Operators

The quantitative information on the spectral gaps for the linearized Boltzmann operator is of primary importance on justifying the Boltzmann model and study of relaxation to equilibrium. In this chapter, for the first time in this field, we provide numerical evidence on the existence of spectral gaps and corresponding approximate values.

4.1 Introduction

The existence of solutions and regularity theory of the BTE in the space inhomogeneous setting have been great mathematical challenges and still remain elusive. Nevertheless, it is well understood that these qualitative properties depend on the intermolecular potential γ and the integrability properties of the angular cross-sections (2.4, 2.5). Indeed, the relaxation to equilibrium has been at the core of kinetic theory ever since the works of Boltzmann. It provides an analytic basis for the second principle of thermodynamics for a statistical physics model of a gas out of equilibrium. The well-known Boltzmann's **H** theorem [36] shows the possible convergence process and equilib-

rium states. However, it's not enough to justify the Boltzmann model with only non-constructive arguments. It is crucial to obtain quantitative informations on the time scale of the convergence. The question of obtainig explicit decay rates in recent new energy methods [74–77, 114–116] also motivates the work on studying spectral gaps and coercivity estimates. Many authors, for instance [8, 10, 18, 31, 33, 39, 71, 72, 81, 99, 104], have made enormous efforts on (non-)constructive estimates for the rate of convergence (we refer to [56] for a review), among which *Cercignani's conjecture* [35] is a great attempt:

For any f and its associated Maxwellian μ, there is an entropy-entropy production relation

$$\mathcal{D}(f) \geq \lambda \left[\mathcal{H}(f) - \mathcal{H}(\mu)\right] ,$$

where $\mathcal{H}(f) = \int f \log(f) dv$ is the (opposite) entropy; $\mathcal{D}(f) = -\frac{d}{dt}\mathcal{H}(f)$ is the dissipation of the entropy, or "entropy-production" functional; $\lambda > 0$ is some "suitable constant". The existence of such inequalities is equivalent to the existence of exponential convergence towards equilibrium.

In the regime very close to equilibrium, the linearized part of the model can actually dominate. In particular, the linearized counterpart of Cercignani's conjecture writes

$$D(F) \geq \lambda \|F - \mathbf{P}F\|_2^2 ,$$

where $D(F) = \langle LF, F\rangle$; $\mathbf{P}$ is the orthogonal projection in L^2 onto the null space $\mathcal{N}(L)$. The definitions of L and $\mathcal{N}(L)$ will be introduced later.

To find the explicit rate λ (if exists) will be our goal. Though there

are very few results on the estimates, we have not seen any numerical results that provide the "exact" rate governing the exponential decay to equilibrium. This will be the first attempt to solve this problem numerically.

4.2 The Linearized Boltzmann Operators and Spectral Gaps

Recall the Boltzmann equation and Boltzmann collision operator

$$\frac{\partial f(x,v,t)}{\partial t} + v \cdot \nabla_x f(x,v,t) = Q_{sym}(f,f)(x,v,t)\,, \tag{4.1}$$

$$f(x,v,0) = f_0(x,v)\,. \tag{4.2}$$

Here, the right-hand side symmetrized Boltzmann bilinear operator

$$Q_{sym}(f,g)(v) = \frac{1}{2}\int\int (f'g'_* + f'_*g' - fg_* - f_*g)B(|u|,\sigma)d\sigma dv_*\,, \tag{4.3}$$

where, for simplicity, denote $f' = f(v')$, $f'_* = f(v'_*)$ and $f_* = f(v_*)$, with pre-collisional velocities v', v'_*. The integration is parametrized in terms of the center of mass and relative velocity. And on the $d-1$ dimensional sphere, integration is done with respect to the unit direction given by the elastic post collisional relative velocity.

The elastic law for pre- and post-collisional velocities obeys (2.3) for $\beta = 1$. and the key terms for the model, the *collision kernel* and *angular cross-sections*, are defined in (2.4, 2.5).

For sake of simplicity, when necessary, we use the symmetrized angular cross-section defined over half sphere,

$$\tilde{b}(\cos\theta) = \frac{1}{2}\big(b(\cos\theta) + b(\cos(\pi-\theta))\big)\,, \quad \theta \in [0, \frac{\pi}{2}]\,. \tag{4.4}$$

And for purpose of analysis in the following, we always assume the angular cross-section $b(\cos\theta)$ satisfies the singularity condition (2.5) when $\theta \sim 0$.

Since our interest focuses on the behavior in a regime very close to equilibrium, we consider the perturbation near equilibrium

$$f = \mu + \mu^{\frac{1}{2}} F \,, \tag{4.5}$$

with $\mu = (2\pi)^{-\frac{d}{2}} e^{-\frac{|v|^2}{2}}$ the normalized equilibrium with mass 1, momentum 0 and temperature 1. Then the linearization of homogeneous Boltzmann equation gives an equation for the perturbation $F(v)$,

$$\partial_t F = -L(F) - \Gamma(F, F) \,.$$

where the *linearized Boltzmann collision operator* L writes

$$L(F) = -2\mu^{-\frac{1}{2}} Q_{sym}(\mu, \mu^{\frac{1}{2}} F) \,, \tag{4.6}$$

and the bilinear operator Γ writes

$$\Gamma(F, F) = \int_{\mathbb{R}^d} \int_{\mathbb{S}^{d-1}} \mu_*^{\frac{1}{2}} \left[F F_* - F' F'_* \right] B(|v - v_*|, \sigma) d\sigma dv_* \,,$$

which will be a negligible term when close to equilibrium.

The *Dirichlet form* associated to the linearized Boltzmann operator is

given by

$$
\begin{aligned}
&\langle L(F), F\rangle := -\int_{\mathbb{R}^d} 2Q_{sym}(\mu, \mu^{\frac{1}{2}}F)F\mu^{-\frac{1}{2}}(v)dv \\
&= \frac{1}{4}\int_{\mathbb{R}^{2d}}\int_{\mathbb{S}^{d-1}} \mu\mu_* \left(\frac{F(v')}{\mu^{\frac{1}{2}}(v')} + \frac{F(v'_*)}{\mu^{\frac{1}{2}}(v'_*)} - \frac{F(v)}{\mu^{\frac{1}{2}}(v)} - \frac{F(v_*)}{\mu^{\frac{1}{2}}(v_*)}\right)^2 \\
&\quad \cdot B(u,\sigma)d\sigma dv_* dv \\
&= -\int_{\mathbb{R}^{2d}}\int_{\mathbb{S}^{d-1}} \left[F(v)\mu^{\frac{1}{2}}(v_*) + F(v_*)\mu^{\frac{1}{2}}(v)\right]\left[F(v')\mu^{\frac{1}{2}}(v'_*) - F(v)\mu^{\frac{1}{2}}(v_*)\right] \\
&\quad \cdot B(u,\sigma)d\sigma dv_* dv \\
&= -\int_{\mathbb{R}^{2d}}\int_{\mathbb{S}^{d-1}} \mu(v)\mu(v_*)\left[g(v) + g(v_*)\right]\left[g(v') - g(v)\right]B(u,\sigma)d\sigma dv_* dv
\end{aligned}
\tag{4.7}
$$

where the second line uses the fact that $\mu\mu_* = \mu'\mu'_*$ and the last line changes $g(v) = \frac{F(v)}{\mu^{1/2}(v)}$. The linear operator L has basic properties [36]:

- It is an unbounded symmetric (self-adjoint) operator on un-weighted $L^2(\mathbb{R}^d)$;
- It is a positive operator, i.e has non-negative real spectrum;
- The null space $F(v) \in \mathcal{N}(L) = \mu^{\frac{1}{2}} \cdot \text{span}\{1, v, |v^2|\}$. Thus 0 is an eigenvalue of multiplicity $d+2$.

Here, in order to study the decay of F for $t \to \infty$, we need to study the eigenvalue problem

$$
Lg = \lambda g\,, \tag{4.8}
$$

for which, we know it has $d+2$ eigen-solutions (collision invariants) for $\lambda = 0$. All the other $\lambda > 0$.

If the eigen-solutions of equation (4.8), $g_\lambda(v)$, can be taken as generalized functions, then it's known that the linearized Boltzmann equation

$$\partial_t F = -LF \tag{4.9}$$

has solutions written as [36, 70]

$$F(v,t) = \int_{\lambda_0}^{\lambda_\infty} e^{-\lambda t} g_\lambda(v) h_\lambda(v) d\lambda + \sum_{i=0}^{d+1} h_i(v)\phi_i(v)\,, \tag{4.10}$$

where $h_\lambda(v)$ is an arbitrary function depending on λ and the integrals extends to all $\lambda \neq 0$ for which $g_\lambda \neq 0$ exists. If some λ's form a discrete set, then the corresponding integral is replaced by the sum $\sum_k e^{-\lambda_k t} g_k(v) h_k(v)$. So, if $\lambda_0 \neq 0$ exists, $F(v)$ decays exponentially into the null space $\mathcal{N}(L)$; while if $\lambda_0 = 0$, the decay is not exponential and depends on initial datum.

Definition (Spectral Gap [97])**.** *Denote by $\sigma(L)$ the spectrum for the operator L. For the case $\sigma(L) \subseteq R^+$ (i.e. non-negative spectrum), the spectral gap is defined as the distance between 0 and $\sigma(L) \setminus \{0\}$.*

Hence, the spectral gap is the solution to the constrained minimization problem:

$$\begin{aligned} \min \quad & \frac{\langle L(F), F\rangle}{\| F \|_{L_2}^2} \\ s.t \quad & F \perp \mathcal{N}(L)\,. \end{aligned} \tag{4.11}$$

The solution to the optimization problem tells us how the entropy production functional, given by the Dirichlet form in (4.7), is bounded by the

relative entropy and hence, gives an estimate on the exponential decay of the solutions to the Boltzmann equation.

Thus, both the theoretical as well as numerical existence of this "spectral gap" is very important to us. We will see that the existence of spectral gaps depends on the types of intermolecular potentials (γ) as well as the integrability of the angular cross-section ($b(\cos(\theta))$). We will look at them separately.

4.2.1 Integrable Angular Cross-section

The study on the spectral properties of the linearized Boltzmann collision operator can be traced back all the way to Hilbert [81]. He suggested the splitting, in the case of hard spheres, between the local and non-local parts of L and proved the compactness of the non-local part. Later, Carleman [33] introduced the use of so-called Weyl's theorem to prove the existence of a spectral gap, and Grad [72] generalized it to hard potentials with cutoff ($0 < \gamma \leq 1$). Recently, Caflisch [31] and Golse and Poupaud [71] proved the non-existence of spectral gap for soft potentials with cutoff, but the existence of a "degenerated" spectral gap. All the above results are non-constructive. The first constructive estimates were given by Baranger and Mouhot [10] for the hard spheres model.

For the integrable angular cross-sections, index $\alpha < 0$ in (2.5). Basically, by splitting, L writes

$$L(F)(t,v) = \nu(v)F(t,v) + (\mathbf{K}F)(t,v)\,, \tag{4.12}$$

where the collision frequency

$$\nu(v) = \iint \mu(v_*) dv_* B(|v - v_*|, \hat{u} \cdot \sigma) d\sigma \,, \tag{4.13}$$

and the integral operator $\mathbf{K}$ with kernel $k(v, \eta)$ can be explicitly given in the next section.

Here, an important feature is that the non-local $\mathbf{K}$ is proven to be a compact bounded integral operator.

4.2.1.1 Carleman Representation and Grad Splitting

Due to the integrability of the angular cross-sections, we can separate terms in the Dirichlet form of L (4.7) and easily obtain the non-local $\mathbf{K}F$, which is given by

$$\begin{aligned} \mathbf{K}F(v) &= \mu^{\frac{1}{2}}(v) \iint \mu^{\frac{1}{2}}(v_*) F(v_*) B(|v - v_*|, \sigma) d\sigma dv_* \\ &\quad - \iint \left[\mu^{\frac{1}{2}}(v_*) \mu^{\frac{1}{2}}(v') F(v'_*) + \mu^{\frac{1}{2}}(v_*) \mu^{\frac{1}{2}}(v'_*) F(v') \right] B(|v - v_*|, \sigma) d\sigma dv_* \\ &= \mathbf{K}_1 F - \mathbf{K}_2 F \,, \end{aligned} \tag{4.14}$$

where it's not hard to observe that the kernel $k_1(v, \xi)$ (here $\xi = v_*$) for the integral operator $\mathbf{K}_1$ is

$$k_1(v, \xi) = \mu^{\frac{1}{2}}(v) \mu^{\frac{1}{2}}(\xi) |v - \xi|^{\gamma} \int_{\mathbb{S}^{d-1}} b((v - \xi) \cdot \sigma) d\sigma \,. \tag{4.15}$$

The remaining part of (4.14) defines $\mathbf{K}_2$. The kernel $k_2(v, \xi)$ will be derived explicitly.

We start from *Carleman Representation*, first introduced by Carleman [32] in 3 dimensions for hard spheres. It is actually a transformation of integrals over spheres to integrals over some orthogonal planes.

Lemma 4.2.1 (Carleman)**.** *The following identity holds for any appropriate test functions* $\phi(z)$*:*$\mathbb{R}^d \to \mathbb{R}$

$$\int_{\mathbb{S}^{d-1}} \phi(\frac{|u|\sigma - u}{2}) d\sigma = 2^{d-1}|u|^{2-d} \int_{\mathbb{R}^d} \phi(z)\delta(|z|^2 + z\cdot u)dz\,, \tag{4.16}$$

where $u \in \mathbb{R}^d$ *is an arbitrary vector and* δ *is the one-dimensional Dirac delta function.*

Then, we take the following changes of variables

$$u = v - v_*\,, \quad z = \frac{1}{2}(|u|\sigma - u)\,, \quad w = -\frac{1}{2}(|u|\sigma + u)\,. \tag{4.17}$$

Then, $u = -(z+w)$, $v_* = v + w + z$, $v'_* = v + w$ and $\xi := v' = v + z$. Also assuming a symmetrized angular cross-section (4.4) and noticing that $|v' - v| = |u|\sin(\theta/2)$ and $|u| = (|\xi - v|^2 + |w|^2)^{\frac{1}{2}}$, we obtain the integral form of $\mathbf{K}_2 F$ given by

$$\begin{aligned}
&\mathbf{K}_2 F(v)\\
&= 2^d \int_{\mathbb{R}^{2d}} \mu^{\frac{1}{2}}(v+w+z)\mu^{\frac{1}{2}}(v+w)F(v+z)|u|^{2-d}B(u, \frac{2z+u}{|u|})\delta(z\cdot(z+u))dzdu\\
&= 2^d \int_{\mathbb{R}^{2d}} \mu^{\frac{1}{2}}(v+w+z)\mu^{\frac{1}{2}}(v+w)F(v+z)\tilde{B}(w,z)\delta(z\cdot w)dzdw\\
&= 2^d \int_{\mathbb{R}^d}\int_{w\perp z} |z|^{-1}\mu^{\frac{1}{2}}(v+w+z)\mu^{\frac{1}{2}}(v+w)F(v+z)\tilde{B}(w,z)dzdw\\
&= \frac{2}{\pi}\int_{\mathbb{R}^d}\int_{w\perp z} F(\xi)\mu^{\frac{1}{2}}(\xi+w)\mu^{\frac{1}{2}}(v+w)|\xi - v|^{-d-\alpha}\left(|w|^2 + |\xi - v|^2\right)^{\frac{\gamma+1+\alpha}{2}} d\xi dw\,,
\end{aligned}$$

where we used the relationship $w \perp z$ and

$$\tilde{B}(w,z) = |w+z|^{2-d} B(-(w+z), \frac{z-w}{|z+w|}) = \frac{1}{2^{d-1}\pi} |z|^{-(d-1)-\alpha} \left(|w|^2 + |z|^2\right)^{\frac{\gamma+1+\alpha}{2}} .$$

Therefore, the explicit kernel $k_2(v,\xi)$ for integral operator $\mathbf{K}_2$ can be extracted, which writes

$$k_2(v,\xi) = \frac{2}{\pi} |\xi - v|^{-d-\alpha} \int_\Pi \mu^{\frac{1}{2}}(\xi + w) \mu^{\frac{1}{2}}(v + w) \left(|w|^2 + |\xi - v|^2\right)^{\frac{\gamma+1+\alpha}{2}} dw \,,$$

where the plane $\Pi := \{w \in \mathbb{R}^d : (\xi - v) \cdot w = 0\}$.

However, we can simplify more, following tricks from [38]. Notice that

$$|v + w|^2 + |\xi + w|^2 = 2|w + \frac{1}{2}(\xi + v)|^2 + \frac{1}{2}|\xi - v|^2 \,, \tag{4.18}$$

and decompose $\frac{1}{2}(\xi + v)$ into parts perpendicular to $\xi - v$ and parallel to $\xi - v$. The projection onto $\xi - v$ is denoted by $\zeta^\perp$, which is

$$\zeta^\perp := \left(\frac{1}{2}(\xi + v) \cdot \frac{\xi - v}{|\xi - v|}\right) \frac{\xi - v}{|\xi - v|} = \left(\frac{1}{2} \frac{|\xi|^2 - |v|^2}{|\xi - v|}\right) \frac{\xi - v}{|\xi - v|} \,. \tag{4.19}$$

Its orthogonal part, denoted by ζ, is in the same plane as w,

$$\zeta := \frac{1}{2}(\xi + v) - \zeta^\perp = \frac{1}{2}(\xi + v) - \left(\frac{1}{2} \frac{|\xi|^2 - |v|^2}{|\xi - v|}\right) \frac{\xi - v}{|\xi - v|} \,. \tag{4.20}$$

Thus, plugging these into k_2 gives

$$\begin{aligned} k_2(v,\xi) = \frac{2}{\pi} (2\pi)^{-\frac{d}{2}} |\xi - v|^{-d-\alpha} \exp(-\frac{1}{8}|\xi - v|^2 - \frac{1}{8} \frac{(|\xi|^2 - |v|^2)^2}{|\xi - v|^2}) \\ \cdot \int_\Pi \exp(-\frac{|w + \zeta|^2}{2}) \left(|\xi - v|^2 + |w|^2\right)^{\frac{\gamma+1+\alpha}{2}} dw \,. \end{aligned} \tag{4.21}$$

Clearly, $k_2(v,\xi)$ is symmetric.

Remark. The kernel $k_2(v,\xi)$ can be further simplified if $\gamma+1+\alpha=0$. For example, in the case of 2-d Maxwell model or 3-d hard sphere model, since ζ is just a shift of w on plane Π and thus the integrations on plane Π can be done analytically,

$$k_2(v,\xi)=2^{\frac{1}{2}}\pi^{-\frac{3}{2}}|\xi-v|^{-(d-1)-1-\alpha}\exp(-\frac{1}{8}|\xi-v|^2-\frac{1}{8}\frac{(|\xi|^2-|v|^2)^2}{|\xi-v|^2})\,. \quad (4.22)$$

Thus,

$$L(F)(v)=\nu(v)F(v)+\mathbf{K}F(v)\,, \quad (4.23)$$

where the kernel for the integral operator $\mathbf{K}$ is explicitly given

$$\begin{aligned} k(v,\xi)&=k_1(v,\xi)-k_2(v,\xi)\\ &=(2\pi)^{-\frac{d}{2}}\exp(-\frac{|v|^2+|\xi|^2}{4})|\xi-v|^{\gamma}\int_{S^{d-1}}b(\sigma)d\sigma\\ &-\frac{2}{\pi}(2\pi)^{-\frac{d}{2}}|\xi-v|^{-d-\alpha}\exp(-\frac{1}{8}|\xi-v|^2-\frac{1}{8}\frac{(|\xi|^2-|v|^2)^2}{|\xi-v|^2})\\ &\cdot\int_{\Pi}\exp(-\frac{|w+\zeta|^2}{2})\left(|\xi-v|^2+|w|^2\right)^{\frac{\gamma+1+\alpha}{2}}dw\,, \end{aligned} \quad (4.24)$$

which makes the integral operator $\mathbf{K}$ compact on $L^2(\mathbb{R}^d)$.

Combining (4.15) and (4.21) yields the explicit definition for $\mathbf{K}$, which can be proven to be a Hilbert-Schmidt integral operator (this needs $\alpha<0$ which is satisfied due to the integrability of angular cross-sections), and thus $k_1(v,\xi)-k_2(v,\xi)$ is L^2 integrable. Starting from Carleman representation, we actually have recovered the results from *Grad splitting* [72].

4.2.1.2 The Geometry of Existence of Spectral Gaps

Now, we have separated the linear operator L into a local part and an non-local compact part. The following *Weyl's* theorem will be used.

Theorem 4.2.2 (Weyl's)**.** *The essential spectrum (here, the continuous spectrum due to the self-adjoint L) is unchanged under a compact perturbation.*

We easily get that the information of continuous spectrum is completely contained in the local part $\nu(v)$. If assuming a normalized angular cross-section, i.e. $\int_{S^{d-1}} b(\hat{u}\cdot\sigma)d\sigma = 1$, then, we get the collision frequency

$$\nu(v) = (2\pi)^{-\frac{d}{2}} \int |v - v_*|^\gamma e^{-\frac{|v_*|^2}{2}} dv_* \,. \tag{4.25}$$

- Hard potential model, $\gamma \geq 0$, we can see the continuous spectrum will range from some positive value to infinity. What's left is the discrete spectrum, i.e the eigenvalues. There will be a smallest positive one, which is the spectral gap;

- Soft potential model, $\gamma < 0$, the continuous spectrum can go all the way down to zero; thus we cannot expect a spectral gap. (But, there will be a "degenerate" one.)

The spectrum can be described with pictures, see Figure 4.1, Figure 4.2 and Figure 4.3. Thus the geometry of the spectrum of linearized Boltzmann operators is clear to us. However, we will revisit the details of splitting in Section 4.2.1.1, since a numerical treatment can be designed based on this property of "splitting", see Section 4.4.

4.2.2 Non-integrable Angular Cross-section

In particular, the above "splitting" property does not hold in the usual way with an non-integrable $b(\cos(\theta))$. Thus the above perturbation theories

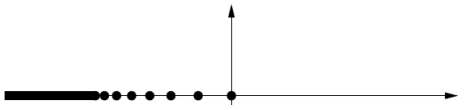


Figure 4.1: Spectrum of $-L$ for variable hard potential with angular cutoff

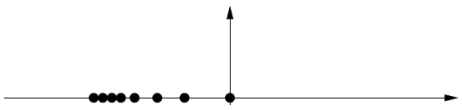


Figure 4.2: Spectrum of $-L$ for Maxwell type with angular cutoff

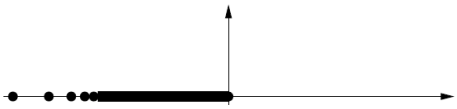


Figure 4.3: Spectrum of $-L$ for soft potential with angular cutoff

may no longer directly apply to the spectrum of non-cutoff linearized Boltzmann. Nevertheless, with a suitable choice of truncated angular domain, which depends on relative velocities, one can still perform some "splitting" and study each term separately. Thus some constructive coercivity estimates for the Dirichlet form can be found and used to characterize the spectral gaps. This is what Mouhot & Strain [99] conjectured and partially proved

Theorem 4.2.3 (Mouhot & Strain [99])**.** *With the collision kernel B specified in (2.4), one has*

- *For any $\epsilon > 0$ there is a constructive constant $C_{B,\epsilon}$, such that the Dirichlet form satisfies:*

$$\langle LF, F\rangle \geq C_{B,\epsilon}\|(F - \mathbf{P}F) < v >^{(\gamma+\alpha-\epsilon)/2} \|^2_{L^2(\mathbb{R}^d)}. \tag{4.26}$$

- *There is a non-constructive constant $C_{B,0}$ such that*

$$\langle LF, F\rangle \geq C_{B,0}\|(F - \mathbf{P}F) < v >^{(\gamma+\alpha)/2} \|^2_{L^2(\mathbb{R}^d)}. \tag{4.27}$$

where $\langle v\rangle = (1+|v|^2)^{\frac{1}{2}}$ and $\mathbf{P}$ is the orthogonal projector onto the null space $\mathcal{N}(L)$. So, it's sufficient to claim that when $\gamma + \alpha \geq 0$, there exists a spectral gap for linearized Botlzmann operator. But they went further and conjectured the necessary part

Theorem 4.2.4 (Mouhot & Strain [99]). *With $\gamma \in (-d, \infty)$ and $\alpha \in [0, 2)$ in B, the linearized Boltzmann collision operator associated to the kernel B (2.4) admits a spectral gap if and only if $\gamma + \alpha \geq 0$. Moreover, this statement is still valid if one includes formally the case of angular cutoff in "$\alpha = 0$", and adds the linearized Landau collision operator as the limit case "$\alpha = 2$".*

Recently the necessary part was answered by Gressman and Strain [73], by proving sharp constructive upper and lower bounds for the linearized collision operator in terms of a geometric fractional Sobolev norm.

4.3 A New Nonconstructive Proof of the Existence of Spectral Gaps for Non-integrable Angular Cross-sections

The sufficient condition on existence of spectral gaps for the linearized Boltzmann operators has been proved [99], in term of L^2 norms and the the necessary part is conjectured. Later, a complete theorem on the existence of spectral gaps is built when proving the existence of global classical solutions for the Boltzmann equation with non-cutoff interactions [73]. However, the latter is developed in terms of a very complicated weighted non-isotropic Sobolev norm.

Our current numerical work aims at verifying the existence of spectral gaps proved in and also answering the conjecture proposed in [99], both in L^2 norms. Thus, we would also like to provide a proof on the existence in L^2 spaces. Here, a non-constructive one is enough, since we will practically compute the spectral gaps. Inspired by the Grad splitting method in Section 4.2.1 for study on the geometry of spectral gaps for integrable angular cross-section, we generalize Grad's derivations to arbitrary intramolecular potential γ and α. This is how our new nonconstructive proof comes out, which will be stated in details below.

We have to carefully treat the singularities in $b(\frac{u\cdot\sigma}{|u|})$ and find out a way to cancel the singularity. This is done by cut the sphere $\mathbb{S}^{d-1}$ into two parts: with a small and fixed parameter $\epsilon > 0$,

$$\Sigma_s = \{\sigma \in \mathbb{S}^{d-1} : |v' - v| \leq \epsilon\}, \qquad \Sigma_r = \{\sigma \in \mathbb{S}^{d-1} : |v' - v| > \epsilon\}, \quad (4.28)$$

where the subscript "s" denotes "singular" part containing the singularity neighborhood; "r" denotes "regular" part with singularity excluded.

Recall the Dirichlet form of the linearized Boltzmann operator L in (4.7)(the last line of formulas). With the same parameter ϵ and cuts of sphere domain (4.28), we define an approximate operator L^ϵ, such that

$$\langle L^\epsilon F, F\rangle := \\ -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v - v_*|^\gamma(g + g_*)\left(\int_{\Sigma_s} \nabla g(v)\cdot(v' - v) + \int_{\Sigma_r}(g' - g)\right) b(\sigma)d\sigma dv dv_*\,, \quad (4.29)$$

with a relationship $g(v) = \mu^{-1/2}(v)F(v)$. Here, for the time being, we assume F or g is smooth up to second order (this is true for the null space of L). Later, we will release the smooth assumption.

Consider the Taylor expansion

$$g(v') - g(v) = \nabla g(v) \cdot (v' - v) + O(|v' - v|^2)\,, \tag{4.30}$$

where $O(|v' - v|^2)$ is bounded by $|v' - v|^2$ times some constant depending on $\nabla^2 g$. It's not difficult to observe that, for $F(v) \in L^2$,

$$\langle L^\epsilon F, F\rangle \to \langle LF, F\rangle\,, \tag{4.31}$$

as $\epsilon \searrow 0$. So, it's enough to study the geometry of the spectral gaps for the approximate operator L^ϵ.

Corresponding to the cuts of sphere domain (4.28), we can also define the singular and regular parts for L^ϵ, namely,

$$\langle L^\epsilon F, F\rangle = \langle \mathbf{S}_1(F), F\rangle + \langle \mathbf{S}_2(F), F\rangle + \langle \mathbf{R}(F), F\rangle\,, \tag{4.32}$$

where $\mathbf{S}_1$ and $\mathbf{S}_2$ are the singular parts coming from integrals over Σ_s; $\mathbf{R}$ is the remaining regular integrals over Σ_r. In particular,

$$\begin{aligned}
\langle \mathbf{S}_1(F), F\rangle &= -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_* |v - v_*|^\gamma g \int_{\Sigma_s} \nabla g(v) \cdot (v' - v) b(\sigma) d\sigma dv dv_*\,,\\
\langle \mathbf{S}_2(F), F\rangle &= -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_* |v - v_*|^\gamma g_* \int_{\Sigma_s} \nabla g(v) \cdot (v' - v) b(\sigma) d\sigma dv dv_*\,,\\
\langle \mathbf{R}(F), F\rangle &= -\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_* |v - v_*|^\gamma (g + g_*) \int_{\Sigma_r} (g' - g) b(\sigma) d\sigma dv dv_*\,.
\end{aligned} \tag{4.33}$$

We will study them separately.

(1) Singular parts $\mathbf{S}_1 F$ and $\mathbf{S}_2 F$.

First, let's look at the angular integrals

$$I_{\Sigma_s} := \int_{\Sigma_s} \nabla g(v) \cdot (v' - v) b(\sigma) d\sigma \,. \tag{4.34}$$

Consider a local spherical coordinate system with u being the polar direction. Then, take a transformation which rotates the polar direction back onto z-axis of the standard Cartesian coordinate system. The orthogonal rotation matrix A can be constructed explicitly, with relative velocity $u = v - v_*$. They are given by

$d = 2$:

$$A = \frac{1}{|u|} \begin{pmatrix} -u_2 & u_1 \\ u_1 & u_2 \end{pmatrix} ; \tag{4.35}$$

$d = 3$:

$$A = \frac{1}{|u|} \begin{pmatrix} \frac{u_1 u_3}{\sqrt{u_1^2+u_2^2}} & \frac{u_2 u_3}{\sqrt{u_1^2+u_2^2}} & -\sqrt{u_1^2+u_2^2} \\ -\frac{u_2|u|}{\sqrt{u_1^2+u_2^2}} & \frac{u_1|u|}{\sqrt{u_1^2+u_2^2}} & 0 \\ u_1 & u_2 & u_3 \end{pmatrix} , \tag{4.36}$$

where we assume $u_1^2 + u_2^2 \neq 0$. Otherwise, the rotation matrix is reduced to the identity matrix (with possibly minus sign).

Then, consider a change of variable $\sigma \leftarrow A^{-1}\sigma = A^T \sigma$, for which the Jacobian is 1. If denote by θ the angle between u and σ, as exactly defined in (2.5), and φ is the corresponding azimuthal angle. Recall post-collisional

velocity $v' = v + \frac{1}{2}(|u|\sigma - u)$, we have

$$v' - v =: z = \begin{cases} \dfrac{|u|}{2} A^T \left(\sin\theta, \cos\theta - 1\right)^T, \text{ if } d = 2 \\ \dfrac{|u|}{2} A^T \left(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta - 1\right)^T, \text{ if } d = 3\,. \end{cases} \tag{4.37}$$

We will take $d = 3$ for example. It's not hard to see that the azimuthal angle φ is integrated out due to its periodicity over $[0, 2\pi]$. With (4.36), (4.37) and a change of variable $t = \sin\frac{\theta}{2}$, we have

$$\begin{aligned} I_{\Sigma_s} &= \int_{|v'-v|\le\epsilon} \nabla g(v) \cdot z b(\sigma) d\sigma \\ &= \int_{|u|\sin\frac{\theta}{2}\le\epsilon} \int_{\varphi=0}^{2\pi} \frac{|u|}{2} A\nabla g(v) \cdot \left(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta - 1\right)^T d\varphi b(\sigma)\sin\theta d\theta \\ &= 2\pi \int_{|u|\sin\frac{\theta}{2}\le\epsilon} \frac{|u|}{2} \big(A\nabla g(v)\big)_3 (\cos\theta - 1) b(\sigma)\sin\theta d\theta \\ &= \frac{1}{2} \int_{|u|\sin\frac{\theta}{2}\le\epsilon} \frac{1}{2} u \cdot \nabla g(v) (-2\sin^2\frac{\theta}{2}) \big(\sin^{-2-\alpha}\frac{\theta}{2}\big) \big(4\sin\frac{\theta}{2}\big) d\sin\frac{\theta}{2} \\ &= -2u \cdot \nabla g(v) \int_0^{\frac{\epsilon}{|u|}} t^{1-\alpha} dt \\ &= -2\frac{u \cdot \nabla g(v)}{2-\alpha} \left(\frac{\epsilon}{|u|}\right)^{2-\alpha}, \end{aligned} \tag{4.38}$$

where $u = v - v_*$ denotes the relative velocity. The formula (4.38) is well-defined, because in the Boltzmann regime, we require $\alpha < 2$.

Plugging (4.38) back into (4.33) and applying the Divergence theorem

on variable v, gives

$$\begin{aligned}
&\langle \mathbf{S}_1(F), F\rangle \\
&= c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v-v_*|^{\gamma+\alpha-2} g\nabla g(v)\cdot(v-v_*)dvdv_* \\
&= -c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \nabla\cdot\left(\mu\mu_* g|v-v_*|^{\gamma+\alpha-2}(v-v_*)\right) g(v)dvdv_* \\
&= -\langle \mathbf{S}_1(F), F\rangle \\
&\quad + c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v-v_*|^{\gamma+\alpha-2}\left(v\cdot(v-v_*) - (\gamma+\alpha-2+d)\right) g^2(v)dvdv_*\,,
\end{aligned} \tag{4.39}$$

where the constant $c = 2\frac{\epsilon^{2-\alpha}}{2-\alpha}$. Thus,

$$\langle \mathbf{S}_1(F), F\rangle = \frac{1}{2}c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v-v_*|^{\gamma+\alpha-2}\left(v\cdot(v-v_*) - (\gamma+\alpha-2+d)\right) g^2(v)dvdv_*\,. \tag{4.40}$$

To further simplify (4.40), we make a change of variables $(v, v_*) \to (v, u)$ in (4.33) and performing integration by parts again w.r.t v. We get

$$\begin{aligned}
&\langle \mathbf{S}_1(F), F\rangle = c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu(v)\mu(v-u)|u|^{\gamma+\alpha-2} g(v)\nabla g(v)\cdot u dvdu \\
&= -c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \nabla\left(\mu(v)\mu(v-u)g(v)\right)\cdot u|u|^{\gamma+\alpha-2} g(v)dvdu \\
&= -\langle \mathbf{S}_1(F), F\rangle + c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu(v)\mu(v-u)|u|^{\gamma+\alpha-2}\left(2v\cdot u|u|^2\right) g^2(v)dvdu\,.
\end{aligned} \tag{4.41}$$

Thus,

$$\langle \mathbf{S}_1(F), F\rangle = \frac{1}{2}c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v-v_*|^{\gamma+\alpha-2}\left(2v\cdot(v-v_*) - |v-v_*|^2\right) g^2(v)dvdv_*\,. \tag{4.42}$$

Taking twice (4.40) and subtracting (4.42) gives

$$\langle \mathbf{S}_1(F), F\rangle = \frac{1}{2}c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_*|v-v_*|^{\gamma+\alpha-2}\left(|v-v_*|^2 - 2(\gamma+\alpha-2+d)\right) g^2(v)dvdv_*\,. \tag{4.43}$$

Similarly, we have

$$\langle \mathbf{S}_2(F), F\rangle = \frac{1}{2}c\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu\mu_* g g_*|v-v_*|^{\gamma+\alpha-2}\left(|v-v_*|^2 - 2(\gamma+\alpha-2+d)\right) dvdv_*\,. \tag{4.44}$$

So, the operators $\mathbf{S}_1$ and $\mathbf{S}_2$ can be defined explicitly by extracting the integral kernels in (4.43) and (4.44). In addition, $\mathbf{S}_1$ acts as a local operator and $\mathbf{S}_2$ is a global one.

(2) Regular parts $\mathbf{R}F$.

With a neighborhood of singularity being removed, the remaining angular cross-section is integrable

$$I_\sigma := \int_{\Sigma_r} b(\sigma)d\sigma = 2\begin{cases} -\log\frac{\epsilon}{|u|} & \alpha = 0; \\ \frac{1}{\alpha}\left(\left(\frac{\epsilon}{|u|}\right)^{-\alpha} - 1\right) & \alpha > 0. \end{cases}\,. \tag{4.45}$$

where $u = v - v_*$ is the relative velocity. Thus, we can split the "gain" and "loss" terms and study them separately, following a similar argument as for integrable angular cross-sections. That is, the reformulating is expected

$$\mathbf{R}(F)(v) = \nu^\epsilon(v)F(v) + \mathbf{K}_1^\epsilon F(v) - \mathbf{K}_2^\epsilon F(v)\,, \tag{4.46}$$

where we denote the "truncated" collision frequency

$$\nu^\epsilon(v) = \int \mu(v_*)|v-v_*|^\gamma I_\sigma dv_*\,, \tag{4.47}$$

the kernel $k_1^\epsilon(v, v_*)$ for $\mathbf{K}_1^\epsilon$

$$k_1^\epsilon(v, v_*) = \mu^{\frac{1}{2}}(v)\mu^{\frac{1}{2}}(v_*)|v - v_*|^\gamma I_\sigma \,, \tag{4.48}$$

and the kernel $k_2^\epsilon(v, \xi)$ for $\mathbf{K}_2^\epsilon$ can be explicitly derived, based on the Carleman representation Lemma 4.2.1 (with a symmetrized angular cross-section). Then, following a quite similar argument in Section 4.2.1.1, we obtain

$$\begin{aligned}&\mathbf{K}_2^\epsilon F(v)\\ &= \frac{2}{\pi}\int_{|z|>\epsilon}\int_{w\perp z} F(\xi)\mu^{\frac{1}{2}}(\xi + w)\mu^{\frac{1}{2}}(v + w)|\xi - v|^{-d-\alpha}\left(|w|^2 + |\xi - v|^2\right)^{\frac{\gamma+1+\alpha}{2}} d\xi dw \,,\end{aligned}$$

With a change of variable $v + z \to \xi$ and same derivations in Section 4.2.1.1, We can finally extract an explicit kernel for $\mathbf{K}_2^\epsilon$

$$\begin{aligned}k_2^\epsilon(v, \xi) = \frac{2}{\pi}(2\pi)^{-\frac{d}{2}}\mathbf{1}_{|\xi-v|>\epsilon}|\xi - v|^{-d-\alpha}\exp(-\frac{1}{8}|\xi - v|^2 - \frac{1}{8}\frac{(|\xi|^2 - |v|^2)^2}{|\xi - v|^2})\\ \cdot \int_{w\perp(\xi-v)} \exp(-\frac{|w + \zeta|^2}{2})\left(|\xi - v|^2 + |w|^2\right)^{\frac{\gamma+1+\alpha}{2}} dw \,,\end{aligned} \tag{4.49}$$

where $\mathbf{1}_{|\xi-v|>\epsilon}$ is the characteristic function over the domain $|\xi - v| > \epsilon$ and ζ is defined in (4.20).

Till now, we have completed the reformulating of the Dirichlet form and thus obtained an approximating formula, L^ϵ, for the linearized operator L. The approximated L^ϵ depends on the parameter ϵ because of the Taylor expansion used in formulating the singular parts. Let's summarize here (changing back $F(v) = \mu^{1/2}(v)g(v)$):

$$L^\epsilon F(v) = (s_1(v) + \nu^\epsilon(v))F(v) + \mathbf{S}_2 F + \mathbf{K}_1^\epsilon F(v) - \mathbf{K}_2^\epsilon F(v) \,, \tag{4.50}$$

where ν^ϵ is defined in (4.47) and the kernel for the local integral operator $\mathbf{S}_1$ is

$$s_1(v) = \frac{\epsilon^{2-\alpha}}{2-\alpha} \int_{\mathbb{R}^d} \mu_* |v - v_*|^{\gamma+\alpha-2} \left(|v - v_*|^2 - 2(\gamma + \alpha - 2 + d) \right) dv_* \,. \quad (4.51)$$

the kernel $s_2(v, \xi)$ for the non-local integral operator $\mathbf{S}_2$ is given by

$$s_2(v, \xi) = \frac{\epsilon^{2-\alpha}}{2-\alpha} \mu(v)\mu(\xi) |v - \xi|^{\gamma+\alpha-2} \left(|v - \xi|^2 - 2(\gamma + \alpha - 2 + d) \right) \,, \quad (4.52)$$

and the kernel $k_1^\epsilon(v, \xi)$ for $\mathbf{K}_1^\epsilon$ is defined in (4.48) (note here, actually $\xi = v_*$ and thus $u = v - \xi$.) And the kernel $k_2^\epsilon(v, \xi)$ for $\mathbf{K}_2^\epsilon$ is given above in (4.49). Please note, following the reformulation of L^ϵ in (4.50), we have actually released the regularity assumption on F.

It's not hard to verify that L^ϵ, defined in (4.50), is self-adjoint on L^2. A very interesting observation is that, $s_1(v) + \nu^\epsilon(v)$ behaves like $\int \exp(-\frac{|v-u|^2}{2}) |u|^{\gamma+\alpha} du$ $\sim (1 + |v|)^{\gamma+\alpha}$. The operator $\mathbf{S}_2 + \mathbf{K}_1^\epsilon + \mathbf{K}_2^\epsilon$, with ϵ fixed, can be seen as a perturbation. If the perturbation could be shown to be compact, then, according to Weyl's theorem, the continuous spectrum for the self-adjoint operator L^ϵ will be characterized only by the kernel $s_1(v) + \nu^\epsilon(v)$. So, when $\gamma + \alpha \geq 0$, $s_1(v) + \nu^\epsilon(v)$ is bounded from below by a positive number and thus we can expect a spectral gap; otherwise, there is no spectral gap. This is our following theorem.

Theorem 4.3.1. *The approximate self-adjoint linear operator L^ϵ, defined in (4.29) or (4.50), with the small parameter ϵ fixed, has a spectral gap if and only if $\gamma + \alpha \geq 0$.*

Proof. Let's only consider the case $d = 3$ and $\alpha > 0$. The case $\alpha = 0$ will follow the same way. First, let's find the lower and upper bound for $s_1(v) + \nu^\epsilon(v)$.

Notice that

$$\int \exp(-\frac{|v|^2}{2})|v - v_*|^s dv_* \sim (1 + |v|)^s \ . \tag{4.53}$$

Thus, from (4.47) and (4.51),

$$\begin{aligned}
&\nu^\epsilon(v) + s_1(v) \\
&= \left(\frac{2\epsilon^{-\alpha}}{\alpha} + \frac{\epsilon^{2-\alpha}}{2-\alpha}\right) \int_{\mathbb{R}^3} \mu(v_*)|v - v_*|^{\gamma+\alpha} dv_* - \frac{2}{\alpha} \int_{\mathbb{R}^3} \mu(v_*)|v - v_*|^\gamma dv_* \\
&- \frac{2\epsilon^{2-\alpha}}{2-\alpha}(\gamma + \alpha + 1) \int_{\mathbb{R}^d} \mu(v_*)|v - v_*|^{\gamma+\alpha-2} dv_* \\
&\sim \left(\epsilon^{-\alpha} + \epsilon^{2-\alpha}\right)(1 + |v|)^{\gamma+\alpha} - (1 + |v|)^\gamma - \epsilon^{2-\alpha}(1 + |v|)^{\gamma+\alpha-2} \\
&= (1 + |v|)^{\gamma+\alpha} \left((\epsilon^{-\alpha} + \epsilon^{2-\alpha}) - (1 + |v|)^{-\alpha} - (1 + |v|)^{-2}\right) \\
&= c(\epsilon, \alpha)(1 + |v|)^{\gamma+\alpha} \ ,
\end{aligned} \tag{4.54}$$

where $c(\epsilon, \alpha) > 0$ is some constant independent of v and bounded from below by a positive constant independent of parameter ϵ. Thus, if $\gamma + \alpha \geq 0$, $\nu^\epsilon(v) + s_1(v)$ is bounded from below by a positive constant; otherwise, if $\gamma + \alpha < 0$, $\nu^\epsilon(v) + s_1(v)$ goes to zero for large $|v|$, so there is a uniform control from below.

Next, we show $\mathbf{S}_2 + \mathbf{K}_1^\epsilon$ and $\mathbf{K}_2^\epsilon$ are both compact operators on $L^2(\mathbb{R}^3)$, by proving that they are actually Hilbert-Schmidt integral operators. That mean, their kernels are L^2 integrable.

Indeed, $k_1^\epsilon(v, \xi) + s_2(v, \xi)$ is L^2 integrable because each term, $\mu^{\frac{1}{2}}(v)\mu^{\frac{1}{2}}(\xi)|v - \xi|^{\gamma+\alpha}$, $\mu^{\frac{1}{2}}(v)\mu^{\frac{1}{2}}(\xi)|v - \xi|^\gamma$ and $\mu^{\frac{1}{2}}(v)\mu^{\frac{1}{2}}(\xi)|v - \xi|^{\gamma+\alpha-2}$, is L^2 integrable. In

addition, the kernel $k_2^\epsilon(v,\xi)$ for operator $\mathbf{K}_2^\epsilon$, defined in (4.49), satisfies the following estimates.

Take a sequence, with index $\delta > 0$,

$$k_2^{\epsilon,\delta}(v,\xi) := k_2^\epsilon(v,\xi)\mathbf{1}_{|\frac{v\cdot(\xi-v)}{|v||\xi-v|}|\geq\delta} \tag{4.55}$$

Since the set of compact operators is closed in $L^2(\mathbb{R}^3)$, it's enough to show $k_2^{\epsilon,\delta}(v,\xi)$ is L^2 integrable.

The integral over the plane $w\perp(\xi-v)$ can be estimated as

$$\begin{aligned}
&\int_{w\perp(\xi-v)} \exp(-\frac{|w+\zeta|^2}{2})|\xi-v+w|^{\gamma+1+\alpha}dw \\
&= \int_{w\perp(\xi-v)} \exp(-\frac{|w|^2}{2})|\xi-v+w-\zeta|^{\gamma+1+\alpha}dw \\
&\leq C\begin{cases}(1+|\zeta|)^{\gamma+1+\alpha}(1+|\xi-v|)^{\gamma+1+\alpha}, \text{ if } \gamma+1+\alpha\geq 0\\ (|\zeta|^2+|\xi-v|^2)^{(\gamma+1+\alpha)/2}, \text{ elsewhere}\end{cases},
\end{aligned} \tag{4.56}$$

where constant C is uniform in γ and α, and the estimate for $\gamma+1+\alpha<0$ uses the fact $\zeta\perp(\xi-v)$, (see (4.20)).

When setting $z=\xi-v$, from (4.20),

$$\begin{aligned}
|\zeta|^2 &= \frac{1}{4}\left[|2v+z|^2-\frac{(|v+z|^2-|v|^2)^2}{|z|^2}\right] \\
&= \frac{1}{4}\left[|2v+z|^2-\frac{(|z|^2+2v\cdot z)^2}{|z|^2}\right] \\
&= \frac{1}{4}\left[|2v+z|^2-\left(|z|+2v\cdot\frac{z}{|z|}\right)^2\right] \\
&= |v|^2\left(1-\left(\frac{v\cdot z}{|v||z|}\right)^2\right).
\end{aligned} \tag{4.57}$$

Plugging back into $k_2^{\epsilon,\delta}(v,\xi)$ and transferring to the spherical coordinate, if $\gamma+1+\alpha\geq 0$, then the following estimate holds

$$
\begin{aligned}
&\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\left(k_2^{\epsilon,\delta}(v,\xi)\right)^2 dvd\xi\\
&\leq C\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\mathbf{1}_{|\xi-v|>\epsilon}|\xi-v|^{-2(3+\alpha)}\exp(-\frac{1}{4}|\xi-v|^2-\frac{1}{4}\frac{(|\xi|^2-|v|^2)^2}{|\xi-v|^2})\\
&\qquad\cdot(1+|\zeta|)^{2(\gamma+1+\alpha)}(1+|\xi-v|)^{2(\gamma+1+\alpha)}dvd\xi\\
&\leq C\int_{|z|=\epsilon}^{\infty}|z|^{-4-2\alpha}(1+|z|)^{2(\gamma+1+\alpha)}\exp(-\frac{|z|^2}{4})\int_{|v|=0}^{\infty}|v|^2\\
&\qquad\cdot\int_0^{\pi}\exp(-\frac{(|z|+2|v|\cos\theta)^2}{4})(1+|v|\sin\theta)^{2(\gamma+1+\alpha)}\sin\theta\mathbf{1}_{|\cos\theta|\geq\delta}d\theta d|v|d|z|\\
&\leq C\int_{|z|=\epsilon}^{\infty}|z|^{-4-2\alpha}(1+|z|)^{2(\gamma+1+\alpha)}\exp(-\frac{|z|^2}{4})\int_{|v|=0}^{\infty}|v|^2(1+|v|)^{2(\gamma+1+\alpha)}\\
&\qquad\cdot\int_0^{\pi}\exp(-\frac{(|z|+2|v|\cos\theta)^2}{4})\sin\theta\mathbf{1}_{|\cos\theta|\geq\delta}d\theta d|v|d|z|\\
&\leq C\int_{|z|=\epsilon}^{\infty}|z|^{-4-2\alpha}(1+|z|)^{2(\gamma+1+\alpha)}\exp(-\frac{|z|^2}{8})d|z|\\
&\qquad\cdot\int_{|v|=0}^{\infty}|v|^2(1+|v|)^{2(\gamma+1+\alpha)}\exp(-\frac{\delta^2|v|^2}{3})d|v|\\
&<C_1^{\epsilon,\delta},
\end{aligned}
\tag{4.58}
$$

where $C_1^{\epsilon,\delta}$ is a constant only depending on parameters ϵ and δ, when γ and α are fixed.

Indeed, in the above estimates for integral w.r.t θ,

$$\begin{aligned}
&|z|^2 + (|z| + 2|v|\cos\theta)^2 \\
&= 2|z|^2 + 4|v|^2\cos^2\theta + 4|z||v|\cos\theta \\
&= 2|z|^2 + 4|v|^2\cos^2\theta + t|z|\cdot\frac{4}{t}|v|\cos\theta \\
&\geq \left(2 - \frac{\tau t^2}{2}\right)|z|^2 + 4\left(1 - \frac{2}{\tau t^2}\right)|v|^2\cos^2\theta \\
&\geq \left(2 - \frac{\tau t^2}{2}\right)|z|^2 + 4\left(1 - \frac{2}{\tau t^2}\right)|v|^2\delta^2\,,
\end{aligned} \tag{4.59}$$

where τ is the parameter in the Young's inequality. So, an estimate with $2 < \tau t^2 < 4$, e.g $t = 2$ and $\tau = \frac{3}{4}$, would serve our purpose.

Similarly, if $\gamma + 1 + \alpha < 0$,

$$\begin{aligned}
&\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\left(k_2^{\epsilon,\delta}(v,\xi)\right)^2 dvd\xi \\
&\leq C\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\mathbf{1}_{|\xi-v|>\epsilon}|\xi - v|^{-2(3+\alpha)}\exp(-\frac{1}{4}|\xi-v|^2 - \frac{1}{4}\frac{(|\xi|^2-|v|^2)^2}{|\xi-v|^2}) \\
&\qquad\cdot|\zeta|^{\gamma+1+\alpha}|\xi-v|^{\gamma+1+\alpha}dvd\xi \\
&\leq C\int_{|z|=\epsilon}^{\infty}|z|^{\gamma-3-\alpha}\exp(-\frac{|z|^2}{8})d|z|\int_{|v|=0}^{\infty}|v|^{\gamma+3+\alpha}\exp(-\frac{\delta^2|v|^2}{3})d|v| \\
&\qquad\cdot\int_0^{\pi}\sin^{\gamma+2+\alpha}\theta\mathbf{1}_{|\cos\theta|\geq\delta}d\theta \\
&< C_2^{\epsilon,\delta}\,,
\end{aligned} \tag{4.60}$$

by noticing that $\gamma + 1 + \alpha > -2$. Here, $C_2^{\epsilon,\delta}$ is also a constant only depending on parameters ϵ and δ, when γ and α are fixed.

At last, for ϵ fixed, we show

$$\lim_{\delta\to 0^+}\int\int\left(k_2^{\epsilon,\delta}(v,\xi) - k_2^{\epsilon}(v,\xi)\right)^2 dvd\xi\,. \tag{4.61}$$

Actually, simply replace $\mathbf{1}_{|\cos\theta|\geq\delta}$ by $\mathbf{1}_{|\cos\theta|<\delta}$ in (4.58) and (4.60), respectively. Observe that, when $\gamma+1+\alpha\geq 0$, from (4.58),

$$\begin{aligned}
&\int_0^\pi \exp(-\frac{(|z|+2|v|\cos\theta)^2}{4})\sin\theta\mathbf{1}_{|\cos\theta|<\delta}d\theta \\
&= \exp(-\frac{|z|^2}{4})\int_0^\pi \exp(-(|v|^2\cos^2\theta+|v|\cos\theta))\sin\theta\mathbf{1}_{|\cos\theta|<\delta}d\theta \qquad (4.62)\\
&\leq \exp(-\frac{|z|^2}{4})\int_{-\delta}^{\delta}\exp(-|v|^2\cos^2\theta)d\cos\theta\to 0 \text{ as } \delta\to 0^+,
\end{aligned}$$

and when $-2<\gamma+1+\alpha<0$, from (4.60),

$$\int_0^\pi \sin^{\gamma+2+\alpha}\theta\mathbf{1}_{|\cos\theta|<\delta}d\theta\to 0 \text{ as } \delta\to 0^+. \qquad (4.63)$$

Finally, according to Weyl's theorem, we conclude that L^ϵ and $(\nu^\epsilon(v)+s_1(v))\mathbf{I}$ have the same essential spectrum (here, the continuous spectrum due to the self-adjointness of L^ϵ). From the above, we can conclude that, $(\nu^\epsilon(v)+s_1(v))\mathbf{I}$ has a continuous spectrum positively bounded from below if and only if $\gamma+\alpha\geq 0$. Thus, there exists a spectral gap.

□

Now Theorem 4.3.1 has shown the existence of spectral gap for the linearized operator L^ϵ under cutoff condition (4.29). However, it still remains an open problem on the existence of spectral gap for the original linearized Boltzmann operator L, in the sense of L^2-norm relaxation of perturbation $F(v)$. We conjecture that, as long as one can prove the uniformity when passing the cutoff parameter ϵ to zero, the following result holds

Conjecture 1. *The linearized Boltzmann operator L, defined in (4.6) or (4.7), has a spectral gap, in the sense that, there exists a positive constant λ such that*

$$\langle LF, F\rangle \geq \lambda \|(F - \mathbf{P}F)\|^2_{L^2(\mathbb{R}^d)}\,, \tag{4.64}$$

if and only if $\gamma + \alpha \geq 0$.

Remark. In order to see the estimate (4.64) in the sense of a norm different than L^2, readers can appeal to the argument in [73].

4.4 The Discontinuous Galerkin Projections

In this section, we introduce how to project the original eigenvalue problem onto a finite approximation space, based on Discontinuous Galerkin methods. The key is the treatment of the angular integrals over the $d-1$ dimensional sphere $\mathbb{S}^{d-1}$. Our DG approximation can handle both integrable and non-integrable angular cross-sections, and has also served a foundation for the development of the deterministic DG solvers for fully nonlinear Boltzmann equations, also done by the author. Particularly, for operators with integrable angular cross-sections, it can be specially reformulated based on so-called "Grad splitting", done in Section 4.2.1.1 and can be easily projected onto our DG meshes.

Albeit the high complexity of DG discretizations, we still prefer the DG scheme, since with little knowledge of eigenfunction behaviors, DG approximations are expected to accommodate various kinds of regular and/or

irregular eigenfunctions and thus to provide more accurate eigenvalues. In order to apply DG, we first need to build a reasonable truncated domain.

4.4.1 Domain and Mesh

Let's recall the Dirichlet form for the operator L (4.7) and the equivalent minimization problem for the spectral gap (4.11). If we employ change of variables, $g(v) = \frac{F(v)}{\mu^{1/2}(v)}$, then, equivalently, the spectral gap problem becomes

$$\begin{aligned} \min \quad & \frac{\langle L(F), F\rangle}{\| g \|^2_{L_2(\mu)}}, \\ \text{s.t.} \quad & g \perp \left(\mu^{-\frac{1}{2}} \cdot \mathcal{N}(L)\right), \end{aligned} \tag{4.65}$$

where $\| \cdot \|_{L_2(\mu)}$ is the weighted L^2 norm with weight $\mu(\cdot)$.

It's not difficult to observe that, $g(v)$ can be restricted onto a truncated domain, $\Omega_v = [-V, V)^d$, which is large enough such that the objective function and constraint in (4.65) will only differ than their real values within small errors, respectively. Besides, since the whole linearization only makes sense at the regime very close to equilibrium, it is still reasonable only to consider perturbations $F(v)$ with the same "compact support" as $\mu(v)$. Thus, in the following, our computational domain is the truncated set Ω_v, for $g(v)$ and/or $F(v)$.

Remark. It is vitally important to pay attention to the domain truncation here. With a velocity cutoff, we are actually dealing with the corresponding cutoff operator

$$L_\Omega = \chi_\Omega L, \tag{4.66}$$

which will definitely possess a spectral gap due to the finite integration domain. Though, see (4.25) and analysis below for example, this will not essentially influence the spectral gap for $\gamma \geq 0$, yet for soft potential case, $\chi_\Omega L$ is expected to have a "spectral gap" bounded by $\chi_\Omega \mu(v)$, up to some constant factors. However, as Ω gets larger, we can expect this "spectral gap" goes to zero. An analytical reasoning is provided in the convergence analysis.

A regular mesh is applied, that is, we divide each direction into N disjoint elements uniformly, such that $[-L, L] = \bigcup_k I_k$, where interval $I_k = [w_{k-\frac{1}{2}}, w_{k+\frac{1}{2}})$, $w_k = -L + (k + \frac{1}{2})\Delta v$, $\Delta v = \frac{2L}{n}$, $k = 0 \dots n-1$ and thus there is a Cartesian partitioning $\mathcal{T}_h = \bigcup_k E_k$, with uniform cubic element $E_k = I_{k_1} \otimes I_{k_2} ... \otimes I_{k_d}$, $k = (k_1, k_2, ..., k_d)$.

Discontinuous Galerkin methods assume piecewisely defined basis functions, that is

$$g(v) = \sum_k \mathbf{u}_k \cdot \Phi(v) \chi_k(v)\,, \tag{4.67}$$

where multi-index $k = (k_1, k_2, ..., k_d)$, $0 \leq |k| < (n-1)^3$; $\chi_k(v)$ is the characteristic function over element E_k; coefficient vector $\mathbf{u}_k = (\mathbf{u}_k^0, ..., \mathbf{u}_k^p)$, where p is the total number of basis functions locally defined on E_k; basis vector $\Phi(v) = (\phi_0(v), ..., \phi_p(v))$. Usually, we choose element of basis vector $\Phi(v)$ as local polynomial in $P^p(E_k)$, which is the set of polynomials of total degree at most p on E_k.

4.4.2 Evaluations of Collision Integrals

For Boltzmann-type equations, the treatment of various collision kernels always remains the most important and challenging part. To demonstrate our scheme, for simplicity, we take piecewise constant basis functions as example, i.e. $p = 0$, such that only the characteristic function $\chi_k(v)$ is applied over each element E_k. Due to the possible singularity in angular cross-section, $b(\cos\theta)$, we keep the "gain-loss" term and will show that this is where the cancelation of singularity occurs.

Plugging (4.67) back into the Dirichlet form (4.7) (the last line of formulas) gives, with change of variables $(v, u) \leftarrow (v, v_*)$, where $u = v - v_*$ is the relative velocity,

$$\langle L(F), F\rangle = \mathbf{u}^T \mathbf{G} \mathbf{u}\,, \tag{4.68}$$

with $\mathbf{G}$ the "collision matrix" with $N \times N$ blocks, each of which is $(p+1)^d \times (p+1)^d$ block defined as

$$\begin{aligned}\mathbf{G}(k,m) = &\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu(v)\mu(v-u)\left(\Phi(v)\chi_k(v) + \Phi(v-u)\chi_k(v-u)\right)\\ &\otimes \int_{\mathbb{S}^{d-1}} \left(\Phi(v')\chi_m(v') - \Phi(v)\chi_m(v)\right) B(u,\sigma) d\sigma du dv\,.\end{aligned} \tag{4.69}$$

Let's only look at the typical term

$$\begin{aligned}&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \mu(v)\mu(v-u)\chi_k(v) \int_{\mathbb{S}^{d-1}} \left(\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)\right) B(u,\sigma) d\sigma du dv\\ &= \sum_{\bar{k}} \int_{v\in E_k}\int_{v-u\in E_{\bar{k}}} \mu(v)\mu(v-u) \int_{\mathbb{S}^{d-1}} \left(\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)\right) B(u,\sigma) d\sigma du dv\,.\end{aligned}$$

The other terms are evaluated in a same way.

Due to the convolution formulation, the integrals w.r.t v, u can be approximated through *Triangular quadratures*. Indeed, along each dimension, if $v_i \in I_{k_i}$, $v_i - u_i \in I_{\bar{k}_i}$, then the pair (v_i, u_i) will form a parallelegram which can be divided into two triangles. The integrals on the sphere take the most efforts, because one has to figure out how the Cartesian cubes intersect with the spheres. Let's extract the angular integrals in the above typical term, denoted by $\mathfrak{g}_{m,i}(v, u)$, and study it separately

$$\mathfrak{g}_{m,i}(v,u) := \int_{\mathbb{S}^{d-1}} \left(\phi_i(v')\chi_m(v') - \phi_i(v)\chi_m(v)\right) b(\frac{u\cdot\sigma}{|u|}) d\sigma \,. \tag{4.70}$$

The treatments for (4.70) follows exactly the same as in Chapter 3, where deterministic DG solvers for nonlinear Boltzmann equations are developed. Please refer to Chapter 3 for more details.

Once $\mathfrak{g}_{m,i}(v, u)$ is done, plugging it back into (4.69), we get the "collision matrix" $\mathbf{G}$.

Finally, we would like to mention that, specially for the Grad splitting formulations in Section 4.2.1.1, the block $\mathbf{G}(k, m)$ can be written down immediately, from (4.23),

$$\begin{aligned}\mathbf{G}(k,m) = \delta_{km} &\int_{E_k} \nu(v)\Phi(v)\otimes\Phi(v)dv \\ &+ \int_{E_k}\int_{E_m} \left(k_1(v,\xi) - k_2(v,\xi)\right)\Phi(v)\otimes\Phi(\xi)dvd\xi \,,\end{aligned} \tag{4.71}$$

where δ_{km} denotes Kronecker delta. Here, the collision matrix $\mathbf{G}$ is symmetric semi-positive definite.

4.5 The Approximate Rayleigh Quotient

Recall the equivalent minimization problem for solving spectral gaps in (4.11) or (4.65). With the approximation above, we can easily rewrite this constrained minimization problem as

$$\begin{aligned} \min \quad & \frac{\mathbf{u}^T\mathbf{G}\mathbf{u}}{\mathbf{u}^T\mathbf{D}\mathbf{u}} \\ s.t \quad & \mathbf{C}\mathbf{u} = \mathbf{0}\,, \end{aligned} \tag{4.72}$$

where the block diagonal matrix $\mathbf{D}$ generated from the tensor product of the basis functions; the constraint matrix $\mathbf{C}$ is of size $(d+2) \times M$ (here $M = N(p+1)^d$ is the number of coefficients), obtained from the constraints.

$$\int F(v)\mu^{\frac{1}{2}}(v)dv = \int F(v)\mu^{\frac{1}{2}}(v)vdv = \int F(v)\mu^{\frac{1}{2}}(v)|v|^2dv = 0\,. \tag{4.73}$$

Since, we need to find the global optimization solution, then, we first find an orthogonal basis $\mathbf{P}$ for the constraint space

$$\mathcal{P} = \{\mathbf{u} \in \mathbb{R}^M : \mathbf{C}\mathbf{u} = 0\}\,. \tag{4.74}$$

This calculation can be done through performing QR factorization for $\mathbf{C}^T$, the last $M-(d+2)$ columns will form the orthogonal (actually, orthonormal) basis $\mathbf{P}$, of size $M \times (M-(d+2))$ and $\mathbf{P}^T\mathbf{P} = \mathbf{I}_{M-(d+2)}$

Then, the minimization problem becomes

$$\min_{0 \neq b \in \mathbb{R}^{M-(d+2)}} \quad \frac{b^T\mathbf{P}^T\mathbf{G}\mathbf{P}b}{b^T\mathbf{P}^T\mathbf{D}\mathbf{P}b}\,, \tag{4.75}$$

which is equivalently to find the smallest singular value from the generalized eigenvalue problem

$$\mathbf{P}^T\mathbf{G}\mathbf{P} = \lambda\mathbf{P}^T\mathbf{D}\mathbf{P}\,. \tag{4.76}$$

In practice, instead of solving (4.75) and (4.76) which requires extra QR decomposition and matrix multiplications, we find out another way to force the constraints (4.73), which is much more efficient and easier to implement. This is done by perturbing the "collision matrix" $\mathbf{G}$ to its "L^2-closest" counterpart, through introducing a "conservation routine". A similar conservation routine has been successfully applied to deterministic conservative solvers for nonlinear Boltzmann equations based on Spectral methods [68] as well as Discontinuous Galerkin methods (see Chapter 3).

Our objective is to enforce the eigenvalues to be zeros whenever the corresponding eigenfunctions fall onto the null space $\mathcal{N}(L)$ of operator L. That is, we solve

Conservation Routine [Functional Level]: Minimize in the Banach space

$$\mathcal{B}^e = \left\{ X \in L^2(\Omega_v) : \int_{\Omega_v} X\mu^{\frac{1}{2}}(v) = \int_{\Omega_v} X\mu^{\frac{1}{2}}(v)v = \int_{\Omega_v} X\mu^{\frac{1}{2}}(v)|v|^2 = 0 \right\},$$

the objective functional

$$\mathcal{A}^e(X) := \int_{\Omega_v} (LF(v) - X)^2 \, dv \,. \tag{4.77}$$

To enforce the conservation, we seek, in L^2-distance, the closest $\mathbf{Q} := \mathbf{Gu}$, which is the minimizer of the following constrained optimization problem:

Conservation Routine [Discrete Level]: *Find* $\mathbf{Q}_c$ *(the subscript c means a conservative correction), the minimizer of the problem*

$$\begin{aligned} &\min \frac{1}{2}(\mathbf{Q}_c - \mathbf{Q})^T \mathbf{D} (\mathbf{Q}_c - \mathbf{Q}) \\ &s.t. \quad \mathbf{C}\mathbf{Q}_c = \mathbf{0} \,, \end{aligned}$$

where the constraint matrix $\mathbf{C}$ *is defined in (3.19).*

This conservation routine has been studied in Section 3.4. Please refer to Section 3.4 for more details.

Finally, we get the minimizer $\mathbf{Q}_c$

$$\mathbf{Q}_c = [\mathbf{I} - \mathbf{D}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T)^{-1}\mathbf{C}]\mathbf{Q}\,, \tag{4.78}$$

where $\mathbf{I}$ is an identity matrix of size $M \times M$. So, $\mathbf{Q}_c$ is a perturbation of $\mathbf{Q}$. So, finally, the perturbed "collision matrix" $\mathbf{G}$ will be

$$\mathbf{G}_c = [\mathbf{I} - \mathbf{D}^{-1}\mathbf{C}^T(\mathbf{C}\mathbf{D}^{-1}\mathbf{C}^T)^{-1}\mathbf{C}]\mathbf{G}\,, \tag{4.79}$$

which is forced to have $d+2$ zero eigenvalues whenever $\mathbf{u} \notin \mathcal{P}$ defined in (4.74).

The $(d+3)$-th eigenvalue of $\mathbf{G}_c$ will be defined as our numerical spectral gap.

4.5.1 Convergence of The Approximate Rayleigh Quotient

We will prove that the above discrete Rayleigh quotient (4.72) will converge to the spectral gap solved from (4.11).

With standard approximation theory, it is not hard to prove that, the above discrete Rayleigh quotient (4.72) converges to the spectral gap (if exists) of the original linearized Boltzmann operator. We summarize it in the following theorem.

Theorem 4.5.1 (Convergence of Rayleigh Quotients)**.** *For the angular integrable (i.e.* $\alpha < 0$ *in (2.5)) linearized Boltzmann operator, defined in the*

Dirichlet form (4.7), with a piecewise polynomial approximation (4.67) for the perturbation $F(v)$, the spectral gap, denoted by $\lambda(G)$, solved from minimized Rayleigh quotient (4.72) approximates the original spectral gap, denoted by $\lambda(L)$, solved from (4.11), in the following way,

- *When $\gamma \geq 0$, $|\lambda(L) - \lambda(G)| \lesssim h^{k+1}$;*
- *When $-d < \gamma < 0$, $|\lambda(L) - \lambda(G)| \lesssim h^{k+1} + e^{-\frac{V^2}{2}}$,*

where $h = \max_{E \in \mathcal{T}_h} diam(E)$ is the mesh size of the regular triangulation, k is the total degree of polynomials in the piecewise polynomial space $\mathbf{P}^k$. The "$\lesssim$" is only upto some constant depending on the truncated domain $\Omega = [-V, V]^d$ as well as eigenfunctions associated with the spectral gap eigenvalue.

Proof. As shown in the Dirichlet form (4.7) of L, the eigenvalue zero is corresponding to the conservation laws for mass, momentum and kinetic energy. Therefore, it is of multiplicity $d + 2$, with eigenfunctions $\phi_0(v) = \mu^{1/2}(v)$, $\phi_i(v) = \mu^{1/2}(v)v_i$ for $i = 1, .., d$ and $\phi_{d+1}(v) = \mu^{1/2}(v)|v|^2$.

Suppose the truncated velocity domain $\Omega = [-V, V]^d$ is large enough. We are indeed dealing with the cutoff operator $L_\Omega = \chi_\Omega L$ applying to $\chi_\Omega(v)F(v)$. That is, the kernel, denoted by k_Ω, for cutoff L_Ω is given by

$$k_\Omega = \chi_\Omega(v)\nu(v)\delta(v - \xi) + \chi_\Omega(v)k(v, \xi)\,, \tag{4.80}$$

where $\delta(v-\xi)$ is short for $\delta(v_1-\xi_1)\cdots\delta(v_d-\xi_d)$, $\nu(v)$ is the collision frequency defined in (4.25) and $k(v, \xi)$ is the kernel for the compact operator $\mathbf{K}$ in (4.14).

However, the null space $\mathcal{N}(\mathcal{L})$ is not invariant under the cutoff. Nevertheless, since $\mathcal{N}(\mathcal{L})$ is spanned by collision invariants weighted with a Guassian distribution, as long as Ω is large enough, the approximation error due to cutoff can be negligible. To save trouble on dealing with null space, we consider the modified linear operator $\bar{L}$, with the null space of L removed

$$\bar{L}F = LF + \sum_{i=0}^{d+1} \phi_i(F, \phi_i)\,, \tag{4.81}$$

where $(F, \phi_i) = \int_{\mathbb{R}^d} F(v)\phi_i(v)dv$. This is to replace the integral kernel $k(v, \xi)$ by

$$\bar{k}(v, \xi) = k(v, \xi) + \sum_{i=0}^{d+1} \phi_i(v)\phi_i(\xi)\,, \tag{4.82}$$

which is still $L^2(\mathbb{R}^d)$ integrable. That is, $\bar{L}$ can be still written as collision frequency $\nu(v)$ plus a compact perturbation.

Thus, the minimum Rayleigh quotient of $\bar{L}$ is the expected spectral gap, if exists. That is, $\lambda(L) = \lambda(\bar{L})$. So, we only need to study the approximations for the Rayleigh quotient of operator $\bar{L}$.

Similarly, we are working with the cutoff operator $\bar{L}_\Omega = \chi_\Omega \bar{L}$ applying to $\chi_\Omega(v)F(v)$. That is, the kernel $\bar{k}_\Omega$ for cutoff $\bar{L}_\Omega$ is given by

$$\bar{k}_\Omega = \chi_\Omega(v)\nu(v)\delta(v - \xi) + \chi_\Omega(v)\bar{k}(v, \xi)\,. \tag{4.83}$$

According to Weyl's theorem, for $\gamma \geq 0$, the spectral gap for the new $\bar{L}$ still exists. And in the case, the cutoff doesn't change the minimum of the Rayleigh quotient of $\bar{L}$. So, the spectral gap stays the same, or $\lambda(\bar{L}) = \lambda(\bar{L}_\Omega)$.

While for the case $-d < \gamma < 0$,

$$\min_{v\in\Omega} \nu(v) \gtrsim e^{-\frac{V^2}{2}} , \tag{4.84}$$

which is the lower bound for the continuum spectrum of $\bar{L}_\Omega$. This implies, the spectral gap for the cutoff operator $\bar{L}_\Omega$ is no larger than $e^{-\frac{V^2}{2}}$ (up to some constant factor), if ever exists. That is, $|\lambda(\bar{L}) - \lambda(\bar{L}_\Omega)| \lesssim e^{-\frac{V^2}{2}}$.

Suppose $\mathcal{T}_h$ is a regular Cartesian partition for Ω, with mesh size $h = \max_{E\in\mathcal{T}_h} \mathrm{diam}(E)$. Please refer to Appendix for some notations and the standard approximation theory.

For any mesh elements E_v and E_ξ, according to the approximation theories (8), it's not hard to prove the following

$$\|F(v)F(\xi) - P_hF(v)P_hF(\xi)\|_{L^2(E_v\times E_\xi)} \le h^{k+1} \left(\|F\|_{H^{k+1}(E_v)} \|F\|_{H^{k+1}(E_\xi)} \right) , \tag{4.85}$$

where P_hF is the L^2 projection defined in (7) in Appendix.

Then, the Dirichlet form is approximated as follows

$$\begin{aligned} &|\langle \bar{L}_\Omega F, F\rangle - \langle \bar{L}_\Omega(P_hF), (P_hF)\rangle| \\ &\le \sum_m \sum_n \|\bar{k}_\Omega\|_{L^2(E_m\times E_n)} \|F(v)F(\xi) - P_hF(v)P_hF(\xi)\|_{L^2(E_m\times E_n)} \\ &\le C(\Omega) h^{k+1} \|F\|^2_{H^{k+1}(\mathcal{T}_h)} , \end{aligned} \tag{4.86}$$

where $C(\Omega)$ is some constant depending on the truncated domain Ω.

And thus, the Rayleigh quotients have the following estimates

$$\begin{aligned}
&\left| \frac{\langle \bar{L}_\Omega F, F\rangle}{\|F\|^2_{L^2(\Omega)}} - \frac{\langle \bar{L}_\Omega (P_h F), (P_h F)\rangle}{\|P_h F\|^2_{L^2(\mathcal{T}_h)}} \right| \\
&= \frac{1}{\|F\|^2_{L^2(\Omega)} \|P_h F\|^2_{L^2(\mathcal{T}_h)}} \big(\langle \bar{L}_\Omega F, F\rangle \big(\|P_h F\|^2_{L^2(\mathcal{T}_h)} - \|F\|^2_{L^2(\Omega)} \big) \\
&\quad + \|F\|^2_{L^2(\Omega)} \big(\langle \bar{L}_\Omega F, F\rangle - \langle \bar{L}_\Omega (P_h F), (P_h F)\rangle \big) \big) \\
&\leq C(\Omega) h^{k+1} \,,
\end{aligned} \tag{4.87}$$

which implies,

$$|\lambda(\bar{L}_\Omega) - \lambda(G)| \leq C(\Omega) h^{k+1} \,, \tag{4.88}$$

where now the generic constant $C(\Omega)$ also depends on the eigenfunction associated with the spectral gap eigenvalue.

Finally, noticing

$$|\lambda(L) - \lambda(G)| \leq |\lambda(\bar{L}) - \lambda(\bar{L}_\Omega)| + |\lambda(\bar{L}_\Omega) - \lambda(G)| \,, \tag{4.89}$$

gives our final estimates. □

4.6 Numerical Results

We will present some raw results for $2d$ as well as $3d$ linearized Bolzmann operators with integrable angular cross-sections.

The computing of weight matrix G is parallelized with MPI [65]. The matrix will be computed and stored in a way of two-dimensional block cyclic distribution [12], on a process grid, as shown in Figure 4.4

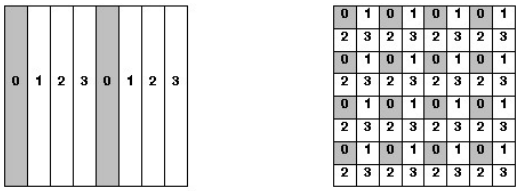


Figure 4.4: The 1d block-cyclic column- and 2d block-cyclic distributions

Some scalable eigensolvers in ScaLAPACK, for example, PDSYGVX and PDSYEVX [12], are called to compute the eigenvalues for the distributed matrix.

At first, we would like to interpret the relationship between our numerical results and the true spectral gaps. Due to the domain truncation and DG approximation, the numerical results may not represent the true spectral gaps; however, the convergence Theorem 4.5.1 for the approximate Raleigh quotients in Section 4.5.1 tells us that, if there exists a spectral gap for the true problem, then as long as the domain is truncated large enough, what matters will be only the DG scheme approximation accuracy. And if there is no spectral gap, then as computing domain gets larger, the numerical "spectral gap" will clearly decay down to zero. This is exactly what Figure 4.5 and Figure 4.6 are showing.

Note: When increasing the lateral size of the truncated velocity domain, we keep the mesh size to be consistent (say, in our tests, Δv=0.5), for sake of comparison. For the case of soft potential, as shown in Figure 4.6 for $\gamma = -1$, some "pseudo spectral gap" in the numerical results might be observed, for example in the segment $V \in [7, 9]$; but such "pseudo spectral gap"

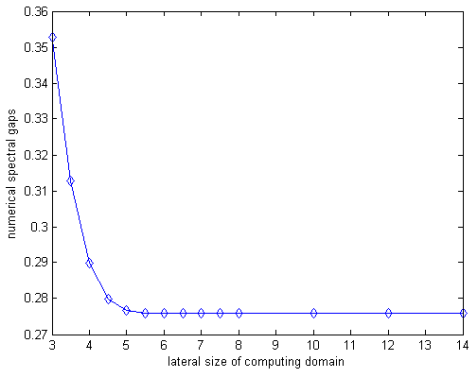


Figure 4.5: The numerical spectral gaps for 2d Maxwell type model, i.e. $\gamma = 0, \alpha = -1$

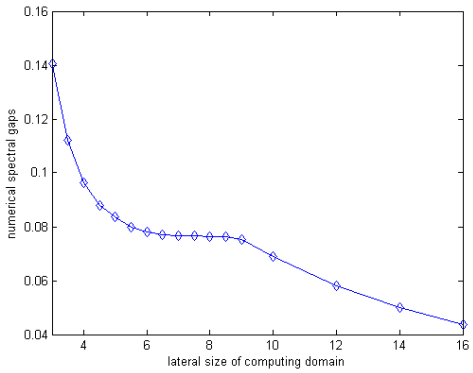


Figure 4.6: The numerical spectral gaps for 2d, $\gamma = -1, \alpha = -1$

will break immediately when increasing the domain size.

Then, we fix a large enough lateral size, increasing the number of mesh elements on each direction. More accurate results can be expected. We can see from Figure 4.7 and Figure 4.8, the numerical values will approach the analytical value $\frac{1}{4}$ (for 2d) and $\frac{1}{3}$ (for 3d) respectively, when finer discretization is applied, as calculating the spectral gap for Maxwell type of interactions ($\gamma = 0$), where the exact eigenvalue for Maxwell-type interactions ($\gamma = 0$) is known and given by [15, 34, 39]:

$$\lambda_{nl} = \int_{S^{d-1}} b(\cos(\theta)) \left[\cos^{2n+1} \frac{\theta}{2} P_l(\cos(\frac{\theta}{2})) + \sin^{2n+1} \frac{\theta}{2} P_l(\sin \frac{\theta}{2}) - 1 - \delta_{l0}\delta_{n0} \right] ,$$

where $P_l(x)$ is the l-th Legendre polynomial; n, l=0,1,....

In particular, by actually solving the nonlinear Boltzmann equation and plotting the evolution of the weighted L^2 norm of the solution, we can expect an exponential decay rate governed by or close to the spectral gap.

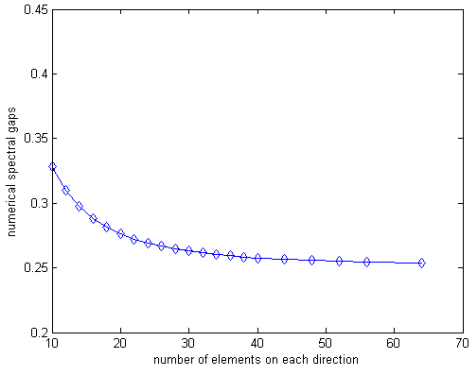


Figure 4.7: The numerical spectral gaps with $\Omega_v = [-5,5]^2$ for 2d Maxwell type model, i.e. $\gamma = 0, \alpha = -1$

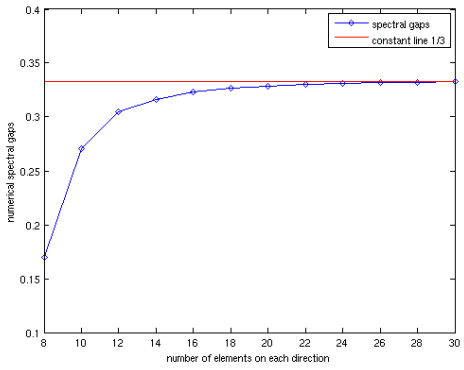


Figure 4.8: The numerical spectral gap with $\Omega_v = [-5,5]^3$ for 3d Maxwell type model, $\gamma = 0, \alpha = -2$

With the same DG discretization, the numerical value of the corresponding spectral gap for $\gamma = 1$ (hars sphere) is 0.72. The numerical solutions for the corresponding nonlinear BE is obtained by conservative DG solver developed also by the authors, see Chapter 3. See Figure 4.9.

Remark. This can only be expected after long time or with an initial state very close to equilibrium, because the spectral gap, as the first non-zero eigenvalue, can only dominate the decay rate when time t is large enough.

We have computed spectral gaps for 2d variable hard potentials with isotropic angular cross-sections, using a moderate domain discretization (piecewise constant basis functions; $V = 5$, $N = 24$) . As seen from Table 4.1, stronger intermolecular potential will force a faster decay to equilibrium.

We also apply piecewise linear basis functions (P^1 polynomials) for

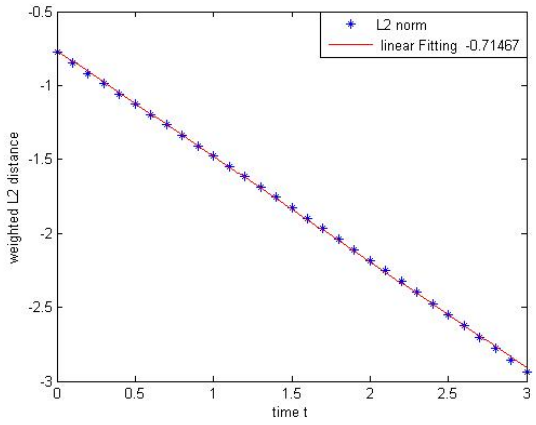


Figure 4.9: The exponential decay for solutions of 2d nonlinear Boltzmann equation with $\gamma = 1, \alpha = -1$

γ	0	0.1	0.25	0.5	0.75	0.9	1
gaps	0.25	0.29	0.34	0.44	0.58	0.67	0.72

Table 4.1: Numerical spectral gaps for 2d variable hard potentials with isotropic angular cross-sections

approximating $F(v)$. Table 4.2 is the comparison with piecewise constant case. from which one can easily see the P^1 basis functions give a much more accurate

gap	(V,N)=(5,20)	(V,N)=(5,24)
P^0	0.383798	0.353494
P^1	0.351826	0.332835

Table 4.2: Comparisons of numerical spectral gaps between P^0 and P^1 basis, for 3d Maxwell model.

approximation than P^0, which is stated in the theorem of convergence.

For the non-cutoff cases, when $\int_{\mathbb{S}^{d-1}} b(\frac{u\cdot\sigma}{|u|})d\sigma$ is unbounded, we also have numerically verified the "conjecture" on the existence of spectral gaps,

i.e. there exists spectral gap if and only if $\gamma + \alpha \geq 0$. Similar to the cutoff case, as we show in Section 4.3, the geometry of the spectral gaps for truncated operator $\chi_\Omega L$ also depends on the truncation of the domain and the discretization resolution. If there exists a spectral gap, as long as the computing velocity domain is large enough, the approximation accuracy only depends on the resolution of the mesh and vice versa; otherwise, if there is no spectral gap, with the lateral size getting larger, the numerical spectral gap is expected to decay to zero, and vice versa. See Figure 4.10 and 4.11.

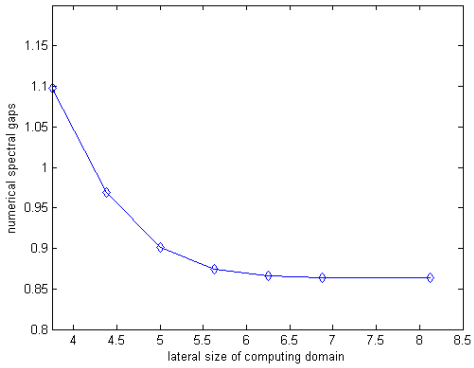


Figure 4.10: The numerical spectral gaps for 3d non-cutoff case, $\gamma = 0, \alpha = 0$

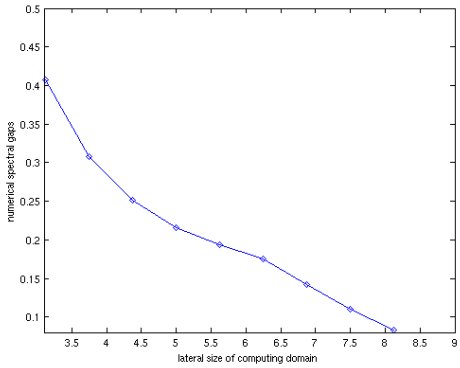


Figure 4.11: The numerical spectral gaps for 3d non-cutoff case, $\gamma = -1, \alpha = 0$

So, once we know there exists a spectral gap, we can fix a large enough truncated velocity domain and apply DG meshes with finer resolutions, then more accurate approximations to the real spectral gap can be expected. See Figure 4.12 for the numerical spectral gaps when $\gamma = 0, \alpha = 0$, where an approximate value 1.0 is achieved.

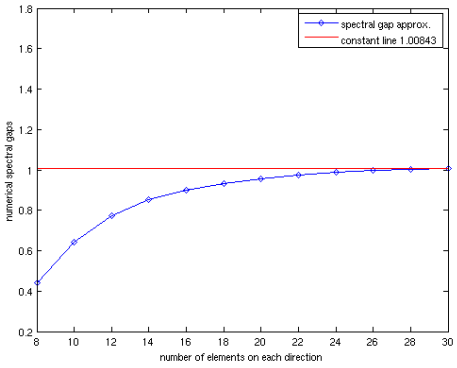


Figure 4.12: The numerical spectral gaps with $\Omega_v = [-5, 5]^3$ for 3d $\gamma = 0, \alpha = 0$

We list the results for 3d variable hard potentials with isotropic angular cross-sections, see Table 4.3. from which we also can tell, as in 2d case, stronger

γ	0	0.25	0.5	0.75	1
gaps	0.33	0.45	0.62	0.83	1.10

Table 4.3: Numerical spectral gaps for 3d variable hard potentials with isotropic angular cross-sections

intramolecular potential imposes faster decay to equilibrium.

A hybrid OpenMP [13] and MPI [65] paralleling computing is implemented to compute the eigenvalues of the conservative corrected "collision matrix". Some routines in package like Scalapck [12] have been called. Our test computations have been distributed among up to 256 nodes and 4000 cores on clusters Lonestar and Stampede affiliated with TACC [103]. As long as memory and computing power allows, one can improve the accuracy of the numerical spectral gaps by choosing larger velocity domain, finer DG meshes

and higher accuracy quadrature rules.

Finally, we stress that the spectral gap information can be a benchmark property for the solution of linearized and nonlinear Boltzmann equation. One can compute the decay of (weighted) L^2 norms of the solutions and observe, when it's close enough to equilibrium, if the decay rate is approximately the given spectral gap.

Chapter 5

A Conservative Scheme for Approximating Collisional Plasmas

We have developed a deterministic conservative solver for the inhomogeneous Fokker-Planck-Landau equations coupled with Poisson equations, which is a rather realistic and primary model for collisional plasmas.

5.1 Introduction

The plasma dynamics is governed by infinite-range interactions, i.e. Coulomb potentials, and thus behaves differently than ordinary molecular gases. At the kinetic level, among various plasma models, the *Vlasov-Poisson* (VP) equations and *Fokker-Planck-Landau* (FPL) equations are the most representative ones describing, respectively, collisionless and collisional plasma systems.

The VP system is a nonlinear kinetic system modeling the transport of charged particles in a collisionless plasma, under the effect of a self-consistent electrostatic field and possibly an externally supplied field. The electrostatic potential is coupled through Poisson equation. Some natural plasmas, as for example solar wind, behaves as collisionless, since the mean free path of a

particle traveling from the Sun to the Earth is of the order of Sun-Earth distance. Because of its comparative simplicity, numerical schemes for VP equations have been not only thoroughly explored but also well developed. One can refer to, for example [41, 49, 78]. The collisionless VP system exhibits a variety of dynamical phenomena. For example, the well-known filamentation (filaments in phase space and steep gradients in v) due to its dispersive nature and Landau damping mechanism for near equilibrium states satisfying some conditions. Readers can refer to [40] for more physical insights.

If collisions are taken into account, particles are scattered and things could be different. To our best knowledge, there is rare work on such models. Thus, we expect to study the numerical behaviors of the inhomogeneous FPL system for multiple species. The transport of probability density for the particle species α is given by

$$\partial_t f_\alpha + v\cdot\nabla_x f_\alpha + F(t,x)\cdot\nabla_v f_\alpha = \sum_\beta a_{\alpha\beta} Q_{\alpha,\beta}(f_\alpha, f_\beta), \qquad v \in \mathbb{R}^{d_v}, x \in \Omega_x \subseteq \mathbb{R}^{d_x}, \tag{5.1}$$

subject to some initial and boundary conditions on f_α. Here, f_α is the distribution for species α, the term $Q_{\alpha,\beta}(f_\alpha, f_\beta)$ is a nonlinear, nonlocal operator in divergence form and models the (α, β) pair collisions (e.g. electron-electron, ion-ion, electron-ion, etc.) and $a_{\alpha\beta}$ are the coupling parameters. In our present work, we take $a_{\alpha\beta} = \frac{1}{\varepsilon}$ to be collision frequency with ε the Knudesen number. The case $a_{\alpha\beta} \to 0$ corresponds to the Vlasov-Poisson system. The force field $F(t,x)$ only depends on time and space position and can be external or self-consistent. If it is self-consistent, it corresponds to the electrostatic force

$qE(t,x)$, where q is the charge and $E(t,x)$ is the self-consistent electrostatic field obtained from the Poisson equation for charges

$$E(t,x) = -\nabla_x \Phi(t,x); \qquad -\Delta_x \Phi = \sum_\beta \int_{\mathbb{R}^3} f_\beta(v) dv \,, \tag{5.2}$$

subject to some boundary condition on Φ.

The FPL transport equation is used to model long-range Coulomb interactions between charged particles (e.g binary collisions occurring in a plasma). It is of primary importance in modeling evolution of collisonal plasma and actually a rather realistic model especially when the magnetic field is very weak. The FPL transport equation can be derived from the general Boltzmann transport equation by taking the so-called binary grazing collision limit, i.e collisions that only result in very small deflections of particle trajectories, as is the case for Coulomb potentials with Rutherford scattering [111]. The original derivation is due to Landau [89]. Readers can refer to Villani [117] and the references therein for some mathematical aspects, and to [86] for a recent calculation of the grazing collision asymptotics in Fourier space.

With the general non-isotropic Landau collision operator Q, the inhomogeneous FPL model gains huge difficulties to handle, both analytically and numerically. The main factors generating such difficulties are the nonlinearity, non-locality and diffusive nature with high dimensionalities. Unlike other kinetic models, for example Boltzmann equations where some non-deterministic methods (DSMC) have been successfully applied, the infinite-range potential interactions greatly limit the applications of Monte Carlo methods. Many

people have tried to develop efficient deterministic solvers for the inhomogeneous FPL equations. However, due to the computational complexity mentioned above, people have turned to some simplified versions of this problem, for example, space homogeneous FPL equations in the isotropic case [29] or cylindrically symmetric problem [90] or 1D Fokker-Planck type operator [108]. Previously, L. Pareschi et al. proposed a spectral method to solve FPL equations [107], by taking truncated Fourier series and extending solutions by periodicity. This method was not intended to preserve moments as desired and introduced unphysical binary collisions. It cannot avoid *aliasing effects*, which will be present whenever a vanishing function is approximated by a periodic one. Later, Filbet and Pareschi [60] applied the spectral method to study inhomogeneous FPL with 1D in space and 2D in velocity. The pure transport equations was further splitted and a finite volume scheme was used. Then, Crouseilles and Filbet [53] proposed a solver for inhomogeneous FPL with 1D in space and 3D in velocity, where the pure transport part was treated with a finite volume scheme and the Landau operator was approximated by averaging of uncentered finite difference operators. However, the solver in [53] only preserved mass and energy at the discrete level (for the uncentered finite difference approximate Landau operator), under some symmetry assumptions on the initial datum.

In our current work, we follow a regular time-splitting scheme, splitting the original inhomogeneous FPL equation into a pure transport problem, i.e Vlasov-Poisson equation for advection and a homogeneous FPL equation for

collisions. These two subproblems can be treated with completely different schemes. For the VP problem, we apply the RKDG method with a piecewise polynomial basis subspace covering all collision invariants, which can be proved to conserve mass, momentum and kinetic energy up to some boundary effects that disappear if the domain is taken large enough. While for the homogeneous FPL equation, different than the one in [107], we extend the spectral method first introduced in [68] for the nonlinear Boltzmann transport equation and propose a conservative spectral method for homogeneous FPL equation, by first extending the solution by zero, representing the collision integral through choosing Fourier modes as the test functions in the weak form and enforcing conservation routines. Since two completely different numerical scheme are applied separately, our challenge is not only to link two different meshes and at the same time, but also to keep the conserved quantities. We have designed a new conservation correction process such that, after projecting the conservative spectral solution onto the DG mesh, the conserved moments are transferred to the DG solution as well.

5.2 The Fokker-Planck-Landau Operator

The FPL operator models binary collisions in a system of single- or multi-species and reads

$$Q_{\alpha,\beta}(f_\alpha, f_\beta) = \nabla_v \cdot \int_{\mathbb{R}^3} \mathbf{S}(v - v_*)(f_\beta(v_*)\nabla_v f_\alpha(v) - f_\alpha(v)\nabla_{v_*} f_\beta(v_*))dv_* \,, \quad (5.3)$$

with the $d \times d$ nonnegative and symmetric projection matrix

$$\mathbf{S}(u) = L|u|^{\gamma+2}(\mathbf{Id} - \frac{u \otimes u}{|u|^2}), \tag{5.4}$$

where $\mathbf{Id}$ is the $d \times d$ identity matrix; $\Pi(u) = \mathbf{Id} - \frac{u \otimes u}{|u|^2}$ is the orthogonal projection upon the space orthogonal to u. It's semi positive definite with eigenvalues 0,1,1. The constant L is a positive(a value related to the logarithm of the dimensionless Debye radius of screening of the Coulomb potential in plasma). For simplicity, we take $L = 1$ in the following.

The inverse-power laws has $\gamma \geq -3$. Similar to Boltzmann equations, different γ categorizes hard potentials for $\gamma > 0$, Maxwellian molecules for $\gamma = 0$ and soft potentials for $\gamma < 0$. But here, we only focus on most interesting case $\gamma = -3$, corresponding to Coulomb interactions.

When $\alpha = \beta$, the operator $Q_{\alpha,\alpha}$ will be a nonlinear (bilinear) integro-differential operator in divergence form. Here and in the following, when talking about single-species distributions, we will drop the subscript α for simplicity. The strong form of this nonlinear partial integrodifferential equation is

$$\partial_t f + v \cdot \nabla_x f + F(t,x) \cdot \nabla_v f = Q_{FPL}(f,f), \qquad v \in \mathbb{R}^3, x \in \Omega_x \subseteq \mathbb{R}^3, \tag{5.5}$$

where the collision kernel is of the form

$$Q_{FPL}(f,f) = \nabla_v \cdot \int_{\mathbb{R}^3} \mathbf{S}(v - v_*)(f(v_*)\nabla_v f(v) - f(v)\nabla_{v_*} f(v_*))dv_* . \tag{5.6}$$

The FPL operator, as a limit of the Boltzmann collision operator, possesses a similar conservation laws and decay of entropy(H-theorem). That

is

$$\int_{\mathbb{R}^3} Q(f,f)(v)\phi(v)dv = 0\,, \tag{5.7}$$

if and only if

$$\phi(v) = 1, v, |v|^2 \tag{5.8}$$

corresponding to the conservation of mass (charge), momentum and kinetic energy. As in the previous chapters, we call the $d+2$ test functions $\phi(v) = 1, \mathbf{v}, |v|^2$ *collision invariants*.

In addition, for any $f(v) > 0$, if set $\phi(v) = \log f(v)$, one can show the following dissipation of entropy

$$\frac{d}{dt}\int_{\mathbb{R}^d} f\log f = \int_{\mathbb{R}^d} Q(f,f)(v)\log f(v)dv \le 0\,, \tag{5.9}$$

which also implied the equilibrium states given by the Maxwellian distribution

$$M(x,v) = \frac{\rho}{(2\pi k_B T)^{\frac{3}{2}}}\exp\left(-\frac{|v-\bar{v}|^2}{2k_B T}\right)\,, \tag{5.10}$$

where k_B is the Boltzmann constant. The local dependence of x is from the mass $\rho(x)$, the mean velocity $\bar{v}(x)$ and the kinetic temperature $T(x)$, given by

$$\rho = \int_{\mathbb{R}^3} f(x,v)dv, \qquad \bar{v} = \frac{\int_{\mathbb{R}^3} f(x,v)vdv}{\rho}, \qquad T = \frac{\int_{\mathbb{R}^3} f(x,v)|v-\bar{v}|^2 dv}{3\rho}\,. \tag{5.11}$$

When $\alpha \neq \beta$, the operator $Q_{\alpha,\beta}$ models collisions between two different species. It is essentially a linear operator and the treatment will be similar and sometimes even much simpler compared with the fully nonlinear one (5.6). We will consider different problems associated to different forms of the operator $Q_{\alpha,\beta}$ in the following sections.

5.3 Spectral Gap For Linearized Landau Operators

Before going to the real full nonlinear Landau operator, we study in this section the spectral gap for the linear Landau operators and thus the rate of convergence of the solution. This work was suggested and initiated by A.V Bobylev, through a personal communication with him in October, 2013.

The linear Landau equation is an equation for a test particle, which collides with equilibriumly distributed "field" particles. It can be obtained by rewriting the homogeneous Landau equation in the form of nonlinear diffusion equations for $f(v,t)$, then replacing the $f(v,t)$ in the integral terms by a constant Maxwellian, say, $M(v) = \exp(-|v|^2)$. Finally, consider the isotropic case when the function $f(v,t)$ is a radial one denoted by $f(x,t)$ with $x = |v|^2$.

More specifically, the generalized linear isotropic Landau equation reads

$$\partial_t f(x,t) = x^\theta \partial_x \left(\mathcal{D}_\theta(x)(\partial_x f(x,t) + f(x,t)) \right), \quad x, t \geq 0\,, \tag{5.12}$$

where

$$\mathcal{D}_\theta(x) = \int_0^x y^\theta e^{-y} dy, \quad 0 \leq \theta \leq \frac{1}{2}\,. \tag{5.13}$$

The 3-D Landau equation (see (5.4), (5.5) and (5.6)) corresponds to $\theta = \frac{1}{2}$. The case $\theta = 0$ can be exactly solvable through the Laplace transform.

If we take

$$f(x,0) = f_0(x). \quad \int_0^\infty x^\theta f_0(x) dx = \Gamma(1+\theta)\,, \tag{5.14}$$

where the Gamma function is defined by

$$\Gamma(z) = \int_0^\infty x^{z-1} e^{-x} dx, \quad z > 0\,, \tag{5.15}$$

then, it is easy to check that, for any $x \geq 0$,

$$\lim_{t\to\infty} f(x,t) = e^{-x}\,. \tag{5.16}$$

Our goal is to study the rate of convergence.

Consider a perturbation around the equilibrium, that is,

$$f(x,t) = e^{-x}(1+\varphi(x,t))\,. \tag{5.17}$$

Plugging this back to the generalized linear Landau equation (5.12) gives an equation for φ

$$x^{\theta}e^{-x}\partial_t\varphi(x,t) = -\mathcal{L}(\varphi)(x,t)\,, \tag{5.18}$$

where the linear operator $\mathcal{L}$ reads

$$\mathcal{L}(\varphi)(x,t) = -\partial_x[\mathcal{D}_\theta(x)e^{-x}\partial_x\varphi(x,t)]\,. \tag{5.19}$$

If define the weighted L^2 norm as

$$\|\varphi\|^2 = \int_0^\infty x^{\theta}e^{-x}\varphi^2 dx\,, \tag{5.20}$$

then,

$$\frac{1}{2}\frac{d}{dt}\|\varphi\|^2 = -\langle\mathcal{L}(\varphi),\varphi\rangle = -\int_0^\infty \mathcal{D}_\theta(x)e^{-x}[\partial_x\varphi(x,t)]^2 dx\,, \tag{5.21}$$

where the *Dirichlet form* $\langle\mathcal{L}(\varphi),\varphi\rangle$ ($\langle\cdot,\cdot\rangle$ denotes the usual unweighted L^2 innner product) is obtained through integration by parts.

Obviously, $\mathcal{L}$ is a positive operator. Its smallest eigenvalue is 0 with multiplicity 1 with its eigensapce spanned by constant functions. In particular,

we can conclude, at the formal level, that

$$\varphi(x,t) = O\left(e^{-\lambda_\theta t}\right), \quad \text{as } t \to \infty\,, \tag{5.22}$$

where λ_θ is the *spectral gap*, if exists, of operator $\mathcal{L}$, defined as the minimized Rayleigh quotient of $\mathcal{L}$

$$\begin{aligned} \lambda_\theta &= \min \frac{\langle \mathcal{L}(\varphi), \varphi \rangle}{\|\varphi\|^2} \\ \text{s.t.} \int_0^\infty & x^\theta e^{-x} \varphi(x) dx = 0\,, \end{aligned} \tag{5.23}$$

that is, φ is orthogonal to the eigensapce of eigenvalue 0.

That means, we can expect an exponential decay when the state is close to equilibrium, where the decay rate is given by $\lambda_\theta > 0$, if exists. In addition, the existence can be analytically proved for $\theta = 0$. For $\theta > 0$, we would like to study it numerically. This is done by taking a finite-dimensional approximation space for φ, and examine the behavior for increasing dimensions the approximate spaces.

To this goal, we introduce an orthogonal basis $\{\varphi_n(x)\}$, $n = 0, 1, \ldots$, for the weighted L^2 space with norm (5.20), such that $\varphi_0(x) = \text{const}$, and

$$\langle \varphi_n, \varphi_m \rangle_w = \int_0^\infty x^\theta e^{-x} \varphi_n(x) \varphi_m(x) dx = \delta_{n,m}\,, \tag{5.24}$$

where $\langle \cdot, \cdot \rangle_w$ denotes the weighted L^2 inner product with weight $w = x^\theta e^{-x}$.

Such requirements are perfectly satisfied by the normalized *generalized Laguerre polynomials*,

$$\varphi_n(x) = \frac{L_n^\theta(x)}{\|L_n^\theta\|}, \quad n = 0, 1, \ldots\,. \tag{5.25}$$

In particular,

$$L_0^\theta = 1, \quad L_1^\theta = 1 + \theta - x, \quad L_n^\theta = \sum_{k=0}^{n} a_k^{(n,\theta)} x^k \,, \tag{5.26}$$

where the coefficients

$$a_k^{(n,\theta)} = \frac{(-1)^k}{k!} \begin{pmatrix} n+\theta \\ n-k \end{pmatrix} = \frac{(-1)^k}{k!} \frac{\Gamma(n+\theta+1)}{\Gamma(n-k+1)\Gamma(\theta+k+1)} \,, \tag{5.27}$$

and the weighted norm of L_n^θ is given by

$$\|L_n^\theta\|^2 = \int_0^\infty x^\theta e^{-x} [L_n^\theta]^2 dx = \frac{\Gamma(n+\theta+1)}{n!} \,. \tag{5.28}$$

Thus, we have found a polynomial approximation for $\varphi(x)$ that well accommodate the constraint in (5.23). For a fixed order of approximation N, we consider the minimization problem (5.23) for

$$\varphi(x) = \sum_{n=1}^{N} u_n \varphi_n(x) \,, \tag{5.29}$$

where u_n are the coefficients. Note that, the summation starts from $n = 1$, because $\langle \varphi_0, 1 \rangle_w = 0$ for any $n > 1$. This automatically fullfill the constraint in (5.23).

Thus, with this approximation, the Dirichlet form in (5.21) can be written as a quadratic form

$$\langle \mathcal{L}(\varphi), \varphi \rangle = \mathbf{u}^T G \mathbf{u} \,, \tag{5.30}$$

where $\mathbf{u} = (u_1, \ldots, u_N)$ is the coefficient vector. The entries G_{nm} of the symmetric weight matrix G is given by

$$\begin{aligned} G_{nm} &= \int_0^\infty \mathcal{D}_\theta(x) e^{-x} \varphi_n'(x) \varphi_m'(x) dx \\ &= \frac{1}{\|L_n^\theta\| \|L_m^\theta\|} \sum_{k=1}^{n} \sum_{l=1}^{m} k l a_k^{(n,\theta)} a_l^{(m,\theta)} (k+l-2)! S(k+l-2) \,, \end{aligned}$$

with $S(p)$ given by

$$S(p) = \sum_{j=0}^{p} \frac{\Gamma(j+\theta+1)}{2^{j+\theta+1} j!} . \tag{5.31}$$

It's not hard to find that with increasing order of approximations, the numerical spectral gap is decreasing. For example, when $N = 1$, the matrix G reduces to one single entry

$$\lambda_\theta = G_{11} = \frac{1}{1+\theta} 2^{-(1+\theta)} . \tag{5.32}$$

It can be computed analytically that for $\theta = 0$, there exists spectral gap $\lambda_0 = \frac{1}{4}$. In particular, the above first order approximation gives a rough approximation $\lambda_0 \approx \frac{1}{2}$. We compute the smallest eigenvalues of G and study its asymptotic behavior with increasing N. Here shows the results for $\theta = 0$ and $\theta = \frac{1}{2}$. For $\theta = 0$, the convergence to the analytical value $1/4$ can be observed; for

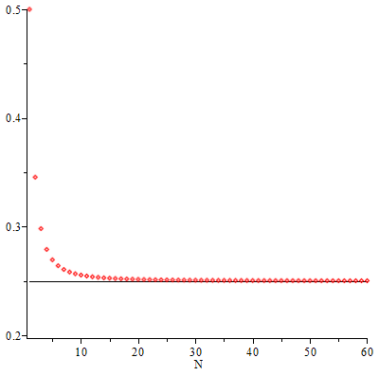


Figure 5.1: The numerical spectral gaps of linear Landau operator for $\theta = 0$ with increasing N

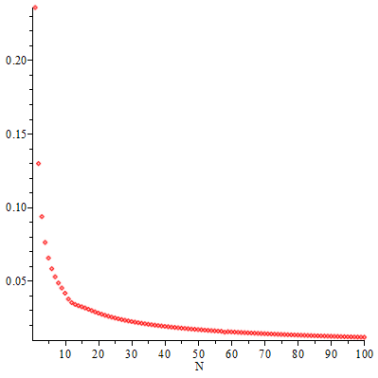


Figure 5.2: The numerical spectral gaps of linear Landau operator for $\theta = 1/2$ with increasing N

$\theta = 0.5$, the "gap" goes all the way down to zero with increased orders of basis

Laguerre polynomials (Only results w.r.t N up to 100 are plotted, but "gap" will continue to decrease towards zero when order $N > 100$), which implies there is no such a spectral gap.

At last, we would like to make the following remarks.

- As for the real (non-isotropic) linearized Landau equation, it has been shown the solution decays with an "almost exponential" (polynomial) rate, see for example [115] and references therein. For the isotropic linearized Landau equation that we studied here, for $\theta = \frac{1}{2}$, heuristic arguments indicate that the solution will behave like $O(e^{-\lambda t^{\beta}})$ with some $0 < \beta < 1$. In future, some numerical simulations will be implemented benchmarked on this analytical result.

- In the definition of the entries of the matrix G, there involves arithmetics among numbers of enormously different magnitudes, say the factorials and Gamma functions. To avoid large error caused by the fixed-precision floating point arithmetic standard in C/C++, we include the package GNU MPFR (for GNU Multiple Precision Floating-Point Reliably [62]), which is a portable C library for arbitrary-precision binary floating-point computation with correct rounding, based on GNU Multi-Precision Library.

5.4 Time Splitting

The main challenges come from the high dimensionality, nonlinearity, diffusive nature, the conservation properties, positivity, etc, which require very careful design of the numerical scheme. We divide and conquer starting from a time splitting method. For zero force field, i.e $F(t,x)=0$, the time-splitting is an efficient and reliable way for conquering inhomogeneous problems; however, we will employ the time-splitting to non-zero force field as well and show that it also works.

We discretize time $t_n = t_0 + n\Delta t$, where Δt is the time step size. Denote $f_n(x,v) = f(t_n,x,v)$. In a time interval $[t_n, t_{n+1}]$, a first order time splitting scheme turn the original problem into two subproblems

(1) The Vlasov (Collisionless) Problem

$$\begin{aligned} \partial_t g(x,v,t) + v\cdot \bigtriangledown_x g(x,v,t) + F(t,x)\cdot \nabla_v g = 0\,, \\ g(0,x,v) = f_n(x,v)\,, \end{aligned} \tag{5.33}$$

and

(2) The Homogenous FPL (Collisional) Problem

$$\begin{aligned} \partial_t \tilde{f}(x,v,t) = \frac{1}{\varepsilon} Q(\tilde{f},\tilde{f})\,, \\ \tilde{f}(0,x,v) = g(\Delta t,x,v)\,. \end{aligned} \tag{5.34}$$

If denote the above solution operators (5.76) and (5.34) by $A_n(\Delta t)$ and $H_n(\Delta t)$, respectively. Then the solution at time step t_{n+1} is given by

$$f_{n+1}(x,v) = H_n(\Delta t)\circ A_n(\Delta t) f_n(x,v)\,. \tag{5.35}$$

Remark. This splitting is first order in time. Higher order time splitting is also possible. For example, one common scheme is *Strang splitting*, which gives second order in time.

The above two steps can be performed with different methods. The collisionless step can be done with finite difference, finite volume or (DG)FEM; while the collisional step requires special techniques to handle the collisional operator. They will be introduced in the following sections.

5.5 The Conservative Spectral Method for Homogeneous FPL Equation

As mentioned in the time splitting scheme above, the collisionless and collisional subproblems can be treated separately with different methods. In the current section, we restrict ourselves to homogeneous FPL for the most interesting Coulombian case $\gamma = -3$, in 3d velocity space.

Different than the one proposed in [107], by taking truncated Fourier series and extending solutions by periodicity, we don't have to introduce non-physical binary collisions and simply extending the solution by zero, by means of the extension operator in Sobolev spaces as done in Chapter 3 for the approximation of the nonlinear Boltzmann collision operator. Conservation of moments are guaranteed by calling a conservation routine.

5.5.1 Domain of Computation

We assume that the distribution function f, the solution of the FPL equation, usually is not compactly supported in v but of negligible mass outside of a finite ball

$$B_L(\bar{v}) = \{v \in \mathbb{R}^3 : |v - \bar{v}| \leq R\}\,,$$

where $\bar{v}$ and R actually depends on x in the inhomogeneous case. However, numerically, in order to find an approximation in a finite domain, we assume f is compactly supported in the above ball.

Consider the cube

$$\Omega_v = \{v \in \mathbb{R}^3 : |v_i - \bar{v}_i| \leq L_v,\, i = 1, 2, 3\}\,,$$

which contains $B_L(\bar{v})$. This cube will be defined as the domain of computation for all velocity variables.

For the sake of simplicity, we assume a uniform discretization over the domain and also $\bar{v} = 0$. Let N be the number of discretizations in each direction of velocity, then the mesh for each direction of velocities is

$$h_v = \frac{2L_v}{N}, \qquad v_i = -L_v + ih_v,\, 0 \leq i < N\,.$$

In order to employ the standard FFT package [63], the corresponding mesh for the Fourier space should satisfy

$$h_v h_\xi = \frac{2\pi}{N}, \qquad L_\xi = \frac{N}{2}h_\xi, \qquad \xi_i = -L_\xi + ih_\xi, \quad 0 \leq i < N\,, \tag{5.36}$$

where h_v and L_v , h_ξ and L_ξ are the mesh size and cube side-length for the velocity and Fourier domain, respectively.

The whole mesh for the cubic domain will be the tensor product of the mesh on each direction.

5.5.2 Spectral Representation

We first look at the weak form of the FPL integrals. Suppose $\varphi(v)$ is smooth over the whole domain and the unknown f has exponentially decaying tails when $|v| \to \infty$ with some rate. For sake of simplicity, we drop the dependence on variable t and x.

Then, the weak form of the FPL operator is

$$\begin{aligned}\int_{\mathbb{R}^3} Q(f,f)\varphi(v)dv &= -\int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \mathbf{S}(v-v_*)(f_*\nabla f - (\nabla f)_* f)\cdot\nabla_v\varphi(v))dv_*dv \\ &= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} (\nabla_{v_*}\varphi(v_*) - \nabla_v\varphi(v))^T\mathbf{S}(v-v_*)f_*\nabla f dv_*dv\,.\end{aligned} \tag{5.37}$$

Let $\varphi(v) = (2\pi)^{-d/2}e^{-i\xi\cdot v}$ be the Fourier multiplier, and $u = v - v_*$.

Then,

$$\begin{aligned}
\widehat{Q}(\xi) &= (2\pi)^{-d/2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\mathbf{S}_{kl}(v-v_*)(\partial_k\varphi(v_*)-\partial_k\varphi(v))f_{v_*}\partial_l f(v)dv_*dv \\
&= (2\pi)^{-d/2}\int_{\mathbb{R}^3}\int_{\mathbb{R}^3}\mathbf{S}_{kl}(v-v_*)(-i\xi_k)e^{-i\xi\cdot v}(e^{-i\xi\cdot(v_*-v)}-1)f_{v_*}\partial_l f(v)dv_*dv \\
&= \int_{\mathbb{R}^3}du\mathbf{S}_{kl}(u)(-i\xi_k)(e^{i\xi\cdot u}-1)((2\pi)^{-d/2}\int_{\mathbb{R}^3}\tau_u f(v)\partial_l f(v)e^{-i\xi\cdot v}dv) \\
&= (2\pi)^{-d/2}\int_{\mathbb{R}^3}\widehat{\tau_u f}*\widehat{\partial_l f}(\xi)\mathbf{S}_{kl}(u)(-i\xi_k)(e^{i\xi\cdot u}-1)du \\
&= \int_{\mathbb{R}^3}d\omega\xi_k\omega_l\widehat{f}(\xi-\omega)\widehat{f}(\omega)((2\pi)^{-d/2}\int_{\mathbb{R}^3}\mathbf{S}_{kl}(u)(e^{i\omega\cdot u}-e^{-i(\xi-\omega)\cdot u})du) \\
&= \xi_k\int_{\mathbb{R}^3}[\widehat{\mathbf{S}}_{kl}(-\omega)-\widehat{\mathbf{S}}_{kl}(\xi-\omega)]\omega_l\widehat{f}(\xi-\omega)\widehat{f}(\omega)d\omega \\
&= \int_{\mathbb{R}^3}\left(\widehat{f}(\xi-\omega)\widehat{f}(\omega)\omega^T\widehat{\mathbf{S}}(\omega)\omega-(\xi-\omega)^T\widehat{\mathbf{S}}(\omega)(\xi-\omega)\widehat{f}(\xi-\omega)\widehat{f}(\omega)\right)d\omega\,,
\end{aligned} \tag{5.38}$$

where there is a summation over the same subscript indices.

Another weak form that is of interest is given by

$$\begin{aligned}
&\int_{\mathbb{R}^3}Q_{FPL}(f,f)\varphi(v)dv = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}(\nabla_{v_*}\varphi(v_*)-\nabla_v\varphi(v))^T\mathbf{S}(v-v_*)f_*\nabla f dv_*dv \\
&= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}ff_*\left(2[\nabla_v\cdot\mathbf{S}(v-v_*)]\cdot\nabla_v\varphi(v)+\mathbf{S}(v-v_*)\colon\nabla_v^2\varphi(v)\right)dvdv_*\,.
\end{aligned} \tag{5.39}$$

In addition, with the same derivation, we have

$$\begin{aligned}
\widehat{Q}(f,f)(\xi) &= \int_{\mathbb{R}^3}\int_{\mathbb{R}^3}ff_*e^{-i\xi\cdot v}G(\xi,u)dvdu \\
&= \int_{\mathbb{R}^3}\widehat{f}(\xi-\omega)\widehat{f}(\omega)\widehat{G}(\xi,\omega)d\omega\,.
\end{aligned} \tag{5.40}$$

where the precomputed weight in Fourier domain $\widehat{G}(\xi,\omega)$ is the same as given by the above (5.38), and the weight in velocity domain is

$$G(\xi,u) = |u|^{-3}\left(i4u\cdot\xi-|u|^2|\xi^\perp|^2\right)\,, \tag{5.41}$$

where $\xi^{\perp} = \xi - (\frac{\xi \cdot u}{|u|})\frac{u}{|u|}$. We point out that (5.41) can be also retrieved from the Fourier transform representation of the Boltzmann collision operator written as a weighted convolution of Fourier transforms. It is recently shown in [86] that the weight corresponding to the Boltzmann collision operator converges to the one for Landau operator, if collisions are grazing and the solutions of the BTE have some regularity and decay for large velocity.

It is easy to see that the above weighted convolution (5.38), since variables ω and $\xi - \omega$ are separable in the weights, leads to an $N^d \log(N)$ scheme (where N is the number of discretizations on each direction), when FFT is applied. In addition, the weights can be pre-computed and only have to be computed once. And, we will derive the above weight analytically, without any extra integral approximations.

Using the same notations to denote the truncated transforms (i.e integrated over some ball $u \in B_R(0)$ instead of the whole domain), we write

$$\widehat{\mathbf{S}_{kl}}(\omega) = (2\pi)^{-d/2} \int_{B_R(0)} \mathbf{S}_{kl}(u) e^{-i\omega \cdot u} du \,. \tag{5.42}$$

In addition, they can be decomposed into

$$\widehat{\mathbf{S}_{kl}}(\omega) = \widehat{\mathbf{S}^1_{kl}}(\omega) - \widehat{\mathbf{S}^2_{kl}}(\omega) \,, \tag{5.43}$$

with

$$\widehat{\mathbf{S}^1_{kl}}(\omega) = (2\pi)^{-d/2} \int_{B_R(0)} |u|^{\gamma+2} \delta_{kl} e^{-i\omega \cdot u} du$$
$$\widehat{\mathbf{S}^2_{kl}}(\omega) = (2\pi)^{-d/2} \int_{B_R(0)} |u|^{\gamma} u_k u_l e^{-i\omega \cdot u} du \,.$$

It is not hard to observe the following symmetry properties of $\widehat{\mathbf{S}_{kl}}(\omega)$

$$\begin{aligned}
&\widehat{\mathbf{S}^2_{11}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{33}}(\pm\omega_2,\pm\omega_3,\pm\omega_1) = \widehat{\mathbf{S}^2_{33}}(\pm\omega_3,\pm\omega_2,\pm\omega_1)\,,\\
&\widehat{\mathbf{S}^2_{22}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{33}}(\pm\omega_1,\pm\omega_3,\pm\omega_2) = \widehat{\mathbf{S}^2_{33}}(\pm\omega_3,\pm\omega_1,\pm\omega_2)\,,\\
&\widehat{\mathbf{S}^2_{12}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{21}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{13}}(\omega_1,\pm\omega_3,\omega_2) = -\widehat{\mathbf{S}^2_{13}}(-\omega_1,\pm\omega_3,\omega_2)\,,\\
&\widehat{\mathbf{S}^2_{23}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{32}}(\omega_1,\omega_2,\omega_3) = \widehat{\mathbf{S}^2_{13}}(\omega_2,\omega_1,\omega_3)\,.
\end{aligned} \tag{5.44}$$

Therefore, we only need to study, say, $\widehat{\mathbf{S}^1_{11}}$, $\widehat{\mathbf{S}^2_{33}}$ and $\widehat{\mathbf{S}^2_{13}}$. See Appendix for detailed derivations.

Then, consider the symmetry properties (5.44)

$$\widehat{\mathbf{S}}(\omega) = \begin{pmatrix}
\widehat{\mathbf{S}^1}(\omega) - \widehat{\mathbf{S}^2_{33}}(\omega_2,\omega_3,\omega_1) & -\widehat{\mathbf{S}^2_{13}}(\omega_1,\omega_3,\omega_2) & -\widehat{\mathbf{S}^2_{13}}(\omega)\\
-\widehat{\mathbf{S}^2_{13}}(\omega_1,\omega_3,\omega_2) & \widehat{\mathbf{S}^1}(\omega) - \widehat{\mathbf{S}^2_{33}}(\omega_1,\omega_3,\omega_2) & -\widehat{\mathbf{S}^2_{13}}(\omega_2,\omega_1,\omega_3)\\
-\widehat{\mathbf{S}^2_{13}}(\omega) & -\widehat{\mathbf{S}^2_{13}}(\omega_2,\omega_1,\omega_3) & \widehat{\mathbf{S}^1}(\omega) - \widehat{\mathbf{S}^2_{33}}(\omega)
\end{pmatrix}, \tag{5.45}$$

we observe that if we write $\widehat{\mathbf{S}}(\omega)$ as

$$\widehat{\mathbf{S}}(\omega) = 2\sqrt{\frac{2}{\pi}}\frac{R|\omega| - \sin(R|\omega|)}{R|\omega|^3}\widetilde{\Pi}(\omega)\,, \tag{5.46}$$

then, $\widetilde{\Pi}(\omega)$ is an orthogonal projection onto ω, i.e. $\widetilde{\Pi}(\omega)\omega = \omega$. And thus the weighted convolution becomes

$$\begin{aligned}
\widehat{\mathbf{Q}}(\widehat{f},\widehat{f}) &= \int_{\Omega_\xi}\Big(\widehat{f}(\xi-\omega)\widehat{f}(\omega)\omega^T\widehat{\mathbf{S}}(\omega)\omega - (\xi-\omega)^T\widehat{\mathbf{S}}(\omega)(\xi-\omega)\widehat{f}(\xi-\omega)\widehat{f}(\omega)\Big)\,d\omega\\
&= 2\sqrt{\frac{2}{\pi}}\int_{\Omega_\xi}\frac{R|\omega| - \sin(R|\omega|)}{R|\omega|}\widehat{f}(\omega)\widehat{f}(\xi-\omega)d\omega\\
&\quad - \int_{\Omega_\xi}(\xi-\omega)^T\widehat{\mathbf{S}}(\omega)(\xi-\omega)\widehat{f}(\xi-\omega)\widehat{f}(\omega)d\omega\,,
\end{aligned} \tag{5.47}$$

where $\Omega_\xi = [-L_\xi, L_\xi]^d$ with L_ξ defined in (5.36), and the first integral in the above last formula is zero if $|\omega| = 0$. Apparantly, it can be computed in $O(N^3 \log(N))$, through FFT.

5.5.3 Conservation Routines

Let $M = N^d$ be the total number of Fourier modes and

$$\mathbf{Q} = (Q_0, \ldots, Q_{M-1})^T \tag{5.48}$$

be the inverse discrete Fourier transform of $\widehat{\mathbf{Q}}$ in (5.47) and

$$\mathbf{F} = (F_1, \ldots, F_{M-1})^T \tag{5.49}$$

be the distribution vector at current time step.

The conservation routine here is actually following a similar argument as for the conservative DGFEM solver for nonlinear BTE. If we use some quadrature, say Trapezoidal rule, to evaluate the integrals over the whole velocity domain, then here the constrained matrix $\mathbf{C}$, analogous to (3.19), is of size $(d+2) \times M$ and defined as

$$\mathbf{C}_{:,j} = \begin{pmatrix} \omega_j \\ v\omega_j \\ |v|^2\omega_j \end{pmatrix}, \tag{5.50}$$

where ω_j is the j-th integration weight of the quadrature rule (say, Trapezoidal rule).

Following a same derivations, we obtain a conservation correction for the original collisional operator.

$$\mathbf{Q}_c = [\mathbf{I} - \mathbf{C}^T(\mathbf{C}\mathbf{C}^T)^{-1}\mathbf{C}]\mathbf{Q}\,. \tag{5.51}$$

Then, the semi-discrete problem (with conservations) is

$$\frac{d\mathbf{F}}{dt} = \mathbf{Q}_c \,. \tag{5.52}$$

5.5.4 Time Discretization

The high dimensionality and nonlinearity would make an implicit iterative time discretization really expensive. Thus, an explicit method is preferred. Due to the diffusive nature of the collision operator, a stiff problem has to be solved, and thus the corresponding stability condition forces the time step to be on the order of the square of the velocity step. We will show this property in the following. The original proof is due to [60] and can easily extend to our spectral method.

What we need to solve is the following problem

$$\frac{d}{dt}\widehat{f}(\xi_k) = F(\widehat{f}(\xi_k)) \,, \tag{5.53}$$

where

$$F(\widehat{f}(\xi_k)) = \frac{1}{\varepsilon}\widehat{\mathbf{Q}}(\widehat{f},\widehat{f})(\xi_k) \tag{5.54}$$

with $\widehat{\mathbf{Q}}(\widehat{f},\widehat{f})$ defined in (5.47).

In practice, we employ a fourth-order explicit Runge-Kutta scheme that achieves high temporal accuracy and at the same time does not ruin the spectral accuracy. Since the Runge-Kutta method is just a convex combination of first order Euler scheme, we only need to consider the first order Euler scheme

$$\widehat{f}^{n+1}(\xi_k) = \widehat{f}^{n}(\xi_k) + \Delta t F(\widehat{f}^{n}(\xi_k)) \,, \tag{5.55}$$

where the superscript n denotes the mode value at the n-th time step. The linear stability theory tells us the stability condition is determined by the eigenvalues of the Jacobian $\mathcal{J}_{k,l} = \frac{\partial F(\widehat{f}(\xi_k))}{\partial \widehat{f}(\xi_l)}$. We need to find an upper bound on the (negative) eigenvalues λ, such that $\lambda \Delta t < 1$.

Then, we have the following proposition

Proposition 5.5.1 (Stability condition for homogeneous FPL)**.** *For the first order Euler scheme, the time step Δt should satisfy the following stability condition,*

$$\Delta t \leq C L_v \varepsilon (\Delta v)^2 \,, \tag{5.56}$$

where L_v is the lateral size of the fixed velocity domain, ε is the Knudesen number and constant C only depends on the L^1 norm of the current solution f.

Proof. We rewrite (5.47) into two convolution forms

$$\widehat{\mathbf{Q}}(\widehat{f}, \widehat{f})(\xi) = \widehat{f} * G(\widehat{f})(\xi) - \sum_{i,j=1}^{d} H_{i,j}(\widehat{f}) * J_{i,j}(\widehat{\mathbf{S}}; \widehat{f})(\xi) \tag{5.57}$$

with, $\xi = (\xi^{(1)}, \xi^{(2)}, \ldots, \xi^{(d)})$ being defined component-wisely,

$$\begin{aligned} G(\widehat{f})(\xi) &:= 2\sqrt{\frac{2}{\pi}} \frac{R|\xi| - \sin(R|\xi|)}{R|\xi|} \widehat{f}(\xi) \,; \\ H_{i,j}(\widehat{f})(\xi) &:= \widehat{f}(\xi) \xi^{(i)} \xi^{(j)} \,; \\ J_{i,j}(\widehat{\mathbf{S}}; \widehat{f})(\xi) &:= \widehat{f}(\xi) \widehat{\mathbf{S}}_{i,j}(\xi) \,. \end{aligned} \tag{5.58}$$

The convolutions in (5.57) will be evaluated by the Trapezoidal quadrature

rule, with the Fourier nodes $\widehat{f}(\xi_k)$ being the quadrature points. That is,

$$\widehat{\mathbf{Q}}(\widehat{f},\widehat{f})(\xi_k) = h_\xi^d \sum_l \omega_l \left[\widehat{f}(\xi_k-\xi_l) G(\widehat{f})(\xi_l) - \sum_{i,j=1}^{d} H_{i,j}(\widehat{f})(\xi_k-\xi_l) J_{i,j}(\widehat{\mathbf{S}};\widehat{f})(\xi_l) \right], \tag{5.59}$$

where h_ξ is the step size in Fourier space as determined by (5.36), and ω_l are quadrature weights.

According to [60], the time step should satisfy

$$\Delta t \leq \frac{1}{\mathrm{Lip}(F(\cdot))}, \tag{5.60}$$

where $\mathrm{Lip}(F(\cdot))$ is the Lipschitz norm of $F(\cdot)$. This can be found through estimating the upper bound on the Jacobian

$$\begin{aligned}
|\mathcal{J}_{k,l}| &= \left| \frac{d}{d\widehat{f}(\xi_l)} F(\widehat{f}(\xi_k)) \right| \\
&\leq \frac{1}{\varepsilon} \frac{C}{L_v^d} \max\left(|\widehat{f}(\xi_k-\xi_l)|, |\widehat{f}(\xi_l)| \right) \\
&\cdot \left[\max_\xi \left| \frac{R|\xi| - \sin(R|\xi|)}{R|\xi|} \right| + |(\xi_k-\xi_l)^T \widehat{\mathbf{S}}(\xi_l)(\xi_k-\xi_l)| + |\xi_l^T \widehat{\mathbf{S}}(\xi_k-\xi_l)\xi_l| \right] \\
&\leq \frac{C}{\varepsilon L_v} |\widehat{f^n}(0)| L_\xi^2 \\
&\leq \frac{C}{\varepsilon L_v} \|f\|_{L^1(\mathbb{R}^d)} \frac{1}{(\Delta v)^2},
\end{aligned} \tag{5.61}$$

where the FFT relationship (5.36) is applied, and it is not hard to observe the following uniform bound estimates

$$|\widehat{\mathbf{S}}(\xi)| \lesssim L_v^{d-1}, \quad |\xi^T \widehat{\mathbf{S}}(\xi)\xi| \lesssim 1, \quad |(\xi-w)^T \widehat{\mathbf{S}}(\xi)(\xi-w)| \lesssim L_v^{d-1} L_\xi^2. \tag{5.62}$$

Therefore, the time step has to satisfy the stability condition

$$\Delta t \leq \frac{C}{\|f\|_{L^1(\mathbb{R}^d)}} L_v \varepsilon (\Delta v)^2 \,. \tag{5.63}$$

□

This proves (5.56). In practice, we employ a fourth-order explicit Runge-Kutta scheme and the conservation routine should be performed ad every intermediate step. Recall our discretization of time $t_n = t_0 + n\Delta t$, where Δt is the time step size. Denote by $\mathbf{F}_n$ the distribution vector at time step t_n. In a time interval $[t_n, t_{n+1}]$, the numerical evolution $\mathbf{F}_n \to \mathbf{F}_{n+1}$ follows

$$\widehat{\mathbf{F}_n} = \mathrm{FFT}(\mathbf{F}_n),\ \widehat{\mathbf{K}_n^1} = \mathrm{Compute}\left(\widehat{\mathbf{Q}}(\widehat{\mathbf{F}_n}, \widehat{\mathbf{F}_n})\right),\ \mathbf{K}_n^1 = \mathrm{IFFT}\left(\widehat{\mathbf{K}_n^1}\right),\ \mathbf{K}_n^1 = \mathrm{Conserve}(\mathbf{K}_n^1)$$

$$\widetilde{\mathbf{F}}_n = \mathbf{F}_n + \Delta t K_n^1;$$

$$\widehat{\widetilde{\mathbf{F}}_n} = \mathrm{FFT}(\widetilde{\mathbf{F}}_n),\ \widehat{\mathbf{K}_n^2} = \mathrm{Compute}\left(\widehat{\mathbf{Q}}(\widehat{\widetilde{\mathbf{F}}_n}, \widehat{\widetilde{\mathbf{F}}_n})\right),\ \mathbf{K}_n^2 = \mathrm{IFFT}\left(\widehat{\mathbf{K}_n^2}\right),\ \mathbf{K}_n^2 = \mathrm{Conserve}(\mathbf{K}_n^2)$$

$$\widetilde{\mathbf{F}}_n = \mathbf{F}_n + \frac{\Delta t}{2} K_n^1 + \frac{\Delta t}{2} K_n^2;$$

$$\widehat{\widetilde{\mathbf{F}}_n} = \mathrm{FFT}(\widetilde{\mathbf{F}_n}),\ \widehat{\mathbf{K}_n^3} = \mathrm{Compute}\left(\widehat{\mathbf{Q}}(\widehat{\widetilde{\mathbf{F}}_n}, \widehat{\widetilde{\mathbf{F}}_n})\right),\ \mathbf{K}_n^3 = \mathrm{IFFT}\left(\widehat{\mathbf{K}_n^3}\right),\ \mathbf{K}_n^3 = \mathrm{Conserve}(\mathbf{K}_n^3)$$

$$\widetilde{\mathbf{F}}_n = \mathbf{F}_n + \frac{\Delta t}{2} K_n^1 + \frac{\Delta t}{2} K_n^3;$$

$$\widehat{\widetilde{\mathbf{F}}_n} = \mathrm{FFT}(\widetilde{\mathbf{F}_n}),\ \widehat{\mathbf{K}_n^4} = \mathrm{Compute}\left(\widehat{\mathbf{Q}}(\widehat{\widetilde{\mathbf{F}}_n}, \widehat{\widetilde{\mathbf{F}}_n})\right),\ \mathbf{K}_n^4 = \mathrm{IFFT}\left(\widehat{\mathbf{K}_n^4}\right),\ \mathbf{K}_n^4 = \mathrm{Conserve}(\mathbf{K}_n^4)$$

$$\mathbf{F}_{n+1} = \mathbf{F}_n + \frac{1}{6}(3\mathbf{K}_n^1 + \mathbf{K}_n^2 + \mathbf{K}_n^3 + \mathbf{K}_n^4).$$

where $\widetilde{\mathbf{F}}_n$ a generic intermediate step; IFFT is the (discrete) fast inverse Fourier transform routine.

5.5.5 Numerical Results and Applications

5.5.5.1 Single Species Charge Carriers

We test our scheme to a sum of two Gaussians in 3D velocity space, to compute the evolution of entropy and moments and thus verify its validity,

$$f_0(v) = \frac{1}{2(2\pi\sigma^2)^{3/2}} \left[\exp\left(-\frac{|v - 2\sigma e|^2}{2\sigma^2} \right) + \exp\left(-\frac{|v + 2\sigma e|^2}{2\sigma^2} \right) \right], \quad (5.64)$$

with parameter $\sigma = \pi/10$ and $e = (1, 0, 0)$.

We select domain $\Omega_v = [-3, 3]^3$, number of modes in each direction $N = 32$, The entropy decays to its equilibrium state fast and keeps stable

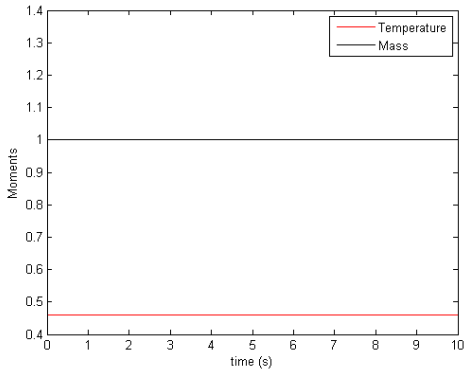


Figure 5.3: The evolution of moments of numerical solution

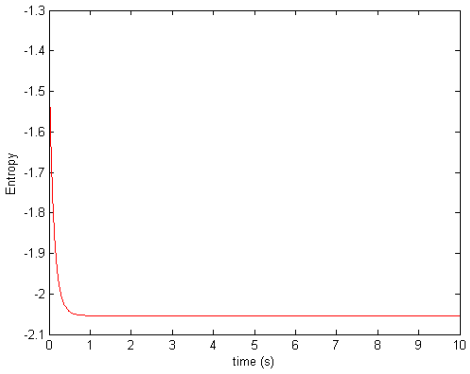


Figure 5.4: The Entropy decay of numerical solution

after that. The whole decay process preserves mass, momentum and kinetic energy. See Figure 5.3 and 5.4.

5.5.5.2 Multi-component Plasmas

In this section, we apply our scheme to a specific example of electro-neutral hydrogen plasma. The dimensionless system of equation writes [23]

$$\begin{aligned} \frac{\partial f_e}{\partial t} &= \frac{1}{2}\left[Q^{(1)}_{FPL}(f_e, f_e) + Q^{(\theta)}_{FPL}(f_e, f_i)\right] \\ \frac{\partial f_i}{\partial t} &= \frac{\theta^2}{2}\left[Q^{(1)}_{FPL}(f_i, f_i) + Q^{(1/\theta)}_{FPL}(f_i, f_e)\right], \end{aligned} \tag{5.65}$$

where $\theta < 1$ is the dimensionless mass ratio of electrons to ions; the subscripts e, i stand for electrons and ions respectively. For any $\theta > 0$,

$$Q^{(\theta)}_{FPL}(f, g) = \nabla_v \cdot \int \mathbf{S}(v - v_*)(f(v_*)\nabla_v g(v) - \theta f(v)\nabla_{v_*} g(v_*))dv_*\,, \tag{5.66}$$

with the projection matrix $\mathbf{S}$ defined in (5.4).

The system of equations (5.65) will be solved with normalized initials, that is, $\int f_e(v, 0)dv = \int f_i(v, 0)dv = 1$.

We can obtain the weak form of $Q^{(\theta)}_{FPL}(f, g)$

$$\begin{aligned} &\int Q^{(\theta)}_{FPL}(f, g)\varphi(v)dv \\ &= \iint f(v)g(v_*)\left[(1+\theta)\nabla_v \cdot \mathbf{S}(v - v_*)\nabla\varphi(v) + \mathbf{S}(v - v_*) : \nabla^2\varphi(v)\right] dv_* dv\,, \end{aligned} \tag{5.67}$$

where the Frobenius inner product $A : B = \text{Trace}(A^T B)$.

By taking $\varphi(v) = (2\pi)^{-d/2} e^{-i\xi\cdot v}$, which is the Fourier multiplier, then following a similar derivation as in (5.38) in the spectral representation of Q_{FPL}, we get

$$\widehat{Q^{(\theta)}_{FPL}}(\widehat{f}, \widehat{g})(\xi) = \int \widehat{f}(\xi - w)\widehat{g}(w)\left[(1+\theta)\xi^T \widehat{\mathbf{S}}(w)w - \xi^T \widehat{\mathbf{S}}(w)\xi\right] dw\,, \tag{5.68}$$

Remark. When $\theta = 1$ and $f = g$ in (5.66), the monotomic case (5.47) is recovered.

Inspired by the work of A.V. Bobylev et al [24], which is on radial (basically one-dimensional) Landau equations, we here study the non-isotropic distribution functions in 3-d velocity space.

We study the relaxation process of the space uniform two-temperature plasma theoretically, by deriving a set of governing ordinary differential equation. Also, the relaxation process will be demonstrated numerically, and at the same time to test our conservative spectral scheme.

First, we define the (dimensionless) time-dependent temperatures for electrons and ions

$$T_e(t) = \frac{1}{3}\int f_e(v,t)|v|^2 dv, \quad T_i(t) = \frac{1}{3\theta}\int f_i(v,t)|v|^2 dv\,. \tag{5.69}$$

While it is impossible to know the exact solution pair $f_{e,i}(v,t)$ for arbitrary initial data and the flow tends to equilibrate for large time, then we can replace them by constructing two Maxwellians, such that they owns the same temperatures, respectively,

$$M_e(v,t) = (2\pi T_e)^{-\frac{3}{2}}\exp\left(-\frac{|v|^2}{2T_e}\right), \quad M_i(v,t) = (2\pi\theta T_i)^{-\frac{3}{2}}\exp\left(-\frac{|v|^2}{2\theta T_i}\right). \tag{5.70}$$

Thus, one can derive

$$\begin{aligned}\frac{dT_e}{dt} &= \frac{1}{3}\frac{d}{dt}\int f_e(v,t)|v|^2 dv \\ &\approx \frac{1}{6}\left[Q^{(1)}_{FPL}(M_e,M_e) + Q^{(\theta)}_{FPL}(M_e,M_i)\right] = \frac{1}{6}Q^{(\theta)}_{FPL}(M_e,M_i)\,.\end{aligned} \tag{5.71}$$

Due to conservation of the kinetic energy of the two-plasma system, there is a constant $\bar{T}$, such that $T_e(t) + T_i(t) \equiv \bar{T}$. We obtain a set of ordinary differential equation governing the relaxation of the two-temperature plasma

$$\begin{aligned} (\theta\bar{T} + (1-\theta)T_e)^{\frac{3}{2}} \frac{dT_e}{dt} &= \frac{4}{3\sqrt{2\pi}}(T_e - T_i)\theta\,, \\ T_e(t) + T_i(t) &= \bar{T}\,, \end{aligned} \tag{5.72}$$

and the temperature difference follows

$$\frac{d(T_i - T_e)}{dt} = -\theta \frac{8}{3\sqrt{2\pi}} \frac{T_i - T_e}{(\theta\bar{T} + (1-\theta)T_e)^{\frac{3}{2}}}\,, \tag{5.73}$$

which implies

$$|T_i - T_e| \to 0, \quad \text{as } t \to \infty\,. \tag{5.74}$$

So, when t is large enough, or when the system approaches equilibrium, $T_e \approx T_i \approx \frac{\bar{T}}{2}$, the difference of temperatures decays "almost" exponentially (note that this is an approximation!)

$$|T_i(t) - T_e(t)| \approx |T_{i,0} - T_{e,0}| \exp\left(-\frac{16}{3\sqrt{\pi}} \frac{\theta}{\left((1+\theta)\bar{T}\right)^{3/2}} t \right)\,. \tag{5.75}$$

We solve the equation system (5.65) by our conservative scheme introduced above and observe the relaxation of temperatures for electrons and ions. The dimensionless mass ratio $\theta = \frac{1}{16}$. The initial states are two Maxwellians for hot ions and cold electrons, say $T_e = \frac{1}{2}$ and $T_i = \frac{3}{2}$ (then $\bar{T} = 2$) in (5.70). Figure 5.5 shows the decay to equilibrium of the 2-plasma system as expected. If we take the logarithm of the the temperature difference in (5.75), we can actually expect to observe the exponential decay rate in (5.75), which is

$-\frac{16}{3\sqrt{\pi}}\frac{\theta}{\left((1+\theta)\bar{T}\right)^{3/2}} = -0.061$ in this example. Figure 5.6 shows the logarithm of the temperature difference (scattered data) when time is large enough (states approaching equilibrium) and its linear fitting, with a slope of -0.066343, which is a rough verification of our analytical prediction.

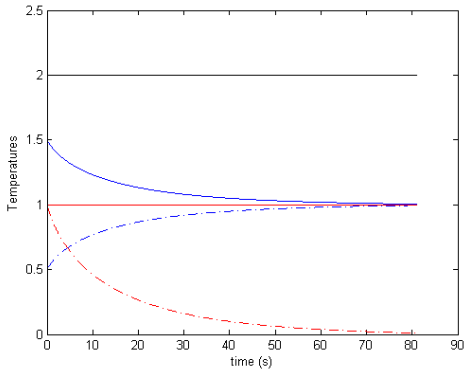


Figure 5.5: The relaxation of temperatures for the 2-plasma system: solid blue line: temperatures of ions; dash-dot blue: temperatures of electrons; top solid black: the total temperature; bottom dash-dot red: temperature difference

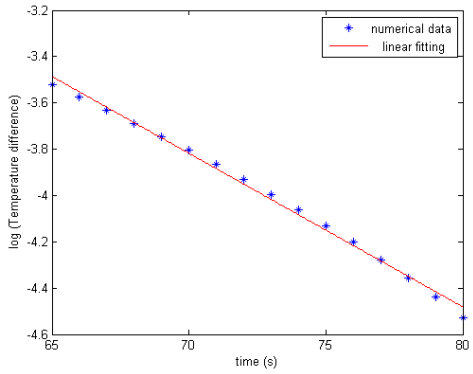


Figure 5.6: The logarithm of temperature difference for large time and its linear fitting

5.6 The RKDG Method for Vlasov-Poisson Equation

The VP system is a nonlinear kinetic system modeling the transport of charged particles in a collisionless plasma, under the effect of a self-consistent electrostatic field and possibly an externally supplied field. The electrostatic potential is coupled through Poisson equation. The collisionless VP exhibits a variety of dynamical phenomena, for example, the well-known filamentation

(filaments in phase space and steep gradients in v) and Landau damping.

With coupling to Poisson equation, the collisionless VP problem becomes

$$\text{The VP (Collisionless) Problem}$$

$$\begin{aligned}
\partial_t g(x,v,t) + v \cdot \bigtriangledown_x g(x,v,t) - \mathbf{E}(t,x) \cdot \nabla_v g = 0\,, \\
\mathbf{E}(t,x) = -\nabla_x \Phi(t,x)\,, \\
-\Delta_x \Phi(t,x) = 1 - \int_{R^3} g(t,x,v) dv\,, \\
\Phi(t,x) = \Phi_B(t,x) \quad x \in \partial\Omega_x\,, \\
g(0,x,v) = f_n(x,v)\,, \qquad (5.76)
\end{aligned}$$

where f_n is the current solution of the homogeneous Landau equation.

5.6.1 The Semi-discrete DG Form

In this section, we introduce a conservative Runge-Kutta Discontinuous Galerkin (RKDG) scheme for the VP equation (5.76), for $(x,v) \in \Omega = \Omega_x \times \Omega_v \subseteq \mathbb{R}^+ \times \mathbb{R}^d$. Or, we restrict the problem to the first spatial dimension $\mathbf{x} = (x,0,0)$, $\mathbf{E} = (E,0,0)$. The conservation properties are proved to be well satisfied if we choose a piecewise polynomial approximation space covering $d+2$ collision invariants.

We first list some notations for the DG method in use. Consider the computing domain $\Omega = \Omega_x \times \Omega_v = [0, L_x] \times [-L_v, L_v]^3$, 1D in x and 3D in v. Denote by $\mathcal{T}_h^x = I_x$ and $\mathcal{T}_h^v = K_v$ the regular partitions of Ω_x and Ω_v,

respectively, with

$$\mathcal{T}_h^x = \bigcup_1^{N_x} I_i = \bigcup_1^{N_x} [x_{i-1/2}, x_{i+1/2})$$

$$\mathcal{T}_h^v = \bigcup_{|j|=1}^{N_v^3} K_j = \bigcup_{j_1,j_2,j_3=1}^{N_v} [v_{j_1-1/2}, v_{j_1+1/2}) \times [v_{j_2-1/2}, v_{j_2+1/2}) \times [v_{j_3-1/2}, v_{j_3+1/2}),$$

with $x_{1/2} = 0$, $x_{N_x+1/2} = L_x$, $v_{1/2} = -L_v$ and $v_{N_v+1/2} = L_v$.

Then, $\mathcal{T}_h = \{E : E = I_x \times K_v, \forall I_x \in \mathcal{T}_h^x, \forall K_v \in \mathcal{T}_h^v\}$ defines a partition of Ω. Denote by ε_x and ε_v be set of edges of $\mathcal{T}_h^x$ and $\mathcal{T}_h^v$, respectively. Then, the edges of $\mathcal{T}_h$ will be $\varepsilon = \{I_x \times e_v : \forall I_x \in \mathcal{T}_h^x, \forall e_v \in \varepsilon_v\} \cup \{e_x \times K_v : \forall e_x \in \varepsilon_x, \forall K_v \in \mathcal{T}_h^v\}$. In addition, $\varepsilon_x = \varepsilon_x^i \cup \varepsilon_x^b$ with ε_x^i and ε_x^b being the interior and boundary edges, respectively. Same for variable velocity domain. The mesh size $h = \max(h_x, h_v) = \max_{E \in \mathcal{T}_h} \operatorname{diam}(E)$, with $h_x = \max_{I_x \in \mathcal{T}_h^x} \operatorname{diam}(I_x)$ and $h_v = \max_{K_v \in \mathcal{T}_h^v} \operatorname{diam}(K_v)$.

Next, we define the following approximation space. (Note that we only have 1D in x):

$$X_h^l = \{f \in L_\Omega^2 : g|_E \in P^l(I_x) \times P^l(K_v), \forall E = I_x \times K_v \in \mathcal{T}_h\}, \tag{5.77}$$

and

$$W_h^l = \{f \in L_\Omega^2 : g|_E \in P^l(I_x) \times Q^l(K_v), \forall E = I_x \times K_v \in \mathcal{T}_h\}, \tag{5.78}$$

where $P^l(K)$ denotes the space of polynomials of total degree at most l on some element K, while Q^l the space of polynomials of degree l in each variable on K. $P^l(K)$ has number of degrees of freedom $(l+1)^d$, while $Q^l(K)$ has degrees of freedom $\sum_{i=0}^{l} \binom{i+d-1}{d-1}$ (here $d = 3$).

Since basis polynomials are piecewise defined over each element, we need to introduce the concepts of jumps and averages. For any test function $\phi_h(x,v) \in X_h^l$ (or, W_h^l), define $(\phi_h)^{\pm}_{i+1/2,v} = \lim_{\epsilon\to 0}\phi_h(x_{i+1/2}\pm\epsilon, v)$, $(\phi_h)^{\pm}_{x,K_v} = \phi_h|_{K_v^{\pm}}$. For any edge $e_x \in \varepsilon_x$, which is actually one end point of intervals, and any edge $e_v \in \varepsilon_v$, with $\mathbf{n}_v^{\pm}$ as the outward unit normal to $\partial K_v^{\pm}$, the jump across e_x and e_v are defined as

$$[\phi_h]_{x_i} = (\phi_h)^+_{i-1/2,v} - (\phi_h)^-_{i-1/2,v}, \quad [\phi_h]_v = (\phi_h)^+_{x,K_v}\mathbf{n}_v^+ + (\phi_h)^-_{x,K_v}\mathbf{n}_v^- . \quad (5.79)$$

and the averages are

$$\{\phi_h\}_{x_i} = \frac{1}{2}((\phi_h)^+_{i-1/2,v} + (\phi_h)^-_{i-1/2,v}), \quad \{\phi_h\}_v = \frac{1}{2}((\phi_h)^+_{x,K_v} + (\phi_h)^-_{x,K_v}) . \quad (5.80)$$

Here and below, we denote by E_h the discrete electric field computed from the Poisson's equation. With proper partitioning, we can assume each direction of v has a single sign.

The DG scheme for the nonlinear VP equation is described as follows. We seek an approximation solution $g_h(x,v) \in X_h^l$ (or, W_h^l), such that, for any test function $\phi_h(x,v) \in X_h^l$ (or, W_h^l),

$$\int_{I_i\times K_j} (g_h)_t \varphi_h dxdv = H_{i,j}(g_h, E_h, \varphi_h) \quad (5.81)$$

where

$$\begin{aligned} &H_{i,j}(g_h, E_h, \varphi_h) \\ &= \int_{I_i\times K_j} v_1 g_h (\varphi_h)_x dxdv - \int_{K_j} (\widehat{v_1 g_h}\varphi_h^-)_{i+\frac{1}{2},v} dv + \int_{K_j} (\widehat{v_1 g_h}\varphi_h^+)_{i-\frac{1}{2},v} dv \\ &- \int_{I_i\times K_j} E_h g_h \partial_{v_1}\varphi_h dxdv + \int_{I_i}\int_{\varepsilon_v} (\widehat{E_h g_h}\varphi_h^-)_{x,j_1+\frac{1}{2}} ds_v dx - \int_{I_i}\int_{\varepsilon_v} (\widehat{E_h g_h}\varphi_h^+)_{x,j_1-\frac{1}{2}} ds_v dx . \end{aligned} \quad (5.82)$$

Here, $j = (j_1, j_2, j_3)$ is the multi-index, corresponding to the three directions of v. The following upwinding fluxes (the trace at the element interfaces) are used,

$$\widehat{v_1 g_h} = \begin{cases} v_1 g_h^-, & \text{if } v_1 \geq 0 \text{ in } K_j; \\ v_1 g_h^+, & \text{if } v_1 < 0 \text{ in } K_j. \end{cases} \tag{5.83}$$

and

$$\widehat{E_h g_h} = \begin{cases} E_h g_h^-, & \text{if } \int_{I_i} E_h dx \leq 0; \\ E_h g_h^+, & \text{if } \int_{I_i} E_h dx > 0. \end{cases} \tag{5.84}$$

The electric field is solved from the Poisson's equation, as is used in [49]. In the one-dimensional case, the exact solution of the Poisson's equation can be obtained through the classical representation of Green's function, if we enforce the periodicity condition $\Phi(0) = \Phi(L_x)$,

$$\Phi_h = \int_0^x \int_0^s \rho_h(z,t) dz dx - \frac{x^2}{2} - C_E x \,, \tag{5.85}$$

where $\rho_h = \int_{\Omega_v} g_h dv$, $C_E = -\frac{L_x}{2} + \frac{1}{L_x} \int_0^{L_x} \int_0^s \rho_h(z,t) dz ds$, and

$$E_h = -\Phi' = C_E + x - \int_0^x \rho_h(z,t) dz \,. \tag{5.86}$$

The above semi-DG problem (5.81) can be solved by coupling with a suitable time discretization, e.g. total variation diminishing (TVD) Runge-Kutta method. The third order TVD-RK method for evolving $t_n \to t_{n+1}$ is

implemented as

$$\begin{aligned}
\int_{I_i\times K_j} g_h^{(1)}\varphi_h dxdv &= \int_{I_i\times K_j} g_h^n\varphi_h dxdv + \Delta t H_{i,j}(g_h^n, E_h^n, \varphi_h)\,, \\
\int_{I_i\times K_j} g_h^{(2)}\varphi_h dxdv &= \frac{3}{4}\int_{I_i\times K_j} g_h^n\varphi_h dxdv + \frac{1}{4}\int_{I_i\times K_j} g_h^{(1)}\varphi_h dxdv + \frac{\Delta t}{4} H_{i,j}(g_h^{(1)}, E_h^{(1)}, \varphi_h)\,, \\
\int_{I_i\times K_j} g_h^{n+1}\varphi_h dxdv &= \frac{1}{3}\int_{I_i\times K_j} g_h^n\varphi_h dxdv + \frac{2}{3}\int_{I_i\times K_j} g_h^{(2)}\varphi_h dxdv + \frac{2\Delta t}{3} H_{i,j}(g_h^{(2)}, E_h^{(2)}, \varphi_h)\,,
\end{aligned} \tag{5.87}$$

where $E_h^{(1)}, E_h^{(2)}$ are also obtained through the exact representation (5.86). Readers can refer to [112] for a detailed introduction to TVD Runge-Kutta methods.

This completes the RKDG scheme for nonlinear VP problem. We propose to apply basis function $\varphi_h|_{K_j} = 1, v, |v|^2$, as is proposed in study of Vlasov-Maxwell equations in [44], hoping that the RKDG scheme can well preserve mass, momentum and energy.

5.6.2 Positivity-preserving Limiters

To ensure a positive DG solution, many authors have successfully applied positivity-preserving limiters in the intermediate time steps. Please refer to [50, 119–122] for full descriptions and applications. We summarize the scheme here. For each intermediate step of Runge-Kutta method,

- On each mesh element $E_{i,j} = I_i \times K_j$, compute $T_{i,j} := \min_{(x,v)\in S_{i,j}} g_h(x,v)$, where $S_{i,j} = \left(S_i^x \otimes \hat{S}_j^v\right) \cup \left(\hat{S}_i^x \otimes S_j^v\right)$, and S_i^x, S_j^v are sets of $(l+1)$ Gauss

quadrature points and $\hat{S}_i^x, \hat{S}_j^v$ sets of $(l+1)$ Gauss-Lobatto quadrature points.

- Compute $\widetilde{f}_h(x,v) = \theta\big(g_h(x,v) - (\overline{g_h})_{i,j}\big) + (\overline{g_h})_{i,j}$ with $(\overline{g_h})_{i,j}$ the average over element $E_{i,j}$ and $\theta = \min\{1, |(\overline{g_h})_{i,j}|/|T_{i,j} - (\overline{g_h})_{i,j}|\}$.
- Update $g_h \leftarrow \widetilde{g_h}$.

The above limiter adjusts the function to be positive while preserving the cell average. Thus, application of such positivity-preserving limiter still achieves conservation of total mass, yet however will deteriorate the conservation of energy. This limiter maybe added when necessary, but for the time being, we would like to highlight the conservation of all desired moments.

5.6.3 Conservation and L^2-Stability

A piecewise polynomial approximation subspace containing all collision invariants will be applied. We will show, the total mass (charge) and momentum is conserved, up to some boundary effects; as for the total energy, the variation relies on the approximation accuracy of the solution together with the projection error of the potential Φ_h. Also, the approximate solution is L^2 stable. The following propositions are extensions of some related results studied in [9, 49, 79] to higher dimensions.

Proposition 5.6.1 (Conservations of total mass and momentum)**.** *The approximate solution $g_h \in X_h^l$ (or, W_h^l) for semi-DG problem (5.81) satisfies*

$$\frac{d}{dt}\int_{\mathcal{T}_h} g_h dxdv = \Theta_{h,1}(g_h, E_h)\,, \tag{5.88}$$

with

$$\Theta_{h,1}(g_h, E_h) = \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h})_{x,N_v+\frac{1}{2}} ds_v dx - \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h})_{x,\frac{1}{2}} ds_v dx \,, \quad (5.89)$$

and

$$\frac{d}{dt} \int_{\mathcal{T}_h} g_h v dx dv = \Theta_{h,2}(g_h, E_h) \,, \quad (5.90)$$

with

$$\Theta_{h,2}(g_h, E_h) = \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h} v)_{x,N_v+\frac{1}{2}} ds_v dx - \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h} v)_{x,\frac{1}{2}} ds_v dx \,. \quad (5.91)$$

Here, boundary effects $\Theta_{h,1}(g_h, E_h)$ and $\Theta_{h,2}(g_h, E_h)$ are negligible if Ω_v is large enough or equal zero if assume a compact support in the velocity space for g_h.

Proof. Take $\varphi_h = 1$, then

$$\begin{aligned} \sum_{i,j} H_{i,j}(g_h, E_h, 1) &= \int_{\mathcal{T}_h^v} \int_{\varepsilon_x} \widehat{v_1 g_h}[1]_x ds_x dv - \int_{\mathcal{T}_h^x} \int_{\varepsilon_v} \widehat{E_h g_h}[1]_{v_1} ds_v dx \\ &= \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h})_{x,N_v+\frac{1}{2}} ds_v dx - \int_{\mathcal{T}_h^x} \int_{\varepsilon_v^b} (\widehat{E_h g_h})_{x,\frac{1}{2}} ds_v dx \,. \end{aligned} \quad (5.92)$$

where considering the periodicity in x.

Take $\varphi_h = v_1$, then

$$\sum_{i,j} H_{i,j}(g_h, E_h, v_1) = \int_{\mathcal{T}_h^v} \int_{\varepsilon_x} \widehat{v_1 g_h}[v_1]_x ds_x dv - \int_{\mathcal{T}_h^x} \int_{\mathcal{T}_h^v} E_h g_h dx dv - \int_{\mathcal{T}_h^x} \int_{\varepsilon_v} \widehat{E_h g_h}[v_1]_{v_1} ds_v dx \,. \quad (5.93)$$

The first term above is zero due to the periodic boundary conditions; the third term is the boundary effect same as above; let's only look at the second term.

Thanks to the exact solver for Poisson equation (5.85) and (5.86),

$$\int_{\mathcal{T}_h^x}\int_{\mathcal{T}_h^v} E_h g_h dxdv = \int_{\mathcal{T}_h^x} \rho_h E_h dx = -\int_{\mathcal{T}_h^x} E_h (E_h)_x dx + \int_{\mathcal{T}_h^x} E_h dx = 0\,. \quad (5.94)$$

The cases for $\varphi_h = v_2$ and $\varphi_h = v_3$ follow a same way. □

Proposition 5.6.2 (Variation of total energy)**.** *The total energy of the approximate solution $g_h \in X_h^l$ (or, W_h^l) for semi-DG problem (5.81) satisfies*

$$\frac{d}{dt}\left(\frac{1}{2}\int_{\mathcal{T}_h} g_h|v|^2 dxdv + \frac{1}{2}\int_{\mathcal{T}_h^x} |E_h|^2 dx\right) = \Theta_{h,3}(g_h, E_h) = \Theta_{h,3}(g_h - g, \Phi_h - \mathbf{P}\Phi_h)\,, \quad (5.95)$$

with

$$\Theta_{h,3}(g_h, E_h) = \int_{\mathcal{T}_h} (\Phi_h)_x g_h v_1 dxdv - \int_{\mathcal{T}_h} \Phi_h (g_h)_t dxdv\,, \quad (5.96)$$

where $\mathbf{P}\Phi_h$ is the projection of Φ_h onto X_h^l (or, W_h^l) and $\mathbf{P}\Phi_h = \Phi_h$ on all interfaces of $\mathcal{T}_h^x$ (such that $\mathbf{P}\Phi_h$ is continuous).

Proof. Take $\varphi_h = \frac{1}{2}|v|^2$, then

$$\sum_{i,j} H_{i,j}(g_h, E_h, \frac{1}{2}|v|^2) = \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \widehat{v_1 g_h}\frac{1}{2}[|v|^2]_x ds_x dv - \int_{\mathcal{T}_h^x}\int_{\varepsilon_v} \widehat{E_h g_h}\frac{1}{2}[|v|^2]_{v_1} ds_v dx - \int_{\mathcal{T}_h} E_h g_h \partial_{v_1}\varphi_h dxdv\,. \quad (5.97)$$

The first term above is zero due to the periodicity; the second term is the boundary effect, which is zero if we assume the solution if compactly supported in Ω_v.

On the other hand, noticing again the exact Poisson solver (5.85) and (5.86),

$$\begin{aligned}\frac{1}{2}\int_{\mathcal{T}_h^x}|E_h|^2dx &= \int_{\mathcal{T}_h^x}E_h(E_h)_t dx = -\int_{\mathcal{T}_h^x}(\Phi_h)_x(E_h)_t dx \\ &= \int_{\mathcal{T}_h^x}\Phi_h(1-\rho_h)_t dx = -\int_{\mathcal{T}_h}\Phi_h(g_h)_t dxdv\,.\end{aligned} \tag{5.98}$$

which gives (5.99).

If take $\varphi_h = \mathbf{P}\Phi_h \in X_h^l$ (or, W_h^l), then we obtain $\Theta_{h,3}(g_h, \mathbf{P}\Phi_h) = 0$, which is also valid for the exact solution g. The exact solution g also obviously conserves total energy, which implies $\Theta_{h,3}(g, \Phi_h - \mathbf{P}\Phi_h) = 0$. Thus, $\Theta_{h,3}(g_h, E_h) = \Theta_{h,3}(g_h - g, \Phi_h - \mathbf{P}\Phi_h)$. □

This proposition means the variation of total energy relies on the numerical error of $g - g_h$ and projection error $\Phi_h - \mathbf{P}\Phi_h$. If the Poisson equation is not solved by exact formula but instead through a local DG method, then with special choice of flux, the total energy on the discrete level is proven to be conserved, see [9]. But here, we focus on the inhomogeneous model coupled with the Landau collision operator, thus the exact Poisson solver is preferred without many extra efforts. Actually, when a relatively fine DG mesh is applied, the variations on total energy are negligible.

Proposition 5.6.3 (L^2-stability)**.** *The approximate solution $g_h \in X_h^l$ (or, W_h^l) for semi-DG problem (5.81) decays enstrophy*

$$\frac{d}{dt}\int_{\mathcal{T}_h} g_h^2 dxdv = \Theta_{h,4}(g_h, E_h) \le 0\,, \tag{5.99}$$

with

$$\Theta_{h,4}(g_h, E_h) = -\frac{1}{2}\int_{\mathcal{T}_h^v}\int_{\varepsilon_x} |v_1|[g_h]_x^2 ds_x dv - \frac{1}{2}\int_{\mathcal{T}_h^x}\int_{\varepsilon_v} |E_h|[g_h]_{v_1}^2 ds_v dx\,. \quad (5.100)$$

Proof. Take $\varphi_h = g_h$, then

$$\begin{aligned}\sum_{i,j} H_{i,j}(g_h, E_h, g_h) &= \int_{\mathcal{T}_h} v_1 g_h (g_h)_x dx dv + \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \widehat{v_1 g_h}[g_h]_x ds_x dv \\ &\quad - \int_{\mathcal{T}_h} E_h g_h (g_h)_{v_1} dx dv - \int_{\mathcal{T}_h^x}\int_{\varepsilon_v} \widehat{E_h g_h}[g_h]_{v_1} ds_v dx \\ &:= a_1 + a_2\,,\end{aligned} \quad (5.101)$$

where, noticing the definition of upwinding flux

$$\widehat{v_1 g_h} = \{v_1 g_h\}_x - \frac{|v_1|}{2}[g_h]_x\,, \quad (5.102)$$

we have,

$$\begin{aligned} a_1 &= \int_{\mathcal{T}_h} v_1 g_h (g_h)_x dx dv + \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \widehat{v_1 g_h}[g_h]_x ds_x dv \\ &= \int_{\mathcal{T}_h} v_1 \left(\frac{g_h^2}{2}\right)_x dx dv + \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \widehat{v_1 g_h}[g_h]_x ds_x dv \\ &= -\int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \frac{1}{2}[v_1 g_h^2]_x ds_x dv + \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \widehat{v_1 g_h}[g_h]_x ds_x dv \\ &= \int_{\mathcal{T}_h^v}\int_{\varepsilon_x} \left(-\frac{1}{2}[v_1 g_h^2]_x + \{v_1 g_h\}_x [g_h]_x - \frac{|v_1|}{2}[g_h]_x^2\right) ds_x dv \\ &= -\frac{1}{2}\int_{\mathcal{T}_h^v}\int_{\varepsilon_x} |v_1|[g_h]_x^2 ds_x dv\,, \end{aligned} \quad (5.103)$$

and similarly

$$\begin{aligned} a_2 &= -\int_{\mathcal{T}_h} E_h g_h (g_h)_{v_1} dxdv - \int_{\mathcal{T}_h^x}\int_{\varepsilon_v} \widehat{E_h g_h}[g_h]_{v_1} ds_v dx \\ &= -\int_{\mathcal{T}_h} E_h \left(\frac{g_h^2}{2}\right)_{v_1} dxdv - \int_{\mathcal{T}_h^x}\int_{\varepsilon_v} \widehat{E_h g_h}[g_h]_{v_1} ds_v dx \\ &= \int_{\mathcal{T}_h^x}\int_{\varepsilon_v} \left(\frac{1}{2}[E_h g_h^2]_{v_1} - \{E_h g_h\}_{v_1}[g_h]_{v_1} - \frac{|E_h|}{2}[g_h]_{v_1}^2\right) ds_v dx \\ &= -\frac{1}{2}\int_{\mathcal{T}_h^x}\int_{\varepsilon_v} |E_h|[g_h]_{v_1}^2 ds_v dx\,. \end{aligned} \tag{5.104}$$

So, $\Theta_{h,4}(g_h, E_h) = a_1 + a_2 \leq 0$.

□

5.7 The Linking Process - Conservative Projection

So far, we have solved two subproblems separately: Vlasov-Poisson equation and homogeneous Fokker-Planck-Landau equation. The next step is to link them together, i.e project the Fourier series solution of the homogeneous FPL equation onto the DG mesh. If denote by $F_n(f)$ the Fourier series solution of the homogeneous FPL equation at the n-th time step, and $P : L^2(\Omega_v) \to X_h^l$ (or, W_h^l) the L^2 projection, then according to the time splitting scheme, the initial condition for $(n+1)$-st Vlasov-Poisson problem (5.33) is

$$g(0, x, v) = P(F_n(f))(x, v)\,. \tag{5.105}$$

During this linking process, conservation of desired moments will be lost if the conservation routine correct the collision operator on discrete level of Fourier modes, as in Section 5.5.3. Thus, such a concern inspires us to

develop a new conservation routine, which conserves moments fully on level of piecewise polynomials in the DG space X_h^l (or, W_h^l). This is shown in the following.

Take $d = 3$ for example. After having $\widehat{\mathbf{Q}}(f, f)(\xi)$, the $Q(f, f)(v)$ will be approximated by a partial sum of Fourier series,

$$Q_N(f, f)(v) = \frac{(2\pi)^{3/2}}{(2L)^3} \sum_{|k| \leq N^3} \widehat{\mathbf{Q}}(\xi_k) e^{i\xi_k \cdot v}\,, \tag{5.106}$$

where $\xi_k = \frac{\pi k}{L}$ are the spectral modes, $k = (k_1, k_2, k_3)$ is the multi-index.

Let $M = N^3$ be the total number of discretizations in the velocity space, i.e. the total number of Fourier modes. We will find corrected mode coefficients $\widehat{\mathbf{Q}}(\xi_k)$, such that

$$\int_{\Omega_v} Q_N(f, f)(v)\phi(v)dv = 0$$

is independent of the quadrature rules. Here $\phi(v)$ are the collision invariants.

Plugging this back into (5.106) gives constraints on the corrected mode coefficients. If denote by $\widehat{\mathbf{Q}}$ the vector of mode coefficients, $\widehat{\mathbf{Q}}_R$, $\widehat{\mathbf{Q}}_I \in \mathbb{R}^M$ the real and imaginary parts, respectively, then

$$\mathbf{C}_R \widehat{\mathbf{Q}}_R - \mathbf{C}_I \widehat{\mathbf{Q}}_I = \mathbf{0}\,, \tag{5.107}$$

where the constraint matrices $\mathbf{C}_R$, $\mathbf{C}_I \in \mathbb{R}^{5 \times M}$, are the real and imaginary parts of the following

$$\mathbf{C}_R(l, k) + i\mathbf{C}_I(l, k) = \frac{1}{(2L)^3} \int_{\Omega_v} e^{i\xi_k \cdot v} \phi_l(v)dv\,, \tag{5.108}$$

where $\phi_l(v) = 1, v, |v|^2$.

Indeed,

$$\begin{aligned}
&\mathbf{C}_R(0,k) = \prod_{i=1}^{3} \mathrm{sinc}(L\xi_{k_i}), \qquad \mathbf{C}_I(1,k) = 0 \\
&\mathbf{C}_R(l,k) = 0, \qquad \mathbf{C}_I(l,k) = \begin{cases} \frac{\mathrm{sinc}(L\xi_{k_l}) - \cos(L\xi_{k_l})}{\xi_{k_l}} \prod_{i\neq l}^{3} \mathrm{sinc}(L\xi_{k_i}) & \xi_{k_l} \neq 0; \\ 0 & \xi_{k_l} = 0 \end{cases}, \quad l = 1,2,3 \\
&\mathbf{C}_R(4,k) = \sum_{l=1}^{3} \left(\prod_{i\neq l}^{3} \mathrm{sinc}(L\xi_i) \right) \cdot \begin{cases} L^2 \mathrm{sinc}(L\xi_l) - 2\frac{\mathrm{sinc}(L\xi_l) - \cos(L\xi_l)}{\xi_l^2} & \xi_l \neq 0; \\ \frac{L^2}{3} & \xi_l = 0 \end{cases}, \\
&\mathbf{C}_I(4,k) = 0
\end{aligned} \tag{5.109}$$

The conservation correction is found by solving the following constrained optimization problem: Find $\widehat{\mathbf{Q}} = [\widehat{\mathbf{Q}}_R^T, \widehat{\mathbf{Q}}_I^T]^T \in \mathbb{R}^{2M}$, the minimizer of the optimization problem

$$\begin{aligned}
&\min \quad \|\widehat{\mathbf{Q}}_o - \widehat{\mathbf{Q}}\|_2^2 \\
&\text{s.t} \quad \mathbf{C}\widehat{\mathbf{Q}} = \mathbf{0}\,,
\end{aligned} \tag{5.110}$$

where $\widehat{\mathbf{Q}}_o$ is the original mode coefficient vector at the current time step; $\mathbf{C} = [\mathbf{C}_R, -\mathbf{C}_I] \in \mathbb{R}^{5\times 2M}$.

Following a same derivation, we obtain the conservative correction $\widehat{\mathbf{Q}_c}$

$$\widehat{\mathbf{Q}_c} = \left[\mathbf{I} - \mathbf{C}^T \left(\mathbf{C}\mathbf{C}^T \right)^{-1} \mathbf{C} \right] \widehat{\mathbf{Q}}_o\,, \tag{5.111}$$

where $\mathbf{I}$ is a $2M \times 2M$ identity matrix.

Thus, in the temporal evolution, the above *CONSERVE* (5.111) and *RECONSTRUCT* (5.106) routines have to be implemented at every intermediate step of Runge-Kutta schemes.

Remark. We expect the whole discrete scheme to be stable and also to be able to construct a priori error estimates. These two goals will be done in a future project.

5.8 Parallelization

One common feature for nearly all realistic kinetic models is the high dimensionality. Plus the higher than linear complexity, it addresses the importance of implementations of parallel computing.

For RKDG schemes for VP problem, the parallelization becomes more natural due to the locality of basis functions. Once all the nodes can access to the information from previous time step, the evolution of each grid point is done independently without communications across computing nodes. After evolution is done for the current time step for all nodes, the information will be gathered together and redistributed to all computing nodes in the community. We will use Message Passing Interface (MPI) [65] to distribute the velocity grid points.

For the spectral solver for homogeneous FPL equation, at each time step, a single grid point only "sees" the particles at the same spatial grid point, through the collision term. Since collisions involve all participating particles and take most of the computation time, to avoid large amount of communicating latency, we restrict all of the information needed for the current time step on the same computing node, and thus only distribute spatial grid points across the computing node community.

To further parallelize the computation, we realize that, for each phase velocity grid point ξ, the computation of $\widehat{Q}(\xi)$ is a weighted sum over all phase velocities w, with no information interrupted by other grid points ξ's. Similar features also apply to the integrations in RKDG method for the VP problem. Thus, the work load will be further shared using OpenMP [13].

As the majority of computations occur in the collision steps, the computational time consumed in collisions will dominate. Since all information needed for collisions will be kept on the same computing node and only spatial grid points are distributed, an almost linear strong scaling efficiency would be expected. We run tests on a typical linear Landau damping problem for the Landau-Poisson system, and record the time consumed for one single time step in Table 5.1. This example is associated with the one in Figure 5.9.2. Tests run on Xeon Intel 3.33GHz Westmere processors (on cluster Lonestar-TACC [103]).

nodes	cores	wall clock time (s)
1	12	1228.18
2	24	637.522
4	48	307.125
8	96	154.385
16	192	80.6144
32	384	41.314

Table 5.1: The wall clock time for one single time step of a typical linear Landau damping problem.

5.9 Numerical Results

Our target is a two-species plasma system of electrons and ions.

5.9.1 Electron Plasma Waves

In most plasmas of interest, the ion temperature is much smaller than the electron temperature. Together with the fact that electrons have much smaller mass, the ions may be assumed to be stationary. If we assume the ions temperature is negligible compared to the electron temperature, i.e $T_i/T_e \sim 0$, we may assume the ions obey a Dirac measure [53], or see [54] for physical derivations,

$$f_i(t,x,v) = \rho_i(t,x)\delta_0(v-\bar{v}_i)\,, \tag{5.112}$$

where ion density ρ_i and mean velocity $\bar{v}_i$ are given or satisfy certain hydrodynamic equations. Then, we get the ion-electron collision operator

$$Q_{e,i}(f_e) = \rho_i \nabla_v \cdot (\mathbf{S}(v-\bar{v}_i)\nabla_v f_i(v))\,, \tag{5.113}$$

which is basically a linear operator w.r.t distribution f_i.

The weak form of (5.113) reads

$$\int_{\mathbb{R}^3} Q_{e,i}(f_e)\varphi(v)dv = -\rho_i \int_{\mathbb{R}^3} (\mathbf{S}(v-\bar{v}_i)\nabla_v f_i(v)) \cdot \nabla_v \varphi(v) dv\,, \tag{5.114}$$

from which it is not difficult to prove that the linear operator (5.113) conserves mass and energy, by noticing that the zero eigen-space of projection matrix $\mathbf{S}(v)$ is spanned by v.

Similar to the spectral representation of the fully nonlinear collision operator (5.38), we can also obtain the spectral representation for (5.113)

$$\begin{aligned}\widehat{Q}_{e,i}(\widehat{f}_e) &= i(2\pi)^{-3/2}\int_{\mathbb{R}^3}\xi^T\mathbf{S}(v)\nabla_v f_e\exp(-iv\cdot\xi)dv \\ &= -(2\pi)^{-3/2}\int_{\mathbb{R}^3}\xi^T\mathbf{S}(w)(\xi-w)\widehat{f}_e(\xi-w)dw\,.\end{aligned} \tag{5.115}$$

Since the conservation routine, see Section 5.7, can force the conservation of any desired moments, we have to adjust it for the linear operator (5.113), which only conserves mass and energy. This is done by choosing a new $2\times 2N^3$ constraint matrix by only extracting the first the fifth (in 3D case) rows of the full $5\times 2N^3$ constraint matrix (5.109).

Then, the final model for electron plasma waves reads

$$\frac{\partial}{\partial t}f_e + v\cdot\nabla_x f_e + E(t,x)\cdot\nabla_v f_e = \frac{1}{\varepsilon}\big(Q_{e,e}(f_e,f_e)+Q_{e,i}(f_e)\big)\,, \tag{5.116}$$

which will be solved by the combined RKDG-Spectral method developed in this chapter.

5.9.2 The Linear Landau Damping

Perhaps, one of the most astonishing theoretical discoveries of plasma physics is the wave damping without energy dissipation by collisions. It is a result of wave-particle interactions. It occurs due to the energy exchange between particles in motion in the plasma and an electromagnetic wave. The velocity of a particle maybe greater or less than the phase velocity of the wave. Thus, there are particles gaining energy from the wave and leading to wave

damping, and also, there are particles losing energy to the wave and resulting in a increase of the wave energy. The Landau damping is studied by perturbing the Maxwellian distribution by a wave. An extremely small wave amplitude will restrict the problem in a linear regime, and thus lead to problem of "linear Landau damping"; however, if the the wave amplitude is relatively large, we are in a regime of "nonlinear Landau damping". In this section, we study the linear damping first.

The initial condition is taken as a small of perturbation of the global equilibrium $M(v) = (2\pi)^{-\frac{3}{2}} \exp(-\frac{|v|^2}{2})$

$$f_0(x, v) = (1 + A\cos(kx))M(v)\,, \tag{5.117}$$

for $(x, v) \in [0, 2\pi/k] \times \mathbb{R}^3$. Such an initial state has been chosen by many authors, see for instance [49, 53], as a benchmark problem for studying damping properties.

To study linear damping, we have to make the amplitude small enough, e.g. $A = 10^{-5}$, to restrict the problem under linear regimes. To well capture the Landau damping, the velocity domain must be large enough. It has to be larger than the phase velocity $v_\phi = \omega/k$, where ω is the frequency approximated by [53]

$$\omega^2 = 1 + 3k^2\,. \tag{5.118}$$

Here, we select $L_v = 5.75$.

The classical Landau theory tells that the square root of the electro-

static energy

$$\frac{1}{2}\int_0^{L_x} |E_h(x)|^2 dx \tag{5.119}$$

is expected to decay exponentially with frequency ω. We will plot the evolution of logarithm of square root of the electrostatic energy and compute its numerical damping rate.

According to [40, 55], the theoretical damping rate can be estimated as

$$\lambda = \lambda_l + \lambda_c \,, \tag{5.120}$$

where λ_l is the damping rate for collisionless plasma and λ_c is the "correction" for collisional case.

$$\lambda_c = -\frac{\nu}{3}\sqrt{\frac{2}{\pi}} \,, \tag{5.121}$$

with $\nu = \frac{1}{\varepsilon}$ denoting the collision frequency. And, λ_l is estimated by

$$\lambda_l = -\sqrt{\frac{\pi}{8}}\frac{1}{k^3}\exp(-\frac{1}{2k^2} - \frac{3}{2}) \,. \tag{5.122}$$

However, as pointed out in [53], (5.122) is more accurate when wave number k is large; so, for small wave numbers, more accurate estimate is available in [92]

$$\lambda_l = -\sqrt{\frac{\pi}{8}}\left(\frac{1}{k^3} - 6k\right)\exp(-\frac{1}{2k^2} - \frac{3}{2} - 3k^2 - 12k^4) \,, \tag{5.123}$$

and frequency

$$\omega = 1 + 3k^2 + 6k^4 + 12k^6 \,. \tag{5.124}$$

We will test with initials (5.117) for both collisionless and collisional cases.

We assume $\rho_i = 1$ and $\bar{v}_i = 0$, and fix wave number $k = 0.3, 0.5$. Since here amplitude A is small enough, the model can be seen in its linear regime and we can compare the numerical damping results against theoretical predictions (5.120). Our numerical results recovered the exponential damping behaviors and show that the damping is stronger if collisions are taking effects. And the damping rate increases with larger wave number k. In collisionless case, i.e $\varepsilon = \infty$, when $k = 0.5$, formula (5.122) gives an estimation -0.151 which agrees well with our numerical result in Figure 5.9.2; but for $k = 0.3$, formula (5.123) gives a more accurate estimate -0.0132 (formula (5.122) gives -0.020). In collisional case, e.g $\varepsilon = 100$, theoretically estimated damping rate for $k = 0.5$ is -0.154, while for $k = 0.3$ is -0.0167. Also, from the damping result, we know larger collision frequency impose a stronger damping.

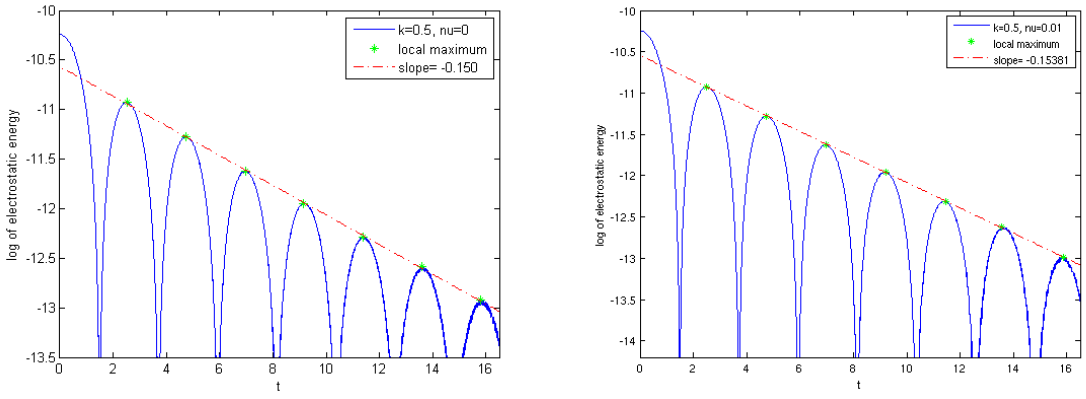


Figure 5.7: Linear Landau damping for wave number $k = 0.5$: $\varepsilon = \infty$ (left), $\varepsilon = 100$ (right)

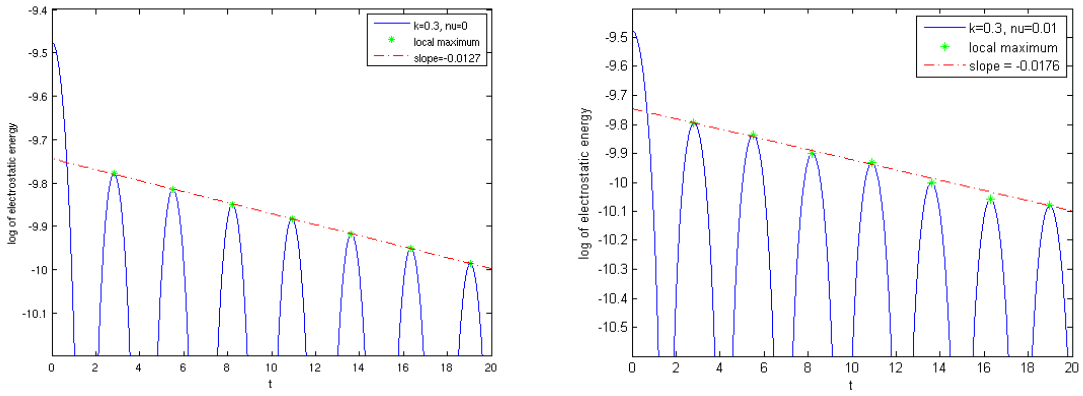


Figure 5.8: Linear Landau damping for wave number $k = 0.3$: $\varepsilon = \infty$ (left), $\varepsilon = 100$ (right)

5.9.3 The Nonlinear Landau Damping

The linear theory regarding plasmas has been relatively well developed (though still many problems remain unsolved). However, the nonlinear phenomena of plasma is much less understood. From last section, we know as long as the wave amplitude A is small enough, a well-developed linear theory is valid. Nevertheless, when the wave amplitude gets larger, many waves in experiments can be no longer described by the linear theory. Some of them are even not trackable through analysis.

One example would be "electron trapping" phenomena. It occurs with the nonlinear Landau damping of the waves. Since the particles travel relative to the wave, a large electric potential together with collisions will trap the electrons in a potential well of the wave. The trapped electrons will be bounced back and forth in the well, causing fluctuating amplitudes of the wave. Thus,

one cannot always expect an exponential damping as in the linear case.

In order to capture the electron trapping, we extract the contours of the following marginal distribution

$$F(t,x,v_x) = \int_{\mathbb{R}^2} f(t,x,v_x,v_y,v_z)dv_y dv_z \,. \tag{5.125}$$

In phase space, $F(t,x,v_x)$ will form peaks whenever there is a potential through. Trapped electrons will move in closed orbits in phase space, since the contours $F(t,x,v_x)$ are also the electron trajectories. Please refer to [40] for more explanations.

In this section, we will study the nonlinear damping with the following initial wave

$$f_0(x,v) = (1 + A\cos(kx))M(v)\,, (x,v) \in [0, 2\pi/k] \times \mathbb{R}^3\,, \tag{5.126}$$

for a relatively large amplitude A such that it is no longer in the linear regime. Here, we choose the Maxwellian

$$M(v) = (2\pi T)^{-\frac{3}{2}} \exp(-\frac{|v|^2}{2T})\,.$$

Figure 5.9 shows the nonlinear damping results for $A = 0.2$, $T = 0.5$, $k = 0.5$ and a large enough velocity domain $L_v = 5$, with different collision frequencies $\nu = 0, 0.05, 0.1$. We choose $N_x = 36$ mesh elements on x-direction, $N_v = 36$ mesh elements on each direction of velocity v for the RKDG VP problem, and $N = 24$ Fourier nodes for the spectral method. We can see

the electric energy, in all cases, decreases exponentially at first. In the collisionless regime, the electric energy then starts to oscillate around a constant, which agrees well with the known property. With collision, the oscillations are weaken. In particular, with the presence of stronger collisions, the amplitude of electric energy will start to form an exponential decay again. Although a relatively large amplitude A is imposed and moderate resolution of mesh is applied, we still obtain good preservation of the total energy, which is even better conserved that [53]. See Figure 5.10 on variations of total energy during the whole process of simulation.

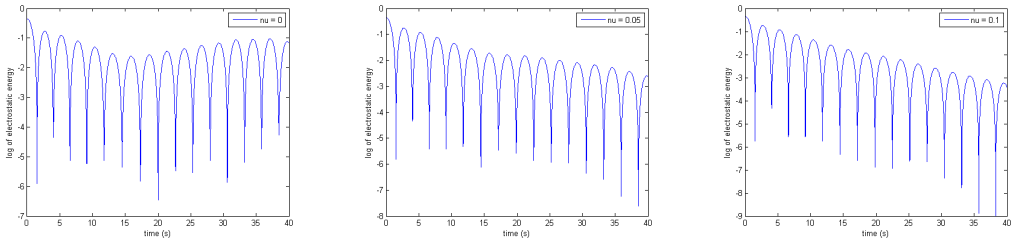


Figure 5.9: Nonlinear damping with $A = 0.2$ for $\nu = 0$ (left), $\nu = 0.05$ (middle) and $\nu = 0.1$ (right)

Figure 5.11 shows the electron trapping effects for much larger amplitude $A = 0.5$, $T = 0.25$, $k = 2\pi/4$ and $L_v = 4$. We choose $N_x = 48$, $N_v = 32$ and $N = 24$. Collision effects range from weak to strong , that is, $\nu = 0, 0.005, 0.2$. one can observe that, without collisions, the much more electrons are trapped in the potential through. While with collisions get stronger, less and less electrons are trapped and a stationary state is reached early.

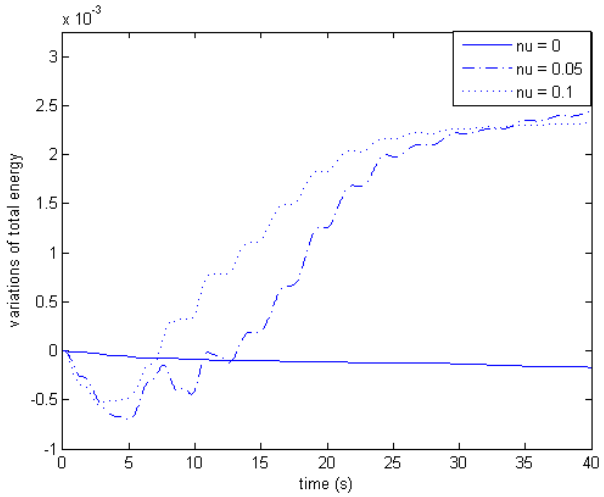


Figure 5.10: Variations of total energy during nonlinear damping simulation with $A = 0.2$ for $\nu = 0, 0.05, 0.1$

5.9.4 Two Stream Flow

This is of primary importance for studying nonlinear effects of plasmas in future. In this section, we consider a plasma with fixed ion background and only consider the electron-electron collisions. We will study how well the above time-splitting and conservative linking process work, by initializing with a non-isotropic two-stream flow.

$$f_0(x, v) = (1 + A\cos(kx))f_{TS}(v)\,, \tag{5.127}$$

where A is the amplitude of the perturbation and k the wave number, and

$$f_{TS}(v) = \frac{1}{2(2\pi\sigma^2)^{3/2}}\left[\exp\left(-\frac{|v - 2\sigma e|^2}{2\sigma^2}\right) + \exp\left(-\frac{|v + 2\sigma e|^2}{2\sigma^2}\right)\right], \tag{5.128}$$

with parameter $\sigma = \pi/10$ and $e = (1, 0, 0)$. We would like it to be far from the linear regime, so a relatively large perturbation is considered $A = 0.5$,

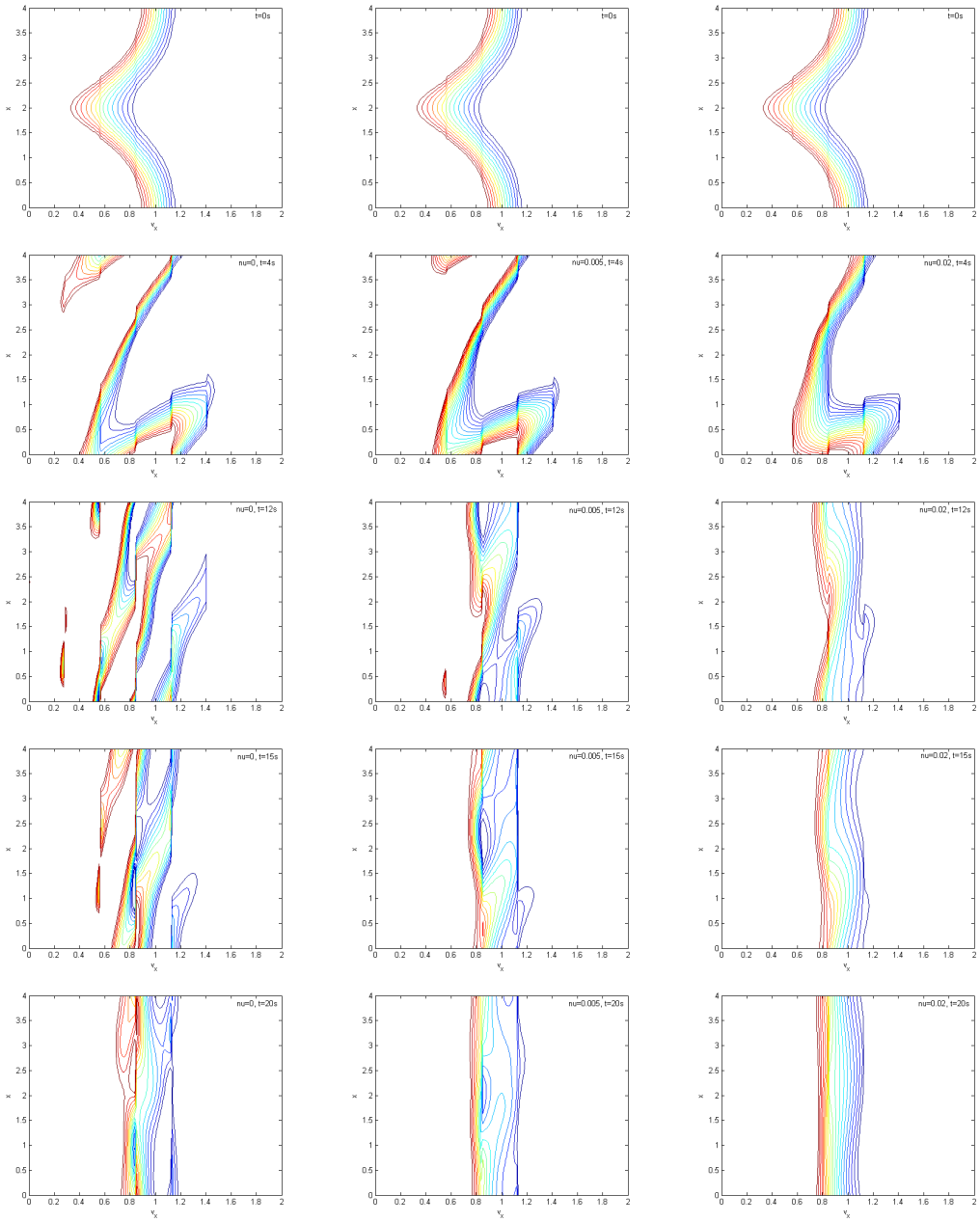


Figure 5.11: Evolution of contours of $F(t, x, v_x)$ for $\nu = 0$ (left), $\nu = 0.005$ (middle) and $\nu = 0.02$ (right)

$k = 2\pi/L_x$ with $L_x = 4$. A large enough velocity domain is selected $L_v = 4.5$. We choose $N_x = 48$ mesh elements on x-direction, $N_v = 32$ mesh elements on each direction of velocity v for the RKDG VP problem, and $N = 24$ Fourier nodes for the spectral method.

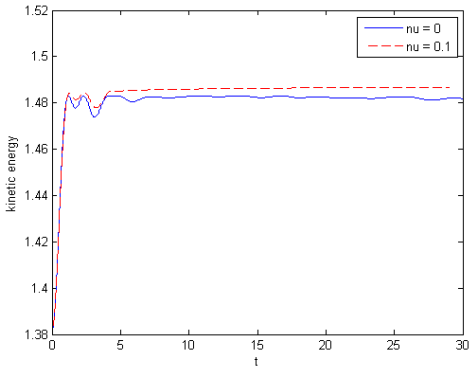


Figure 5.12: The evolution of kinetic energy for the two-stream flow

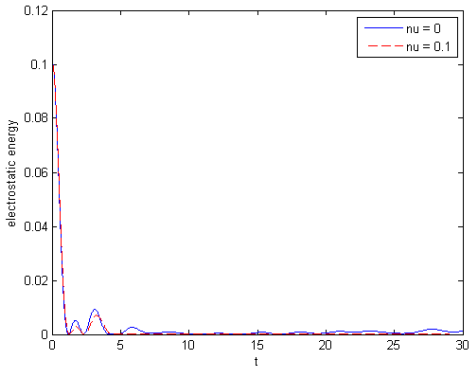


Figure 5.13: The evolution of electrostatic energy for the two-stream flow

Relatively stronger collision effects are considered by taking a relatively large collision frequency $\nu = 0.1$ (relatively small Knudesen number $\varepsilon = 10$). Results are also compared to collisionless case, i.e. $\nu = 0$. In Figure (5.12), (5.13) and (5.14), the total energy initially comes from both the kinetic and electrostatic energy, but with time forwarding, the electrostatic energy decays with oscillations down to zero and the total energy at the end all comes from pure kinetic motions, which means the system has reached at its global equilibrium. During the whole process, the total energy is well preserved only with negligible variations. In addition, from Figure (5.12) and Figure (5.13) one can observe that, since the Landau operator is essentially a diffusive operator,

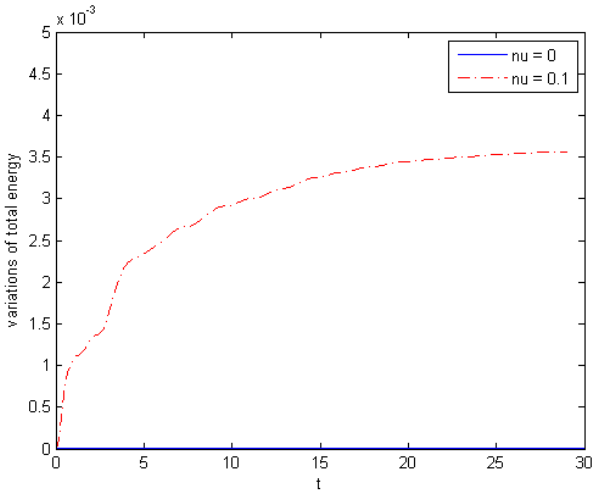


Figure 5.14: The variations of total energy for the two-stream flow

the oscillations generated by coupling with the Poisson equations damps with collisions, and thus the state reaches stationary in a much earlier stage.

Chapter 6

Summary and Future Work

The Boltzmann-type equation, as the keystone of kinetic theories, bridges the gap between microscopic world and the macroscopic world. It studies the behaviors of atomic structures in a mesoscopic level and is supposed to provide information that cannot be contained in one another and reveals the relationships between different scales of models.

The study of the collision operators for Botlzmann-type equations is the key for study of properties of kinetic problems. And also, it's always the most challenging part both analytically and numerically. My studies include, but not limited to, the numerical treatments for the nonlinear collisional operators and the spectral gaps for the linearized operators.

For the full nonlinear Boltzmann equations, we evaluated the collisional operators under a Discontinuous Galerkin method framework. Due to its high dimensionality and the complex collision kernels, to our best knowledge, there is barely any work on this topic. During the evaluation, we take the conservation laws into consideration and design a conservation routine to enforce the conservation of desired moments. This will be the base for future development of DG finite element methods for kinetic equations and applications to inho-

mogeneous transport equations for problems of non-smooth density functions as well as irregular spatial domains. The high dimensionality, and thus high computation and storage complexity, is well reduced, by digging the "shifting symmetry" properties of the collisional integrals, the sparsity of "collision matrix" as well as the parallelizability.

The study on relaxation to equilibrium for the solutions of kinetic equations has been the keystone of kinetic theories ever since the work of Boltzmann. There are extensive researches, results and conjectures on it. When close to equilibrium, the relaxation properties are dominated by its linearized counterpart. Liouville's theorems tell us that the spectral gaps, if exists, will control the exponential decay of the solutions. The existence and values of such gaps are of great significance for studying the hydrodynamic limits as well as the validity of Botlzmann models. There are some work on study the existence and estimates (very rarely constructive) of the spectral gaps, but none of them gave a numerical approximation. My work computed the spectral gap numerically, to verify the conjectures on the existence of spectral gaps, especially for non-integrable angular cross-section, as well as giving numerical approximations.

As a grazing limit of the Boltzmann equation, the Fokker-Planck-Landau equation is a very important mathematical model for collisional plasmas. Similar to the Boltzmann equations, the FPL collision operator also remains the most challenging and important. We studied the inhomogeneous FPL equations, coupled with Poisson equations which governing the self-consistent elec-

tric field. The complicated inhomogeneous problem is splitted into two sub-problems, by time-splitting scheme. We applied two different methods for treating the pure transport Vlasov problem and the pure collisional homogeneous FPL equation. The former is solved by RKDG method, which has achieved its success in many other kinetic problems; while the latter is treated using conservative spectral method. The conservative spectral method was well developed for solving Boltzmann equations and we extended the method to the FPL problems and apply to study the multi-component plasams. The temperature relaxation of the multi-component plasma is studied both analytically and numerically. To link the two differnent methods, or computing grids, we developed a new conservation routine which can guaranttee no loss of moments when projecting the Fourier solution onto DG meshes. All desired moments are preserved only with error of DG approximations. The whole scheme has been applied to study the well-known Landau damping problems, whose results agree well with theoretical estimates, and to two stream flows.

All the projects are implemented with parallelization, hybrid MPI [65] and OpenMP [13].

In the future, we plan to dig more on speeding-up of the comptuations for the nonlinear Boltzmann collisional operators as well as the time evoltuion for Boltzmann equations. We hope to apply the conservative DG solver that we developed to 3D inhomogeneous Boltzmann equations. The computations on spectral gaps can achieve more accuracy by employing finer velocity meshes. We also hope to solve the linear Landau equations to study

the "pseudo-exponential" decay, conjectured in Section 5.3. For the collisional plasma problem, we would like to apply non-periodic boundary conditons on the Poisson equations and thus to study more nonlinear effects, for instance plasma sheath problems, which is of primary importance for Aerospace Engineering.

Appendix

0.1 Tools For Asymptotic Behavior Study of DG Conservative Solver

The classical Sobolev spaces are defined as

$$W^{\alpha,p}(\Omega) = \{f \in L^p(\Omega) : D^\beta f \in L^p(\Omega) \quad \text{for all multi-indices } \beta \text{ such that } |\beta| \leq \alpha\}$$
$$H^\alpha(\Omega) = W^{\alpha,2}(\Omega) \tag{1}$$

and they are equipped with the norms

$$\|f\|_{W^{\alpha,p}(\Omega)} = \textstyle\sum_{|\beta|\leq\alpha} \|D^\beta f\|_{L^p(\Omega)}, \qquad \text{if } p < \infty$$
$$\|f\|_{W^{\alpha,\infty}(\Omega)} = \max_{|\beta|\leq\alpha} \|D^\beta f\|_{L^\infty(\Omega)}, \qquad \text{if } p = \infty \tag{2}$$

The *weighted Sobolev spaces* H^m_α are H^m spaces weighted with $\langle v\rangle^\alpha = \left(1 + |v|^2\right)^{\gamma\alpha/2}$. That is,

$$\|f\|_{H^m_\alpha(\Omega)} = \sum_{|\alpha|\leq m} \|D^\alpha f \langle v\rangle^\alpha\|_{L^2(\Omega)} \tag{3}$$

Here, please note that specially for the asymptotic error analysis for DG solver, we include the intermolecular potential parameter γ in here. The *broken Sobolev spaces* for the partition of Ω are defined as

$$W^{\alpha,p}(\mathcal{T}_h) = \{f \in L^p(\Omega) : f|_E \in W^{\alpha,p}(E) \text{ for all } E \in \mathcal{T}_h\}$$
$$H^\alpha(\mathcal{T}_h) = W^{\alpha,2}(\mathcal{T}_h) \tag{4}$$

and the corresponding norms

$$\|f\|_{W^{\alpha,p}(\mathcal{T}_h)} = \textstyle\sum_{E\in\mathcal{T}_h} \|f\|_{W^{\alpha,p}(E)}, \qquad \text{if } p < \infty \tag{5}$$
$$\|f\|_{W^{\alpha,\infty}(\mathcal{T}_h)} = \max_{E\in\mathcal{T}_h} \|f\|_{W^{\alpha,\infty}(E)}, \qquad \text{if } p = \infty \tag{6}$$

Then, we define the standard d-dimensional L^2 projection $P_h : f \mapsto P_h f$ by

$$\int_E P_h f(v)\phi(v)dv = \int_E f(v)\phi(v)dv, \quad \forall \phi \in \mathbf{P}^l|_E \tag{7}$$

By Poincare's inequality and Sobolev embedding theorems, we can prove the following approximation theory

$$\|f - P_h f\|_{L^2(\mathcal{T}_h)} \lesssim h^{\alpha+1}\|f\|_{H^{\alpha+1}(\Omega)}, \quad \forall f \in H^{\alpha+1}(\Omega)$$

$$\|f - P_h f\|_{L^\infty(\mathcal{T}_h)} \lesssim h^{\alpha+1}\|f\|_{W^{\alpha+1,\infty}(\Omega)}, \quad \forall f \in W^{\alpha+1,\infty}(\Omega)$$

$$\|P_h f\|_{L^p(\mathcal{T}_h)} \lesssim \|f\|_{L^p(\Omega)}, \quad \forall f \in L^p(\Omega), \quad 1 \le p \le \infty$$

where $h = \max_{E\in\mathcal{T}_h} \mathrm{diam}(E)$.

0.1.1 Extension Operators

For fixed $\alpha_0 \ge 0$, there exists an extension operator $\mathrm{E} : L^2(\Omega_v) \to L^2(\mathbb{R}^d)$ such that for any $\alpha \le \alpha_0$ one has additionally $\mathrm{E} : H^\alpha(\Omega_v) \to H^\alpha(\mathbb{R}^d)$. The construction of such operator is well known and has the properties [113]:

1. Linear and bounded with

$$\|\mathrm{E}f\|_{H^\alpha(\mathbb{R}^d)} \le C_\alpha \|f\|_{H^\alpha(\Omega_v)} \quad \text{for } \alpha \le \alpha_0.$$

2. $\mathrm{E}f = f$ a.e. in Ω_v.

3. Outside Ω_v the extension is constructed using a reflexion of f near the boundary $\partial\Omega_v$. Thus, for any $\delta \ge 1$ we can choose an extension with support in $\delta\Omega_v$, the dilation of Ω_v by δ, and

$$\|\mathrm{E}f\|_{L^p(\delta\Omega_v\setminus\Omega_v)} \le C_0 \|f\|_{L^p(\Omega_v\setminus\delta^{-1}\Omega_v)} \quad \text{for } 1 \le p \le 2,$$

where the constant C_0 is independent of the support of the extension.

4. In particular, properties 2. and 3. imply that for any $\delta \geq 1$ there is an extension such that

$$\|\mathrm{E}f\|_{L^p_k(\mathbb{R}^d)} \leq 2C_0\delta^{2k}\|f\|_{L^p_k(\Omega_v)} \quad \text{for } 1 \leq p \leq 2,\ k \geq 0.$$

0.1.2 Lemmas For Asymptotic Behavior Study

Following the arguments in [5], we have

Lemma (Elastic Lagrange Estimate)**.** *The problem* (4.77) *has a unique minimizer given by*

$$X^* = Q_u(f)(v) - \frac{1}{2}\Big(\gamma_1 + \sum_{j=1}^{d}\gamma_{j+1}v_j + \gamma_{d+2}|v|^2\Big),$$

where γ_j, for $1 \leq j \leq d+2$, are Lagrange multipliers associated with the elastic optimization problem. Furthermore, they are given by

$$\begin{aligned}
\gamma_1 &= O_d\rho_u + O_{d+2}e_u\,,\\
\gamma_{j+1} &= O_{d+2}\mu_u^j\,, \quad j = 1, 2, \cdots, d,\\
\gamma_{d+2} &= O_{d+2}\rho_u + O_{d+4}e_u\,.
\end{aligned}$$

The estimate constants $O_r := O(L^{-r})$ only depends inversely on $|\Omega_v|$. The parameters $\rho_u, , \mu_u^j, e_u$ are density, momentum and kinetic energy associated with the unconserved collision operator $Q_{uc}(f_h)$.

In particular, for dimension $d = 3$, the minimized objective function is

given by

$$\mathcal{A}^e(X^*) = \|Q_u(f) - X^*\|^2_{L^2(\Omega_v)} = 2\gamma_1^2 L^3 + \frac{2}{3}(\gamma_2^2+\gamma_3^2+\gamma_4^2)L^5 + 4\gamma_1\gamma_5 L^5 + \frac{38}{15}\gamma_5^2 L^7 \,. \tag{8}$$

The minimizer is the expected conservation correction, i.e. $Q_c(f_h) = X^*$. So the elastic case (conservation up to kinetic energy), the conserved projection operator $Q_c(f_h)$ is a perturbation of $Q_{uc}(f_h)$ by a second order polynomial.

In the sequel we denote the moments of a function f by

$$m_k(f) := \int_{\mathbb{R}^d} \big|f(v)\big|\, |v|^{\gamma k}\, dv.$$

and

$$Z_k(f) := \sum_{j=0}^{k-1} \left(\begin{array}{c} k \\ j \end{array} \right) m_{j+1} m_{k-j}. \tag{9}$$

Besides the above lemma, we list several other results necessary for the final convergence and error estimate, most of which are generalized from the work [5].

Lemma (Conservation Correction Estimate)**.** *Fix $f \in L^2(\Omega_v)$, then the accuracy of the conservation minimization problem is proportional to the spectral accuracy. That is, for any $k, k' \geq 0$ and $\delta > 1$, there exists some extension operator E, such that*

$$\begin{aligned} \| \left(Q_c(f) - Q_u(f)\right) |v|^k \|_{L^2(\mathcal{T}_h)} \leq\ & \frac{C}{\sqrt{(k+d)}} L^{\gamma k} \|Q(Ef, Ef) - Q_u(f)\|_{L^2(\mathcal{T}_h)} \\ & + \frac{\delta^{2\gamma k'}}{\sqrt{(k+d)}} O_{d/2+\gamma(k'-k)}\big(m_{k'+1}(f) m_0(f) + Z_{k'}(f)\big)\,, \end{aligned}$$

where C is a universal constant.

To prove our final convergence estimate, we need another theorem in the L^2-theory of the collision operator, which is the *Sobolev Bound Estimate*.

Lemma (Sobolev Bound Estimate)**.** *Let $\mu > \frac{d}{2} + \gamma$. For $f, g \in H^{\alpha}_{k+\mu}$, the collision operator satisfies*

$$\|Q(f,g)\|^2_{H^{\alpha}_k} \leq C \sum_{j \leq \alpha} \binom{\alpha}{j} (\|f\|^2_{H^{\alpha-j}_{k+\gamma}} \|g\|^2_{H^{j}_{k+\mu}} + \|f\|^2_{H^{\alpha-j}_{k+\mu}} \|g\|^2_{H^{j}_{k+\gamma}}), \tag{10}$$

where the dependence of the constant is $C := C(d, \beta, \alpha, \|b\|_1)$.

And also, we need the following H^{α}_k-norm propagation properties of the solutions.

Lemma (H^{α}_k-norm Propagation)**.** *Assume $f_{h,0} \in H^{\alpha}_{k+1+\alpha}(\Omega_v)$, then there exists an extension operator E_α, for any time T, we can choose a lateral size $L_0(f_0, k, \alpha)$ for the truncated domain Ω_v, such that for any $L \geq L_0$ there exists a small mesh size $h_0 = \max_{E_v \in \mathcal{T}_h} diam(E_v)$,*

$$\sup_{t \in [0,T]} \|f_h\|_{H^{\alpha}_k(\Omega_v)} \leq \max\left\{ \|f_{h,0}\|_{H^{\alpha}_{k+1+\lambda}(\Omega_v)}, C_k(m_{k'}(g_0)) \right\}, \quad h \leq h_0$$

where $k' \geq k$ is a finite number of moments. Additionally, C_k is independent of the parameters L.

0.2 Calculations of $\widehat{\mathbf{S}}$

(1). $\widehat{\mathbf{S}^1_{11}}(\omega)$.

This is done immediately.

$$\begin{aligned}\widehat{\mathbf{S}^1}(\omega) &= (2\pi)^{-3/2}\int_{B_R(0)}\frac{1}{|u|}e^{-i\omega\cdot u}du \\ &= \sqrt{\frac{2}{\pi}}\frac{1}{|\omega|^2}[1-\cos(R|\omega|)]\end{aligned} \tag{11}$$

And, if $|\omega|=0$, $\widehat{\mathbf{S}^1}(\omega)=\sqrt{\frac{1}{2\pi}}R^2$.

(2). $\widehat{\mathbf{S}^2_{33}}(\omega)$.

$$\begin{aligned}\widehat{\mathbf{S}^2_{33}}(\omega) &= (2\pi)^{-3/2}\int_{B_R(0)}\frac{u_3^2}{|u|^3}e^{-i\omega\cdot u}du \\ &= (2\pi)^{-3/2}\int_0^R r\int_{S^2}\sigma_3^2 e^{-ir\omega\cdot\sigma}d\sigma dr\end{aligned} \tag{12}$$

Suppose $\omega=|\omega|(\sin(\theta)\cos(\phi),\sin(\theta)\sin(\phi),\cos(\theta))^T$, and consider the orthogonal rotation matrices

$$R_y(\theta)=\begin{pmatrix}\cos(\theta) & 0 & \sin(\theta)\\ 0 & 1 & 0\\ -\sin(\theta) & 0 & \cos(\theta)\end{pmatrix}\quad\text{and}\quad R_z(\phi)=\begin{pmatrix}\cos(\phi) & \sin(\phi) & 0\\ -\sin(\phi) & \cos(\phi) & 0\\ 0 & 0 & 1\end{pmatrix} \tag{13}$$

which rotates the vectors about $y-$ and $z-$axis, respectively.

Then,

$$R_y^T(\theta)R_z(\phi)\omega=(0,0,|\omega|)^T:=\tilde{\omega}$$

Denote $A=R_y^T(\theta)R_z(\phi)$, then A is also an orthogonal rotation matrix

$$A=\frac{1}{|\omega|}\begin{pmatrix}\frac{\omega_1\omega_3}{\sqrt{\omega_1^2+\omega_2^2}} & \frac{\omega_2\omega_3}{\sqrt{\omega_1^2+\omega_2^2}} & -\sqrt{\omega_1^2+\omega_2^2}\\ -\frac{\omega_2|\omega|}{\sqrt{\omega_1^2+\omega_2^2}} & \frac{\omega_1|\omega|}{\sqrt{\omega_1^2+\omega_2^2}} & 0\\ \omega_1 & \omega_2 & \omega 3\end{pmatrix} \tag{14}$$

where we assume $\omega_1^2 + \omega_2^2 \neq 0$; otherwise, matrix A is reduced to the identity matrix.

Then

$$\begin{aligned}
&\int_{S^2} \sigma_3^2 e^{-ir\omega\cdot\sigma} d\sigma \\
&= \frac{1}{|\omega|^2}\big(4\pi(\omega_1^2+\omega_2^2)\frac{\sin(r|\omega|) - r|\omega|\cos(r|\omega|)}{(r|\omega|)^3} \\
&\quad + 4\pi\omega_3^2\frac{((r|\omega|)^2-2)\sin(r|\omega|) + 2r|\omega|\cos(r|\omega|)}{(r|\omega|)^3}\big)
\end{aligned}$$

So, plugging back into $\widehat{\mathbf{S}_{33}^2}$ (12) gives

$$\begin{aligned}
\widehat{\mathbf{S}_{33}^2}(\omega) &= (2\pi)^{-3/2}\int_0^R r\int_{S^2}\sigma_3^2 e^{-ir\omega\cdot\sigma}d\sigma dr \\
&= \sqrt{\frac{2}{\pi}}\frac{1}{|\omega|^4}\big((\omega_1^2+\omega_2^2)\frac{R|\omega| - \sin(R|\omega|)}{R|\omega|} \\
&\quad - \omega_3^2\frac{R|\omega| + R|\omega|\cos(R|\omega|) - 2\sin(R|\omega|)}{R|\omega|}\big)
\end{aligned} \tag{15}$$

And, if $|\omega| = 0$, $\widehat{\mathbf{S}_{33}^2}(\omega) = \sqrt{\frac{1}{2\pi}}\frac{R^2}{3}$.

(3). $\widehat{\mathbf{S}_{13}^2}(\omega)$.

$$\begin{aligned}
\widehat{\mathbf{S}_{13}^2}(\omega) &= (2\pi)^{-3/2}\int_{B_R(0)}\frac{u_1u_3}{|u|^3}e^{-i\omega\cdot u}du \\
&= (2\pi)^{-3/2}\int_0^R r\int_{S^2}\sigma_1\sigma_3 e^{-ir\omega\cdot\sigma}d\sigma dr
\end{aligned} \tag{16}$$

Following the same change of variables as above,

$$\begin{aligned}
&\int_{S^2}\sigma_1\sigma_3 e^{-ir\omega\cdot\sigma}d\sigma dr = \int_{S^2}(A^T\sigma)_1(A^T\sigma)_3 e^{-ir\tilde{\omega}\cdot\sigma}d\sigma \\
&= 4\pi\frac{\omega_1\omega_3}{|\omega|^2}\frac{((r|\omega|)^2-3)\sin(r|\omega|) + 3r|\omega|\cos(r|\omega|)}{(r|\omega|)^3}
\end{aligned}$$

So,

$$\begin{aligned}\widehat{\mathbf{S}_{13}^2}(\omega) &= (2\pi)^{-3/2}\int_0^R r\int_{S^2}\sigma_1\sigma_3 e^{-ir\omega\cdot\sigma}d\sigma dr \\ &= -\sqrt{\frac{2}{\pi}}\frac{\omega_1\omega_3}{|\omega|^4}\frac{2R|\omega|+R|\omega|\cos(R|\omega|)-3\sin(R|\omega|)}{R|\omega|}\end{aligned} \tag{17}$$

And, if $|\omega| = 0$, $\widehat{\mathbf{S}_{13}^2}(\omega) = 0$.

Bibliography

[1] The sun's magnetic field. `http://www.angelfire.com/rnb/pp0/sun5.html`.

[2] A. Alekseenko and E. Josyula. Deterministic solution of the boltzmann equation using discontinuous galerkin discretizations in velocity space. *Journal of Computational Physics*, submitted.

[3] R. Alexandre, L. Desvillettes, C. Villani, and B. Wennberg. Entropy dissipation and long-range interactions. *Arch. Rat. Mech. Anal.*, 152:327–355, 2000.

[4] R. Alonso, J. Canizo, I. Gamba, and C. Mouhot. A new approach to the creation and propagation of exponential moments in the boltzmann equation. *Comm. Part. Diff. Equat.*, 38(1):155–169, 2013.

[5] R. Alonso, I.M. Gamba, and S.H. Tharkabhushaman. Convergence of the lagrangian based conservative spectral method for space-homogeneous non-linear boltzmann equation for hard potentials. Submitted 2014.

[6] R.J. Alonso and I.M. Gamba. $l^1 - l^\infty$-maxwellian bounds for the derivatives of the solution of the homogeneous boltzmann equation. *Journal de Mathematiques Pures et Appliquees*, 89(6):575–595, 2008.

[7] V. V. Aristov. *Direct methods for solving the Boltzmann equation and study of nonequilibrium flows*. Kluwer Academic Publishers, Dordrecht, 2001.

[8] L. Arkeryd. Stability in l^1 for the spatially homogeneous boltzmann equation. *Arch. Rational Mech. Anal.*, 103:151–167, 1988.

[9] B. Ayuso, J.A. Carrillo, and C.-W. Shu. Discontinuous galerkin methods for the one-dimensional vlasov-poisson system. *Kinetic and Related Models*, 4:955–989, 2011.

[10] C. Baranger and C. Mouhot. Explicit spectral gap estimates for the linearized boltzmann and landau operators with hard potentials. *Rev. Mat. Iberoam*, 21:819–841, 2005.

[11] G.A. Bird. *Molecular Gas Dynamics*. Clarendon Press, Oxford, 1994.

[12] L. S. Blackford, J. Choi, A. Cleary, E. D'Azevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1997.

[13] OpenMP Architecture Review Board. OpenMP application program interface version 3.0, May 2008. `http://www.openmp.org/mp-documents/spec30.pdf`.

[14] A. V. Bobylev. Exact solutions of the nonlinear Boltzmann equation and the theory of relaxation of a Maxwellian gas. *Translated from Teoreticheskaya i Mathematicheskaya Fizika*, 60:280–310, 1984.

[15] A. V. Bobylev. The theory of the nonlinear spatially uniform boltzmann equation for maxwell molecule. *Sov.Sci.Rev.C.Math.Phys*, 7:111–233, 1988.

[16] A. V. Bobylev, J. A. Carrillo, and I. M. Gamba. On some properties of kinetic and hydrodynamic equations for inelastic interactions. *Journal of Statistical Physics.*, 98:743–773, 2000.

[17] A. V. Bobylev and C. Cercignani. Discrete velocity models without non-physical invariants. *Journal of Statistical Physics.*, 97:677–686, 1999.

[18] A. V. Bobylev and C. Cercignani. On the rate of entropy production for the bolzmann equation. *Journal of Statistical Physics.*, 94:603–618, 1999.

[19] A. V. Bobylev, C. Cercignani, and I. M. Gamba. Generalized kinetic maxwell type models of granular gases. In P. Giovine G. Capriz and P. M. Mariano, editors, *Mathematical models of granular matter*, number 1937 in Lecture Notes in Mathematics. Springer, 2008.

[20] A. V. Bobylev, C. Cercignani, and I. M. Gamba. On the self-similar asymptotics for generalized non-linear kinetic Maxwell models. arXiv:math-ph/0608035 2006.

[21] A. V. Bobylev and I. M. Gamba. Boltzmann equations for mixtures of maxwell gases: Exact solutions and power like tails. *Journal of Statistical Physics.*, 124:497–516, 2006.

[22] A. V. Bobylev, I. M. Gamba, and V. Panferov. Moment inequalities and high-energy tails for Boltzmann equations with inelastic interactions. *Journal of Statistical Physics.*, 116:1651–1682, 2004.

[23] A. V. Bobylev, S. A. Karpov, and I. F. Potapenko. Dsmc methods for multicomponent plasmas. In *28th International Symposium on Rarefied Gas Dynamics 2012*. American Institute of Physics, 9–13 July 2012.

[24] A. V. Bobylev, I.F. Potapenko, and P.H. Sakanaka. Relaxation of two-temperature plasma. *Phyical Review E*, 56(2):2081–2093, 1997.

[25] A. V. Bobylev and S. Rjasanow. Difference scheme for the Boltzmann equation based on the Fast Fourier Transform. *European journal of mechanics. B, Fluids*, 16(22):293–306, 1997.

[26] A. V. Bobylev and S. Rjasanow. Fast deterministic method of solving the Boltzmann equation for hard spheres. *European journal of mechanics. B, Fluids*, 18(5):869–887, 1999.

[27] A. V. Bobylev and S. Rjasanow. Numerical solution of the Boltzmann equation using fully conservative difference scheme based on the Fast Fourier Transform. *Transport Theory Statist. Phys.*, 29:289–310, 2000.

[28] J. E. Broadwell. Study of rarefied shear flow by the discrete velocity method. *J. Fluid Mech.*, 19:401–414, 1964.

[29] C. Buet and S. Cordier. Conservative and entropy decaying numerical scheme for the isotropic fokker-planck-landau equation. *Journal of Computational Physics*, 145:228–245, 1998.

[30] H. Cabannes. Global solution of the initial value problem for the discrete Boltzmann equation. *Comm. Math. Phys*, 74:71–95, 1980.

[31] R. E. Caflisch. The boltzmann equation with a soft potential. i. linear, spatially homogeneousd. *J. Fluid Mech.*, 19:401–414, 1964.

[32] T. Carleman. Sur la théorie de léquation intégrodifférentielle de boltzmann. *Acta Math.*, 60:91–146, 1933.

[33] T. Carleman. *Problèmes Mathématiques dans la Théorie Cinétique des Gaz*. Almqvist & Wiksell, 1957.

[34] C. Cercignani. *Mathematical Methods in Kinetic Theory*. Plenum Pres, New York, 1969.

[35] C. Cercignani. H-theorem and trend to equilibrium in the kinetic theory of gases. *Arch. Mech.*, 34:231–241, 1982.

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

[37] C. Cercignani and H.Cornille. Shock waves for a discrete velocity gas mixture. *Journal of Statistical Physics.*, 99:115–140, 2000.

[38] C. Cercignani, R. Illner, and M. Pulvirenti. *The mathematical theory of dilute gases*. Springer-Verlag, New York, 1994.

[39] C. S. Wang Chang and G. E. Uhlenbeck. *On the propagation of Sound in Monoatomic Gases*. Univ. of Michigan Press, Ann Arbor, MI, 1956.

[40] F.F. Chen. *Introduction to Plasma Physics and controlled fusion, 2nd ed.* Plenum Press, New York and London.

[41] Y. Cheng and I.M. Gamba. Numerical study of vlasov-poisson equations for infinite homogeneous stellar systems. *Communications in Nonlinear Science and Numerical Simulation*, 17(5):2052–2061, 2012.

[42] Y. Cheng and I.M. Gamba. Numerical study of vlasov-poisson equations for infinite homogeneous stellar systems. *Communications in Nonlinear Science and Numerical Simulation*, 17(5):2052–2061, 2012.

[43] Y. Cheng, I.M. Gamba, and P. Morrison F. Li. Discontinuous galerkin methods for the vlasov-maxwell equations. *Journal of Statistical Physics.*, arXiv:1302.2136v2 [math.NA], 2013.

[44] Y. Cheng, I.M. Gamba, F. Li, and P. Morrison. Discontinuous galerkin methods for vlasov-maxwell equations. *Journal of Computational Physics*, arXiv:1302.2136v2 [math.NA], 2014, to appear.

[45] Y. Cheng, I.M. Gamba, A. Majorana, and C-W Shu. A discontinuous galerkin solver for boltzmann poisson systems in nano devices. *Comput. Methods Appl. Mech. Engrg*, 198(37–40):3130–3150, 2009.

[46] Y. Cheng, I.M. Gamba, A. Majorana, and C-W Shu. A brief survey of the discontinuous galerkin method for the boltzmann-poisson equations. *SEMA J*, 54:47–64, 2011.

[47] Y. Cheng, I.M. Gamba, A. Majorana, and C-W Shu. Discontinuous galerkin methods for the boltzmann-poisson systems in semiconductor device simulations. *AIP Conference Proceedings*, 1333:892–895, 2011.

[48] Y. Cheng, I.M. Gamba, A. Majorana, and C-W Shu. A discontinuous galerkin solver for full-band boltzmann-poisson models. *13th International Workshop on Computational Electronics Proceedings*, pages 211–214, 2011.

[49] Y. Cheng, I.M. Gamba, and P. Morrison. Study of conservation and recurrence of runge-kutta discontinuous galerkin schemes for vlasov-poisson system. *Journal of Scientific Computing*, 56(2):319–349, 2013.

[50] Y. Cheng, I.M. Gamba, and J. Proft. Positivity-preserving discontinuous galerkin schemes for linear vlasov-boltzmann transport equations. *Mathematics of Computation*, 81:153–190, 2012.

[51] B. Cockburn and C.-W. Shu. Runge-kutta discontinuous galerkin methods for convection-dominated problems. *Journal of Scientific Comput-*

ing, 16:173–261, 2001.

[52] NASA/ISS Expedition 28 Crew. http://earthobservatory.nasa.gov/IOTD/view.php?id=514

[53] N. Crouseilles and F. Filbet. Numerical approximation of collisional plasmas by high order methods. *Journal of Computational Physics*, 201:546–572, 2004.

[54] A. DeCoster, B. Perthame, and P. Marcowich. Modeling of collisions. In *Series in Applied Mathematics*. 1998.

[55] J.P. Delcroix and A. Bers. *Physique des plasmas*. Savoirs Actuels, CNRS Editions, 1994.

[56] L. Desvillettes, C. Mouhot, and C. Villani. Celebrating cercignani's conjecture for the boltzmann equation. arXiv:1009.4006 [math.AP], 2010.

[57] M.H. Ernst. *Exact solutions of the nonlinear Boltzmann equation and related kinetic models, Nonequilibrium Phenomena, I, Stud. Statist. Mech. 10*. North-Holland, Amsterdam, 1983.

[58] M. Galassi et al. Gnu scientific library reference manual (3rd ed.). `http://www.gnu.org/software/gsl/`.

[59] F. Filbet, C. Mouhot, and L. Pareschi. Solving the Boltzmann equation in nlogn. *SIAM J. Sci. Comput.*, 28:1029–1053, 2006.

[60] F. Filbet and L. Pareschi. A numerical method for the accurate solution of the fokker-planck-landau equation in the nonhomogeneous case. *Journal of Computational Physics*, 179:1–26, 2002.

[61] F. Filbet and G. Russo. High order numerical methods for the space non homogeneous Boltzmann equation. *Journal of Computational Physics*, 186:457–480, 2003.

[62] Laurent Fousse, Guillaume Hanrot, Vincent Lefèvre, Patrick Pélissier, and Paul Zimmermann. MPFR: A multiple-precision binary floating-point library with correct rounding. *ACM Transactions on Mathematical Software*, 33(2):13:1–13:15, June 2007.

[63] Matteo Frigo and Steven G. Johnson. The design and implementation of FFTW3. *Proceedings of the IEEE*, 93(2):216–231, 2005. Special issue on "Program Generation, Optimization, and Platform Adaptation".

[64] E. Gabetta, L. Pareschi, and G. Toscani. Relaxation schemes for nonlinear kinetic equations. *SIAM J. Numer. Anal.*, 34:2168–2194, 1997.

[65] E. Gabriel, G.E. Fagg, G. Bosilca, T. Angskun, J. J. Dongarra, J. M. Squyres, V. Sahay, P. Kambadur, B. Barrett, A. Lumsdaine, R. H. Castain, D. J. Daniel, R. L. Graham, and T. S. Woodall. Open MPI: Goals, concept, and design of a next generation MPI implementation. In *Proceedings, 11th European PVM/MPI Users' Group Meeting*, pages 97–104, Budapest, Hungary, September 2004.

[66] I.M. Gamba, V. Panferov, and C. Villani. On the Boltzmann equation for diffusively excited granular media. *Communications in Mathematical Physics*, 246(39):503–541, 2004.

[67] I.M. Gamba, V. Panferov, and C. Villani. Upper Maxwellian bounds for the spatially homogeneous Boltzmann equation. *Arch.Rat.Mech.Anal*, 194(1):253–282, 2009.

[68] I.M. Gamba and Sri H. Tharkabhushaman. Spectral - lagrangian based methods applied to computation of non - equilibrium statistical states. *Journal of Computational Physics*, 228:2012–2036, 2009.

[69] I.M. Gamba and Sri H. Tharkabhushaman. Shock and boundary structure formation by spectral-lagrangian methods for the inhomogeneous boltzmann transport equation. *Jour. Comp. Math*, 28(4):430–460, 2010.

[70] I.M. Gel'fand and co authors. *Generalized Functions*. Academic Press, New York, 1964. English translation from the Russian.

[71] F. Golse and F. Poupaud. Stationary solutions of the linearized boltzmann equation in a half-space. *Math. Methods Appl. Sci.*, 11:483–502, 1989.

[72] H. Grad. Asymptotic theory of the boltzmann equation, ii. In *Rarefied Gas Dynamics (Proc. 3rd Internat. Sympos., Palais de l'UNESCO, Paris, 1962), Vol.I*, pages 26–59. Academic Press, New York, 1963.

[73] P. T. Gressman and R. Strain. Global classical solutions of the boltzmann equation with long-range interactions. *Proc. Natl. Acad. Sci. USA*, 107(13):5744–5749, 2010.

[74] Y. Guo. The vlasov-poisson-boltzmann system near maxwellians. *Comm. Pure Appl. Math*, 55:1104–1135, 2002.

[75] Y. Guo. Classical solutions to the boltzmann equation for molecules with an angular cutoff. *Arch. Rational Mech. Anal.*, 169:305–353, 2003.

[76] Y. Guo. The vlasov-maxwell-boltzmann system near maxwellians. *Invent. Math*, 153:593–630, 2003.

[77] Y. Guo. The boltzmann equation in the whole space. *Indiana Univ. Math. J.*, 53:1081–1094, 2004.

[78] R. E. Heath, I. M. Gamba, P. J. Morrison, and C. Michler. A discontinuous galerkin method for the vlasov-poisson system. *Journal of Computational Physics*, 231(4):1140–1174, 2012.

[79] R.E. Heath. *Numerical analysis of the discontinuous Galerkin method applied to plasma physicsa*. PhD thesis, ICES, The University of Texas at Austin, Austin, Texas, 2007.

[80] M. Herty, L. Pareschi, and M. Seaid. Discrete-velocity models and relaxation schemes for traffic flows. *SIAM J. Sci. Comput.*, 28:1582–1596, 2006.

[81] D. Hilbert. Grundzüge einer allgemeinen theorie der linearen integralgleichungen. In *Math. Ann. 72 (1912)*. Chelsea Publ., New York, 1953.

[82] W. Hoitinga and E.H. van Brummelen. A discontinuous galerkin finite-element method for a 1d prototype of the boltzmann equation. *Journal of Computational Physics*, 230(15):6115–6135, 2011.

[83] I. Ibragimov and S. Rjasanow. Numerical solution of the Boltzmann equation on the uniform grid. *Computing*, 69:163–186, 2002.

[84] R. Illner. On the derivation of the time-dependent equations of motion for an ideal gas with discrete velocity distribution. *J. de Mecanique*, 17:781–796, 1978.

[85] Plasma Wall Interaction. `http://plasma.physik.hu-berlin.de/psi/wall/pwi.html`.

[86] J.R.Haack and I.M. Gamba. Conservative deterministic spectral boltzmann solver near the grazing collisions limit. *28th Rarefied Gas Dynamics Conference (2012)*, AIP Conference Proceedings (2012).

[87] J.R.Haack and I.M. Gamba. High performance computing with a conservative spectral boltzmann solver. *28th Rarefied Gas Dynamics Conference (2012)*, AIP Conference Proceedings (2012).

[88] S. Kawashima. Global solution of the initial value problem for a discrete velocity model of the Boltzmann equation. *Proc. Japan Acad. Ser. A Math. Sci.*, 57:19–24, 1981.

[89] L.D. Landau. Die kinetische gleichung für den fall coulombscher vechselwirkung. *Phys. Z. Sowjet*, 154, 1936.

[90] M. Lemou. Numerical algorithms for axisymmetric fokker-planck-landau operators. *Journal of Computational Physics*, 157:762–786, 2000.

[91] A. Majorana. A numerical model of the boltzmann equation related to the discontinuous galerkin method. *Kinetic and Related Models*, 4(1):139–151, 2011.

[92] C.J. McKinstrie, R.E. Giacone, and E.A. Startsev. Accurate formula for the landau damping rates of electrostatic waves. *Phys. Plasmas*, 6:463–466, 1999.

[93] L. Mieussens. Discrete-velocity models and numerical schemes for the Boltzmann-bgk equation in plane and axisymmetric geometries. *Journal of Computational Physics*, 162:429–466, 2000.

[94] A.B. Morris, P.L. Varghese, and D.B. Goldstein. Improvement of a discrete velocity boltzmann equation solver with arbitrary post-collision velocities. In *26th International Symposium on Rarefied Gas Dynamics 2010*, pages 458–463. AIP Conf. Proc. 1084, 2009.

[95] A.B. Morris, P.L. Varghese, and D.B. Goldstein. Monte carlo solution of the boltzmann equation via a discrete velocity model. *Journal of Computational Physics*, 230:1265–1280, 2011.

[96] A.B. Morris, P.L. Varghese, and D.B. Goldstein. Variance reduction for a discrete velocity gas. In *27th International Symposium on Rarefied Gas Dynamics 2010*. AIP Conf. Proc. 1333, 2011.

[97] C. Mouhot. Quantitative linearized study of the boltzmann collision operator and application. *Comm. Math. Sci*, pages 73–86, suppl. 1 2007.

[98] C. Mouhot and L. Pareschi. Fast algorithms for computing the Boltzmann collision operator. *Math. Comp.*, 75:1833–1852, 2006.

[99] C. Mouhot and R. Strain. Spectral gap and coercivity esitmates for linearized boltzmann collision operators without angular cutoff. *J. Math. Pures Appl*, 87:515–535, 2007.

[100] K. Nanbu. Direct simulation scheme derived from the Boltzmann equation in monocomponent gases. *J. Phys. Soc. Japan*, 52:2042–2049, 1983.

[101] NASA Science News. The day the solar wind disappeared. `http://science.nasa.gov/science-news/science-at-nasa/1999/ast13dec99_1/`.

[102] T. Van Noije and M. Ernst. Velocity distributions in homogeneously cooling and heated granular fluids. *Granular Matter*, 1(2):57–64, 1998.

[103] The University of Texas at Austin. Texas advanced computing center (tacc). `http://www.tacc.utexas.edu`.

[104] Y. P. Pao. Boltzmann collision operator with inverse-power intermolecular potentials. i & ii. *Comm. Pure Appl. Math*, 27:407–428, 559–581, 1974.

[105] L. Pareschi and B. Perthame. A Fourier spectral method for homogenous Boltzmann equations. *Transport Theory Statist. Phys.*, 25:369–382, 2002.

[106] L. Pareschi and G. Russo. Numerical solution of the Boltzmann equation. i. spectrally accurate approximation of the collision operator. *SIAM J. Numerical Anal. (Online)*, 37:1217–1245, 2000.

[107] L. Pareschi, G. Russo, and G. Toscani. Fast spectral methods for the fokker-planck-landau collision operator. *Journal of Computational Physics*, 165:216–236, 2000.

[108] O. Pezzi, F. Valentini, D. Perrone, and P. Veltri. Eulerian simulations of collisional effects on electrostatic plasma waves. *Phys. Plasmas*, 20(9), 2013.

[109] S. Rjasanow and W. Wagner. A stochastic weighted particle method for the Boltzmann equation. *Journal of Computational Physics*, 124(2):243–253, 1996.

[110] S. Rjasanow and W. Wagner. *Stochastic Numerics for the Boltzmann Equation*. Springer, Berlin, 2005.

[111] E. Rutherford. *Philosophical Magazine*, 21, 1911.

[112] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77:439–471, 1988.

[113] E. Stein. *Singular integrals and differentiability properties of functions*. Princeton University Press, Princeton, N.J., 1970.

[114] R.M. Strain. The vlasov-maxwell-boltzmann system in the whole space. *Comm. Math. Phys.*, 268:543–567, 2006.

[115] R.M. Strain and Y. Guo. Almost exponential decay near maxwellian. *Comm. Partial Differential Equaitons*, 31:417–429, 2006.

[116] R.M. Strain and Y. Guo. Exponential decay for soft potentials near maxwellian. *Arch. Rational Mech. Anal.*, 187(2):287–339, 2008.

[117] C. Villani. A review of mathematical topics in collisional kinetic theory. In *Handbook of Fluid Mechanics*. 2003.

[118] ORNL visualization. http://blogs.knoxnews.com/munger/2010/01/fields-for-fusion.html.

[119] X. Zhang and C.-W. Shu. On maximum-principle-satisfying high order schemes for scalar conservation laws. *Journal of Computational Physics*, 229:3091–3120, 2010.

[120] X. Zhang and C.-W. Shu. On positivity preserving high order discontinuous galerkin schemes for compressible euler equations on rectangular meshes. *Journal of Computational Physics*, 229:8918–8934, 2010.

[121] X. Zhang and C.-W. Shu. Positivity-preserving high order discontinuous galerkin schemes for compressible euler equations with source terms. *Journal of Computational Physics*, 230:1238–1248, 2011.

[122] X. Zhang, Y. Xia, and C.-W. Shu. Maximum-principle-satisfying and positivity-preserving high order discontinuous galerkin schemes for conservation laws on triangular meshes. *Journal of Scientific Computing*, 50(1):29–62, 2012.

[123] Y. Zheng and H. Struchtrup. A linearization of Mieussens's discrete velocity model for kinetic equations. *Eur. J. Mech. B Fluids*, 26:182–192, 2007.