Modeling and Computational Methods for Multi-scale Quantum Dynamics and Kinetic Equations

By

Qin Li

A dissertation submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

(Mathematics)

at the

UNIVERSITY OF WISCONSIN – MADISON

2013

Date of final oral examination: May 6, 2012

The dissertation is approved by the following members of the Final Oral Committee:

Professor Shi Jin, Professor, Mathematics

Professor Alex Kiselev, Professor, Mathematics

Professor Saverio Spagnolie, Assistant Professor, Mathematics

Professor Samuel Stechmann, Assistant Professor, Mathematics

Professor Jean-Luc Thiffeault, Associate Professor, Mathematics

Abstract

This dissertation consists of two parts: quantum transitions (Part 1) and hydrodynamic limits of kinetic equations (Part 2). In both parts, we investigate the inner mathematical connections between equations for different physics at different scales, and use these connections to design efficient computational methods for multi-scale problems.

Despite its numerous applications in chemistry and physics, the mathematics of quantum transition is not well understood. Using the Wigner transformation, we derive semi-classical models in phase space for two problems: the dynamics of electrons in crystals near band-crossing points; surface hopping of quantum molecules when the Born-Oppenheimer approximation breaks down. In both cases, particles may jump between states with comparable energies. Our models can capture the transition rates for such processes. We provide analytic analysis of and numerical methods for our models, demonstrated by explicit examples.

The second part is to construct numerical methods for kinetic equation that are efficient in the hydrodynamic regime. Asymptotically, the kinetic equations reduce to fluid dynamics described by the Euler or Navier-Stokes equations in the fluid regime. Numerically the Boltzmann equation is still hard to handle in the hydrodynamic regime due to the stiff collision term. We review the theoretical work that links the two sets of equations, and present our asymptotic-preserving numerical solvers for the Boltzmann equation that naturally capture the asymptotic limits in the hydrodynamic regime. We also extend our methods to the case of multi-species systems.

Acknowledgements

My deepest thanks go to my advisor, Professor Shi Jin. Without three and a half years of discussions with him, as well as his advice and support, this dissertation would not exist. Thank you!

I would also like to thank all my committee members for reading this work and providing their invaluable comments and suggestions.

I owe a debt of gratitude to my collaborators: Prof. Lorenzo Pareschi, Prof. Xu Yang and Dr. Lihui Chai, and I look forward to our discussions in the future. I am grateful to Professors Jose Carrillo, Laurent Desvillettes, Alex Kiselev, Jian-Guo Liu, Paul Milewski, James Rossmanith and Eitan Tadmor who have been constant sources of help and encouragement for me. Thanks to Bokai Yan, Jingwei Hu and Li Wang for sharing their insights.

Thanks to Weiwei & Canran, Suky & Tommy, Jo, Keith, Lalit, Lee, Liu, Mike, Sole, Yongqiang and Zhennan for making these years count. I could never capture in this space what you each mean to me, but I do appreciate you guys for being a part of my life.

One of the hardest things about these four years has been being 1500 kilometers from Yang-Le. Waiting for your phone calls has been very painful in most of the days. I was not able to see you much, but your passion for both science and life has always been with me, and your love has made my life immeasurably better. Finally, and most importantly, I thank my parents for making any of this possible. I don't know how you figured it out but you have been and will alway be brilliant parents.

List of Figures

2.3.1 The eigenvalues of the Mathieu Model, $V(x) = \cos x$. 38

2.3.2 Periodic potential problem. Linear $U(x)$. $t = 0.5$. The left and right column are for the position density ρ^ε, and c.d.f. γ^ε respectively. The solid line, the dash line and the dotted line are the numerical solutions to the Schrödinger, the Liouville-A and the Liouville-S respectively. 45

2.3.3 Periodic potential problem. Linear $U(x)$. Time evolution of $m_1(t)$ defined in (2.3.10). 46

2.3.4 Periodic potential problem. Linear $U(x)$. Err^ε as function of ε at $t = 0.5$. . . . 47

2.3.5 Periodic potential problem. Example 1. $t = 0.5$. The left and right columns show the position density ρ^ε, and the c.d.f. γ^ε respectively. 48

2.3.6 Periodic potential problem. Example 1. Time evolution of $m_1(t)$ defined in (2.3.10). 49

2.3.7 Periodic potential problem. Example 1. The L^1 error Err^ε between Liouville-A and the Schrödinger solution at $t = 0.5$. 50

2.3.8 Periodic potential problem. Example 2. At time $t = 0.5$. The left and the right columns show the position density ρ^ε, and the c.d.f. γ^ε respectively. 52

2.3.9 Periodic potential problem. Example 2. The L^1 error Err^ε as a function of ε at $t = 0.5$. 53

3.3.1 Surface hopping problem. Example 1. The left column is for $\rho^+_{schr/liou}$, the density on the upper band, and the right column gives the results of $\rho^-_{schr/liou}$. $\Delta = \sqrt{\varepsilon}/2$. 80

3.3.2 Surface hopping problem. Example 1. Time evolution of the population (defined in (3.3.1) and (3.3.2)) on the upper band $P^+_{schr/liou}$. $\Delta = \sqrt{\varepsilon}/2$. 81

3.3.3 Surface hopping problem. Example 1. Err^ε decreases at $O(\varepsilon)$. $t = 10\sqrt{\varepsilon}$. 82

3.3.4 Surface hopping problem. Example 2. $\Delta = \sqrt{\varepsilon}/2$. The left column is the density on the upper band $\rho^+_{schr/liou}$ and the right column is $\rho^-_{schr/liou}$. 83

3.3.5 Surface hopping problem. Example 2. Err^ε as a function of ε at $t = 10\sqrt{\varepsilon}$. . . . 84

3.3.6 Surface hopping problem. Example 3. Long time behavior of P^+. 84

3.3.7 Surface hopping problem. Example 4. The left column is for $\rho^+_{schr/liou}$, the density on the upper band, and the right column gives the results of $\rho^-_{schr/liou}$. $\delta = \sqrt{\varepsilon}/2$. 85

3.3.8 Surface hopping problem. Example 4. Time evolution of the population on the upper band $P^+_{schr/liou}$. $\delta = \sqrt{\varepsilon}/2$. The legend "Schrödinger" represents the solution of the Schrödinger equation, "Liouville-j" represents the solution of the Liouville system by the domain decomposition method with $\Delta x = \Delta p = h$ in the classical regions and $\Delta x = \Delta p = h/2$ in then semi-classical region, where $h = \sqrt{\varepsilon}/2^{j-1}$, and $j = 1, 2, 3, 4$ 86

3.3.9 Surface hopping problem. Example 4. Err^ε at $t = 10\sqrt{\varepsilon}$. The Liouville system by the domain decomposition method with $\Delta x = \Delta p = \sqrt{\varepsilon}/8$ in the classical regions and $\Delta x = \Delta p = \sqrt{\varepsilon}/16$ in then semi-classical region. 87

4.1.1 Velocity change after the collision. 93

5.2.1 ExpRK method. Convergence rate test. In each picture, 4 lines are plotted: the lines with dots, circles, stars and triangles on them are given by results of ExpRK2-F, ExpRK2-V, ExpRK3-F and ExpRK3-V respectively. The left column is for Maxwellian initial data, and the right column is for initial data away from Maxwellian (5.2.37). Each row, from the top to the bottom, shows results of $\epsilon = 1, 0.1, 10^{-3}, 10^{-6}$ respectively. 125

5.2.2 Exp-RK method. The Sod problem. Left column: $\varepsilon = 0.01$. The solid line is given by explicit scheme with dense mesh, while dots and circles are given by ExpRK2-F and ExpRK2-V respectively, both with $N_x = 100$. $h = \Delta x/20$ satisfies the CFL condition with CFL number being 0.5. Right column: For $\varepsilon = 10^{-6}$, both methods capture the Euler limit. The solid line is given by the kinetic scheme for the Euler equation, while the dots and circles are given by ExpRK2-F and ExpRK2-V. They perform well in rarefaction, contact line and shock. 127

5.2.3 Mixing Regime: $\varepsilon(x)$ 128

5.2.4 ExpRK method. The mixing regime problem. The left column shows comparison of RK2 and RK3 using the ExpRK-V. The solid line is the reference solution with a very fine mesh in time and $\Delta x = 0.0025$, the dash line is given by RK3 and the dotted line is given by RK2, both with $N_x = 50$ points. The right column compare two methods, both given by RK3, with the reference. The dash line is given by ExpRK-V, and the dotted line is given by ExpRK-F. $N_x = 50$ for both. h is chosen to satisfy CFL condition, in our case, the CFL number is chosen to be 0.5. 130

6.2.1 Distribution function for disparate masses system. The distribution function for heavy species is much narrower than that of the light species. 149

6.2.2 Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. As $\varepsilon \to 0$, solution of the Boltzmann equation goes to the Euler limit. $t = 0.1$. 155

6.2.3 Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. $\varepsilon = 0.1$, $t = 0.1$. The dashed line is given by the AP scheme, and the solid line is given by the forward Euler with a fine mesh: $\Delta x = 0.01$ and $h = 0.0005$. 156

6.2.4 The stationary shock problem. $\varepsilon = 10^{-5}$, $\delta f = f_1 - \overline{M}_1$ diminishes as time evolves. 156

6.2.5 Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. u_1 and u_2 at $x = -0.5$ on the left and T_1, T_2 on the right, as functions of time, for $\varepsilon = 10^{-2}$ and 10^{-5} respectively. Note the different time scales for the two figures. 157

6.2.6 Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. Indifferentiability. $t = 0.1$. ρ, u and T are computed using two species model ("o") and one species model (".") 157

6.2.7 Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $t = 0.1$. As $\varepsilon \to 0$, the numerical results go to the Euler limit. Solid lines give the Euler limit computed by using [142]. 158

6.2.8 Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $t = 0.1$. For $\varepsilon = 0.1, 0.01$, we compare the results of the AP scheme, given by the circled lines, and the results of the forward Euler with a fine mesh, given by the solid lines. 159

6.2.9 Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $\varepsilon = 10^{-4}$. $\delta f = f_1 - \overline{M}_1$. 159

6.2.10 Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. The three figures show velocities u_1 and u_2 at $x = -0.3$ as functions of time for $\varepsilon = 0.1, 0.01,$ and 10^{-5} respectively. Note different time scales for three figures. 160

6.2.11 Multi-species Boltzmann equation. BKG-penalization method. The disparate masses problem. The left figure shows the initial distributions and the right figure show the time evolutions of f_H and f_L. f_H is put at the top and f_L is at the bottom. At $t = 0.007$, f_L is close to $\overline{M}_L$ while f_H is still far away from the equilibrium. Note the different scales for v. 160

6.2.12 Multi-species Boltzmann equation. BKG-penalization method. The disparate masses problem. The velocities for the two species converge to each other. . . . 161

6.3.1 Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. We compare the numerical results at $t = 0.1$ by the Exp-RK method (the dotted line), the BGK penalization method (the circled line), and the reference solution (the solid line). The left figures are for $\varepsilon = 1$, and the right ones are for $\varepsilon = 0.1$. 167

6.3.2 Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. As ε goes to zero, the results get to the Euler limit, i.e. a stationary shock. 168

6.3.3 Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. The velocities converge to the mean velocity at different rates for different ε's. The circles and dots stand for the velocities of the first and second species respectively. The smaller ε gives the faster convergence rate. 168

6.3.4 Multi-species Boltzmann equation. ExpRK method. The Sod problem. Indistinguishability. The solid lines are the results of treating the system as single species, and they agree with the results given by the dashed lines, which are the results of a system that has two species with the same mass. 169

6.3.5 Multi-species Boltzmann equation. ExpRK method. The Sod problem. At $t = 0.1$, we compare the results give by the Exp-RK method (the dotted line), the BGK penalization method (the circled line) and the reference solution (the solid line). The left figures are for $\varepsilon = 1$, and the right ones are for $\varepsilon = 0.1$. . . 169

6.3.6 Multi-species Boltzmann equation. ExpRK method. The Sod problem. As ε goes to zero, the numerical results by the Exp-RK method converge to the Euler limit. 169

6.3.7 Multi-species Boltzmann equation. ExpRK method. The Sod problem. The velocities converge to the mean velocity at different rates for different ε's. The circled line stands for species one, and the dotted line stands for species two. The smaller ε gives the faster convergence rate. 170

6.3.8 Multi-species Boltzmann equation. ExpRK method. The Sod problem. $f - M$ for $\varepsilon = 1$ and $\varepsilon = 10^{-6}$ at $t = 0.1$. 170

Contents

Abstract **i**

Acknowledgements **ii**

I Quantum Transition **1**

1 Semi-classical models **3**

1.1 Introduction 3

1.1.1 WKB approximation 5

1.1.2 Gaussian beam method 7

1.2 Wigner transform 8

1.2.1 Properties 10

1.2.2 Wigner equation 11

2 Semi-classical models for crystals **13**

2.1 Dynamics of electrons in crystals 14

2.1.1 Non-dimensionalization 15

2.1.2 Spectrum analysis 16

2.2 Semi-classical model in phase space 20

2.2.1 Model derivation 21

2.2.2 Model derivation – symmetric case 30

2.2.3 Main properties of the model 32

2.2.4 Numerical method 36

2.3 Numerical example 37
2.3.1 Linear $U(x)$ 38
2.3.2 A domain decomposition computation 44

3 Semi-classical description of molecular dynamics 54

3.1 Born-Oppenheimer approximation and problem background 57
3.1.1 Equation set up 58
3.1.2 Basic properties 62
3.2 Semi-classical models 63
3.2.1 Weyl quantization and Wigner transform 64
3.2.2 Adiabatic and non-adiabatic models 67
3.2.3 Properties and numerical methods 73
3.3 Numerical examples 76
3.3.1 Example 1: 1D, pure state initial data 77
3.3.2 Example 2: 1D, mixed state initial data 77
3.3.3 Long time behavior 78
3.3.4 Example 4: 2D pure initial data 78

II Kinetic Theory 88

4 Kinetic theory and the hydrodynamic limit 91

4.1 Boltzmann equation 91
4.1.1 Properties of the collision term 94
4.1.2 Non-dimensionalization and rescaling process 96
4.2 Asymptotic limit 98

5 Asymptotic preserving numerical methods **101**

5.1 BKG-penalization method 104

5.2 Exponential Runge-Kutta method 106

5.2.1 Numerical methods 107

5.2.2 Properties of ExpRK schemes 112

5.2.3 Numerical example 122

6 Multi-species systems **131**

6.1 The multi-species system 131

6.1.1 The multi-species Boltzmann equation 132

6.1.2 Properties of the multi-species Boltzmann equation 133

6.1.3 A BGK model 136

6.2 BGK-penalization method 138

6.2.1 Numerical method 139

6.2.2 The AP property of the time discretization 143

6.2.3 Disparate masses 148

6.2.4 Numerical examples 151

6.3 ExpRK method for the multi-species Boltzmann equation 154

6.3.1 Exponential Runge-Kutta method 154

6.3.2 Positivity and asymptotic preserving properties 163

6.3.3 Numerical example 166

A Quantum system **172**

A.1 Some basic analysis of the semi-classical Liouville systems 172

A.1.1 Weak convergence 172

A.1.2 Strong convergence: for constant b 173

A.2 The integration method of a simple model system 175

A.3 The derivations of (3.2.18) 177

B Kinetic theory **180**

B.1 Positivity of the mass density in ExpRK-V 180

B.2 $\|P(f) - P(g)\| \leq \|f - g\|$ for single species 182

B.2.1 In d_2 norm 182

B.2.2 In L_2 norm 184

B.3 $\|P(f)\| < C\|f\|$ multi-species 185

B.3.1 In d_2 norm 185

B.3.2 In L_2 norm 187

Bibliography **188**

Part I

Quantum Transition

In this part, we study quantum systems with discrete spectrum, and explore semi-classical models and numerical methods to compute quantum transitions between energy states. We start with first principles: given a Schrödinger equation with an external potential, one first solves the eigenvalue problem of the Hamiltonian operator. The eigenfunctions are usually referred to as molecular orbitals in quantum chemistry, and the eigenvalues are called energy bands in physics. The solution to the Schrödinger equation is a superposition of all these eigenstates, with the coefficients evolving in time. Most of the time, the energy states are decoupled, which means their coefficients evolve independently. However, in some cases, eigenvalues get very close to each other, or even cross. In physics, when this happens, one could not distinguish two energy states that share the same energy value. And mathematically, the projection coefficients need to be modified in response to this quantum phenomena. Quantum transition is a very important issue because it is on the pathway for most chemical reactions. Qualitatively how transition functions and gets involved in chemical reactions has been studied in depth, but not much is known quantitatively due to the lack of mathematical tools available. We address such issues from the perspective of semi-classical limits in phase space, and correspondingly we design efficient numerical methods that give accurate results. Quantum transition occurs in many quantum problems, but we only deal with the following two: the first is band crossing problem that appears in the dynamics of electrons in crystals and the second is the surface hopping problem in quantum molecular dynamics.

In the following, we firstly briefly go over the basic ideas of semi-classical limit in chapter 1, with an outline of the currently dominating methods. We will focus on the Wigner transform, the one we rely on in the derivation of our model. Chapter 2 and 3 are devoted to the two problems mentioned above respectively.

Chapter 1

Semi-classical models

In this chapter we briefly go over the standard methods developed in the past that give the semi-classical limit for quantum systems. Techniques on dealing with quantum transitions are not going to be addressed in this chapter.

1.1 Introduction

The Schrödinger equation is a canonical dispersive equation, whose waves of different wavelength propagate at different speeds. Consider the Cauchy problem for the Schrödinger equation:

$$\begin{aligned} \mathrm{i}\varepsilon\psi^\varepsilon &= -\frac{\varepsilon^2}{2}\Delta\psi^\varepsilon + V\psi^\varepsilon, \quad (t,x) \in \mathbb{R}^+ \times \mathbb{R}^d \\ \psi^\varepsilon(x, t=0) &= \psi^\varepsilon_{in}(x) \end{aligned} \tag{1.1.1}$$

$\psi^\varepsilon(t,x)$ is the wave function, depending on time and space. It can be either a scalar or vector. The space is d dimensional. ε denotes the semi-classical parameter, usually indicating the microscopic/macroscopic ratio. It is assigned different physical meanings depending on different problems in different regimes, as will be clear later on, and for now we only assume it is a dimensionless small parameter. V is the potential term. Associated with the dimension of the wave function, it could be either a scalar or a matrix. Most of the time it relies on both time and space, but in this thesis, its time dependence is not going to be discussed. When

V depends on ψ, nonlinear effects also appear, and the most classical model in this case is $V = \pm|\psi|^2$. This situation will be omitted from this thesis as well.

For later use, we define the following:

- Hamiltonian operator:

$$H = -\frac{\varepsilon^2}{2}\Delta + V(x). \tag{1.1.2}$$

  This Hamiltonian is consistent with the classical one $H = \frac{p^2}{2} + V(x)$, whose p is replaced by the momentum operator $p = -\mathrm{i}\varepsilon\nabla_x$.

- Physical observables:

$$\text{particle density:} \quad n^\varepsilon(x,t) = |\psi^\varepsilon|^2, \tag{1.1.3a}$$

$$\text{current density:} \quad J^\varepsilon(x,t) = -\frac{\mathrm{i}\varepsilon}{2}\left(\bar{\psi}^\varepsilon\nabla_x\psi^\varepsilon - \psi^\varepsilon\nabla_x\bar{\psi}^\varepsilon\right) \tag{1.1.3b}$$

$$\text{energy density:} \quad e^\varepsilon(x,t) = \frac{1}{2}|\varepsilon\nabla_x\psi^\varepsilon|^2 + Vn^\varepsilon \tag{1.1.3c}$$

  In practice only these physical observables could be measured and are cared about. Simple derivation shows

$$\partial_t n^\varepsilon(x,t) + \nabla_x \cdot J^\varepsilon(x,t) = 0 \tag{1.1.4}$$

In the absence of V, its analytical solution is plane wave:

$$\psi^\varepsilon(t,x) = \int C(\xi)\exp\left(\frac{\mathrm{i}}{\varepsilon}\left(\xi\cdot x - \frac{t}{2}|\xi|^2\right)\right)d\xi \tag{1.1.5}$$

where $C(\xi)$ is the projection onto the mode $\xi \in \mathbb{R}^d$ and they are determined by the initial data. Apparently the solution features ε oscillations in both time and space. For fixed ε, the development of numerical methods is very mature, both for time-dependent and independent case, however, in the semi-classical regime, as ε diminishes, very fine mesh is required according

to the typical numerical analysis, and thus leads to extremely heavy computational cost.

To overcome this difficulty, we study the asymptotic behavior of the solution and include the asymptotic mathematical analysis while designing numerical schemes to save some numerical cost. Several methods are in order: WKB expansion, Gaussian beam method, the Hagedorn expansion, and the Wigner transformation. Basically all these methods assume a certain ansatz for the equation that take the oscillations into account, and one retrieves the leading order and sometimes the second order term as $\varepsilon \to 0$. We briefly review these asymptotic methods in the following, and one could find more details in [115] and the references therein. Wigner transform method belongs to this asymptotic analysis category as well, but it will be discussed in details in the following section.

1.1.1 WKB approximation

WKB method, also known as WKBJ method sometimes, is named after Wentzel-Kramers-Brillouin (Jeffreys when J is included). It is a general asymptotic method for linear high order differential equations. In case of time-dependent Schrödinger equation, the ansatz writes as:

$$\psi^\varepsilon = a^\varepsilon(t,x)\exp\left(S^\varepsilon(t,x)\right) \tag{1.1.6}$$

Assume the two real-valued functions: the amplitude $a^\varepsilon(x,t)$ and $S^\varepsilon(x,t)$ have asymptotic expansion:

$$a^\varepsilon = a + \varepsilon a_1 + \varepsilon^2 a_2 + \cdots \tag{1.1.7}$$

$$S^\varepsilon = S + \varepsilon S_1 + \varepsilon^2 S_2 + \cdots \tag{1.1.8}$$

Plugging this expansion into (1.1.1), by matching orders, one could sequentially solve for the approximation up to the order desired. To the leading order, S satisfies a Hamiltonian-Jacobi

equation with energy given by $\frac{\dot{x}^2}{2} + V(x)$, while a is governed by a transport equation:

$$\partial_t S + \frac{1}{2}|\nabla_x S|^2 + V(x) = 0, \quad \partial_t a + \nabla_x S \nabla_x a + \frac{a}{2}\Delta S = 0. \tag{1.1.9}$$

The second equation on a could also be interpreted in term of number density $\rho = |a|^2$:

$$\partial_t \rho + \nabla \cdot (\rho \nabla_x S) = 0. \tag{1.1.10}$$

Given smooth enough S, ρ is governed by a conservation law and could be computed in a classical way. In this way, the WKB method gives an easy solution to the Schrödinger equation up to the leading order.

To compute the equation for S is much tricker. In fact, the WKB approximation is only valid for finite time, since the Hamiltonian-Jacobi equation for S develops singularity even with smooth initial data: mathematically S could be regarded as integration of some solution to the Burger's equation, and it is well-known that even C_∞ initial data leads to shocks. By following characteristics, one solves the following ODE:

$$\dot{x} = \xi, \quad \dot{\xi} = -\nabla_x V. \tag{1.1.11}$$

and often in time, these characteristics cross, meaning the two particles starting from different location bump into each other, generating "caustics".

This phenomenon only occurs for WKB, and not for the original Schrödinger equation, and thus the "caustics" is artificial and embedded in the definition of WKB ansatz. To fix the problem beyond the caustics, many techniques can be applied. We mention here the multiphase WKB approximation, which writes the solution as summation of a series of WKB form [113, 80], with a phase shift called Keller-Maslov index [124]. For theoretical study, refer

to [56]. Level set method is a general way in computing multi-values appeared in Hamiltonian-Jacobian equation [161], and is adopted here as well [116, 114].

1.1.2 Gaussian beam method

The Gaussian beam method, developed for high frequency linear waves, is an efficient approximate method that allows accurate computation of the wave amplitude around caustics [168]. It is widely used in geophysics community as well [102]. The validity of its construction has been mathematically studied by Ralston in [171]. Its usage in quantum system can be found in [121, 120, 107, 200].

The idea is the following. Assume at the initial time step, the data is given in the form of:

$$\psi_0^\varepsilon(t,x,y) = A(t,y)e^{\mathrm{i}T(t,x,y)/\varepsilon} \tag{1.1.12}$$

where both the amplitude and the phase function have a new variable y indicating the center of the beam, as will be more clear soon. T is complex valued function and is Taylor expanded around y as:

$$T(t,x,y) = S(t,y) + p(t,y)\cdot(x-y) + \frac{1}{2}(x-y)^T M(t,y)(x-y) + \mathcal{O}(|x-y|^3) \tag{1.1.13}$$

To form a Gaussian profile, the imaginary part of the Hessian matrix M is positive definite. In the leading order expansion in both ε and $y-x$, one gets the ODE system:

$$\frac{dy}{dt} = p, \quad \frac{dp}{dt} = -\nabla_y V$$

$$\frac{dM}{dt} = -M^2 - \nabla_y^2 V \tag{1.1.14}$$

$$\frac{dS}{dt} = \frac{1}{2}|p|^2 - V, \quad \frac{dA}{dt} = -\frac{1}{2}(Tr(M))A \tag{1.1.15}$$

The equation for M is called Riccati equation, and by following the y-trajectory in this Lagrangian type of approach, one could check several important properties:

1. M is always symmetric and its imaginary part remains being positive definite;

2. The Hamiltonian $H = \frac{|p|^2}{2} + V(x)$ is conserved;

3. $A(t, y)$ does not blow up.

The first property guarantees that the profile is always a Gaussian function, while the second and the third incorporate the physical background.

Given an arbitrary initial data, even if it does not have the form of (1.1.12), it can always be approximated by the summation of several Gaussian packets, but the number could be as big as $\varepsilon^{-1/2}$. In case of the WKB initial data, the approximation could be made in a particular way to reach $\varepsilon^{1/2}$ accuracy [181, 182]. Given these functions, one needs to follow the trajectory for each packets and evolve them according to (1.1.14). They are summed up at the ending time. Also see high order extension in [122], and frozen Gaussian beam method in [149, 150].

1.2 Wigner transform

In this section we mainly focus on the Wigner transform developed by Wigner in [193]. Unlike most other asymptotic method, Wigner transform studies the behavior of the solution in phase space, and helps to establish the link between the quantum wavefunction and the probability distribution function in classical mechanics. The definition has the uncertainty principle built in, and thus provides us a powerful tool to untangle the coherence between the momentum and position in phase space [155]. In representation theory, it is also linked to Weyl quantization. The latter was developed in 1927 in [192] and is a general framework that transforms the real phase space functions to its associated Hermitian operators. More discussion will be found in section 3.2.1.

Wigner transform is defined as:

$$W^{\varepsilon}(t,x,k)=\int_{\mathbb{R}^d}\frac{\mathrm{d}y}{(2\pi)^d}\mathrm{e}^{\mathrm{i}ky}\psi^{\varepsilon}\left(t,x-\frac{\varepsilon}{2}y\right)\bar{\psi}^{\varepsilon}\left(t,x+\frac{\varepsilon}{2}y\right). \tag{1.2.1}$$

where $\bar{\psi}^{\varepsilon}$ is the complex conjugate of ψ^{ε}, with ψ^{ε} solves the Schrödinger equation (1.1.1). The multiplication of two wave functions $\psi^{\varepsilon}(x-\frac{\varepsilon y}{2})\bar{\psi}^{\varepsilon}(x+\frac{\varepsilon y}{2})$ is called the density matrix, and the transform is simply the inverse Fourier transform of it. Note that ψ and its complex conjugate are evaluated at different position with deviation $\mathcal{O}(\varepsilon)$, intrinsically reflecting the uncertainty principle. In bra-ket language, it could be expressed as:

$$W^{\varepsilon}(t,x,k)=\int_{\mathbb{R}^d}\frac{\mathrm{d}y}{(2\pi)^d}\langle x+\frac{\varepsilon}{2}y|\hat{\rho}|x-\frac{\varepsilon}{2}y\rangle e^{\mathrm{i}ky}. \tag{1.2.2}$$

The definition is to characterize the probability of finding a particle at time t, location x with momentum k, as a resemblance to the probability distribution function in statistical mechanics.

Remark 1.2.1. *Note that both two formulas above are written in x-presentation, consistent with the analysis and numerical computation that will be carried below. p-presentation could be used as well. One only needs to replace the states $|x\pm\frac{\varepsilon}{2}y\rangle$ by $|p\pm\frac{\varepsilon}{2}q\rangle$ and integrate over q, the deviation in the momentum space. The discussion for that is neglected from here.*

As a resemblance to the distribution function on the phase space in statistical mechanics, this Wigner function is introduced to characterize the probability of finding a particle at time t, location x with momentum k. However, a classical particle has a definite position and momentum, and hence it is represented by a point in phase space, in the quantum case, due to the uncertainty principle, this restriction breaks down, and the Wigner function can only be understood in a quasi-probability manner, and thus present some properties that are not shared by conventional probability distribution.

1.2.1 Properties

Several properties are in order.

1. W^ε is a real-valued function.

   Take the conjugate of the definition (1.2.1) it is easily seen that $W^\varepsilon = \bar{W}^\varepsilon$.

2. Time symmetry.

$$\psi^\varepsilon \to \bar{\psi}^\varepsilon \Rightarrow W(t,x,k) \to W(t,x,-k). \tag{1.2.3}$$

   It is called time symmetry simply because k indicates the momentum and $W^\varepsilon(t,x,-k)$ is the wave $W^\varepsilon(t,x,k)$ travelling backward.

3. ψ^ε cannot be fully recovered.

   By definition, W^ε has a quadratic form, and if one shift the phase of ψ^ε by a constant θ, i.e. $\psi^\varepsilon \to \psi^\varepsilon e^{i\theta}$, W^ε is unchanged.

4. The physical observables can be recovered from the moments:

$$\text{particle density:} \quad n^\varepsilon(x,t) = \int_{\mathbb{R}^d} W^\varepsilon \mathrm{d}k \tag{1.2.4a}$$

$$\text{current density:} \quad J^\varepsilon(x,t) = \int_{\mathbb{R}^d} kW^\varepsilon \mathrm{d}k \tag{1.2.4b}$$

$$\text{energy density:} \quad e^\varepsilon = \int_{\mathbb{R}^d} \left(\frac{|k|^2}{2} + V(x) \right) W^\varepsilon \mathrm{d}k \tag{1.2.4c}$$

5. Wigner function does not preserve positivity.

   (a) This property distinguishes the Wigner function from the conventional density distribution function. The reason behind it is the uncertainty principle. A classical particle has a definite position and momentum, and hence can be represented by a certain point on the phase space, but in the quantum case, a particle could only be

specified in space and momentum simultaneously with an error of $\mathcal{O}(\hbar)$ (ε in this thesis).

(b) If the Wigner function is smoothed through a filter with size bigger than ε, positivity could still be obtained. The most classical way is Husimi transform [129] where a Gaussian function with width of $\mathcal{O}(\varepsilon)$ is used to convolute with the density matrix.

1.2.2 Wigner equation

By plugging the Schrödinger equation into the definition of Wigner function, one has the Wigner equation:

$$\partial_t W^\varepsilon + k \cdot \nabla_x W^\varepsilon = \frac{1}{\mathrm{i}\varepsilon} \int \frac{\mathrm{d}\omega}{(2\pi)^d} e^{\mathrm{i}\omega x} \hat{V}(\omega) \left[W^\varepsilon(x, k - \frac{\varepsilon\omega}{2}) - W^\varepsilon(x, k + \frac{\varepsilon\omega}{2}) \right]. \tag{1.2.5}$$

Here $\hat{V}$ is the Fourier transform of the potential V. Assume $V(x)$ is a smooth enough function with $\nabla_x V = \mathcal{O}(1)$, and by performing asymptotic expansion for $W^\varepsilon = W_0 + \varepsilon W_1 + \cdots$, formally one obtains the classical Liouville equation:

$$\partial_t W_0 + k \cdot \nabla_x W_0 - \nabla_x V \cdot \nabla_k W_0 = 0 \tag{1.2.6}$$

This has been proved in [151, 147]:

Theorem 1.2.1. *Suppose ψ^ε is uniformly bounded in $L^2(\mathbb{R}^d)$ with respect to ε:*

$$sup_{0<\varepsilon\leq 1} |\psi^\varepsilon|_{L^2} < \infty, \quad \forall t, \tag{1.2.7}$$

then the associated Wigner function $W^\varepsilon \in S'(\mathbb{R}^d \times \mathbb{R}^d)$ is weak- compact. There exists one*

sequence such that:

$$W^{\varepsilon} \xrightarrow{\varepsilon \to 0} W_0 \ in \ L^{\infty}\left([0,T]; S'\left(\mathbb{R}^d \times \mathbb{R}^d\right)\right) \ weak\text{-}* \tag{1.2.8}$$

The limit W_0 is called the Wigner measure. It is positive and follows the Liouville equation (1.2.6).

Remark: One way to prove the positivity of this limit is to link it with the limit of Husimi measure, which keeps being positive for all ε [151].

Follow the characteristics of the equation (1.2.6), the resemblance to the classical mechanics is easily seen:

$$\dot{x} = k, \quad \dot{k} = -\nabla_x V(x). \tag{1.2.9}$$

This tells us that k is the velocity of a particle, and $\dot{k}$, the acceleration is governed by the derivative of the potential $V(x)$, which is exactly the Newton's 2nd Law. In this sense, this two different types of physics are linked with an error of $\mathcal{O}(\varepsilon)$. Compared with the WKB approximation, this approach unfold the singularities (caustics) and does not give the exact wave function ψ^{ε}, and the phase is globally well-defined.

Wigner transform is the tool we use to investigate the two problems described in the following two chapter.

Chapter 2

Semi-classical models for crystals

In this chapter we investigate the dynamics of electrons in crystals in the semi-classical regime. We explore the mathematical description for the quantum transition between energy bands, and develop efficient semi-classical numerical solver.

This problem is difficult because in the regime we consider, the lattice potential oscillates at the same scale of the rescaled Planck constant, making the numerical computation very expensive. Using the Bloch-Floquet theory, one could obtain the spectrum of the Hamiltonian by solving the associated eigenvalue problem. When the eigenvalues are very well separated from each other, it is proved that in the semi-classical limit, in phase space, the projection onto different bands are decoupled, and each coefficient evolves independently according to a simple transport equation. However, when the eigenvalues are degenerate, little theory on the quantum transition is known.

We solve this problem by performing the Wigner transform on the Schrödinger equation and study the resulting Wigner equation in phase space. To better interpret the structure of the phase space, we also perform the Wigner transform onto the Bloch eigenfunctions. This gives us the orthogonal basis in phase space. We project the Wigner function onto them, and discovered that the off-diagonal entries of the obtained Wigner matrix naturally reflect the quantum transition rates. The associated set of equations are obtained. This new model is asymptotically consistent with the classical model under the assumption that the energy bands are non-degenerate. Numerically, the story has two sides: on the one hand, the new model gives

full information both around and far away from the band-crossing region, but its oscillatory nature generates expensive numerical cost, and on the other hand, the classical model, though could not provide transition rate around the crossing point, gives sufficient information away from it. With this observation, we construct a domain-decomposition numerical method: we decompose the phase space into separate domains, and compute proper equations in different regions and hybridize them by appropriate boundary conditions. In this way we capture the correct transition rate without sacrificing much numerical expense.

In section 2.1, we briefly go over the setup of the problem and introduce its basic properties. The standard technique to tackle the problem is the Bloch-Floquet theory. In section 2.2 we go over the classical Wigner transform approach, and around the energy band crossing point where the classical model breaks down, we develop a new model that captures the quantum transition. Its basic properties will be presented in subsection 2.2.3 and the associated numerical methods are designed in subsection 2.2.4. For simplicity both the analysis and the computation are done in one dimension but they could be easily extended to higher dimensions. We show several numerical examples in section 2.3.

2.1 Dynamics of electrons in crystals

A crystal is a solid material whose atoms are arranged in an ordered pattern extending in spatial domain. In our problem we only care about the dynamics of electrons, and all bigger molecules are considered at rest, leaving electrons a static potential. Based on this set up, the Schrödinger equation used contains two potential terms: the lattice periodic potential generated by the ionic cores, and the experimentally imposed external potential that is usually introduced by electromagnetic fields. The latter typically vary on much larger spatial scales.

Denote l as the lattice constant, we rewrite (1.1.1) as:

$$\mathrm{i}\hbar\partial_t\psi = -\frac{\hbar^2}{2}\Delta\psi + V\left(\frac{x}{l}\right)\psi + U(x)\psi. \tag{2.1.1}$$

Here the $V(x)$ part in the potential term reflects the lattice oscillation and $U(x)$ is the smooth external potential. $\hbar$ is the Planck constant.

Two small parameters are involved in the equation: the Planck constant $\hbar$ and the lattice constant l. Depending on the regimes one is interested in, the parameters need to be rescaled and balanced, resulting different behavior of the solution. We firstly perform the nondimensionalization in section 2.1.1, and pick a regime where resonance of lattice potential could be seen. Basic properties, especially the Bloch-Floquet theory will be introduced in 2.1.2.

2.1.1 Non-dimensionalization

Define $\tau = l^2/\hbar$ as the quantum time scale, and denote T and L as macroscopic time and length scale. Rescale the parameters [10, 169, 57]:

$$\tilde{x} = \frac{x}{L}, \quad \tilde{t} = \frac{t}{T}, \quad \varepsilon = \frac{l}{L}, \quad h = \frac{\hbar T}{L^2}, \tag{2.1.2}$$

and the two potential functions:

$$\tilde{V}(\frac{x}{l}) = \frac{\tau^2}{ml^2}V(x), \quad \tilde{U}(\frac{x}{L}) = \frac{T^2}{L^2}U(x). \tag{2.1.3}$$

After dropping the tildes, one gets:

$$\mathrm{i}h\partial_t\psi = -\frac{h^2}{2}\Delta\psi + \frac{h^2}{\varepsilon^2}V\left(\frac{x}{\varepsilon}\right)\psi + U(x)\psi. \tag{2.1.4}$$

This equation is nondimensionalized, and it shows the competition between two parameters h and ε. As one can see, if ε is big compared to h, one is in a regime that no periodic structure of the crystal cannot be seen, and the system is roughly classical; if ε is comparably much smaller, the homogenization limit $\varepsilon \to 0$ of the periodic structure is performed faster than the classical limit $h \to 0$. In fact it is proved that this limit yields a semi-classical Vlasov equation with parabolic bands defined by some effective mass tensor [169] when energy bands are assumed to be separated. The most interesting case is when ε and h is at the same order. If so, the periodic potential $V(x)$ and the external potential $U(x)$ are of the same order, and one could see the $\mathcal{O}(1)$ impact of the periodic structure to the system and expect some resonance. In this case, write $h = \varepsilon$ and the equation becomes:

$$\mathrm{i}\varepsilon\partial_t\psi^\varepsilon = -\frac{\varepsilon^2}{2}\Delta\psi^\varepsilon + V\left(\frac{x}{\varepsilon}\right)\psi + U(x)\psi^\varepsilon, \tag{2.1.5}$$

$$\psi^\varepsilon(t=0,x) = \psi_I^\varepsilon(x). \tag{2.1.6}$$

For simplicity, we assume $V(x)$ is 2π periodic, and $U(x)$ is smooth and of $\mathcal{O}(1)$.

Asymptotic analysis has been extensively studied in, e.g., [162, 89, 164, 10], relying on different analytical tools, and numerical methods are also developed [79, 81, 82, 105, 106]. We also mention some papers that study related models is: Free crystal electrons, i.e. $U(x) = 0$ is studied in [151, 73]. Homogenization limit is studied in [169] with its extension to nonlinear Coulomb potential in [9]. Stratified potential with periodicity appearing in only one direction is studies in [11].

2.1.2 Spectrum analysis

It is the periodic potential that oscillates at the scale of ε that generates a great amount of difficulty in analyzing the equation. To fully make use of the periodicity, the classical way is to use Bloch-Floquet theory. The theory dates back to Floquet who discovered the underlying

mathematics in 1883 [69], and later on Bloch realized its application in the lattice potential case [17]. The whole theory was summarized in [194]. The idea is to combine the $V(\frac{x}{\varepsilon})$ with the Laplacian, and write it as a new Hamiltonian:

$$H_0 = -\frac{\varepsilon^2}{2}\Delta_x + V(\frac{x}{\varepsilon}) = -\frac{1}{2}\Delta_z + V(z). \tag{2.1.7}$$

Here we define $z = \frac{x}{\varepsilon}$ as the fast variable and study the eigenvalue problem:

$$\begin{aligned} -\frac{1}{2}\Delta_z\Psi(z,p) + V(z)\Psi(z,p) &= E(p)\Psi(z,p), && \text{(2.1.8a)} \\ \Psi(z+\nu,p) &= \mathrm{e}^{\mathrm{i}p\nu}\Psi(z,p), \ \ \forall\,\nu \in L = 2\pi\mathbb{Z}, && \text{(2.1.8b)} \\ \frac{\partial\Psi}{\partial z}(z+\nu,p) &= \mathrm{e}^{\mathrm{i}p\nu}\frac{\partial\Psi}{\partial z}(z,p), \ \ \forall\,\nu \in L. && \text{(2.1.8c)} \end{aligned}$$

The boundary conditions for this eigenvalue problem (2.1.8b)-(2.1.8c) is called Bloch boundary condition. Given a specific p, the solution is equipped with pseudo-periodic boundary condition, with the absolute value of Ψ_m unchanged but the phase is shifted by $2\pi p$ when z changes by 2π. With each fixed p, one solves (2.1.8) for a list of discrete eigenvalues $E_m(p)$ and the corresponding Bloch eigenfunctions, marked as $\Psi_m(z,p)$. Here we assume the multiplicity is 1, i.e. each eigenvalue corresponds to one eigenfunction. Note that Ψ_m is a function of z with p regarded either as a parameter or a variable.

We list a few properties as preparation for later analysis[160, 194, 10]:

(a) The eigenvalues $E_m(p)$ are 1-periodic in p. Denote $L^* = \mathbb{Z}$ as the lattice for p. The analysis will be carried on the first Brillouin zone $p \in \left(-\frac{1}{2}, \frac{1}{2}\right)$. Order the eigenvalues as $E_1(p) \le E_2(p) \le ... \le E_m(p) \le ....$ It can be proved that $E_m(p) \to \infty$ as $m \to \infty$, uniformly in p, and that strict inequality can be achieved outside a zero measure set [194].

(b) Pick one period for p, called the first Brillouin zone $\mathcal{B} = \left(-\frac{1}{2}, \frac{1}{2}\right)$, $\{\Psi_m(\cdot,p)\}$ forms a

complete orthonormal basis in $L^2(0, 2\pi)$, i.e.

$$(\Psi_m, \Psi_n) := \int_0^{2\pi} \frac{\mathrm{d}z}{2\pi} \Psi_m(z,p)\bar{\Psi}_n(z,p) = \delta_{mn}. \tag{2.1.9}$$

(c) For all $\phi \in L^2(\mathbb{R})$, one has the following Bloch decomposition:

$$\phi(x) = \sum_m^{\infty} \int_{\mathcal{B}} c_m(p)\Psi_m(x,p)\,\mathrm{d}p \tag{2.1.10}$$

where c_m is the Bloch coefficient: $c_m(p) = \int_{\mathbb{R}} \phi(x)\bar{\Psi}_m(x,p)\,\mathrm{d}x$.

Note: Another way to study this problem is to investigate the following function:

$$\Phi_m^{\alpha}(z,p) = \mathrm{e}^{-\mathrm{i}pz}\,\Psi_m^{\alpha}(z,p). \tag{2.1.11}$$

This definition gives a function that is 2π-periodic in z and 1-periodic in p. $\Phi_m(\cdot,p)$ also form a complete basis in $L^2(\mathbb{R})$ space.

As one can see in the property (c), any L^2 function can be expanded upon the eigenfunctions $\Psi_m(z,p)$. In the absence of the external potential $U(x)$, one could proceed as the following:

- 1st step: Use the expansion (2.1.10) and define the projection onto the mth band: $c_m(t,p) = \int \psi(t,z)\bar{\Psi}_m(z,p)\mathrm{d}z$;
- 2nd step: Multiply $\bar{\Psi}_m(z,p)$ on both sides of the equation (2.1.5) and integrate over z, one gets the evolution of $c_m(t,p)$:

$$\mathrm{i}\varepsilon\partial_t c_m = \int H_0\psi\bar{\Psi}_m\mathrm{d}z = E_m(p)c_m. \tag{2.1.12}$$

  The solution could be explicitly written down: $c_m(t,p) = c_m(0,p)e^{\frac{\mathrm{i}E_m t}{\varepsilon}}$.

- 3rd step: Collect all the c_m at the ending time.

Generally speaking one projects the initial data onto different bands. Since the equation is linear, the projection coefficients evolve independently in time. c_m could be explicit computed, and thus no heavy computation is needed. However, the method above is only true for zero external potential $U(x)$. With a smooth nonzero potential $U(x)$, the analysis could be much more tedious. In fact, in space, we have two scales: x and z, and thus the solution ψ also contains the slow modes in x and the fast fluctuation in z. So we manually split the two scales:

- 1st step: Separate the scaling and write the solution as $\psi = \psi(t, x, z)$. The Schrödinger equation thus is revised:

$$\mathrm{i}\varepsilon\partial_t\psi = \left(-\frac{\Delta_z}{2} + V(z)\right)\psi^\varepsilon + U(x)\psi^\varepsilon - \left(\frac{\varepsilon^2\Delta_x}{2} + \varepsilon\nabla_x\cdot\nabla_z\right)\psi^\varepsilon. \tag{2.1.13}$$

  Note that here we change ∇_x into $\nabla_x + \frac{1}{\varepsilon}\nabla_z$.

- 2nd step: Use the expansion (2.1.10) and define the projection onto the mth band: $c_m(t, p, x) = \int \psi(t, x, z)\bar{\Psi}_m(z, p)\mathrm{d}z$;

- 3rd step: Multiply $\bar{\Psi}_m(z, p)$ on both sides of the equation (2.1.5) and integrate over z, one gets the evolution of $c_m(t, p, x)$:

$$\mathrm{i}\varepsilon\partial_t c_m = -\frac{\varepsilon^2\Delta_x}{2}c_m + \left(E_m(p) + U(x)\right)c_m + \varepsilon\sum_n \int \nabla_x c_n \int_{\mathbb{R}^d} \Psi_n\partial_z\Psi_m\mathrm{d}z\mathrm{d}p. \tag{2.1.14}$$

  The evolution in time has two parts on the right: the first part is given by the first two terms. They resemble the Schrödinger Hamiltonian, containing both the kinetic energy $-\frac{\varepsilon^2\Delta_x}{2}$ and the potential $E_m(p) + U(x)$. One thing to be noted here is that different energy bands provide different potential terms – i.e. different strength of acceleration. The second part is the third term on the right: it is of $\mathcal{O}(\varepsilon)$, and has the Berry connection term involved: $\int \Psi_n\nabla_z\Psi_m\mathrm{d}z$ [12]. It shows the connection between different bands.

- 4th step: Solve (2.1.14) for c_m and collect all the modes at the ending time.

Apparently (2.1.14) is no easier than the original Schrödinger equation: it is not only oscillatory in both time and space, but also shows strong nonlinear coupling among bands.

All in all, when $U(x)$ is present, the solution has two scalings and the standard Bloch-Floquet theory does not simplify the computation. Nevertheless this theory provides us one way to decompose the L^2 function space, and will be used in our Wigner approach.

2.2 Semi-classical model in phase space

In this section we derive a model for the crystal problem in the framework of Wigner transformation. In fact if one follows the standard Wigner transform steps, and define the Wigner equation as in (1.2.1), it is easy to obtain the following Wigner equation:

$$\begin{aligned} \partial_t W^\varepsilon + k\partial_x W^\varepsilon &= \frac{1}{\mathrm{i}\varepsilon}\sum_{\mu\in L^*} \mathrm{e}^{\mathrm{i}\mu x/\varepsilon}\hat{V}(\mu)\left[W^\varepsilon(x,k-\frac{\mu}{2}) - W^\varepsilon(x,k+\frac{\mu}{2})\right] \\ &\quad + \frac{1}{\mathrm{i}\varepsilon}\int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega x}\hat{U}(\omega)\left[W^\varepsilon(x,k-\frac{\varepsilon\omega}{2}) - W^\varepsilon(x,k+\frac{\varepsilon\omega}{2})\right], \end{aligned} \tag{2.2.1}$$

where $\hat{U}(\omega)$ and $\hat{V}(\mu)$ are defined as:

$$\hat{U}(\omega) = \int_{\mathbb{R}} \mathrm{d}y\, \mathrm{e}^{-\mathrm{i}\omega y}U(y), \quad \hat{V}(\mu) = \frac{1}{2\pi}\int_0^{2\pi} \mathrm{d}y\, \mathrm{e}^{-\mathrm{i}\mu y}V(y). \tag{2.2.2}$$

with $\omega \in \mathbb{R}$, and $\mu \in L^*$. Note $V(x)$ is periodic and thus it only has Fourier series instead of Fourier transform.

On the right hand side, the second term is approximated by $\partial_x U \partial_k W$. The first term, however, can be much trickier. This term is a series summation instead of an integral; the deviation μ in k is of $\mathcal{O}(1)$; and the whole term is multiplied with $\frac{1}{\mathrm{i}\varepsilon}$, making it impossible for simplification with standard techniques.

2.2.1 Model derivation

To overcome this difficulty, we use the following two new ideas: 1. we separate the fast variable from the slow one in x, and write function $W(t,x,p)$ as $W(t,x,z,p)$; 2. we found a basis for W to be expand upon in phase space. The model is carried out with asymmetric Wigner transform, and the symmetric version will be briefly summarized in the next section. Define the Wigner function W:

$$W_\varepsilon(t,x,k)=\int_{\mathbb{R}}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}ky}\psi^\varepsilon\left(t,x-\varepsilon y\right)\bar{\psi}^\varepsilon(t,x). \tag{2.2.3}$$

It is easy to obtain the Wigner equation (with the fast variable $z=\frac{x}{\varepsilon}$):

$$\begin{aligned}&\frac{\partial}{\partial t}W_\varepsilon+k\left(\frac{\partial}{\partial x}+\frac{1}{\varepsilon}\frac{\partial}{\partial z}\right)W_\varepsilon+\frac{\mathrm{i}\varepsilon}{2}\left(\frac{\partial}{\partial x}+\frac{1}{\varepsilon}\frac{\partial}{\partial z}\right)^2W_\varepsilon\\=&\frac{1}{\mathrm{i}\varepsilon}\sum_{\mu\in L^*}\mathrm{e}^{\mathrm{i}\mu x/\varepsilon}\hat{V}(\mu)\left[W_\varepsilon(x,k-\mu)-W_\varepsilon(x,k)\right]\\&+\frac{1}{\mathrm{i}\varepsilon}\int_{\mathbb{R}}\frac{\mathrm{d}\omega}{2\pi}\mathrm{e}^{\mathrm{i}\omega x}\hat{U}(\omega)\left[W_\varepsilon(x,k-\varepsilon\omega)-W_\varepsilon(x,k)\right].\end{aligned} \tag{2.2.4}$$

Under the assumption that $U(x)$ is smooth, high order terms are thrown away, leaving:

$$\partial_tW_\varepsilon+k\partial_xW_\varepsilon-\partial_xU\partial_kW_\varepsilon+\mathrm{i}\partial_{xz}^2W_\varepsilon=-\frac{1}{\varepsilon}\mathcal{L}W_\varepsilon, \tag{2.2.5}$$

where the skew symmetric operator $\mathcal{L}$ is given by

$$\mathcal{L}f(z,k)=k\frac{\partial f}{\partial z}+\frac{\mathrm{i}}{2}\frac{\partial^2f}{\partial z^2}-\frac{1}{\mathrm{i}}\sum_{\mu\in L^*}\mathrm{e}^{\mathrm{i}\mu x/\varepsilon}\hat{V}(\mu)\left[f(x,k-\mu)-f(x,k)\right].$$

We use the classical scaling separation argument, asymptotically expand W^ε as:

$$W^\varepsilon(t,x,k)=W_0\left(t,x,\frac{x}{\varepsilon},k\right)+\varepsilon W_1\left(t,x,\frac{x}{\varepsilon},k\right)+\cdots \tag{2.2.6}$$

Plug this ansatz into (2.2.4). By balancing both the $\mathcal{O}(\frac{1}{\varepsilon})$ and $\mathcal{O}(1)$ term, one gets:

$$\mathcal{L}W_0 = 0\,, \tag{2.2.7a}$$

$$\frac{\partial W_0}{\partial t} + k\frac{\partial W_0}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial W_0}{\partial k} + \mathrm{i}\frac{\partial^2 W_0}{\partial x \partial z} = -\mathcal{L}W_1\,. \tag{2.2.7b}$$

The first equation indicates that W_0 is in the kernel of $\mathcal{L}$. To specify the kernel space, and in particular to seek for a good basis of $\ker\mathcal{L}$, Bloch-Floquet theory is adopted [3].

Given two Bloch eigenfunctions Ψ_m and Ψ_n, we define the following:

$$Q_{mn}(z,k) = Q_{mn}(z,\mu_k,p_k) = \int_0^{2\pi} \frac{\mathrm{d}y}{2\pi} \mathrm{e}^{\mathrm{i}ky}\Psi_m(z-y,p_k)\bar{\Psi}_n(z,p_k), \tag{2.2.8}$$

where k is an arbitrary real number and is decomposed as:

$$k = p_k + \mu_k, \quad p_k \in \mathcal{B},\ \mu_k \in L^*. \tag{2.2.9}$$

Note that this definition is very similar to Wigner transformation (1.2.1), but notice the two difference: 1. y is integrated over one period 2π; 2. in the wave function z and y have the same scaling. With this definition, one could prove:

1. Q_{mn} form a basis for W to be expanded upon:

$$W(t,x,p) = \sum_{mn} \sigma_{mn}(t,x,p)Q_{mn}(z,p). \tag{2.2.10}$$

It is a direct corollary of the following lemma:

Lemma 2.2.1. *Define the inner product* $\langle\cdot,\cdot\rangle$*:*

$$\langle f,\, g\rangle := \sum_{\mu\in L^*}\int_0^{2\pi}\frac{\mathrm{d}z}{2\pi} f(z,\mu)\bar{g}(z,\mu), \quad f,\, g \,\in\, L^2\big((0,2\pi),\ell^2(L^*)\big),$$

then for any $p \in (-1/2,\, 1/2)$, $\{Q_{mn}(\cdot,\cdot,p)\}$ *forms a complete orthonormal basis in* $L^2\big((0,2\pi)\times \ell^2(L^*)\big)$.

Proof. The orthonormal condition

$$\langle Q_{mn}, Q_{jl}\rangle \;=\; \delta_{mj}\,\delta_{nl}, \tag{2.2.11}$$

can be proved by simply using (2.1.9). To prove the completeness, it is sufficient to show that: if there exists an $f \in L^2\big((0,2\pi)\times \ell^2(L^*)\big)$, such that $\langle f, Q_{mn}\rangle = 0$ for all $m,n\in\mathbb{N}$, then $f(z,\mu)\equiv 0$. Assume $\langle f, Q_{mn}\rangle = 0$ for all $m,n\in\mathbb{N}$. By the definition of Q_{mn},

$$\sum_{\mu}\int_0^{2\pi}\int_0^{2\pi}\frac{\mathrm{d}y\mathrm{d}z}{(2\pi)^2} f(z,\mu)\mathrm{e}^{\mathrm{i}\mu y}\Phi_m(z-y,p)\bar{\Phi}_n(z,p)=0,\quad \forall m,n\in\mathbb{N}.$$

Since $\{\Phi_n(\cdot,p)\}$ forms a complete orthonormal basis in $L^2(0,2\pi)$, the above equality implies that

$$\sum_{\mu}\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi} f(z,\mu)\mathrm{e}^{\mathrm{i}\mu y}\Phi_m(z-y,p)=0,\quad \forall m\in\mathbb{N},$$

thus

$$\sum_{\mu} f(z,\mu)\mathrm{e}^{\mathrm{i}\mu y}\equiv 0,$$

which implies that $f(z,\mu)\equiv 0$. □

2. A straightforward computation gives:

$$\mathcal{L}Q_{mn}(z,k)=\mathcal{L}Q_{mn}(z,\mu,p)=\mathrm{i}(E_m(p)-E_n(p))Q_{mn}(z,\mu,p)\,. \tag{2.2.12}$$

Apparently the kernel of $\mathcal{L}$ highly relies on the structure of energy bands. Before carrying out any further derivations, we make the following assumption.

Assumption 2.2.1. *Eigenvalues are strictly apart from each other everywhere in p, namely* $E_1(p) < E_2(p) < ... < E_j(p) < ....$

This assumption may be valid or not, and depending on its validity, we call the scenario adiabatic case or non-adiabatic case, and derive two different models respectively, as shown below. Associated with it, we also compute the initial data.

Adiabatic case

We derive the model for adiabatic case in this part. With the assumption, (2.2.12) clearly indicates:

$$\ker\mathcal{L} = \text{span}\{Q_{mm}, m = 1, 2, \cdots\}. \tag{2.2.13}$$

According to the leading order equation (2.2.7a), it is easy to see that W_0 should be in the kernel of $\mathcal{L}$, and thus is simply an expansion of Q_{mm}:

$$W_0(t, x, z, k) = \sum_m \sigma_{mm}(t, x, p) Q_{mm}(z, \mu, p). \tag{2.2.14}$$

Here σ_{mm} representing the expansion coefficients. To close it, we plug the expansion (2.2.14) back into the second order equation (2.2.7b)), and take the inner product with Q_{mm} on both sides. The right hand side vanishes due to the skew symmetry of $\mathcal{L}$ and (2.2.12):

$$\langle -\mathcal{L}W_1, Q_{mm}\rangle = \langle -W_1, \mathcal{L}Q_{mm}\rangle = 0. \tag{2.2.15}$$

The left hand side, on the other hand, gives:

$$\begin{aligned}
& \langle \partial_t W_0 + k\partial_x W_0 - \partial_x U \partial_k W_0 + \mathrm{i}\partial_{xz} W_0,\, Q_{mm} \rangle \\
= \;& \sum_n \partial_t \sigma_{nn} \langle Q_{nn},\, Q_{mm} \rangle + \sum_{nn} \partial_x \sigma_{nn} \langle k Q_{nn},\, Q_{mm} \rangle \\
& -\partial_x U \left(\sum_n \partial_p \sigma_{nn} \langle Q_{nn},\, Q_{mm} \rangle + \sum_{nn} \sigma_m \langle \partial_p Q_{nn},\, Q_{mm} \rangle \right) \\
= \;& \partial_t \sigma_{mm} - \mathrm{i} \left(\partial_z \Psi_n,\, \Psi_m \right) \partial_x \sigma_{mm} - \partial_x U \partial_p \sigma_{mm} - \partial_x U \sum_n \big(\left(\partial_p \Phi_n,\, \Phi_m \right) + \left(\Phi_n,\, \partial_p \Phi_m \right) \big) \\
= \;& \partial_t \sigma_{mm} + \partial_p E_m \partial_x \sigma_{mm} - \partial_x U \partial_p \sigma_{mm} \,. \qquad (2.2.16)
\end{aligned}$$

By combining (2.2.15) and (2.2.16), one gets transport equation on each band:

$$\partial_t \sigma_{mm} + \partial_p E_m \partial_x \sigma_{mm} - \partial_x U \partial_p \sigma_{mm} = 0. \qquad (2.2.17)$$

In the derivation of (2.2.16), the following equalities were used:

$$\langle k Q_{mn}, Q_{jl} \rangle \;=\; -\frac{\mathrm{i}}{2} \big(\delta_{nl} \left(\partial_z \Psi_m, \Psi_j \right) \;+\; \delta_{mj} \left(\partial_z \Psi_l, \Psi_n \right) \big), \qquad (2.2.18a)$$

$$\langle \partial_p Q_{mn}, Q_{jl} \rangle \;=\; \delta_{nl} \left(\partial_p \Phi_m, \Phi_j \right) \;+\; \delta_{mj} \left(\Phi_l, \partial_p \Phi_n \right), \qquad (2.2.18b)$$

$$\langle \partial_z Q_{mn}, Q_{jl} \rangle \;=\; \delta_{nl} \left(\partial_z \Psi_m, \Psi_j \right) \;-\; \delta_{mj} \left(\partial_z \Psi_l, \Psi_n \right), \qquad (2.2.18c)$$

$$\partial_p E_m \delta_{mj} + \left(\partial_p \Phi_m \,,\, \Phi_j \right) \left(E_m - E_j \right) = -\mathrm{i} \left(\partial_z \Psi_m \,,\, \Psi_j \right), \qquad (2.2.18d)$$

$$\left(\partial_p \Phi_m, \Phi_m \right) + \left(\Phi_m, \partial_p \Phi_m \right) = 0. \qquad (2.2.18e)$$

These are simple derivations and the details are omitted.

(2.2.17) is a very important resemblance to the classical Liouville equation (1.2.6) as it also shows clear connection towards the classical mechanics. The Hamiltonian is $H = H_0 + U(x)$, and on each band, the effective potential is $E_m(p) + U(x)$. Following the trajectory, one gets $\dot{x} = \partial_p H = \partial_p E_m(p)$ and $\dot{k} = -\partial_x H = -\partial_x U(x)$. The configuration of the energy bands

apparently controls the momentum.

Remark 2.2.1. *Results for the case that has no external potential were rigorously proved in [151, 74] in the framework of Wigner series, instead of Wigner transform.*

One could also write $W = \sum_{mn} \sigma_{mn} Q_{mn}$, but Q_{mn} with $m \neq n$ only appears in W_1, meaning that $\sigma_{mn} \sim \mathcal{O}(\varepsilon)$, negligible in the semi-classical regime.

Non-adiabatic case

In this part we derive the model for the system when **Assumption 2.2.1** is turned off. In fact, when two adjacent energy get very close $E_m - E_{m-1} < \mathcal{O}(\varepsilon)$, or even $\sim \mathcal{O}(\varepsilon)$, the argument above is no longer valid. With $\mathcal{L}Q_{m,m-1} = \mathcal{O}(\varepsilon)$ in (2.2.6), the previous scaling separation (2.2.7) is no longer correct, and $\sigma_{m,m-1}$ should not be eliminated in the leading order expansion. When σ_{mn} with $m \neq n$ acquires $\mathcal{O}(1)$ value, we say quantum tunnelling is observed.

To get the evaluation of these quantities, ones use the full expansion (2.2.10). We plug (2.2.10) into (2.2.5) and take the inner product with Q_{mn}. The derivation is tedious, and here we only list the results.

Without loss of generality, we tackle a two-band problem. Define $p_c = \arg\min_p\{|E_1 - E_2|\}$ and assume $|E_1(p_c) - E_2(p_c)| = \mathcal{O}(\varepsilon)$ at $p_c = 0$.

$$\begin{aligned} & \frac{\partial \sigma_{11}}{\partial t} + \frac{\partial E_1}{\partial p}\frac{\partial \sigma_{11}}{\partial x} + \frac{1}{\mathrm{i}}\left(\frac{\partial \Psi_1}{\partial z}\,,\,\Psi_2\right)\frac{\partial \sigma_{12}}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial \sigma_{11}}{\partial p} \\ = \; & \frac{\partial U}{\partial x}\left[\left(\frac{\partial \Phi_2}{\partial p}\,,\,\Phi_1\right)\sigma_{21} + \left(\Phi_1\,,\,\frac{\partial \Phi_2}{\partial p}\right)\sigma_{12}\right], \end{aligned}$$

$$\begin{aligned} & \frac{\partial \sigma_{22}}{\partial t} + \frac{\partial E_2}{\partial p}\frac{\partial \sigma_{22}}{\partial x} + \frac{1}{\mathrm{i}}\left(\frac{\partial \Psi_2}{\partial z}\,,\,\Psi_1\right)\frac{\partial \sigma_{21}}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial \sigma_{22}}{\partial p} \\ = \; & \frac{\partial U}{\partial x}\left[\left(\frac{\partial \Phi_1}{\partial p}\,,\,\Phi_2\right)\sigma_{12} + \left(\Phi_2\,,\,\frac{\partial \Phi_1}{\partial p}\right)\sigma_{21}\right], \end{aligned}$$

$$
\begin{aligned}
&\frac{\partial \sigma_{12}}{\partial t} + \frac{\partial E_2}{\partial p}\frac{\partial \sigma_{12}}{\partial x} + \frac{1}{\mathrm{i}}\left(\frac{\partial \Psi_2}{\partial z}\,,\,\Psi_1\right)\frac{\partial \sigma_{11}}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial \sigma_{12}}{\partial p} + \mathrm{i}\frac{E_1 - E_2}{\varepsilon}\sigma_{12} \\
= \;& \frac{\partial U}{\partial x}\left[\left(\Phi_2\,,\,\frac{\partial \Phi_1}{\partial p}\right)\sigma_{11} + \left(\frac{\partial \Phi_2}{\partial p}\,,\,\Phi_1\right)\sigma_{22} + \left(\Phi_2\,,\,\frac{\partial \Phi_2}{\partial p}\right)\sigma_{12} + \left(\frac{\partial \Phi_1}{\partial p}\,,\,\Phi_1\right)\sigma_{12}\right],
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\partial \sigma_{21}}{\partial t} + \frac{\partial E_1}{\partial p}\frac{\partial \sigma_{21}}{\partial x} + \frac{1}{\mathrm{i}}\left(\frac{\partial \Psi_1}{\partial z}\,,\,\Psi_2\right)\frac{\partial \sigma_{22}}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial \sigma_{21}}{\partial p} + \mathrm{i}\frac{E_2 - E_1}{\varepsilon}\sigma_{21} \\
= \;& \frac{\partial U}{\partial x}\left[\left(\Phi_1\,,\,\frac{\partial \Phi_2}{\partial p}\right)\sigma_{22} + \left(\frac{\partial \Phi_1}{\partial p}\,,\,\Phi_2\right)\sigma_{11} + \left(\Phi_1\,,\,\frac{\partial \Phi_1}{\partial p}\right)\sigma_{21} + \left(\frac{\partial \Phi_2}{\partial p}\,,\,\Phi_2\right)\sigma_{21}\right].
\end{aligned}
$$

This system can be written in vector form as:

$$
\partial_t \boldsymbol{\sigma} + A\partial_x \boldsymbol{\sigma} + B\partial_p \boldsymbol{\sigma} = -BC\boldsymbol{\sigma} + \frac{\mathrm{i}D}{\varepsilon}\boldsymbol{\sigma} \tag{2.2.19a}
$$

where

$$
\boldsymbol{\sigma} = (\;\sigma_{11}\quad \sigma_{12}\quad \sigma_{21}\quad \sigma_{22}\;)^T \tag{2.2.19b}
$$

$$
B = -\partial_x U\,\mathbb{I}\,,\; D = \mathrm{diag}\,(\,0,\,E_2 - E_1,\,E_1 - E_2,\,0\,) \tag{2.2.19c}
$$

$$
A = \begin{bmatrix}
\partial_p E_1 & \frac{1}{2}\psi_{12} & \frac{1}{2}\psi_{21} & 0 \\
\frac{1}{2}\psi_{21} & \partial_p \frac{E_1+E_2}{2} & 0 & \frac{1}{2}\psi_{21} \\
\frac{1}{2}\psi_{12} & 0 & \partial_p \frac{E_1+E_2}{2} & \frac{1}{2}\psi_{12} \\
0 & \frac{1}{2}\psi_{12} & \frac{1}{2}\psi_{21} & \partial_p E_2
\end{bmatrix}, \tag{2.2.19d}
$$

$$
C = \begin{bmatrix}
0 & -\phi_{12} & \phi_{21} & 0 \\
-\phi_{21} & \phi_{11} - \phi_{22} & 0 & \phi_{21} \\
\phi_{12} & 0 & \phi_{22} - \phi_{11} & -\phi_{12} \\
0 & \phi_{12} & -\phi_{21} & 0
\end{bmatrix}, \tag{2.2.19e}
$$

$$\psi_{mn} = -\mathrm{i}\left(\partial_z \Psi_m, \Psi_n\right), \text{ and } \phi_{mn} = \left(\partial_p \Phi_m, \Phi_n\right), \tag{2.2.19f}$$

Superscript T stands for matrix transpose and $\mathbb{I}$ is the identity matrix. Generally, $\psi_{12} = \bar{\psi}_{21}$, $\phi_{12} = -\bar{\phi}_{21}$ are complex-valued quantities, and ϕ_{11}, ϕ_{22} are purely imaginary, so $A = A^\dagger$ is Hermitian, and $C = -C^\dagger$ is anti-Hermitian. The superscript $\dagger$ denotes the matrix conjugate transpose.

Initial data for the semi-classical model

One also needs to equip it with appropriate initial condition. Choose the initial data of the Schrödinger equation as two wave packets along the two Bloch bands in the following form [31, 105]:

$$\phi_I = a_1(x)\,\Phi_1\left(\frac{x}{\varepsilon}, \partial_x S_0(x)\right)\,\mathrm{e}^{\mathrm{i}S_0(x)/\varepsilon} + a_2(x)\,\Phi_2\left(\frac{x}{\varepsilon}, \partial_x S_0(x)\right)\,\mathrm{e}^{\mathrm{i}S_0(x)/\varepsilon}. \tag{2.2.20}$$

Then, the initial data of the Wigner function, for $\varepsilon \ll 1$, has the approximation:

$$\begin{aligned} W_I(x,z,k) \quad \sim \quad & |a_1(x)|^2 W_{11}(z,k) + |a_2(x)|^2 W_{22}(z,k) \\ & + a_1(x)a_2(x)\big(W_{12}(z,k) + W_{21}(z,k)\big), \end{aligned} \tag{2.2.21}$$

with

$$W_{mn}(z,k) = \int_{\mathbb{R}} \frac{\mathrm{d}y}{2\pi} \mathrm{e}^{\mathrm{i}ky} \Phi_m(z-y, \partial_x S_0(x-\varepsilon y)) \bar{\Phi}_n(z, \partial_x S_0(x)) \mathrm{e}^{\mathrm{i}(S_0(x-\varepsilon y) - S_0(x))/\varepsilon}.$$

Using Taylor expansion on $S_0(x-\varepsilon y) - S_0(x)$ and $\Phi_m(z-y, \partial_x S_0(x-\varepsilon y))$, one gets

$$W_{mn}(z,k) = \int_{\mathbb{R}} \frac{\mathrm{d}y}{2\pi} \mathrm{e}^{\mathrm{i}(k - \partial_x S_0(x))y} \Phi_m(z-y, \partial_x S_0(x)) \bar{\Phi}_n(z, \partial_x S_0(x)) + \mathcal{O}(\varepsilon),$$

then by ignoring the $\mathcal{O}(\varepsilon)$ term, and using the periodicity of $\Phi_m(z,p)$ on z, one can change the integral into a summation of integrals from 0 to 2π:

$$W_{mn}(z,k)=\sum_{\nu\in L}\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}(k-\partial_x S_0(x))(y+\nu)}\Phi_m(z-y,\partial_x S_0(x))\bar{\Phi}_n(z,\partial_x S_0(x)).$$

Applying the equality

$$\sum_{\nu\in L}\mathrm{e}^{\mathrm{i}k\nu}=\sum_{\mu\in L^*}\delta(k+\mu),$$

one gets

$$\begin{aligned}
W_{mn}(z,k) &= \sum_{\mu\in L^*}\delta(k+\mu-\partial_x S_0)\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}(k-\partial_x S_0)y}\Phi_m(z-y,\partial_x S_0)\bar{\Phi}_n(z,\partial_x S_0)\\
&= \sum_{\mu\in L^*}\delta(p_k+\mu-\partial_x S_0)\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}ky}\Psi_m(z-y,\partial_x S_0)\bar{\Psi}_n(z,\partial_x S_0)\\
&= \sum_{\mu\in L^*}\delta(p_k+\mu-\partial_x S_0)\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}ky}\Psi_m(z-y,p_k+\mu)\bar{\Psi}_n(z,p_k+\mu)\\
&= \sum_{\mu\in L^*}\delta(p_k+\mu-\partial_x S_0)\int_0^{2\pi}\frac{\mathrm{d}y}{2\pi}\mathrm{e}^{\mathrm{i}ky}\Psi_m(z-y,p_k)\bar{\Psi}_n(z,p_k).\\
&= \sum_{\mu\in L^*}\delta\big(p_k+\mu-\partial_x S_0\big)\,Q_{mn}(z,\mu_k,p_k). \qquad (2.2.22)
\end{aligned}$$

In the derivation above, from the second line to the third line, we use the fact that

$$\int\delta(p-p_0)f(p_0)g(p)\,\mathrm{d}p=f(p_0)g(p_0)=\int\delta(p-p_0)f(p)g(p)\,\mathrm{d}p$$

to replace the argument $\partial_x S_0$ by $p_k+\mu$.

Without loss of generality, we assume that $\partial_x S_0\in(-1/2,\,1/2)$, then (2.2.22) becomes

$$W_{mn}(z,k)=\delta\big(p_k-\partial_x S_0\big)\,Q_{mn}(z,\mu_k,p_k). \qquad (2.2.23)$$

Compare (2.2.23) with (2.2.21), one has the initial data for $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma}(0,x,p) = \delta\left(p - \partial_x S_0(x)\right) \begin{pmatrix} a_1^2 & a_1 a_2 & a_1 a_2 & a_2^2 \end{pmatrix}^T. \tag{2.2.24}$$

2.2.2 Model derivation – symmetric case

This section follows the routine in section 2.2, with the standard symmetric Wigner transform:

$$W_\varepsilon^s(t,x,k) = \int_{\mathbb{R}} \frac{\mathrm{d}y}{2\pi} \mathrm{e}^{\mathrm{i}ky} \phi^\varepsilon\left(t, x - \frac{\varepsilon y}{2}\right) \bar{\phi}^\varepsilon\left(t, x + \frac{\varepsilon y}{2}\right).$$

The derivation is similar, thus we skip the details, and give a list of the results:

1. The Wigner equation corresponding (2.2.4) is

$$\begin{aligned} \frac{\partial W_\varepsilon^s}{\partial t} + k\frac{\partial W_\varepsilon^s}{\partial x} =& \frac{1}{\mathrm{i}\varepsilon} \sum_{\mu \in L^*} \mathrm{e}^{\mathrm{i}\mu x/\varepsilon} \hat{V}(\mu) \left[W_\varepsilon^s\left(x, k - \frac{\mu}{2}\right) - W_\varepsilon^s\left(x, k + \frac{\mu}{2}\right)\right] \\ &+ \frac{1}{\mathrm{i}\varepsilon} \int_{\mathbb{R}} \frac{\mathrm{d}\omega}{2\pi} \mathrm{e}^{\mathrm{i}\omega x} \hat{U}(\omega) \left[W_\varepsilon^s\left(x, k - \frac{\varepsilon\omega}{2}\right) - W_\varepsilon^s\left(x, k + \frac{\varepsilon\omega}{2}\right)\right]. \end{aligned}$$

2. Corresponding to the asymptotic Wigner equation for asymmetrical transformation (2.2.5), one has:

$$\frac{\partial W_\varepsilon^s}{\partial t} + k\frac{\partial W_\varepsilon^s}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial W_\varepsilon^s}{\partial k} = -\frac{1}{\varepsilon}\mathcal{L}^s W_\varepsilon^s, \tag{2.2.25}$$

where the skew symmetric operator $\mathcal{L}^s$ is given by

$$\mathcal{L}^s f(z,k) = k\frac{\partial f}{\partial z} - \frac{1}{\mathrm{i}} \sum_{\mu \in L^*} \mathrm{e}^{\mathrm{i}\mu x/\varepsilon} \hat{V}(\mu) \left[f(x, k - \mu/2) - f(x, k + \mu/2)\right].$$

3. Corresponding to (2.2.7), one has the $\mathcal{O}\left(\frac{1}{\varepsilon}\right)$ and $\mathcal{O}(1)$ expansions:

$$\mathcal{L}^s W_0^s = 0, \tag{2.2.26a}$$

$$\frac{\partial W_0^s}{\partial t} + k\frac{\partial W_0^s}{\partial x} - \frac{\partial U}{\partial x}\frac{\partial W_0^s}{\partial k} = -\mathcal{L}^s W_1^s \,. \tag{2.2.26b}$$

4. Same as in (2.2.8), one has the following symmetrical definition for Q_{mm}

$$\begin{aligned} Q_{mn}^s(z,k) &= Q_{mn}^s(z,\mu,p) \\ &= \int_0^{2\pi} \frac{\mathrm{d}y}{2\pi} \mathrm{e}^{\mathrm{i}(p+\mu)y} \Psi_m\left(z-\frac{y}{2},p\right) \bar{\Psi}_n\left(z+\frac{y}{2},p\right). \end{aligned}$$

They are eigenfunctions of $\mathcal{L}^s$

$$\mathcal{L}^s Q_{mn}^s = \mathrm{i}(E_m - E_n) Q_{mn}^s \,.$$

5. If the eigenvalues $\{E_n\}$ are well separated, i.e. $E_m \neq E_n$, for $m \neq n$, the solution to (2.2.26a) is:

$$W_0^s = \sum_m \sigma_{mm}^s Q_{mm}^s.$$

By taking the inner product with Q_{mm}^s on both sides of (2.2.26b), one obtains the same classical Liouville equations for σ_{mm}^s as in (2.2.17)

$$\partial_t \sigma_{mm}^s + \partial_p E_m \partial_x \sigma_{mm}^s - \partial_x U \partial_p \sigma_{mm}^s = 0.$$

6. If some bands touch at point p_c, the solution to (2.2.25) is given by:

$$W_\varepsilon^s = \sum_m \sigma_{mm} Q_{mm}^s + \sum_{m\neq n} \sigma_{mn} Q_{mn}^s.$$

In the two-band case, $m, n = 1, 2$, then σ^s_{mn} is governed by,

$$\partial_t \boldsymbol{\sigma}^s + A^s \partial_x \boldsymbol{\sigma}^s + B^s \partial_p \boldsymbol{\sigma}^s = -B^s C^s \boldsymbol{\sigma}^s + \frac{iD^s}{\varepsilon} \boldsymbol{\sigma}^s, \tag{2.2.27}$$

where $\boldsymbol{\sigma}^s = (\, \sigma^s_{11}\ \sigma^s_{12}\ \sigma^s_{21}\ \sigma^s_{22}\,)^T$, then $B^s = B$, $C^s = C$ and $D^s = D$ are the same as the asymmetric case, while A^s is given by

$$A^s = \begin{bmatrix} \partial_p E_1 & \frac{1}{2}\psi_{12} & \frac{1}{2}\psi_{21} & 0 \\ \frac{1}{2}\psi_{21} & \partial_p \frac{E_1+E_2}{2} & 0 & \frac{1}{2}\psi_{21} \\ \frac{1}{2}\psi_{12} & 0 & \partial_p \frac{E_1+E_2}{2} & \frac{1}{2}\psi_{12} \\ 0 & \frac{1}{2}\psi_{12} & \frac{1}{2}\psi_{21} & \partial_p E_2 \end{bmatrix}$$

and ψ_{mn}, ϕ_{mn} are given by (2.2.19f). Noted that $A^s = (A^s)^\dagger$.

We call this new system obtained by the symmetric Wigner transform (2.2.27) the Liouville-S system. Apparently the only difference from the previous one, named Liouville-A lies in the transport matrices $A \neq A^s$. Despite that, it could be easily shown that the two share the same weak limit as $\varepsilon \to 0$ (see Appendix A.1 for detail). This formally suggests that the behavior of $\sigma_{11},\ \sigma_{22}$ and $\sigma^s_{11},\ \sigma^s_{22}$ in the two models is about the same. This observation is confirmed by numerical results in Section 2.3.1.

2.2.3 Main properties of the model

The system for the non-adiabatic case we derived (2.2.19) has two transport coefficients A and B, and C and D are the two coupling terms. We show below how Berry phase get involved in this system, its hyperbolicity, and its connection with the model for the adiabatic case (2.2.17).

Berry phase

In this part we aim at explaining the role played by Berry phase in our model. It has been known for long that the second order correction to the semi-classical trajectory should be related to Berry phase, see physics argument in [21, 180, 197, 198, 12] and mathematical justification in [163, 162, 57]. Our model incorporate this phenomenon in a very subtle way as well: the Berry connection term explicitly appears in the coupling term on the right hand side. To compute the Berry phase, we fix one particle and follow its trajectory in momentum space. Adiabatic theory claims that the particle moves along the energy band it starts on, as well as it is away from the crossing zone, and thus the wave function associated to this particle should simply be $\Psi_m(p(t))$ with a to be determined phase shift. Here $p(t)$ is the trajectory with $p(0) = p_i$ [180]:

$$\psi_m = \Psi_m(p_i) \to \psi_m(t) = e^{\mathrm{i}\theta_m(t)}\Psi_m(p(t)), \tag{2.2.28}$$

with

$$\theta_m(t) = -\frac{1}{\varepsilon}\int_0^t E_m(p(t'))\mathrm{d}t' + \mathrm{i}\int_0^t \langle\Psi_m(p(t'))|\partial_t|\Psi_m(p(t'))\rangle\mathrm{d}t'. \tag{2.2.29}$$

The first term is called dynamical phase factor and the second term is the so-called Berry phase term. We change the variable and obtain:

$$\gamma_m = \mathrm{i}\int_0^t \langle\Psi_m(p(t'))|\partial_t|\Psi_m(p(t'))\rangle\mathrm{d}t = \mathrm{i}\int_{p_i}^p \langle\Psi_m(p')|\partial_p|\Psi_m(p')\rangle\mathrm{d}p' = \mathrm{i}\int \phi_{mm}\mathrm{d}p'. \tag{2.2.30}$$

Remark 2.2.2. *Note the phase shifting here is related to but not the same thing as c_m appeared before. c_m is a function on time given Ψ_m is static function on x. In this picture, however, we follow the trajectory of a particle and at each time point, we only evaluate Ψ_m at one p, and thus Ψ_m is treated as a time-dependent function. In fact, in this case, it is named as instantaneous eigenstates.*

It is easy to check, the phase change of the off-diagonal element of the density matrix is determined by this Berry phase term, that is, we write out the density matrix:

$$\bar{\psi}_m(p(t))\psi_n(p(t)) = e^{-\mathrm{i}(\theta_m(t)-\theta_n(t))}\bar{\Psi}_m(p(t))\Psi_n(p(t)), \tag{2.2.31}$$

and the time change of the phase shift is:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\theta_m - \theta_n) = -\mathrm{i}\partial_x U(\phi_{mm} - \phi_{nn}) - \frac{1}{\varepsilon}(E_m - E_n). \tag{2.2.32}$$

This is exactly the natural frequency of the off-diagonal term appeared in the coupling term in (2.2.19).

Remark 2.2.3. *Note that if $m = n$, $\theta_m - \theta_n = 0$, and this piece of information is not seen.*

Hyperbolicity

The semi-classical Liouville system (2.2.19) is hyperbolic. To prove that we use the following compact form:

$$\partial_t \boldsymbol{\sigma} + A\partial_x \boldsymbol{\sigma} - U_x \partial_p \boldsymbol{\sigma} = S\boldsymbol{\sigma} \tag{2.2.33}$$

where

$$S = C + \frac{\mathrm{i}D}{\varepsilon}\ ,$$

and $\boldsymbol{\sigma}$, A, C, D are defined in (2.2.19). Hyperbolicity is easily seen since:

- $A = A^\dagger$ is a Hermitian matrix, and can be diagonalized;
- $S = -S^\dagger$ is a skew-Hermitian matrix.

Connection of the two models

The two models derived for adiabatic case (2.2.17) and non-adiabatic case (2.2.19) in 2.2.1 are intrinsically connected. In spite of the fact that these two models are for two different scenarios, the underlying methodology is the same: expand Wigner function in terms of Q_{mn} and derive the associated equation. In fact, (2.2.19) is obtained upon projecting W onto the entire basis, and thus could be regarded as the full system for a general setting. When one applies **Assumption 2.2.1** onto it, and studies the behavior of the solution away from p_c where E_1 and E_2 are well-separated, it should provide a picture that is consistent with (2.2.17). In fact we could formally show that the $\frac{\mathrm{i}}{\varepsilon}$ terms for the transition coefficients σ_{12} and σ_{21} in the system lead to high oscillations, and weakly, as $\varepsilon \to 0$, the system formally goes to the classical one (2.2.17).

To argue that the transition coefficients are weakly zero, and that they have zero influence on σ_{mm} in the weak sense, we assume that the initial data for the transition coefficients are all zero, and that U_x does not change sign in time for all x, take $-U_x > 0$ for example, then:

Case 1. If $p \ll -C_0\sqrt{\varepsilon}$ for $C_0 = \mathcal{O}(1)$, then σ_{12} and σ_{21} are of $o(\sqrt{\varepsilon})$;

Case 2. If $-C_0\sqrt{\varepsilon} < p < C_0\sqrt{\varepsilon}$, then σ_{12} and σ_{21} are of $\mathcal{O}(\sqrt{\varepsilon})$, and σ_{12} and σ_{21} are slowly varying, i.e. $\partial_t\sigma_{12} \ll \mathcal{O}\left(\frac{1}{\varepsilon}\right)$, and $\partial_t\sigma_{21} \ll \mathcal{O}\left(\frac{1}{\varepsilon}\right)$;

Case 3. If $p \gg C_0\sqrt{\varepsilon}$, σ_{12} and σ_{21} are highly oscillatory with mean 0.

Apparently Case 1 and 3 naturally unveil the connection between the two models.

The justification is tedious, and we show the proof for a simpler model that maintains the essential properties of the system in Appendix A.1.

Remark 2.2.4. *The assumption $\sigma_{mn}(t=0) = 0$ $(m \neq n)$ is a reasonable assumption. In fact, given arbitrary initial condition, one could check that away from p_c, the weak limit of σ_{mn} is*

always zero, as $\varepsilon \to 0$, as can be seen in Appendix A.1. So numerically we treat the initial data for both σ_{12} and σ_{21} zero, given the initial velocity $\partial_x S_0(x)$ away from the crossing point, i.e.

$$\boldsymbol{\sigma}(0,x,p) \approx \delta\left(p - \partial_x S_0(x)\right) \begin{pmatrix} a_1^2 & 0 \\ 0 & a_2^2 \end{pmatrix}^T, \quad \text{if} \quad \partial_x S_0(x) \neq p_c. \tag{2.2.34}$$

This assumption is intuitive and empirical, but it does give us some convenience in solving the Liouville model numerically. In fact, the numerical examples provided later indeed show that the band-to-band transition is captured very well with initial data (2.2.34).

2.2.4 Numerical method

With the models and their properties shown above, we are capable designing numerical methods. The full system (2.2.19) is comparably much more difficult to compute than (2.2.17) but the latter one fails to be valid in the transition zone. Thus one needs a numerical method that takes benefits from both, so that the right transition rates could be captured with low computational cost. To do that, we make use of the fact shown above that the two models are intrinsically connected:

- the classical Liouville is an approximation (in weak limit) to the full Liouville system away from the crossing point;
- σ_{12} and σ_{21} are slowly varying in a neighborhood of $p_c = 0$, with a small amplitude before the characteristic hitting p_c, and a rapid oscillation after that.

Heuristically we propose a domain decomposition method, which is: away from p_c, when the classical Liouville equation (2.2.17) is a good approximation, we solve this set of equations, but around p_c, in the $\mathcal{O}(\varepsilon)$ neighbourhood, we switch back to the full model. The gain is obvious: numerically it is much easier and more efficient to solve the classical adiabatic system, and thus this approach saves a great amount of computational cost than solving the full system

everywhere. The details are the following:

Given a fixed spatial point x, the sign of $-U_x$ determines the traveling direction of wave in p. Assume $-U_x > 0$, i.e. the wave we study is right-going:

Classical regions: $p < -C_0\sqrt{\varepsilon}$ and $p > C_0\sqrt{\varepsilon}$: In this region, coarse mesh independent on ε is used to solve the classical Liouville system (2.2.17), and σ_{12} and σ_{21} are set to be zero.

Semi-classical region: $p \in \mathcal{B} \setminus \{\text{Classical region}\}$: Solve the full Liouville system (2.2.19). The incoming boundary conditions for σ_{12} and σ_{21} are set to be zero, and the incoming boundary condition for σ_{11} and σ_{22} are the inflow boundary condition. A fine mesh is used with Δx and Δt much less than $\sqrt{\varepsilon}$.

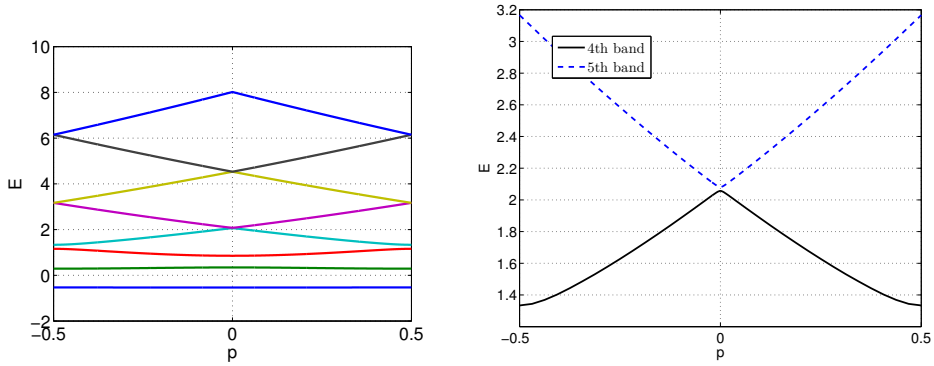
In case of $-U_x < 0$, and the wave if left-going, boundary condition can be set up in the same way. The domain decomposition is carried out on p space, and thus U_x can be regarded as constant for each x.

Remark 2.2.5. *Our analysis is based on the regularity of the coefficient matrix C in the Liouville system. But usually, the value of C's element can be of $\mathcal{O}(1/\delta)$ where $\delta = \min_p |E_1(p) - E_2(p)|$ is the minimal band gap. So C will be large if the minimal band gap δ is small, and the numerical discretization in the semi-classical region should resolve this small parameter δ. In the interested regime $\varepsilon \sim \delta^2$, $o(\sqrt{\varepsilon})$ mesh is enough.*

2.3 Numerical example

In this section we show some numerical examples. Section 2.3.1 is for linear external potential. In this special case, the solution to (2.2.19) could be explicitly expressed, and thus numerically is treated differently. In section 2.3.2 we deal with a nonlinear external potential $U(x)$ using the method invented above in section 2.2.4.

In both examples we use the Mathieu model, i.e. the periodic potential is $V(z) = \cos z$. The first eight Bloch eigenvalues are shown in Figure 2.3.1. Apparently, some eigenvalues get



(a) (b)

Figure 2.3.1: The eigenvalues of the Mathieu Model, $V(x) = \cos x$.

very close to each other around $p = 0,\ \pm 0.5$.

We will focus on the 4th and 5th bands[1]. Denote Ψ_1 and Ψ_2 as the Bloch functions corresponding to the 4th and 5th bands respectively.

For comparison, we will compare the numerical results to the ones given by the original Schrödinger equation, computed through the methods given in [105] with mesh size and time step much smaller than ε.

2.3.1 Linear $U(x)$

We deal with the linear external potential in this section:

$$U(x) = U_0 - \beta x.$$

[1]The minimum gap between the 4th and 5th bands is 0.0247, located at $p = 0$. The gap is small enough so that the quantum effect can be seen for the ε being used.

Then the Schrödinger equation is:

$$\mathrm{i}\varepsilon\psi_t^\varepsilon = -\frac{\varepsilon^2}{2}\psi_{xx}^\varepsilon + \left[\cos\left(\frac{x}{\varepsilon}\right) + (U_0 - \beta x)\right]\psi^\varepsilon, \tag{2.3.1}$$

with the initial data given as a wave packet along the 4th band:

$$\psi_I = a_0(x)\,\Phi_1\left(\frac{x}{\varepsilon}, \partial_x S_0(x)\right)\,\mathrm{e}^{\mathrm{i}S_0(x)/\varepsilon},\ \text{with } S_0(x) = p^* x. \tag{2.3.2}$$

Correspondingly, the Liouville-A becomes:

$$\partial_t \boldsymbol{\sigma} + \beta\partial_p \boldsymbol{\sigma} = \mathcal{R}\boldsymbol{\sigma}, \tag{2.3.3a}$$

$$\boldsymbol{\sigma}(0,x,p) = \boldsymbol{\sigma}_I(x,p) = |a_0(x)|^2\delta\left(p - \partial_x S_0(x)\right)\begin{pmatrix}1 & 0 & 0 & 0\end{pmatrix}^T, \tag{2.3.3b}$$

with $\mathcal{R}$ given by:

$$\mathcal{R} = -\beta C + \frac{\mathrm{i}D}{\varepsilon} - A\partial_x. \tag{2.3.4}$$

To numerically compute it, the difficulty is two fold. Firstly, the solution is highly oscillatory, and typical numerical method calls for very dense mesh that generates a great amount of computational cost; and secondly, the initial data is a delta function and numerically it can be tricky to solve. To resolve the first issue, we rely on the fact that β is a constant and characteristic could be explicitly expressed, following which, the main part of the numerical integration, the solution to the equation, could be repeated used, and is thus evaluated only once at the beginning of computation, as will be more clear later on. To better interpret the delta function in the initial data, instead of following the classical approach: i.e. using a Gaussian function with small variance to approximate the delta function, with the error is related to this variance, we will adopt a singularity decomposition idea invented in [114].

The computation is done for the Liouville-A system. Liouville-S can be computed in exactly

the same way.

A Fourier transform based integration method

We define the Fourier transform of $f(t,x,p)$ with respect to x as:

$$\hat{f}(t,\eta,p) = \int_{\mathbb{R}} e^{-\mathrm{i}\eta x} f(t,x,p)dx.$$

Perform this transform on Liouville-A (2.3.3a), one gets:

$$\partial_t \hat{\boldsymbol{\sigma}} + \beta\, \partial_p \hat{\boldsymbol{\sigma}} = \left(-\beta C + \frac{\mathrm{i}D}{\varepsilon} - \mathrm{i}\eta A\right) \hat{\boldsymbol{\sigma}} =: \hat{\mathcal{R}}\hat{\boldsymbol{\sigma}}$$

In this special case, β is a constant so the characteristic line can be obtained analytically. As a consequence one can avoid using the dense mesh that is required by high oscillation introduced from $\frac{\mathrm{i}D}{\varepsilon}$ term. We take the first time step $t \in [0, \Delta t]$ for example. Along the characteristic line $p(t) = p_0 + \beta t$, one evaluates $\hat{\boldsymbol{\sigma}}$ and $\hat{\mathcal{R}}$ at $(t, p(t), \eta)$ and has:

$$\frac{\mathrm{d}\hat{\boldsymbol{\sigma}}}{\mathrm{d}t} = \hat{\mathcal{R}}\, \hat{\boldsymbol{\sigma}}.$$

The solution to this ODE system could be explicitly written down:

$$\begin{aligned} \hat{\sigma}_{11}(t) =& \hat{\sigma}_{11}(0) - \int_0^t \mathrm{i}\eta\, \partial_p E_1\, \hat{\sigma}_{11}(t)\mathrm{d}t \\ &+ \int_0^t \big(\beta(\phi_{12}\hat{\sigma}_{12} - \phi_{21}\hat{\sigma}_{21}) - \mathrm{i}\eta\, \psi_{12}\hat{\sigma}_{12}\big)\mathrm{d}t\,, \end{aligned} \tag{2.3.5a}$$

and

$$\begin{aligned} \hat{\sigma}_{12}(t) =& \mathrm{e}^{\int_0^t K^\varepsilon(\tau)\mathrm{d}\tau}\hat{\sigma}_{12}(0) \\ &+ \int_0^t \mathrm{e}^{\int_s^t K^\varepsilon(\tau)\mathrm{d}\tau}\big(G(s)\hat{\sigma}_{11}(s) - H(s)\hat{\sigma}_{22}(s)\big)\mathrm{d}s\,, \end{aligned} \tag{2.3.5b}$$

where:

$$K^\varepsilon = \frac{\mathrm{i}}{\varepsilon}(E_2 - E_1) - \mathrm{i}\eta\,\partial_p E_2 - \beta(\phi_{11} - \phi_{22}),$$

and

$$G = -\beta\bar{\phi}_{12} - \mathrm{i}\eta\psi_{21}, \qquad H = \beta\bar{\phi}_{12}.$$

For t small, one obtains an approximation to (2.3.5):

$$\begin{aligned}
\hat{\sigma}_{11}(t) &\approx \hat{\sigma}_{11}(0) - \mathrm{i}\eta t\partial_p E_1\big(p(t)\big)\,\hat{\sigma}_{11}(t) \\
&\quad + \big(\beta\phi_{12}(0) - \mathrm{i}\eta\psi_{12}(0)\big)\int_0^t \hat{\sigma}_{12}\mathrm{d}t - \beta\phi_{21}(0)\int_0^t \hat{\sigma}_{21}\mathrm{d}t\,, \qquad (2.3.6a)\\
\hat{\sigma}_{12}(t) &\approx \hat{\sigma}_{12}(0)\mathrm{e}^{\int_0^t K^\varepsilon(\tau)\mathrm{d}\tau} \\
&\quad + \big(G(0)\hat{\sigma}_{11}(0) - H(0)\hat{\sigma}_{22}(0)\big)\int_0^t \mathrm{e}^{\int_s^t K^\varepsilon(\tau)\mathrm{d}\tau}\mathrm{d}s\,. \qquad (2.3.6b)
\end{aligned}$$

Note that this approximation introduces $(\Delta t)^2$ numerical error. Plug (2.3.6b) into (2.3.6a), to evaluate $\hat{\sigma}_{11}(\Delta t)$ and $\hat{\sigma}_{12}(\Delta t)$, one needs to compute:

$$\begin{aligned}
F_0 &:= \mathrm{e}^{\int_0^t K^\varepsilon(\tau)\mathrm{d}\tau}, \\
F_1 &:= \int_0^{\Delta t} \mathrm{e}^{\int_0^t K^\varepsilon(\tau)\mathrm{d}\tau}\,\mathrm{d}t, \\
F_2 &:= \int_0^{\Delta t}\int_0^t \mathrm{e}^{\int_s^t K^\varepsilon(\tau)\mathrm{d}\tau}\,\mathrm{d}s\,\mathrm{d}t.
\end{aligned}$$

It is not a easy job to compute these three since their integrands are highly oscillatory. However, if one chooses Δt such that $|\beta|\Delta t = \Delta p$, then at each time step the grid points of $p(t)$ lie on the characteristics, and these three could be repeated used. In fact, we compute them only *once* (with a highly resolved calculation) at the very beginning. The analysis for stability could be tricky and we do that for a simpler model problem in Appendix A.2.

A singularity decomposition idea

To handle the delta function in the initial condition (2.3.3b), one usually approximates it with a Gaussian function with small variance, and numerical error was determined by the width of the Gaussian. What is more, the Gaussian may lose its accuracy by expanding along the evolution and the analysis can be nasty. As stated before, in some special cases, this error could be avoided. In fact, as we can see, (2.3.3a) is linear, and the solution keeps being a delta function if it starts as one, in p direction, and one only needs to figure out the position of the delta function and its amplitude. This observation obviously suggests a singularity decomposition idea, which is also used in [114]. Write the ansatz of $\hat{\boldsymbol{\sigma}}_{mn}(t,x,p)$ as:

$$\boldsymbol{\sigma}(t,x,p) = \boldsymbol{\omega}(t,x,p)\,\delta\big(\theta(t,p)\big) \tag{2.3.7}$$

in which:

- $\theta(t,p) = p - (p^* + \beta t)$, which solves the Liouville equation

$$\partial_t\theta + \beta\,\partial_p\theta = 0\,,$$

- $\boldsymbol{\omega}$ satisfies the same equation as $\boldsymbol{\sigma}$:

$$\partial_t\boldsymbol{\omega} + \beta\,\partial_p\boldsymbol{\omega} = \mathcal{R}\boldsymbol{\omega}\,. \tag{2.3.8}$$

These can be proved by simple derivations. Formally, one has

$$\begin{aligned}\frac{\partial \boldsymbol{\sigma}}{\partial t} &= \frac{\partial}{\partial t}\big(\boldsymbol{\omega}\,\delta(\theta)\big) = \frac{\partial \boldsymbol{\omega}}{\partial t}\,\delta(\theta) + \boldsymbol{\omega}\,\delta'(\theta)\frac{\partial \theta}{\partial t} \\ &= \mathcal{R}\boldsymbol{\omega}\,\delta(\theta) - \beta\frac{\partial \boldsymbol{\omega}}{\partial p}\,\delta(\theta) - \beta\boldsymbol{\omega}\,\delta'(\theta) \\ &= \mathcal{R}\boldsymbol{\omega}\,\delta(\theta) - \beta\frac{\partial \boldsymbol{\omega}}{\partial p}\,\delta(\theta) - \beta\boldsymbol{\omega}\,\delta'(\theta)\frac{\partial \theta}{\partial p} \\ &= \mathcal{R}\boldsymbol{\omega}\,\delta(\theta) - \beta\frac{\partial}{\partial p}\big(\boldsymbol{\omega}\,\delta(\theta)\big) \\ &= \mathcal{R}\boldsymbol{\sigma} - \beta\frac{\partial \boldsymbol{\sigma}}{\partial p}.\end{aligned}$$

The equalities above should be understood in the distributional sense. The decomposition (2.3.7) enables one to solve for $\boldsymbol{\omega}$ and θ separately with good (bounded) initial data $|a_0(x)|^2$ and $\partial_x S_0(x)$ respectively. The equation $\boldsymbol{\omega}$ satisfies is the same as the one for $\boldsymbol{\sigma}$, thus the numerical method introduced in Section 2.3.1 can be used. In the final output, one needs to get back to $\boldsymbol{\sigma}$ using (2.3.7), so a discrete delta approximation is only needed at the output time, not during time evolution.

Numerical experiments

We implement the numerical method with the following data

$$\beta = 1, \qquad p^* = -0.25, \qquad a_0(x) = \exp\left(-\frac{25(x-\pi)^2}{2}\right). \tag{2.3.9}$$

To verify the accuracy, we compare our numerical methods with the one given by directly compute the original Schrödinger equation. The output is the density, the cumulative density function (c.d.f.) and mass in the 1st band, defined respectively by:

$$\rho^\varepsilon = |\psi^\varepsilon|^2, \qquad \gamma^\varepsilon = \int_{-\infty}^{x} \rho^\varepsilon(y)dy, \qquad m_1(t) = \int |\mathbb{P}_1\psi^\varepsilon(t,x)|^2 \mathrm{d}x\,, \tag{2.3.10}$$

where $\mathbb{P}_n$ is the projection onto the nth band:

$$\mathbb{P}_n\psi(x) = \int_B \mathrm{d}p \int_{\mathbb{R}} \mathrm{d}y\psi(y)\bar{\Psi}_n(y,p)\Psi_n(x,p), \quad \phi \in L^2(\mathbb{R}),\ m \in \mathbb{N}\ .$$

The two integrals in (2.3.10) are calculated by the midpoint quadrature rule numerically.

Figure 2.3.2 shows the density and c.d.f. computed for the Schrödinger equation, the Liouville-A and the Liouville-S respectively at $t = 0.5$. The results match quite well.

Figure 2.3.3 shows the evolution of m_1 as a function of time t. One can see the total mass on the first band jumps down at around $t = 0.25$, when the momentum p reaches $p_c = 0$, reflecting the 4th-to-5th band transition. The experiment also shows that smaller ε gives smaller transition rate. Note that some small oscillations occur around the crossing region. They are related to the interference phenomena, and are usually called the Stueckelberg oscillation [38, 175, 184].

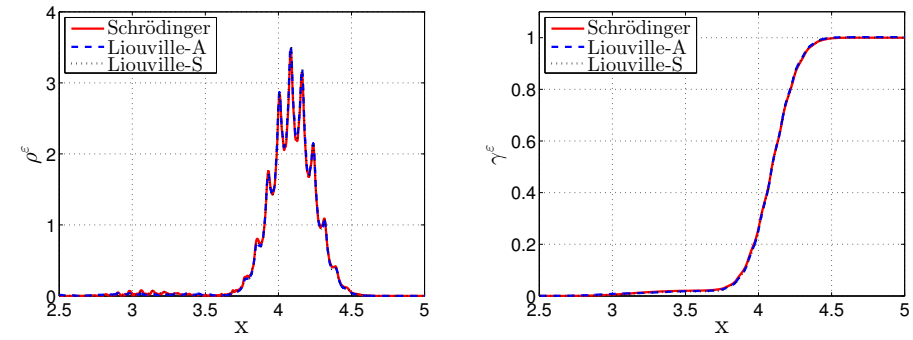
Define L^1 error in the cumulative distribution function (c.d.f.) [125, 82]:

$$\mathrm{Err}^{\varepsilon}(t) \ = \ \int_{\mathbb{R}} \left| \int_{-\infty}^{x} \big(\rho_S^{\varepsilon}(t,z) - \rho_L^{\varepsilon}(t,z)\big)\mathrm{d}z \right| \mathrm{d}x, \tag{2.3.11}$$

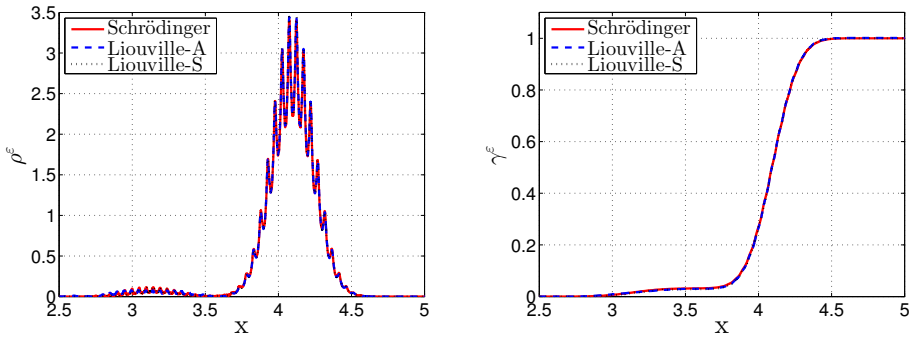
where ρ_S^{ε} and ρ_L^{ε} denote the density calculated by the Schrödinger equation and the Liouville system respectively. Numerically we compute (2.3.11) using the midpoint quadrature rule. Figure 2.3.4 shows this at time $t = 0.5$. As $\varepsilon \to 0$, the Liouville system gets more accurate, and the error decreases with the speed of $\mathcal{O}(\varepsilon)$.

2.3.2 A domain decomposition computation

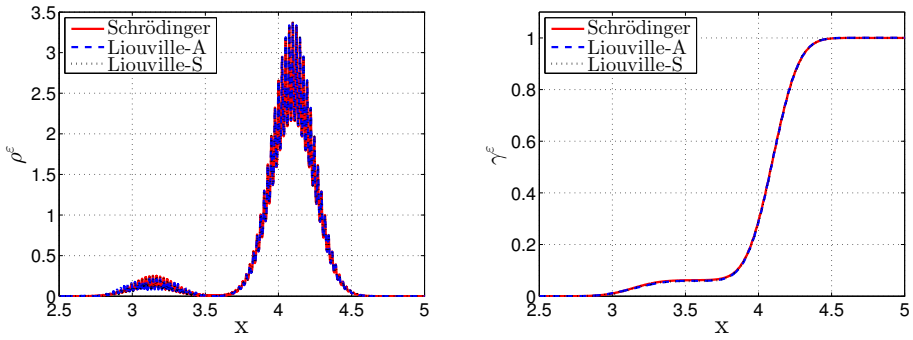
This section shows examples with nonlinear $U(x)$. For this general case, the p-characteristic is no longer a straight line, and we do not have the fast solver in the previous section any more. To numerically solve Liouville-A, we use the domain decomposition method. The convection



(a) $\varepsilon = \frac{1}{80}$

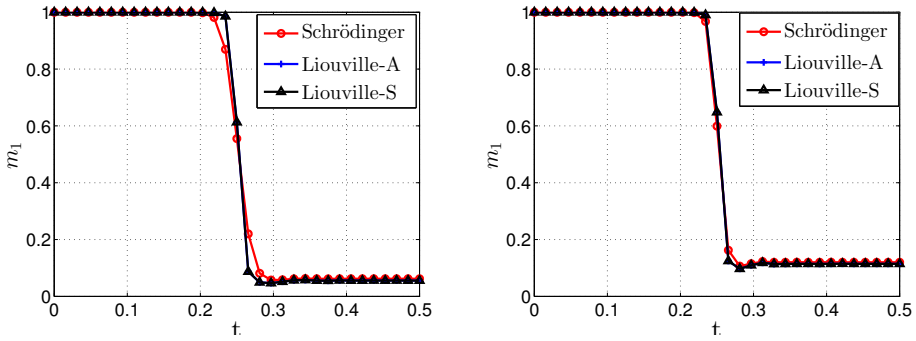


(b) $\varepsilon = \frac{1}{128}$



(c) $\varepsilon = \frac{1}{256}$

Figure 2.3.2: Periodic potential problem. Linear $U(x)$. $t = 0.5$. The left and right column are for the position density ρ^ε, and c.d.f. γ^ε respectively. The solid line, the dash line and the dotted line are the numerical solutions to the Schrödinger, the Liouville-A and the Liouville-S respectively.



(a) $\varepsilon = 2^{-8}$ (b) $\varepsilon = 2^{-9}$

(c) $\varepsilon = 2^{-10}$ (d) $\varepsilon = 2^{-11}$

Figure 2.3.3: Periodic potential problem. Linear $U(x)$. Time evolution of $m_1(t)$ defined in (2.3.10).

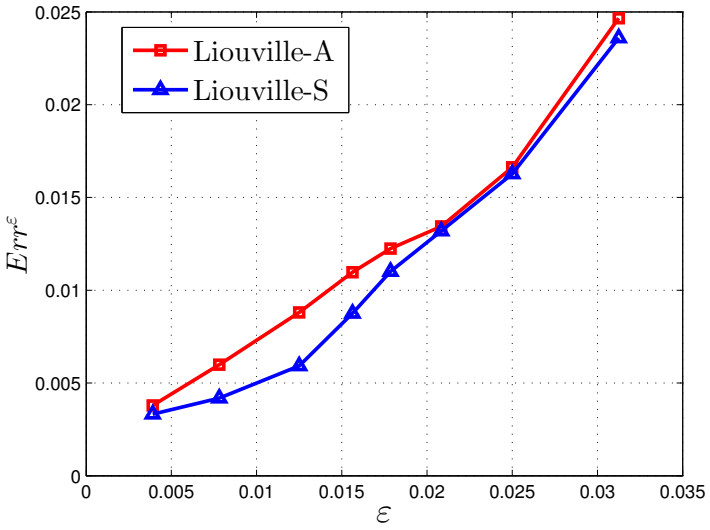


Figure 2.3.4: Periodic potential problem. Linear $U(x)$. Err^{ε} as function of ε at $t = 0.5$.

terms are handled by classical finite volume method.

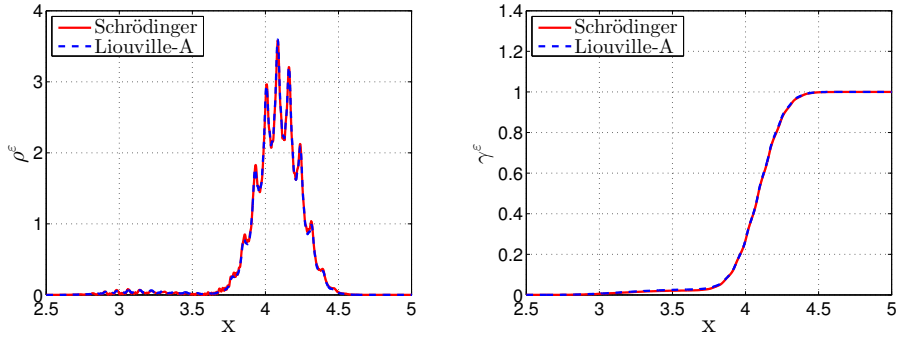
We compute the Liouville-A system with both a pure and a mixed state initial data with:

$$U(x) = -\frac{x - \sin x}{2}\,.$$

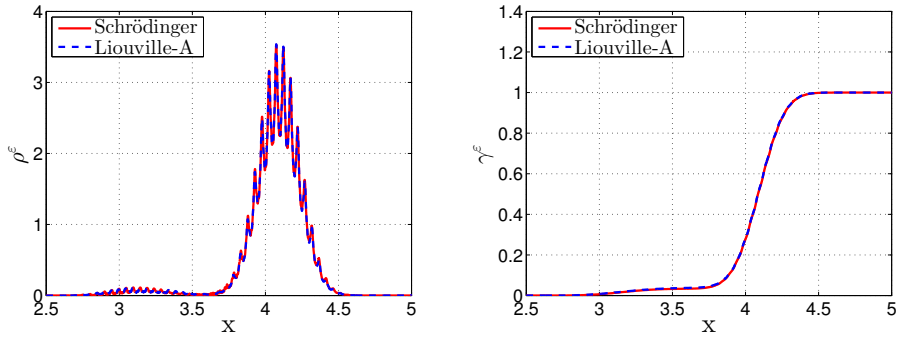
Example 1: A pure state initial data

In this example, we use the same pure state initial data as in the previous example (2.3.2),(2.3.9). Correspondingly, the initial data for the Liouville-A system is given by (2.3.3b). Numerically a Gaussian function centered at p^* with variance of $\varepsilon/16$ is used to approximate the $\delta-$function.

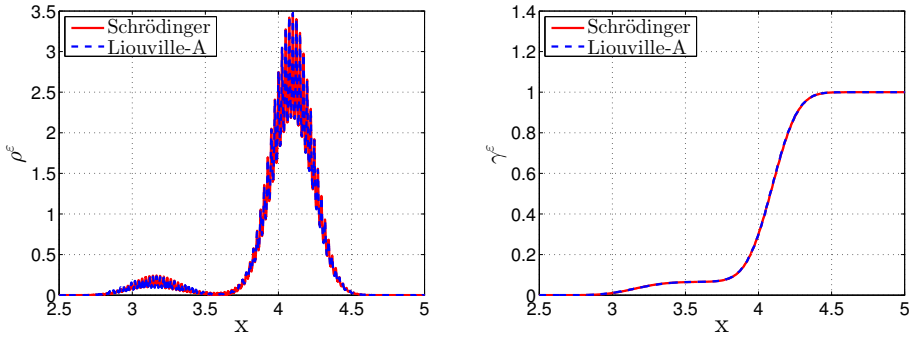
Figures 2.3.5 and 2.3.6 show the density, the c.d.f at $t = 0.5$ and evolution of m_1 (2.3.10) computed for both the Schrödinger equation and the Liouville-A system. The numerical results for the two systems agree well. Figure 2.3.7 gives decay of $\mathrm{Err}^{\varepsilon}$ (2.3.11) with respect to ε.



(a) $\varepsilon = \frac{1}{80}$

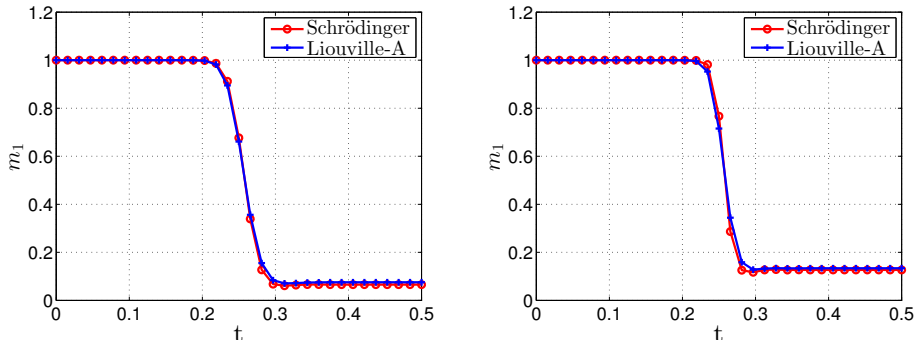


(b) $\varepsilon = \frac{1}{128}$



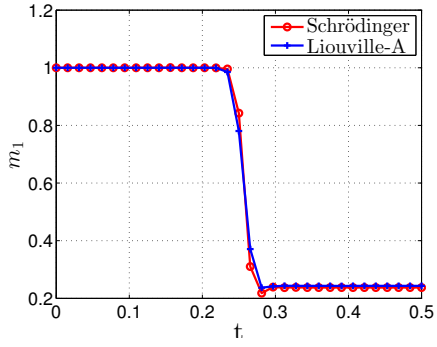
(c) $\varepsilon = \frac{1}{256}$

Figure 2.3.5: Periodic potential problem. Example 1. $t = 0.5$. The left and right columns show the position density ρ^ε, and the c.d.f. γ^ε respectively.

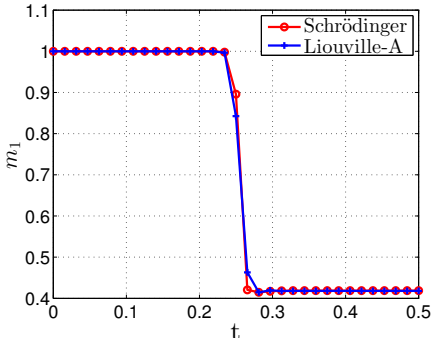


(a) $\varepsilon = 2^{-8}$

(b) $\varepsilon = 2^{-9}$



(c) $\varepsilon = 2^{-10}$



(d) $\varepsilon = 2^{-11}$

Figure 2.3.6: Periodic potential problem. Example 1. Time evolution of $m_1(t)$ defined in (2.3.10).

The numerical results show that the hybrid model can capture the band-to-band transition

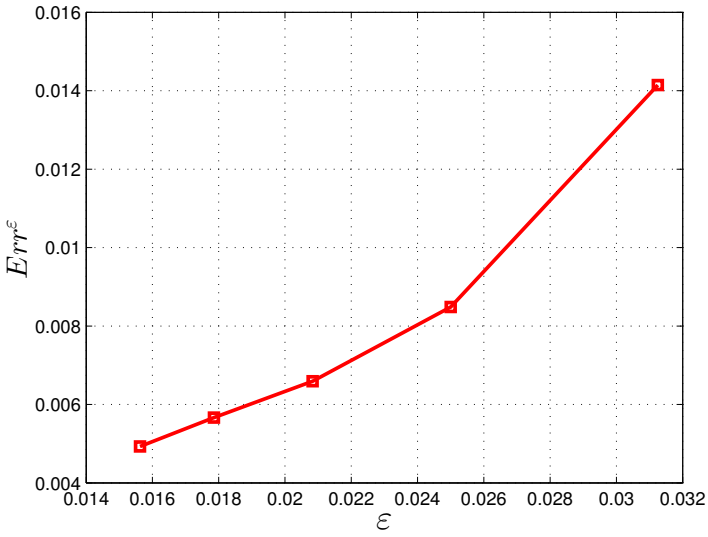


Figure 2.3.7: Periodic potential problem. Example 1. The L^1 error Err^{ε} between Liouville-A and the Schrödinger solution at $t = 0.5$.

phenomena, and the error decays like $\mathcal{O}(\varepsilon)$.

Remark 2.3.1. *As U_x varies with x, the wave packet becomes de-coherent. This weakens the interference phenomenon[38, 175, 184]. As one can see in Figure 2.3.6, the Stueckelberg oscillations around the crossing region is much weaker than those in the previous example.*

Example 2: A mixed state initial data.

This example is for the case when the initial data is a mixed state:

$$\psi_I \;=\; a_0(x)\,\left[\Phi_1\left(\frac{x}{\varepsilon},p^*\right)\,\mathrm{e}^{\mathrm{i}p^*x/\varepsilon}+\Phi_2\left(\frac{x}{\varepsilon},p^*\right)\,\mathrm{e}^{\mathrm{i}p^*x/\varepsilon}\right],\quad p^*=-0.25.$$

Correspondingly, the initial data for the semi-classical Liouville system should be:

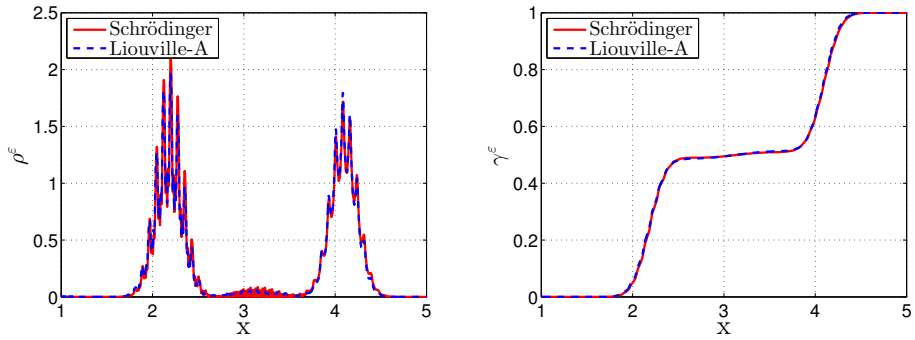
$$\boldsymbol{\sigma} = a_0^2(x)\,\delta(p-p^*)\;[1,1,1,1]^T\;.$$

Since p^* is away from the crossing point and σ_{12} and σ_{21} weakly converge to zero as $\varepsilon \to 0$, numerically, we regard them as zero and use

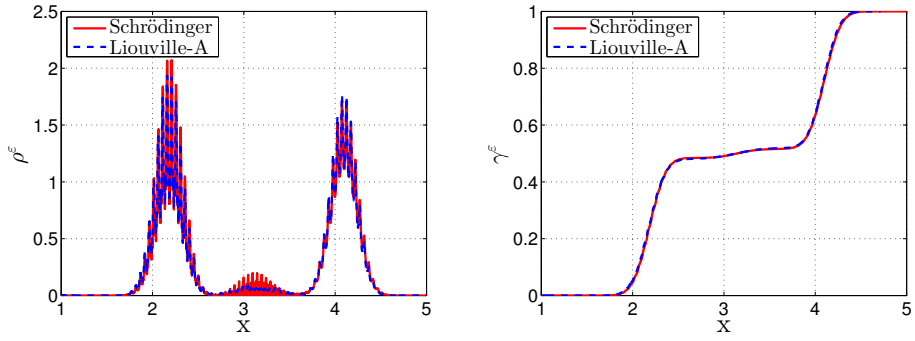
$$\boldsymbol{\sigma} \approx a_0^2(x)\,\delta_G(p-p^*)\;[1,0,0,1]^T$$

as the initial condition, where $\delta_G(p)$ is a Gaussian function centered at zero.

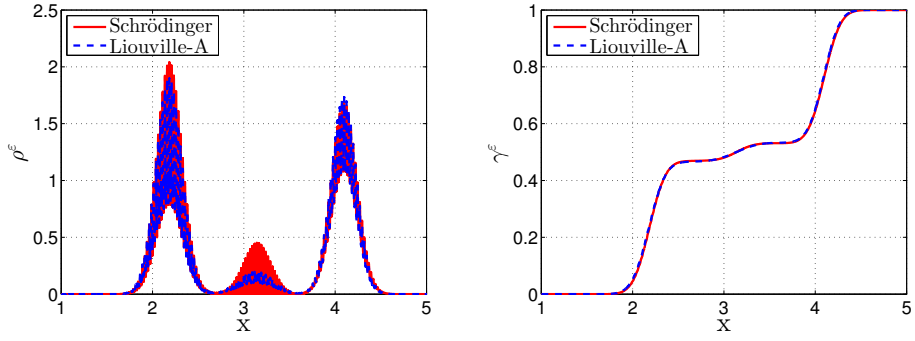
The density and the c.d.f. are computed for the Liouville-A and the Schrödinger, compared in Figure 2.3.8. $\mathrm{Err}^{\varepsilon}$ as a function of ε is shown in Figure 2.3.9.



(a) $\varepsilon = \frac{1}{80}$



(b) $\varepsilon = \frac{1}{128}$



(c) $\varepsilon = \frac{1}{256}$

Figure 2.3.8: Periodic potential problem. Example 2. At time $t = 0.5$. The left and the right columns show the position density ρ^ε, and the c.d.f. γ^ε respectively.

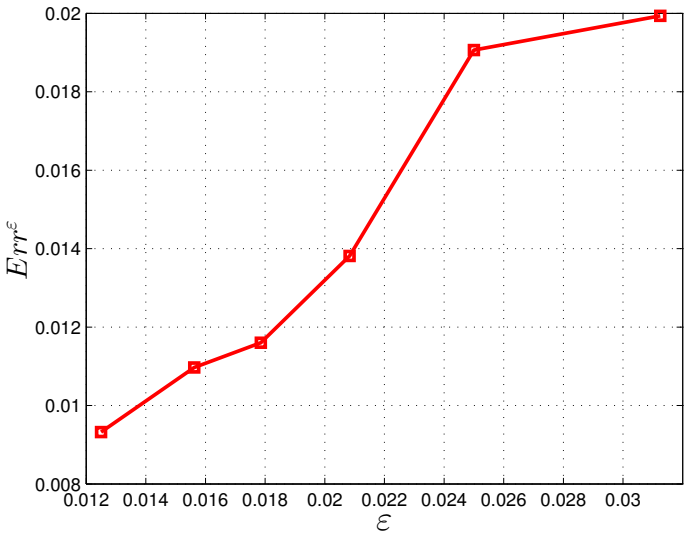


Figure 2.3.9: Periodic potential problem. Example 2. The L^1 error Err^{ε} as a function of ε at $t = 0.5$.

Chapter 3

Semi-classical description of molecular dynamics

In this chapter we study a many body problem with N nuclei and n electrons. Semi-classically, nuclei could jump across different energy surfaces. Around the intersection point where two or more energy surfaces get close or even touch, such quantum transitions can be influential. This quantum molecular dynamics problem plays an important role in analyzing chemical reaction system, and has many interesting applications in chemistry, physics and biology.

An interesting story is about bioluminescence of fireflies [179]. In this luciferin-luciferase system, the luciferase-bound dioxetanone climbs up to an excited energy state through intersection point, producing excited-state luciferase-bound oxyluciferin, who falls back later on to the electronic ground state, with the emission of a visible photon. Most other bioluminescence systems undergo the same mechanism, and similar schematic models are used to interpret them [195]. In general, most chemical reaction could be explained by such energy transition in the quantum regime. Many other applications can also be found in [157, 158, 203]. Also see a review in [28], and [196].

Despite its importance, the understanding to most such problems is very limited. Many models in interpreting the real chemical reaction process are very much simplified to be practical, but even for the simplified model, very little quantitative analysis is achieved. The

numerical developments for this surface hopping problem around intersection points are far from being satisfactory either. Usually to deal with the many body problems in which two types of particles have disparate masses, people use the Born-Oppenheimer approximation. The idea is that the heavy particles are much more reluctant to move than the light ones and thus one could separate the motion of the two types of particles. In our problem to be specific, energy band structure for the electrons is solved first with fixed configuration of nuclei as they are considered at rest. However, this approximation is only valid when energy surfaces are very well-separated, and when they get close to each other, the transition between eigenfunctions should be taken care of.

In chemistry community, direct computation on carefully designed orthogonal basis has been used, namely discrete variable representation. The algorithm dates back to [51, 99] and was popularized in a series of work in 80s, see a review paper in [145]. However, the orthogonal basis for the wave function to is usually defined on the grid points in the physical domain, leading to high computational cost, even after fast solvers are applied [16]. Another approach is the probabilistic algorithms that rely on the Landau-Zener formula, derived in [131, 202], and proved with better accuracy in [91, 13, 108, 127]. The transition rate given by the formula presents the total probability of a particle jumping from one band to another, whose trajectory passes through the entire domain. With some simplification, it is shown that in the leading order, the rate only depends on the smallest gap between the energy bands. This approach is adopted in computation, called surface hopping method, initiated in [185] and studied more thoroughly in [184, 189, 136] on the physical space, as summarized in [55]. In [130, 136] the authors implemented the same idea in the phase space using particle method, under the framework of the Wigner transform, and the underlying analysis is carried out in [137, 62]. Eulerian framework was adopted later on in [118, 117] for uniform global accuracy. However, in all of these methods, it is assumed that the transition takes place locally in the spatial space

– around the intersection point, based on the argument that the transition away from it is exponentially small and thus is treated as zero. Numerically it simplifies the computation a great deal, however, due to the brutal physical simplification, part of the transition is for sure missing.

Our approach is to look for semi-classical models in phase space that capture the correct transition rate. We are going to derive a set of equations in phase space using Wigner transformation and Weyl quantization. These two counterparts link pseudo-operators with phase space functions in different directions. In our approach, we do not cumulate all the transitions into one number, but instead the transition rates are functions in phase space, that are naturally characterized by the off diagonal entries in the Wigner matrix. We find the equations that govern their evolution in time, and we also compare our model with the classical model for adiabatic system where bands are assumed to be well-separated. In the adiabatic regions, these two models are consistent. Numerically we use domain decomposition idea, and take advantage of both models: based on the degeneracy of eigenvalues, we apply appropriate models in different regions, so that the model could be computed cheaply with the transition rate captured correctly.

In section 3.1 we give a brief review of the background of the problem. Our model will be presented in section 3.2: in subsection 3.2.1 we introduce the Weyl quantization and its associated operations including Moyal product, that will be used to derive the equations for both adiabatic and non-adiabatic cases in subsection 3.2.2. The connection between the two systems will be shown in subsection 3.2.3. Based on the conclusion obtained there, numerical method are constructed. Numerical examples will be shown in section 3.3.

3.1 Born-Oppenheimer approximation and problem background

The Schrödinger equation for the quantum molecular dynamics is given by:

$$\mathrm{i}\hbar\partial_t\Phi(t,\mathbf{x},\mathbf{y}) = H\Phi(t,\mathbf{x},\mathbf{y}), \tag{3.1.1}$$

$$\Phi(0,\mathbf{x},\mathbf{y}) = \Phi_0(\mathbf{x},\mathbf{y}). \tag{3.1.2}$$

Here $\hbar$ is the Planck constant, Φ is the wave function that depends on time t, position of nuclei $\mathbf{x} \in \mathbb{R}^{3N}$ and position of electrons $\mathbf{y} \in \mathbb{R}^{3n}$ with N and n standing for the number of particles and 3 is the dimension. H is the Hamiltonian, and in this circumstance, it is given by:

$$H = -\sum_{j=1}^{N}\frac{\hbar^2}{2M_j}\Delta_{\mathbf{x}_j} - \sum_{j=1}^{n}\frac{\hbar^2}{2m_j}\Delta_{\mathbf{y}_j} + \sum_{j<k}\frac{1}{|\mathbf{y}_j-\mathbf{y}_k|} + \sum_{j<k}\frac{Z_jZ_k}{|\mathbf{x}_j-\mathbf{x}_k|} - \sum_{j=1}^{N}\sum_{k=1}^{n}\frac{Z_j}{|\mathbf{x}_j-\mathbf{y}_k|}.$$

The first two terms represent kinetic energy for the two particles, and the last three are the Coulomb potential. m_i, M_i and Z_i stand for mass of electron, nucleus and electric charge respectively. We combine the last four terms all together and call it electronic Hamiltonian, denoted by H_e.

This equation is a general setting for describing chemical systems, but is cursed by its high dimensionality for practical use. For instance the benzene molecule consists of 12 nuclei and 42 electrons, and thus has 162 spacial variable – generating very high computation cost. A standard way to reduce the degree of freedom is to use Born-Oppenheimer approximation [22]. The approximation was derived for the Schrödinger equation for molecular dynamics and was used to describe chemical reactions [148]. We explain in detail about the set up of the problem in 3.1.1 and discuss its basic properties in 3.1.2.

3.1.1 Equation set up

The Born-Oppenheimer approximation is based on the observation that the mass ratio between nucleus and electron is very high, and intuitively the motion of the nuclei should be much slower than that of electrons. As argued in the original paper [22], the electrons quickly adjust themselves in response to the slow motion of nuclei, and they stay on the n-th energy band if they start on it, determined as though the nuclei were not moving. The impact from the electrons to the nuclei is taken as an effective potential, and due to the disparity of the masses, it is seen that the nuclei obey semi-classical motion. This property makes it possible for us to separate the two motions and solve them one by one. Mathematically, one could break the wavefunction of a molecule into the electronic and nuclear component, i.e. we write the solution to the original Schrödinger equation as a separable function with one element dealing with electrons and the other dealing with nuclei. We solve the part for electron first. Since nuclei move at much slower speed, in a short time period, their configurations are regarded fixed, and the spectrum for the electronic part with these fixed parameters are solved. In this step, only the spatial variables of electrons are involved. The second step is to solve the nuclei part with the spectrum, in which only nuclei are computed. In both of these two steps, only a portion of variables are involved, which highly reduced the degree of freedom.

The whole idea starts with the marvelous intuition of Born and Oppenheimer. In their original paper, it is assumed that the mass ratio is going to infinity, so asymptotic expansion could be formally carried out. In the leading order, the nuclei are at their optimal configuration, namely, the minimal point of the given energy surface. The original paper was done using Taylor expansion assuming that the energy surfaces are smooth enough. The understanding that the vibration and rotation could be properly separated is achieved in [58]. The rigorous mathematical verification on its validity was not carried out till about half a century later [41, 90, 178]. For Gaussian wave packet, Hagedorn proved the validity of this approximation to high

orders in [93, 94], for smooth potential and the more realistic Coulomb potential respectively, and the exponentially small error is acquired in [97]. All the analysis is carried out under the assumption that energy surfaces are very well separated.

We start with nondimensionalization, and analyze the validity of the approximation.

Non-dimensionalization

Denote L, T and Ω as the typical scale for length, time and mass, and we rescale the parameters:

$$\tilde{t} = \frac{t}{T}, \quad \tilde{\mathbf{x}} = \frac{\mathbf{x}}{L}, \quad \tilde{\mathbf{y}} = \frac{\mathbf{y}}{L}, \quad \tilde{m} = \frac{m}{\Omega}, \quad \tilde{M} = \frac{M}{\Omega}, \tag{3.1.3}$$

The Hamiltonian is nondimensionalized too:

$$\tilde{H}_e(\tilde{\mathbf{y}}, \tilde{\mathbf{x}}) = \frac{\tilde{m}T^2}{\Omega L^2} H_e(\mathbf{y}, \mathbf{x}). \tag{3.1.4}$$

Dropping the tildes the equation becomes:

$$\mathrm{i}\delta m \partial_t \Phi = -\sum_{j=1}^{N} \frac{\delta^2 m}{2M} \Delta_{\mathbf{x}_j} \Phi + H_e \Phi, \tag{3.1.5}$$

with $\delta = \frac{\hbar T}{\Omega L^2}$. In the scale when $\delta = 1$ and $\hat{t} = \frac{\tilde{t}}{\sqrt{mM}}$, we define $\varepsilon = \sqrt{\frac{m}{M}}$ and have the following:

$$\mathrm{i}\varepsilon \partial_t \Phi = -\sum_{j=1}^{N} \frac{\varepsilon^2}{2} \Delta_{\mathbf{x}_j} \Phi + H_e \Phi. \tag{3.1.6}$$

In the equation (3.1.6), H_e is in the electronic Hamiltonian. As argued above, one has two steps in the computation:

1st step: We solve for the spectrum of H_e in term of $\mathbf{y}_j$, with $\mathbf{x}_j$ regarded as fixed parameters by

computing the following eigenvalue problem:

$$H_e(\mathbf{y},\mathbf{x})\phi_k(\mathbf{y};\mathbf{x}) = E_k(\mathbf{x})\phi_k(\mathbf{y};\mathbf{x}), \quad k = 1, 2, \cdots . \tag{3.1.7}$$

E_k are called energy surfaces, and they depend on $\mathbf{x}$. Here we assume H_e has complete set of orthonormal eigenfunctions ϕ_k:

$$\langle \phi_i, \phi_j \rangle = \int \phi_i^* \phi_j \mathrm{d}\mathbf{y} = \delta_{ij}, \tag{3.1.8}$$

with δ_{ij} being the Kronecker delta and "$*$" is the conjugate transpose. Since ϕ_i gives a basis we expand Φ on it, and the projection coefficients, denoted by χ_i, depend on time t and $\mathbf{x}$:

$$\Phi(t, x, y) = \sum_k \chi_k(t, \mathbf{x})\phi_k(\mathbf{y};\mathbf{x}). \tag{3.1.9}$$

2nd step: Solve for χ_k by plugging the ansatz (3.1.9) into (3.1.6) and perform $\langle \cdot, \phi_k \rangle$. The equation we get is:

$$\mathrm{i}\varepsilon\chi_k = \left[-\sum_j \frac{\varepsilon^2}{2}\Delta_{\mathbf{x}_j} + E_k(\mathbf{x}) \right] \chi_k + \sum_l C_{kl}\chi_l, \tag{3.1.10}$$

with

$$C_{kl} = \langle \phi_k, -\sum_j \frac{\varepsilon^2}{2}\Delta_{\mathbf{x}_j}\phi_l \rangle_{\mathbf{y}} - \sum_j \varepsilon^2 \langle \phi_k, \nabla_{\mathbf{x}_j}\phi_l \rangle \cdot \nabla_{\mathbf{x}_j}.$$

In a compact form, it is equivalent to:

$$\mathrm{i}\varepsilon\chi = \left[-\frac{\varepsilon^2}{2}\left(\nabla_{\mathbf{x}} + A_{geo}\right)^2 + E \right] \chi, \tag{3.1.11}$$

where A_{geo} is a matrix with its ij-th component being a Berry connection $\langle \Phi_i | \nabla_{\mathbf{x}} \Phi_j \rangle$. E is a diagonal matrix with $E_{ij} = E_i(\mathbf{x})\delta_{ij}$.

The Born-Oppenheimer approximation

In equation (3.1.10), it is the C term that complicates the system (also the A_{geo} term in (3.1.11)). Depending on whether it is negligible, the Born-Oppenheimer approximation could be valid or not, and whenever it breaks down, we perform the diabatic representation, to be specific:

- It is shown that when energy surfaces E_k are very well separated, i.e. $E_i - E_j = \mathcal{O}(1), \quad \forall i \neq j$, A_{geo} is considerably negligible compared to that of $\nabla_{\mathbf{x}}$. We throw it away, and (3.1.11) becomes the Born-Oppenheimer approximation:

$$\mathrm{i}\varepsilon\chi = \left(-\frac{\varepsilon^2}{2}\Delta_{\mathbf{x}} + E\right)\chi, \tag{3.1.12}$$

- When two adjacent energy bands get close to each other, for example $E_j - E_{j-1} \ll \mathcal{O}(1)$, A_{geo} can be as large as $\frac{1}{|E_j - E_{j-1}|}$. In this case, (3.1.12) no longer holds. However, mathematically (3.1.12) has a nicer form, and we seek for similar expression by performing diabetic presentation [2] as below.

  Define a new basis by a unitary transformation $\tilde{\Phi}_k(t, \mathbf{x}) = S(\mathbf{x})\Phi_k(t, \mathbf{x})$. Upon this new basis, the equation becomes:

$$\mathrm{i}\varepsilon\partial_t\tilde{\chi} = \left[-\frac{\varepsilon^2}{2}\left(\nabla_{\mathbf{x}} + \tilde{A}_{geo}\right)^2 + \tilde{E}\right]\tilde{\chi}, \tag{3.1.13}$$

  where $\tilde{A}_{geo} = S^\dagger A_{geo} S + S^\dagger(\nabla_{\mathbf{x}} S)$ and $\tilde{E} = S^\dagger E S$ is no longer diagonal, but is still symmetric. Select S in a smart way to make $\tilde{A}_{geo} = 0$ (for possibilities of such operations, see [152]):

$$\mathrm{i}\varepsilon\tilde{\chi} = \left[-\frac{\varepsilon^2}{2}\Delta_{\mathbf{x}} + \tilde{E}(\mathbf{x})\right]\tilde{\chi}. \tag{3.1.14}$$

  Note $\tilde{E}$ is no loner diagonalized.

In both of these two cases, we arrive at a similar expression with E as a function of $\mathbf{x}$, either diagonal (when BO approximation is valid) or not (when it breaks down).

For both simplicity and consistency, we adjust the notation as following:

$$\mathrm{i}\varepsilon\frac{\partial\psi^\varepsilon}{\partial t}(t,\mathbf{x}) = \left(-\frac{\varepsilon^2}{2}\Delta_{\mathbf{x}} + \tilde{V}(\mathbf{x})\right)\psi^\varepsilon(t,\mathbf{x}), \qquad (t,\mathbf{x}) \in \left(\mathbb{R}^+, \mathbb{R}^d\right) \tag{3.1.15}$$

$$\psi^\varepsilon(0,\mathbf{x}) = \psi_0^\varepsilon(\mathbf{x}) \tag{3.1.16}$$

Here ψ is a vector and $\tilde{V}(x)$ is a symmetric matrix. Assume we are studying a 2-bands system (with easy extension to multi-bands) and the potential reads as:

$$\tilde{V}(\mathbf{x}) = \frac{1}{2}\mathrm{tr}\tilde{V}(\mathbf{x}) + \begin{pmatrix} u(\mathbf{x}) & v(\mathbf{x}) \\ v(\mathbf{x}) & -u(\mathbf{x}) \end{pmatrix}. \tag{3.1.17}$$

The trace part is denoted by $U(\mathbf{x})$ and the remainder as $V(\mathbf{x})$. $\psi^\varepsilon \in \mathbb{C}^2$ in this case.

3.1.2 Basic properties

Apparently, the Hamiltonian operator of the system is:

$$\hat{H} = \left(-\frac{\varepsilon^2}{2}\Delta_{\mathbf{x}} + U(\mathbf{x})\right)\mathbb{I} + V(\mathbf{x}), \tag{3.1.18}$$

with $\mathbb{I}$ as the 2×2 identity matrix. Diagonalize the potential matrix $V(\mathbf{x})$, one has:

$$\Theta V \Theta^\dagger = \Lambda_V, \tag{3.1.19}$$

with:

$$\Theta(\mathbf{x}) = \frac{1}{\sqrt{2\left(1+\frac{u(\mathbf{x})}{\sqrt{u(\mathbf{x})^2+v(\mathbf{x})^2}}\right)}} \begin{pmatrix} 1+\frac{u(\mathbf{x})}{\sqrt{u(\mathbf{x})^2+v(\mathbf{x})^2}} & \frac{v(\mathbf{x})}{\sqrt{u(\mathbf{x})^2+v(\mathbf{x})^2}} \\ -\frac{v(\mathbf{x})}{\sqrt{u(\mathbf{x})^2+v(\mathbf{x})^2}} & 1+\frac{u(\mathbf{x})}{\sqrt{u(\mathbf{x})^2+v(\mathbf{x})^2}} \end{pmatrix}. \tag{3.1.20}$$

Here:

$$\Lambda_V = \operatorname{diag}\left(\lambda_V^+, \lambda_V^-\right) = \operatorname{diag}\left(\sqrt{u^2(\mathbf{x})+v^2(\mathbf{x})}, -\sqrt{u^2(\mathbf{x})+v^2(\mathbf{x})}\right), \tag{3.1.21}$$

is a diagonal matrix with the two eigenvalues denoted as $\lambda_V^{\pm}$. We also use $\chi^{\pm}$ to denote the two columns of $\Theta^{\dagger}$ standing for normalized eigenvectors associated with $\lambda_V^{\pm}$.

If the two eigenvalues are very well separated for all $\mathbf{x}$, one could diagonalize the system and compute two separate Schrödinger equations. However, sometimes the two eigenvalues degenerate, as $u(\mathbf{x}) = v(\mathbf{x}) = 0$ here; and when this happens, the eigenvector space associated with the single eigenvalue is 2 dimensional, and one could not distinguish two quantum states. In fact, ambiguity emerges in defining the right eigenfunctions: limit for (3.1.20) does not even exist as $u(\mathbf{x})$ and $v(\mathbf{x})$ go to zero. The standard mathematical decomposition method breaks down at the surface hopping point, and it is our main goal to find the right mathematical description of the system that keeps its validation even at the intersection.

3.2 Semi-classical models

In this section we use the Wigner transform approach to derive a semi-classical model, for both adiabatic and non-adiabatic cases. As defined in (1.2.1), the Wigner function is the inverse Fourier transform of the density matrix:

$$F^{\varepsilon}(\mathbf{x},\mathbf{p}) = \frac{1}{(2\pi)^d}\int \rho^{\varepsilon}\left(\mathbf{x}-\frac{\varepsilon\mathbf{y}}{2}, \mathbf{x}+\frac{\varepsilon\mathbf{y}}{2}\right) \mathrm{e}^{\mathrm{i}\mathbf{p}\cdot\mathbf{y}}\,\mathrm{d}\mathbf{y}, \tag{3.2.1}$$

where ρ^ε is the density matrix:

$$\rho^\varepsilon(\mathbf{x},\mathbf{x}') = \psi^\varepsilon(\mathbf{x}) \otimes \overline{\psi^\varepsilon}(\mathbf{x}'). \tag{3.2.2}$$

$\overline{\psi^\varepsilon}$ is the complex conjugate of ψ^ε. As in (1.2.2), one could express the formula in bra-ket language [201]:

$$F^\varepsilon(\mathbf{x},\mathbf{p}) = \int_{\mathbb{R}^d} \frac{\mathrm{d}\mathbf{y}}{(2\pi)^d} \langle \mathbf{x} - \frac{\varepsilon}{2}\mathbf{y}|\hat{\rho}|\mathbf{x} + \frac{\varepsilon}{2}\mathbf{y}\rangle e^{\mathrm{i}\mathbf{p}\cdot\mathbf{y}}, \tag{3.2.3}$$

with $\hat{\rho} = |\psi^\varepsilon\rangle\langle\psi^\varepsilon|$ standing for the density operator.

Unlike in the scalar case, here the Wigner function is a 2×2 matrix, and the collision term is much more complicated than the one in (1.2.5). Simple derivation gives, if ψ^ε solves the Schrödinger equation (1.1.1), then F solves the Wigner equation:

$$\partial_t F^\varepsilon + \mathbf{p}\cdot\nabla_{\mathbf{x}} F^\varepsilon + \Xi[U\mathbb{I} + V]F^\varepsilon = 0, \tag{3.2.4}$$

where $\Xi[V]$ is an operator that, when acted on the density matrix F^ε, gives:

$$\Xi[V]F^\varepsilon = \int_{\mathbb{R}^{2d}} \frac{V\left(\mathbf{x} - \frac{\varepsilon\mathbf{y}}{2}\right) F^\varepsilon(\mathbf{x},\mathbf{p}') - F^\varepsilon(\mathbf{x},\mathbf{p}')V\left(\mathbf{x} + \frac{\varepsilon\mathbf{y}}{2}\right)}{-\mathrm{i}(2\pi)^d\varepsilon} \mathrm{e}^{\mathrm{i}(\mathbf{p}'-\mathbf{p})\cdot\mathbf{y}} \mathrm{d}\mathbf{p}'\,\mathrm{d}\mathbf{y}.$$

Note that unlike the usual case in (1.2.6) where $V(\mathbf{x} - \frac{\varepsilon\mathbf{y}}{2}) - V(\mathbf{x} + \frac{\varepsilon\mathbf{y}}{2})$ gives $\nabla_{\mathbf{x}} V(\mathbf{x})$ in the limiting regime, here both F and V are matrices, and it is no longer guaranteed that F and V commute. The techniques we use to overcome this difficulty is the Weyl quantization to be introduced in the following subsection.

3.2.1 Weyl quantization and Wigner transform

The Weyl quantization was found in [192] in 1927, and was brought into the field in [74]. It was firstly adopted in [153] for non-adiabatic situation, and later on extended to graphene

controlled by the Dirac equation [154]. The computation for the avoided crossing is done in [36].

The Weyl quantization maps functions defined in phase space to pseudo-operators. As will be shown later, the Wigner transform could be regarded as the inverse Weyl quantization. The restrictions for the Wigner equation do not appear in the pseudo-operator space, and our strategy is to perform diagonalization on this space and map it back. We mention here that the way to bridge the two spaces is not unique [100] and all these mapping techniques are equivalent as summarized in [40].

Weyl Quantization

Given one function $A(x,p)$, called a symbol, in phase space, the Weyl quantization gives its associated operator $\hat{A} = \mathcal{W}(A)$ such that, for an arbitrary function $h(\mathbf{x},\mathbf{y}) \in \mathcal{S}(\mathbb{R}^d \times \mathbb{R}^d)$,

$$\mathcal{W}(A)[h](\mathbf{x}) = \hat{A}[h](\mathbf{x}) = \frac{1}{(2\pi\varepsilon)^d} \int\int A\left(\frac{\mathbf{x}+\mathbf{y}}{2}, \mathbf{p}\right) h(\mathbf{x},\mathbf{y})\, \mathrm{e}^{\frac{\mathrm{i}}{\varepsilon}(\mathbf{x}-\mathbf{y})\cdot\mathbf{p}}\, \mathrm{d}\mathbf{p}\, \mathrm{d}\mathbf{y}. \tag{3.2.5}$$

We list two examples here, both of which can be justified by simple derivation:

1. if $A(\mathbf{x},\mathbf{p})$ is the Wigner function F^ε, then

$$\hat{F}[h](\mathbf{x}) = \frac{1}{(2\pi\varepsilon)^d} \int \rho(\mathbf{x},\mathbf{x}')\, h(\mathbf{x}')\, \mathrm{d}\mathbf{x}'. \tag{3.2.6}$$

   In bra-ket language, $\hat{F}$ is exactly $\hat{\rho} = |\psi\rangle\langle\psi|$, and $\hat{F}[h](x)$ is simply $\langle x|\psi\rangle\langle\psi|h\rangle$. Therefore, in this case, the Weyl quantization can be regarded as the inverse Wigner transform.

2. if $A(\mathbf{x},\mathbf{p})$ is the Hamiltonian $H(\mathbf{x},\mathbf{p}) = \left(\frac{\mathbf{p}^2}{2} + U(\mathbf{x})\right)\mathbb{I} + V(\mathbf{x})$, then:

$$\hat{H}[h](\mathbf{x}) = \left(\frac{\varepsilon^2}{2}\Delta_{\mathbf{x}} + U(\mathbf{x})\mathbb{I}\right) h(\mathbf{x}) + V(\mathbf{x})h(\mathbf{x}) \tag{3.2.7}$$

is the Hamiltonian operator.

We also list a few important properties [103]:

(a) the mapping is one to one, and the inverse mapping is denoted by $A = \mathcal{W}^{-1}(\hat{A})$;

(b) $\mathcal{W}(A\#B) = \mathcal{W}(A)\mathcal{W}(B)$; where the operation # is called the Moyal product defined as:

$$A\#B \quad := \quad A\mathrm{e}^{\frac{\mathrm{i}\varepsilon}{2}\left(\overleftarrow{\nabla}_{\mathbf{x}}\cdot\overrightarrow{\nabla}_{\mathbf{p}}-\overleftarrow{\nabla}_{\mathbf{p}}\cdot\overrightarrow{\nabla}_{\mathbf{x}}\right)}B, \tag{3.2.8}$$

where the arrows indicate on which symbol the gradients act. Another less useful form for Moyal product is:

$$A\#B = \frac{1}{(2\pi)^{2d}}\int A\left(\mathbf{x}-\frac{\varepsilon}{2}\eta, \mathbf{p}+\frac{\varepsilon}{2}\mu\right)B(\mathbf{x}', \mathbf{p}')e^{\mathrm{i}(\mathbf{x}-\mathbf{x}')\cdot\mu+\mathrm{i}(\mathbf{p}-\mathbf{p}')\cdot\eta}\mathrm{d}\mu\mathrm{d}\mathbf{x}'\mathrm{d}\eta\mathrm{d}\mathbf{p}'.$$

Both these two formulas could be verified through simple derivation.

(c) if A is a function only on x, then $\hat{A} = A$;

von Neumann equation

With the Schrödinger equation (3.1.15), one could derive the equation satisfied by $\hat{F}$: the von Neumann equation [27, 174]:

$$\mathrm{i}\varepsilon\hat{F} = [\hat{H}, \hat{F}] \tag{3.2.9}$$

where $[\hat{A}, \hat{B}] = \hat{A}\hat{B} - \hat{B}\hat{A}$ is the commutation operation and $\hat{H}$ is the Hamiltonian defined in (3.1.18).

Remark 3.2.1. (3.2.9) *is in fact very easy to be derived in bra-ket language. We start with the Schrödinger equation* $\mathrm{i}\varepsilon\partial_t|\psi^\varepsilon\rangle = \hat{H}|\psi^\varepsilon\rangle$ *, and its conjugate transpose* $-\mathrm{i}\varepsilon\partial_t\langle\psi^\varepsilon| = \langle\psi^\varepsilon|\hat{H}$.

The von Neumann equation is their simple combination:

$$\begin{aligned}\mathrm{i}\varepsilon\partial_t\hat{F} &= \mathrm{i}\varepsilon\partial_t\left(|\psi^\varepsilon\rangle\langle\psi^\varepsilon|\right) = \mathrm{i}\varepsilon\left(\partial_t|\psi^\varepsilon\rangle\right)\langle\psi^\varepsilon| + \mathrm{i}\varepsilon|\psi^\varepsilon\rangle\partial_t\langle\psi^\varepsilon| \\ &= \hat{H}|\psi^\varepsilon\rangle\langle\psi^\varepsilon| - |\psi^\varepsilon\rangle\langle\psi^\varepsilon|\hat{H} = [\hat{H},\hat{F}].\end{aligned}$$

Here the symbol associated to $\hat{H}$ is not diagonalized. By the structure given from (3.1.19) to (3.1.21), we define $\hat{\Theta}$ and $\hat{\Theta}^\dagger$ as the Weyl quantization of Θ and $\Theta^\dagger$ respectively and multiply them on two sides of (3.2.9). Symbolically we get:

$$\mathrm{i}\varepsilon\frac{\partial\hat{F}'}{\partial t} = \left[\hat{H}',\hat{F}'\right] = \left[-\frac{\Delta_{\mathbf{x}}}{2} + U(\mathbf{x})\mathbb{I},\hat{F}'\right] + \left[\hat{\Lambda}_V,\hat{F}'\right]. \tag{3.2.10}$$

Here we use new density operator $\hat{F}' = \hat{\Theta}\,\hat{F}\,\hat{\Theta}^\dagger$, new Hamiltonian $\hat{H}' = \hat{\Theta}\,\hat{H}\,\hat{\Theta}^\dagger$. According to the definition of Θ, $\hat{\Theta}\hat{V}\hat{\Theta}^\dagger = \hat{\Lambda}_V$.

Remark 3.2.2. *Note the new Hamiltonian is not diagonalized, but the potential part is. The symbol for the new Hamiltonian is*

$$\left(\frac{\mathbf{p}^2}{2} + U(x)\right)\mathbb{I} + \Lambda_V.$$

We will show computation for the exact expression of H' later on in (3.2.17).

3.2.2 Adiabatic and non-adiabatic models

In this section we make use of the tools developed above to derive the model for both the adiabatic and non-adiabatic situations. The computation is done only formally.

Adiabatic case

Define the crossing set:

$$S = \{\mathbf{x} : \lambda^+(\mathbf{x}) = \lambda^-(\mathbf{x})\}. \tag{3.2.11}$$

In this section we deal with adiabatic case and the energy bands are assumed to be very well separated, i.e. the following assumption holds:

Assumption 3.2.1. *The eigenvalues of the system do not degenerate, i.e. the two energy bands are very well separated. In our case, where the potential term is of the form* (3.1.17)*, it is simply* $\lambda_V^- \neq \lambda_V^+ \neq 0$*, and* $u^2(\mathbf{x}) + v^2(\mathbf{x})$ *is never zero. In other words,* S *is empty.*

Following [74, 147], we define the matrices:

$$\Pi_\pm(\mathbf{x}, \mathbf{k}) = \chi^\pm(\mathbf{x}, \mathbf{k}) \otimes \chi^\pm(\mathbf{x}, \mathbf{k}).$$

They reflect the orthogonal projection onto the eigen-spaces correspond to $\lambda^\pm(\mathbf{x}, \mathbf{k})$ respectively. With these, we study the following limit of the Wigner function, defined as:

$$\lim_{\varepsilon \to 0} F(t, \mathbf{x}, \mathbf{p}) = F^0(t, \mathbf{x}, \mathbf{p}).$$

According to Theorem 6.1 in [74], outside the crossing set, the Wigner measure can be decomposed as:

$$F^0(t, \cdot) = \Pi_+ F^0(t, \cdot)\Pi_+ + \Pi_- F^0(t, \cdot)\Pi_- \tag{3.2.12a}$$

$$= f^+(t, \cdot)\Pi_+ + f^-(t, \cdot)\Pi_-, \tag{3.2.12b}$$

where $f^\pm = \mathrm{Tr}(\Pi_\pm F^0(t, \cdot))$ are scalar functions standing for the phase space probability densities on the associated energy levels. The first equivalence comes from the fact that F^0

commutes with the projection operator $\Pi_\pm$ as proved in the theorem, and (3.2.12b) is because the eigen-spaces are one-dimensional. The equation that controls the propagation of $f^\pm$ are:

$$\begin{aligned}
&\partial_t f^\pm + \nabla_{\mathbf{p}}\lambda^\pm \cdot \nabla_{\mathbf{x}} f^\pm - \nabla_{\mathbf{x}}\lambda^\pm \cdot \nabla_{\mathbf{p}} f^\pm = 0, \quad t > 0,\ \mathbf{x} \in \mathbb{R}^d \setminus S,\ \mathbf{p} \in \mathbb{R}^d, \\
&f^\pm(t=0,\mathbf{x},\mathbf{p}) = \mathrm{Tr}(\Pi_\pm F^0(t=0,\mathbf{x},\mathbf{p}))
\end{aligned} \tag{3.2.13}$$

Noted that the theory above is based on the adiabatic assumption, so that degeneracy is avoided and one is able to distinguish the projections onto different functions.

Quantum transition in non-adiabatic case

In this subsection we investigate the situation when the assumption breaks down and energy bands get close to each other. This situation cannot be avoided as long as the trajectories of $f^\pm$ come across S. In fact, dates back to [188], people have realized that the crossing set, if ever exists, is of measure zero, however, its influence is significant. In history, to fix this new problem on the degeneracy, people have tried adopting Landau-Zener formula in both Lagrangian and Eulerian approach with finite volume or Gaussian beam numerical methods [185, 136, 118, 117]. Our idea is based on the von Neumann equation in (3.2.9) [153, 154], upon which we apply the inverse Weyl map. As one can see later, this approach naturally gives the functions that represent the transition rate. It will also be shown that away from S, the transition rate is weakly zero and makes no impact on the whole system. This is consistent with (3.2.13). In this sense, we consider (3.2.13) as the weak limit of the new model.

As in (3.2.10), we have obtained a diagonalized equation on the symbolic side. To obtain its counterpart in phase function space, we perform the inverse Weyl mapping, and the new evolution equation reads as:

$$\mathrm{i}\varepsilon \frac{\partial F'}{\partial t} = \left[H', F'\right]_{\#}, \tag{3.2.14}$$

where the commutator $[A,B]_{\#} = A\#B - B\#A$ with $\#$ being the Moyal product defined in (3.2.8). and H' and F' are symbols associated to $\hat{H}'$ and $\hat{F}'$ respectively:

$$\begin{cases} H'(\mathbf{x},\mathbf{p}) = \Theta(\mathbf{x})\#H(\mathbf{x},\mathbf{p})\#\Theta(\mathbf{x})^\dagger \\ F'(\mathbf{x},\mathbf{p}) = \Theta(\mathbf{x})\#F(\mathbf{x},\mathbf{p})\#\Theta(\mathbf{x})^\dagger \end{cases} . \tag{3.2.15}$$

To further derive the equation, we truncate the Moyal product to its second order:

$$A\#B = AB - \frac{\mathrm{i}\varepsilon}{2}\{A,B\} + \mathcal{O}(\varepsilon^2), \tag{3.2.16}$$

where the Poisson bracket is $\{A,B\} = \nabla_{\mathbf{p}}A\cdot\nabla_{\mathbf{x}}B - \nabla_{\mathbf{x}}A\cdot\nabla_{\mathbf{p}}B$. Thus the symbol H' becomes:

$$H'(\mathbf{x},\mathbf{p}) = \Lambda(\mathbf{x},\mathbf{p}) + \mathrm{i}\varepsilon\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta(\mathbf{x})\Theta^\dagger(\mathbf{x}) + \frac{\varepsilon^2}{2}\nabla_{\mathbf{x}}\Theta(\mathbf{x})\cdot\nabla_{\mathbf{x}}\Theta^\dagger(\mathbf{x}). \tag{3.2.17}$$

Plug it into (3.2.14), equation for F' is obtained, and to the first order of ε, the semi-classical limit gives (derivations are omitted here and details could be found in Appendix A.3):

$$\begin{cases} \frac{\partial f^+}{\partial t} = -\mathbf{p}\cdot\nabla_{\mathbf{x}}f^+ + \nabla_{\mathbf{x}}\big(U+E_g\big)\cdot\nabla_{\mathbf{p}}f^+ + 2\xi\Re\{f^i\} \\ \frac{\partial f^-}{\partial t} = -\mathbf{p}\cdot\nabla_{\mathbf{x}}f^- + \nabla_{\mathbf{x}}\big(U-E_g\big)\cdot\nabla_{\mathbf{p}}f^- - 2\xi\Re\{f^i\} \\ \frac{\partial f^i}{\partial t} = -\mathbf{p}\cdot\nabla_{\mathbf{x}}f^i + \nabla_{\mathbf{x}}U\cdot\nabla_{\mathbf{p}}f^i + \xi(f^- - f^+) + \frac{2E_g}{\mathrm{i}\varepsilon}f^i \end{cases} . \tag{3.2.18}$$

Here $\Re$ stands for the real part, and the energy gap is $2E_g = 2\sqrt{u^2+v^2}$. ξ is defined by:

$$\xi = \frac{1}{2E_g}\mathbf{p}\cdot\left(\nabla_{\mathbf{x}}v(\mathbf{x})\cos\phi - \nabla_{\mathbf{x}}u(\mathbf{x})\sin\phi\right), \tag{3.2.19}$$

with ϕ being the angle of the potential: $\cos\phi = \frac{u(\mathbf{x})}{E_g}$ and $\sin\phi = \frac{v(\mathbf{x})}{E_g}$. The three scalar

functions are the four components of F' as in:

$$F' = \begin{pmatrix} f^+ & f^i \\ \bar{f}^i & f^- \end{pmatrix}. \tag{3.2.20}$$

We split f^i into the real part and imaginary part by defining $f^R = \frac{\sqrt{2}}{2}(f^i + \overline{f^i})$ and $f^I = \frac{\sqrt{2}}{2\mathrm{i}}(f^i - \overline{f^i})$. Then the system becomes:

$$\frac{d}{dt}\vec{f} = -\mathbf{p} \cdot \nabla_x \vec{f} + A \cdot \nabla_p \vec{f} + B \cdot \vec{f}, \tag{3.2.21}$$

with:

$$A = \mathrm{diag}(\nabla_x(U + E_g), \nabla_x(U - E_g), \nabla_x U, \nabla_x U),$$

$$B = \begin{pmatrix} 0 & 0 & \sqrt{2}\xi & 0 \\ 0 & 0 & -\sqrt{2}\xi & 0 \\ -\sqrt{2}\xi & \sqrt{2}\xi & 0 & \frac{2E_g}{\varepsilon} \\ 0 & 0 & -\frac{2E_g}{\varepsilon} & 0 \end{pmatrix}.$$

B is anti-symmetric, and thus the evolution operator is unitary, so the L_2 norm of this hyperbolic system remains unchanged. In this setting, $f^{\pm}$ reveal the projection coefficients onto different bands, and f^i naturally arise representing the transition rate between two bands. Our numerical method is designed for this equation.

Initial condition for the Wigner function in the semi-classical limit

In this subsection we derive the initial condition in the semi-classical regime. In particular, we concern about the transform of the following initial data for the Schrödinger system (1.1.1):

$$\psi^\varepsilon(t=0,\mathbf{x}) = \psi_0^\varepsilon(\mathbf{x}) = g_0^\varepsilon(\mathbf{x})\big(a^+\chi^+(\mathbf{x}) + a^-\chi^-(\mathbf{x})\big), \tag{3.2.22}$$

where, g_0^ε is an ε-scaled Gaussian packet:

$$g_0^\varepsilon(\mathbf{x}) = \frac{1}{(\pi\varepsilon)^{d/4}} \exp\left\{-\frac{1}{2\varepsilon}|\mathbf{x}-\mathbf{x_0}|^2 + \frac{\mathrm{i}}{\varepsilon}\,\mathbf{p_0}\cdot(\mathbf{x}-\mathbf{x_0})\right\}. \tag{3.2.23}$$

Plug it into the definition of the associated Wigner function (3.2.1), we get:

$$\begin{aligned} F^\varepsilon(t=0,\mathbf{x},\mathbf{p}) \quad \sim \quad & \frac{1}{(\pi\varepsilon)^d} \exp\left\{-\frac{1}{\varepsilon}|\mathbf{x}-\mathbf{x_0}|^2 - \frac{1}{\varepsilon}|\mathbf{p}-\mathbf{p_0}|^2\right\} \\ & \times \left((a^+)^2\,\chi^+\otimes\chi^+ + (a^-)^2\,\chi^-\otimes\chi^-\right). \end{aligned} \tag{3.2.24}$$

and in the zero limit of ε, it goes to:

$$\delta(\mathbf{x}-\mathbf{x_0},\mathbf{p}-\mathbf{p_0})\left((a^+)^2\,\chi^+\otimes\chi^+ + (a^-)^2\,\chi^-\otimes\chi^-\right), \quad \text{as } \varepsilon\to 0.$$

The corresponding diagonalized symbol is F' defined in (3.2.15):

$$\begin{aligned} F'(t=0,\mathbf{x},\mathbf{p}) \quad &= \quad \Theta(\mathbf{x})\#F(t=0,\mathbf{x},\mathbf{p})\#\Theta(\mathbf{x})^\dagger, \\ &\xrightarrow{\varepsilon\to 0} \quad \Theta(\mathbf{x})F(t=0,\mathbf{x},\mathbf{p})\Theta(\mathbf{x})^\dagger. \end{aligned} \tag{3.2.25}$$

To be specific, simple calculation gives:

$$\begin{cases} f^+(t=0,\mathbf{x},\mathbf{p}) = \frac{(a^+)^2}{(\pi\varepsilon)^d}\exp\left\{-\frac{1}{\varepsilon}|\mathbf{x}-\mathbf{x_0}|^2-\frac{1}{\varepsilon}|\mathbf{p}-\mathbf{p_0}|^2\right\} \\ f^+(t=0,\mathbf{x},\mathbf{p}) = \frac{(a^-)^2}{(\pi\varepsilon)^d}\exp\left\{-\frac{1}{\varepsilon}|\mathbf{x}-\mathbf{x_0}|^2-\frac{1}{\varepsilon}|\mathbf{p}-\mathbf{p_0}|^2\right\} \\ f^R(t=0,\mathbf{x},\mathbf{p}) = f^I(t=0,\mathbf{x},\mathbf{p}) \ = \ 0 \end{cases} . \tag{3.2.26}$$

3.2.3 Properties and numerical methods

It is easily seen that the system is hyperbolic, with a source term that is anti-symmetric, and thus the L_2 norm of the whole system is conserved.

As in the precious chapter, we intend to design an efficient numerical method based on the possible link between the set of equations (3.2.13) that is easier to compute and the hard one (3.2.21). The argument is very similar to what we had in the periodic potential case. In fact, as we will see later, (3.2.21) contains the full information, whether the Born-Oppenheimer approximation holds or not, and away from the crossing zone S when Assumption 3.2.1 holds, this full system automatically converges to (3.2.13). Numerically since the adiabatic model (3.2.13) is much easier to be computed, and thus one should stick to it whenever possible. Therefore, we decompose the domain into different regions depending on the gap between energy bands, and apply different models in different regions, in particular (3.2.21) around S and (3.2.13) away from it. The idea was adopted from [35], and implemented in [36]. There are many types of intersection, and we focus our attention to the case of avoided crossing [92].

Avoided crossing

Avoided crossing is one type of band structure that was introduced in [92]. For this type of band structure, the energy gap is a function not solely dependent on the spacial variable

$\mathbf{x}$, but also relies on a parameter $\delta \geq 0$: the energy gap E is always small but gets to zero if and only if $\delta = 0$. In recent years avoided crossing is extensively studied: one could find the technical analysis on the relationship between δ and ε in [172] where it is shown that the most interesting case would be when these two parameters are at the same order. This is to say, the gap between the levels shrinks at the same order the nucleus increase the mass. For different type of avoided crossings, Hagedorn and Joye were able to evaluate the propagation of the Gaussian wave packet through avoided crossing by constructing the inner and outer domain and asymptotically match the results on two sides [95, 96]. More general initial data were considered in [62, 61, 60] in the Wigner space. See also a review on both theory and numerics in [24].

One typical example for avoided crossing is given in 2D. We write $\mathbf{x} = (x, y)$ and $\mathbf{p} = (p, q)$, and use the following potential:

$$u(\mathbf{x}, \delta) = x, \quad v(\mathbf{x}, \delta) = \sqrt{y^2 + \delta^2}. \tag{3.2.27}$$

The trace $U(\mathbf{x})$ is set to be 0. So the equation (3.2.21) becomes:

$$\begin{cases} \frac{\partial f^+}{\partial t} = -\mathbf{p} \cdot \nabla_{\mathbf{x}} f^+ + \nabla_{\mathbf{x}} E_g \cdot \nabla_{\mathbf{p}} f^+ + \sqrt{2}\, \xi\, f^R \\ \frac{\partial f^-}{\partial t} = -\mathbf{p} \cdot \nabla_{\mathbf{x}} f^- - \nabla_{\mathbf{x}} E_g \cdot \nabla_{\mathbf{p}} f^- - \sqrt{2}\, \xi\, f^R \\ \frac{\partial f^R}{\partial t} = -\mathbf{p} \cdot \nabla_{\mathbf{x}} f^R + \frac{2}{\varepsilon} E_g f^I + \sqrt{2}\, \xi\, (f^- - f^+) \\ \frac{\partial f^I}{\partial t} = -\mathbf{p} \cdot \nabla_{\mathbf{x}} f^I - \frac{2}{\varepsilon} E_g f^R \end{cases} \tag{3.2.28}$$

where in this case:

$$E_g = \sqrt{x^2 + y^2 + \delta^2}, \quad \xi = \frac{1}{2(x^2 + y^2 + \delta^2)} (qx \frac{y}{\sqrt{y^2 + \delta^2}} - p\sqrt{y^2 + \delta^2}).$$

The potential matrix is marked as $V(x,\delta)$ with its two eigenvalues:

$$\lambda^{\pm}(\mathbf{x},\mathbf{p},\delta)=\frac{\mathbf{p}^2}{2}\pm E_g, \tag{3.2.29}$$

With zero δ, it becomes conical crossing with the two eigenvalues touching at the origin, i.e. $\lambda^{+}(\mathbf{x}_c,\mathbf{p},\delta)=\lambda^{-}(\mathbf{x}_c,\mathbf{p},\delta)$ when and only when $\delta=0$. $\mathbf{x}_c$ represents the crossing point, and is the origin in this case.

Weak limit and domain decomposition

Numerically, the main difficulty comes from the high oscillations introduced by the small ε. When ε is extremely small, with $O(1)$ $\sqrt{x^2+y^2}$, f^R and f^I present very rapid oscillations with the wavelength at the scale of ε. However, one could check, that when E is nonzero, both of these two oscillatory functions f^R and f^I have zero weak limit as $\varepsilon\to 0$, and formally, the system goes to the adiabatic formulation:

$$\begin{cases}\frac{\partial f^+}{\partial t}=-\mathbf{p}\cdot\nabla_{\mathbf{x}}f^+ +\nabla_{\mathbf{x}}E_g\cdot\nabla_{\mathbf{p}}f^+,\\ \frac{\partial f^-}{\partial t}=-\mathbf{p}\cdot\nabla_{\mathbf{x}}f^- -\nabla_{\mathbf{x}}E_g\cdot\nabla_{\mathbf{p}}f^-.\end{cases} \tag{3.2.30}$$

The situation here is very similar to the periodic case, where we adopt the domain decomposition idea and solve appropriate set of equation in different region: in the regions where the oscillation is strong, we compute the weak limit (3.2.13) as a good approximation, and in the regions where the function varies slowly, we stick to the original equation (3.2.28). Two main problems remain to be resolved: one is where to set up the interface and how to give the boundary condition. They will be answered after the following argument about the convergence.

Take $\mathbf{p}>0$ and $\mathbf{x}_0<0$ for example:

Case 1. If $E_g \gg C_0\sqrt{\varepsilon}$ before hitting the crossing zone, then f^R and f^I are of $o(\sqrt{\varepsilon})$;

Case 2. If $-C_0\sqrt{\varepsilon} < E_g < C_0\sqrt{\varepsilon}$, then f^R and f^I are of $\mathcal{O}(\sqrt{\varepsilon})$, and f^R and f^I are slowly varying, i.e. $\partial_t f^i \ll \mathcal{O}\left(\frac{1}{\varepsilon}\right)$;

Case 3. If $E_g \gg C_0\sqrt{\varepsilon}$ after hitting the crossing zone, f^R and f^I are highly oscillatory with mean 0.

As such, we propose the following decomposition:

Classical regions: $E_g > C_0\sqrt{\varepsilon}$: In this region, a coarse mesh independent of ε is used to solve the adiabatic Liouville system (3.2.30). f^R and f^I are set to be zero.

Semi-classical region: $E_g \leq C_0\sqrt{\varepsilon}$: Solve (3.2.28). The incoming boundary conditions for f^R and f^I are set to be zero, and the incoming boundary condition for f^+ and f^- are the inflow boundary condition. A fine mesh is used with Δx and Δt much less than $\sqrt{\varepsilon}$.

3.3 Numerical examples

In this section we show several numerical results and compare with the solution to the original Schrödinger equation. Let $\psi^\varepsilon(t, \mathbf{x})$ be the solution of (3.1.15) with initial data given by (3.2.22). For the lower(-) and upper(+) bands, we define the position density $\rho^{\pm}_{schr}$ and the population $P^{\pm}_{schr}$:

$$\rho^{\pm}_{schr}(t, \mathbf{x}) = |\Pi_{\pm}\psi^\varepsilon(t, \mathbf{x})|^2, \quad \text{and} \quad P^{\pm}_{schr} = \int_{\mathbb{R}^d} \rho^{\pm}_{schr}(t, \mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{3.3.1}$$

Correspondingly, we recover these quantities from the semi-classical Liouville system (3.2.21) by

$$\rho^{\pm}_{liou}(t, \mathbf{x}) = \int_{\mathbb{R}^d} f^{\pm}(t, \mathbf{x}, \mathbf{k})\mathrm{d}\mathbf{k}, \quad \text{and} \quad P^{\pm}_{liou} = \int_{\mathbb{R}^d} \rho^{\pm}_{liou}(t, \mathbf{x})\, \mathrm{d}\mathbf{x}. \tag{3.3.2}$$

Define L^1 error in the cumulative distribution function (c.d.f.) [82, 125]:

$$\mathrm{Err}^{\varepsilon} = \int_{\mathbb{R}^d} \left\{ \left| \int_{\Omega_x} \left(\rho^+_{schr}(\mathbf{y}) - \rho^+_{liou}(\mathbf{y})\right) \mathrm{d}\mathbf{y} \right| + \left| \int_{\Omega_x} \left(\rho^-_{schr}(\mathbf{y}) - \rho^-_{liou}(\mathbf{y})\right) \mathrm{d}\mathbf{y} \right| \right\} \mathrm{d}\mathbf{x},$$

where, $\Omega_{\mathbf{x}} = \left\{ \mathbf{y} \in \mathbb{R}^d \ : \ y_i \leq x_i, \ i = 1, \cdots, d \right\}$ for $\mathbf{x} = (x_1, \cdots, x_d) \in \mathbb{R}^d$.

3.3.1 Example 1: 1D, pure state initial data

In this example, we deal with the problem in 1D with a pure initial data. Set u and v as

$$u(x) = x, \quad \text{and} \quad v(x) \equiv \delta = \frac{\sqrt{\varepsilon}}{16}. \tag{3.3.3}$$

δ is of $\mathcal{O}(\sqrt{\varepsilon})$, larger than 0 and is independent of the spatial variable x. The difference between the two energy bands is $\Delta E = 2E$ that gets its minimum 2δ at $x = 0$. The initial data for the Schrödinger equation is given in (3.2.22-3.2.23) with $a^+ = 1$, $a^- = 0$, $x_0 = 5\sqrt{\varepsilon}$, and $p_0 = -1$. For comparison, we compute both the Schrödinger equation as a reference, and the Liouville system. The Schrödinger equation is computed using the classical time-splitting spectral method, with $\Delta x = \varepsilon/8$ and $\Delta t = 5\varepsilon^{\frac{3}{2}}$ and Liouville semi-classical system is done through the domain decomposition method. In Figure 3.3.1 we compare the results $\rho^{\pm}$ given by the two systems, and in Figure 3.3.2 we check the evolution of the population on the first band P^+ along the time. In Figure 3.3.3, it is showed that the domain decomposition modeling error (3.3) decreases as $\mathcal{O}(\varepsilon)$.

3.3.2 Example 2: 1D, mixed state initial data

In this example, we have mixed states as the initial data in one dimension. u and v are defined in the same way as in (3.3.4). Initial wave packet for the Schrödinger equation is given in (3.2.22-3.2.23) with $a^+ = a^- = 1/\sqrt{2}$, $x_0 = 5\sqrt{\varepsilon}$, and $p_0 = -1$, thus is mixed states. In

Figure 3.3.4 we compare the numerical results on $\rho^{\pm}$ and P^{+} of the Schrödinger equation and the Liouville system, and in Figure 3.3.5, the cumulative error is given as a function of ε. The numerical results show $\mathcal{O}(\varepsilon)$ decay in error.

3.3.3 Long time behavior

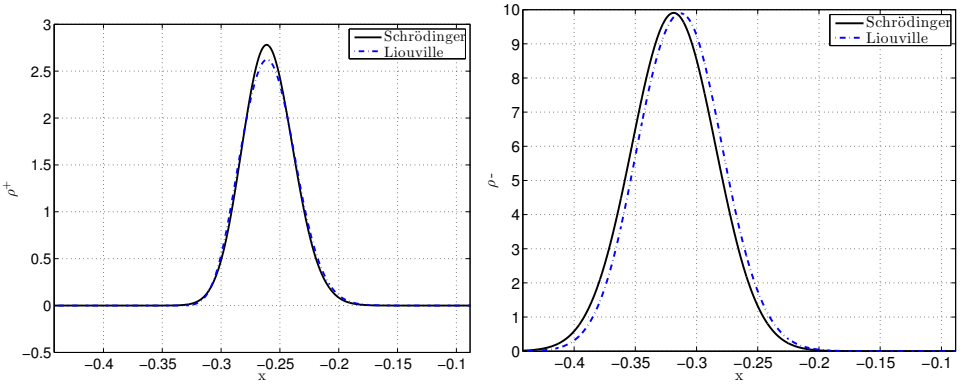
In this subsection we consider the evolution that is long enough so that the data goes through two surface hopping processes. The numerical results are provided for both the original Schrödinger equation and its semi-classical Liouville system. Initial data is a Gaussian packet as in (3.2.22)-(3.2.23) with $a^{+} = 1$, $a^{-} = 0$, $x_0 = 0.3125$, and $p_0 = -1$. ε is set as 2^{-10}. In Figure 3.3.6 we show the evolution of P^{+} with respect to time. Initially, the profile is located to the right of the crossing point, the origin in this case, with negative average velocity, it goes to the origin and around $t = 0.25$ when the center of the packet roughly hits the crossing point, more than half of the portion jump over to the lower energy level, so that we have two packets along the two separate bands after that, and both of them keep moving towards left. The packet on the lower energy level keeps accelerating as the potential diminishes, so it never comes back. But the one on the higher energy level gradually decelerates since its potential gets higher and higher, and at some point, the velocity decreases to zero and the packet starts bouncing back towards the origin. Around $t = 0.275$, it gets to the crossing point for the second time, and undergoes another hopping process. The results for the two systems match quite well.

3.3.4 Example 4: 2D pure initial data

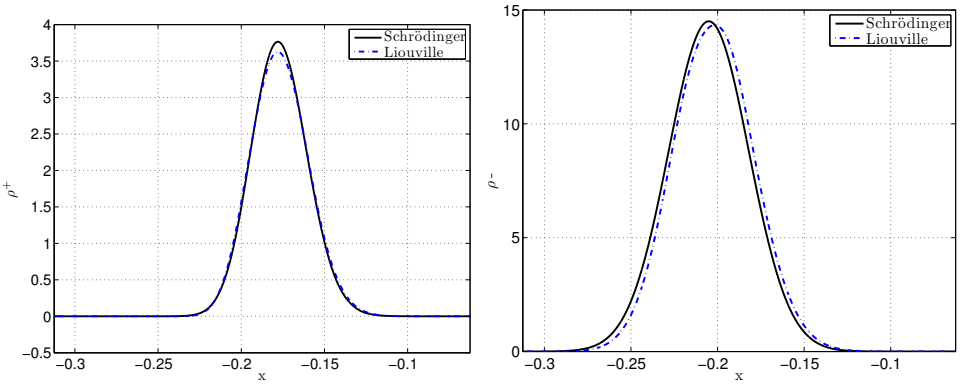
In this example, we deal with the problem in 2D with a pure initial data. Set u and v as

$$u(x) = x, \quad \text{and} \quad v(x) = \sqrt{y^2 + \delta^2}. \tag{3.3.4}$$

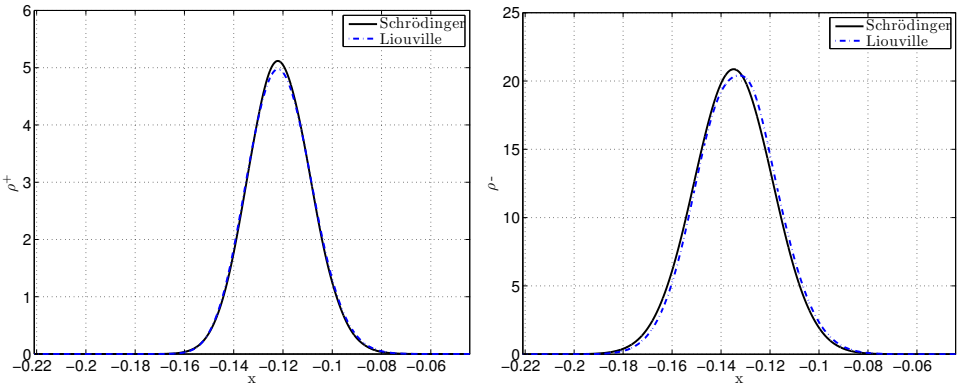
δ is chosen at $\sqrt{\varepsilon}/2$, larger than 0 and is independent of the spatial variable x. The difference between the two energy bands is $\Delta E = 2E$ that gets its minimum 2δ at $x = y = 0$. The initial data for the Schrödinger equation is given in (3.2.22-3.2.23) with $a^+ = 1$, $a^- = 0$, $x_0 = 5\sqrt{\varepsilon}$, $y_0 = 0$, $p_0 = -1$, and $q_0 = 0$. For comparison, we compute both the Schrödinger equation as a reference, and the Liouville system. The Schrödinger equation is computed using the classical time-splitting spectral method, with $\Delta x = \Delta y = \varepsilon/8$ and $\Delta t = 5\varepsilon^{\frac{3}{2}}$ and Liouville semi-classical system is done through the domain decomposition method with $\Delta x = \Delta p = h$ in the classical regions and $\Delta x = \Delta p = h/2$ in then semi-classical region, where $h = \mathcal{O}(\sqrt{\varepsilon})$. In Figure 3.3.7 we compare the results $\rho^{\pm}$ given by the two systems, and in Figure 3.3.8 we check the evolution of the population on the first band P^+ along the time. In Figure 3.3.9, it is shown that the domain decomposition modeling error (3.3) decreases as $\mathcal{O}(\sqrt{\varepsilon})$.



(a) $\varepsilon = 2^{-9}$

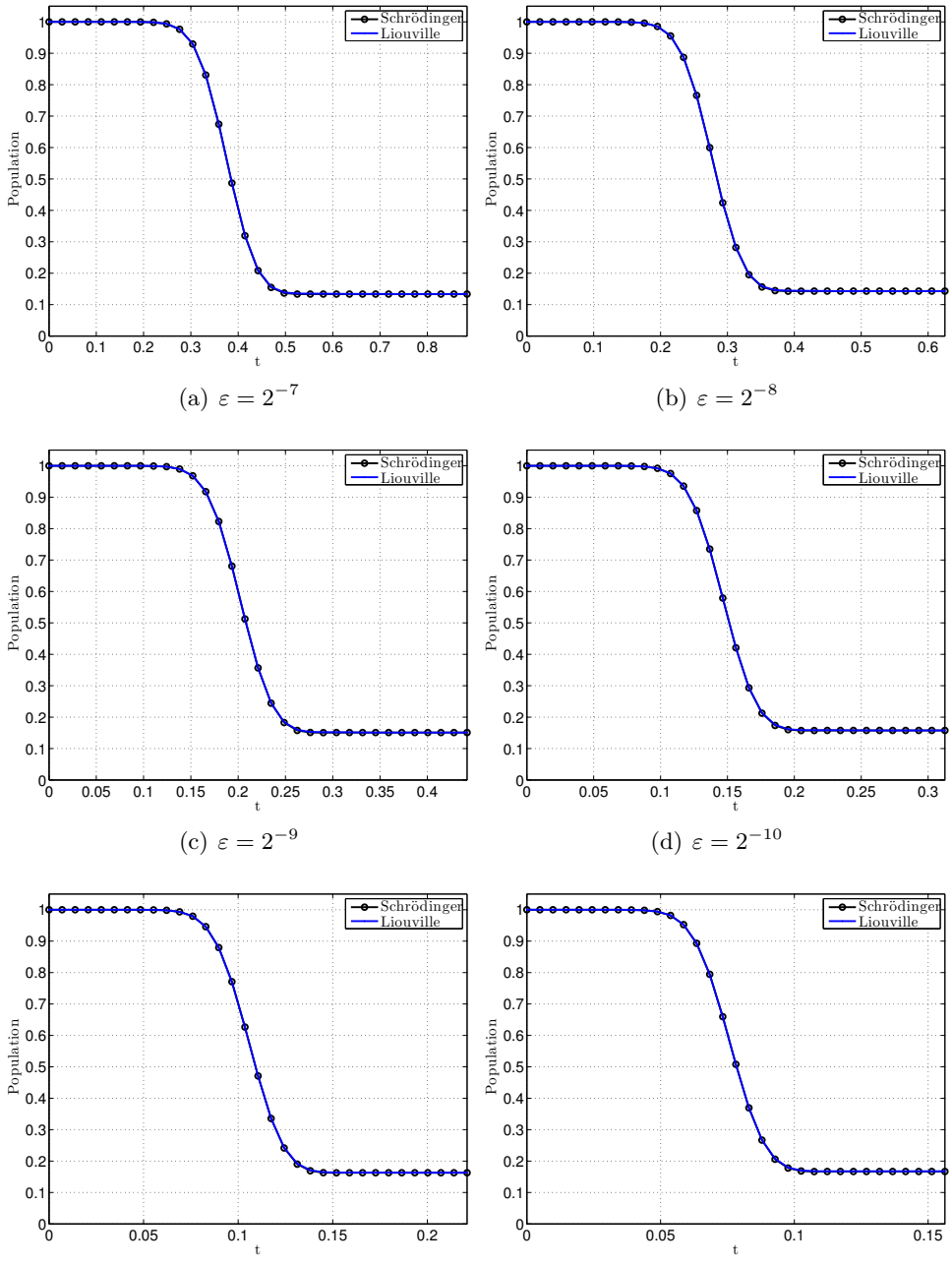


(b) $\varepsilon = 2^{-10}$



(c) $\varepsilon = 2^{-11}$

Figure 3.3.1: Surface hopping problem. Example 1. The left column is for $\rho^{+}_{schr/liou}$, the density on the upper band, and the right column gives the results of $\rho^{-}_{schr/liou}$. $\Delta = \sqrt{\varepsilon}/2$.



(a) $\varepsilon = 2^{-7}$ (b) $\varepsilon = 2^{-8}$

(c) $\varepsilon = 2^{-9}$ (d) $\varepsilon = 2^{-10}$

(e) $\varepsilon = 2^{-11}$ (f) $\varepsilon = 2^{-12}$

Figure 3.3.2: Surface hopping problem. Example 1. Time evolution of the population (defined in (3.3.1) and (3.3.2)) on the upper band $P^{+}_{schr/liou}$. $\Delta = \sqrt{\varepsilon}/2$.

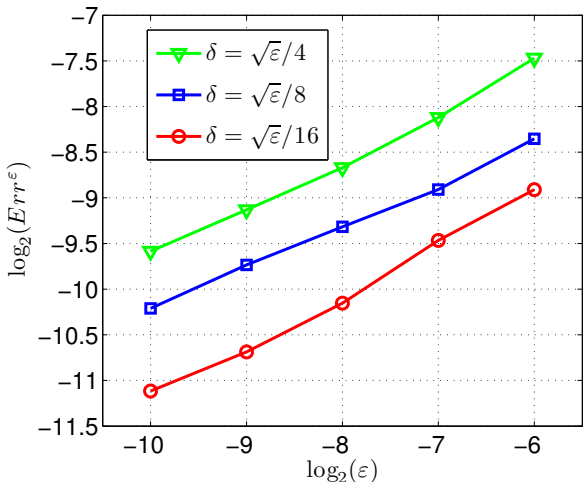
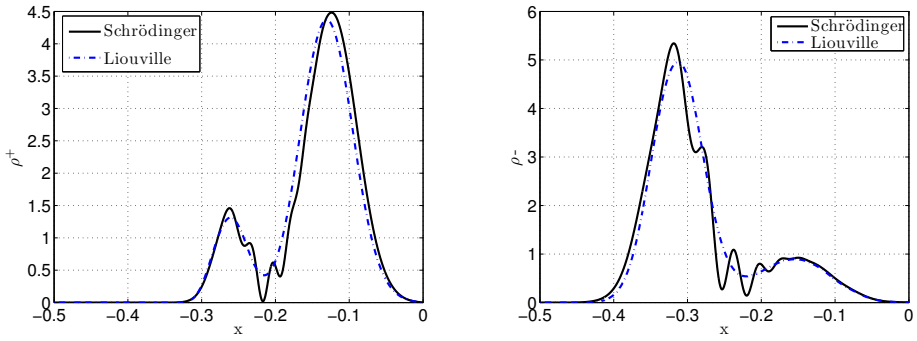
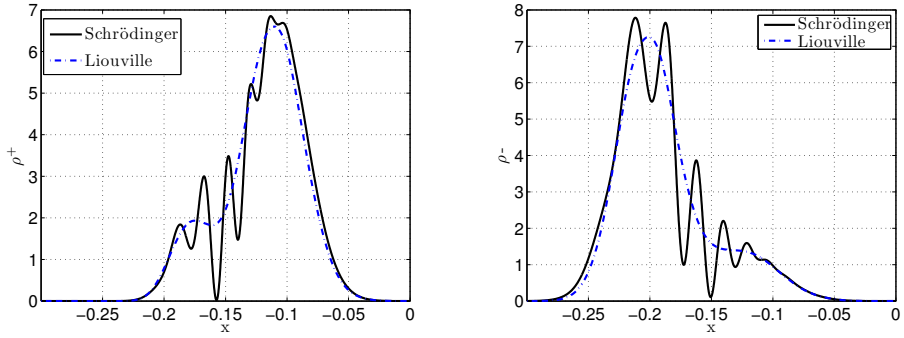


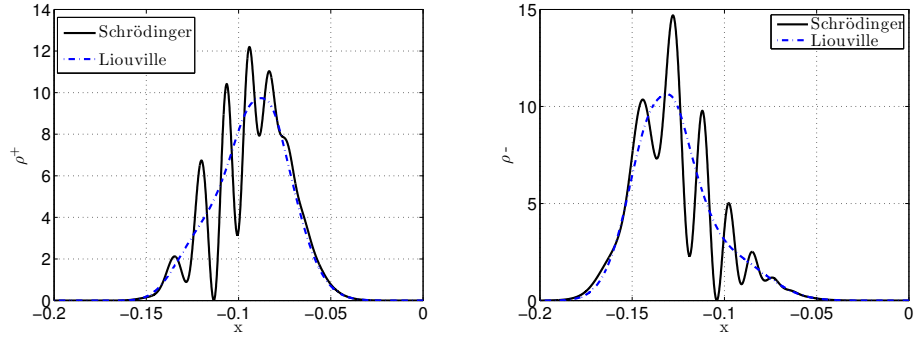
Figure 3.3.3: Surface hopping problem. Example 1. Err^ε decreases at $O(\varepsilon)$. $t = 10\sqrt{\varepsilon}$.



(a) $\varepsilon = 2^{-9}$



(b) $\varepsilon = 2^{-10}$



(c) $\varepsilon = 2^{-11}$

Figure 3.3.4: Surface hopping problem. Example 2. $\Delta = \sqrt{\varepsilon}/2$. The left column is the density on the upper band $\rho^+_{schr/liou}$ and the right column is $\rho^-_{schr/liou}$.

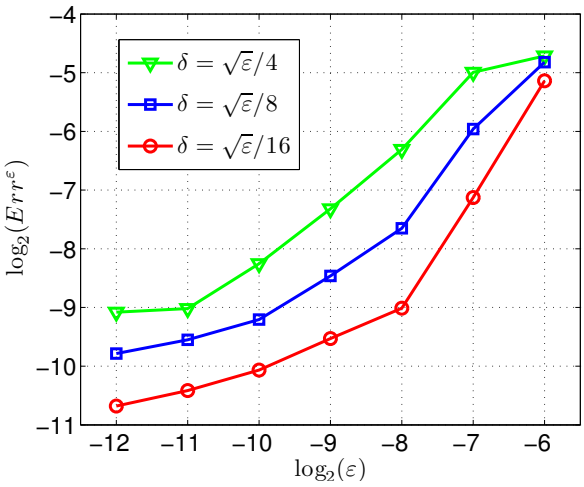


Figure 3.3.5: Surface hopping problem. Example 2. Err^ε as a function of ε at $t = 10\sqrt{\varepsilon}$.

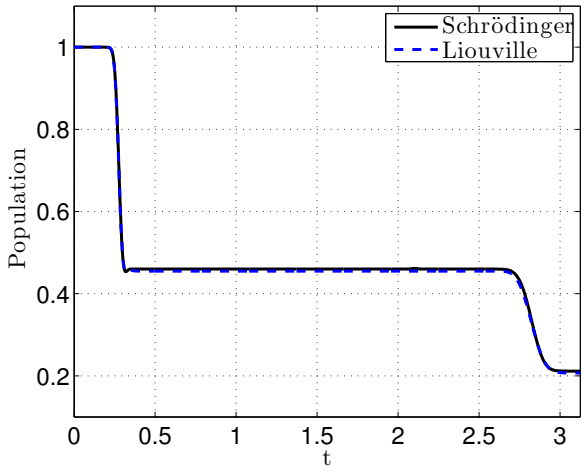
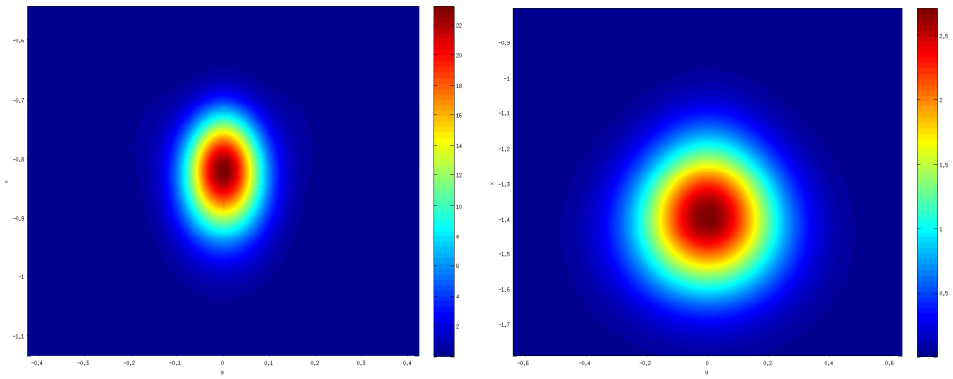
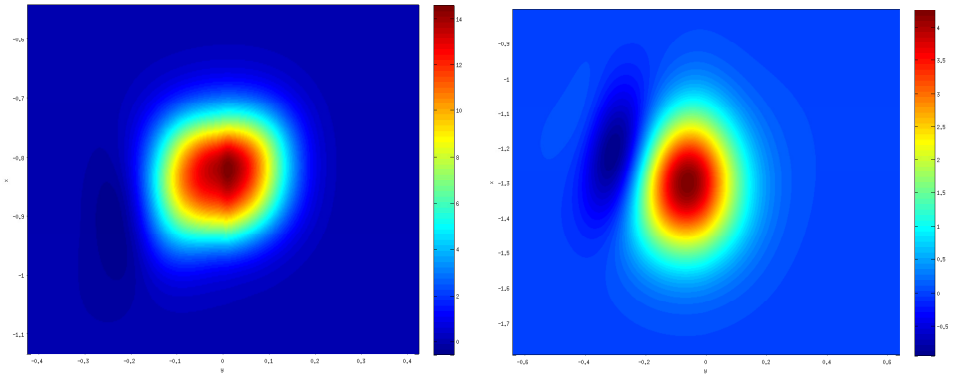


Figure 3.3.6: Surface hopping problem. Example 3. Long time behavior of P^+.

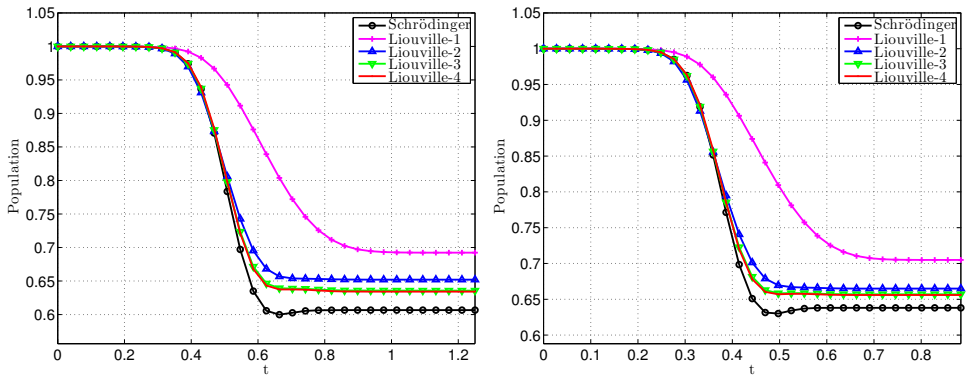


(a) Schrödinger Solutions



(b) Liouville Solutions

Figure 3.3.7: Surface hopping problem. Example 4. The left column is for $\rho^{+}_{schr/liou}$, the density on the upper band, and the right column gives the results of $\rho^{-}_{schr/liou}$. $\delta = \sqrt{\varepsilon}/2$.



(a) $\varepsilon = 2^{-6}$ (b) $\varepsilon = 2^{-7}$

(c) $\varepsilon = 2^{-8}$ (d) $\varepsilon = 2^{-9}$

Figure 3.3.8: Surface hopping problem. Example 4. Time evolution of the population on the upper band $P^{+}_{schr/liou}$. $\delta = \sqrt{\varepsilon}/2$. The legend "Schrödinger" represents the solution of the Schrödinger equation, "Liouville-j" represents the solution of the Liouville system by the domain decomposition method with $\Delta x = \Delta p = h$ in the classical regions and $\Delta x = \Delta p = h/2$ in then semi-classical region, where $h = \sqrt{\varepsilon}/2^{j-1}$, and $j = 1, 2, 3, 4$

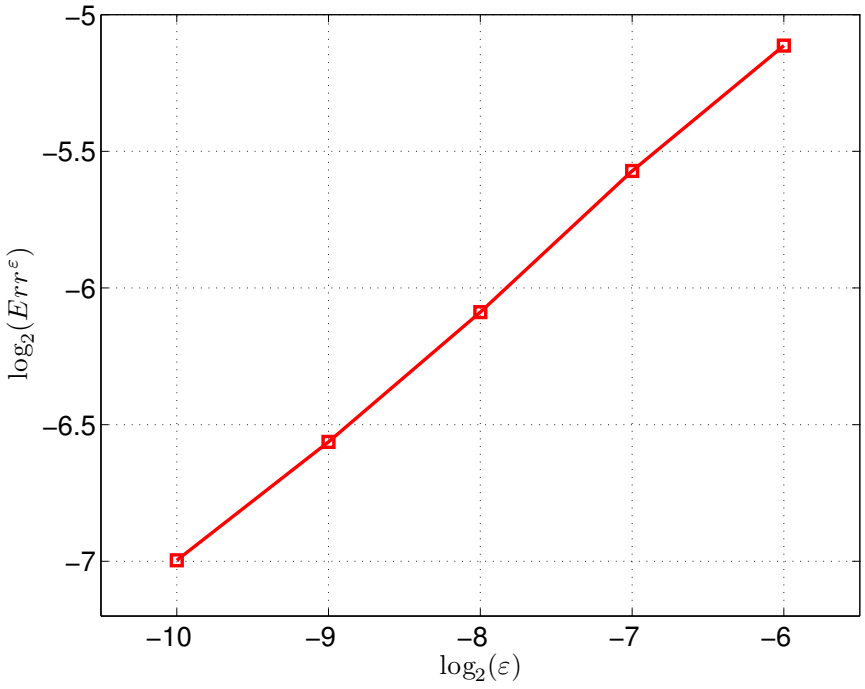


Figure 3.3.9: Surface hopping problem. Example 4. Err^ε at $t = 10\sqrt{\varepsilon}$. The Liouville system by the domain decomposition method with $\Delta x = \Delta p = \sqrt{\varepsilon}/8$ in the classical regions and $\Delta x = \Delta p = \sqrt{\varepsilon}/16$ in then semi-classical region.

Part II

Kinetic Theory

In this part we study the hydrodynamic limit of kinetic theory. We are especially interested in the compressible Euler limit of the Boltzmann equation, and focus on designing proper numerical algorithms for it.

The Boltzmann equation describes the evolution of the probability density distribution of rarefied gas. When the typical length and time scale is big enough, mathematically the Knudsen number, a parameter that is the ratio of local mean free path and the typical domain length, is small, and the rarefied gas system could be treated as continuum. We are interested in finding the connection between these two levels of descriptions on the numerical level.

The efforts to connect statistical mechanics and fluid dynamics date back to Boltzmann, who tried to develop "mathematically the limiting processes" that "lead from the atomistic view to the laws of motion of continua". Mathematical treatment started booming since Hilbert marked it as his 6th problem in 1900. Since then, it has attracted a great amount of attention: people study this problem out of mathematical curiosity, and out of practical concern. Mathematically, the problem is two fold: how to study the entropy decay, and how to link the kinetic equation to the fluid dynamics. See an early book [34] by Cercignani and the most advanced analysis collected in [186, 173] by Villani and Saint-Raymond respectively.

On the theoretical level, the picture is rather clear. The formal connection between the (in)compressible Euler or Navier-Stokes and the Boltzmann equation has been established, and the rigorous proof is complete for the link to the incompressible Navier-Stokes equation. Our target to present the same connection on the numerical level: as the Euler equations are the hydrodynamic limit of the Boltzmann equation, we are seeking accurate numerical methods for the Boltzmann equation that capture the solution to the Euler equations automatically in this limit.

To this end, we introduce the concept, "asymptotic preserving". As will be clearer later on, the whole area explores possibilities of relaxing discretization from the strong requirement

of the scaling parameter. Typically, an asymptotic preserving method is a cheap solver that preserves the asymptotic limit, on the numerical level. The idea has been very widely used in implementing multi-scale problems, but in this thesis, we only focus our attention on one of its many facets: which is the link between the Boltzmann equation and the compressible Euler equations.

In chapter 4, as a introduction, we briefly review the established analytical background of the Boltzmann equation. In chapter 5 we go over two major asymptotic preserving methods currently available, to be specific, the BGK penalization method and the Exponential Runge-Kutta method. Both of the two methods were initially designed for single species systems, and we explore their extension to multi-species systems in chapter 6.

Chapter 4

Kinetic theory and the hydrodynamic limit

In this chapter we briefly go over the basics of kinetic theory and show the hydrodynamic limit of the Boltzmann equation.

4.1 Boltzmann equation

Kinetic theory aims at describing a system constituted of a large number of particles on a mesoscopic level. We usually model the system by a distribution function f in phase space. In case of monatomic gas, the distribution function is $f = f(t, x, v)$, that depend on time $t > 0$, position $x \in \Omega$ and velocity $v \in \mathbb{R}^3$, presenting the number of particles in an infinitesimal volume $\mathrm{d}x\mathrm{d}v$ around the point (x, v) at time t.

Remark 4.1.1. *The Boltzmann equation is not ab initio, and its derivation, though not within our focus in this thesis, is a remarkable thing by itself. The whole process was pioneered by Grad in 1958 in [85], and the rigorous proof on the validity of the Boltzmann equation is later on obtained in [132] for very short time, also see a review [133]. The general approach is that, one starts with a collection of N particles, and traces trajectory of each particle. The particle obeys Newton's 2nd law, so the system is time-reversible. A distribution function f^N is a delta function that lives on $6N$ dimensional space $\mathbb{R}^{3N} \times \mathbb{R}^{3N}$ ($3D$ for position and $3D$ for momentum for each particle in the system). f is seen as the marginal distribution*

function of one particle of f^N. From f^N to f, some assumptions are to be made: N is big enough and particles are considered identical, so that f^N is sufficiently close to tensor product of N one particle distribution function fs (this is the molecular chaos assumption). With these assumptions, in the so-called Grad-Boltzmann limit, one gets the Boltzmann equation. However, for the Boltzmann equation, time has an arrow and the dynamics is irreversible, unlike the Hamiltonian equation that governs f^N. This inconsistency reflects the fact that atomistic dynamics is not the only mechanism that should be taken into account, and in fact all the other influences are subtly hidden in the assumptions made in the derivation, the details of which is another topic that we will not discuss here.

The Boltzmann equation reads:

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f, f),\ t \geq 0,\ (x, v) \in \mathbb{R}^d \times \mathbb{R}^d. \tag{4.1.1}$$

d is the dimension. Here $\partial_t f + v \cdot \nabla_x f$ part is to formulate the free transport: particles move in a straight line with velocity v. The Q is called the collision term, and in general has the following form:

$$Q(f, f) = Q^+ - fQ^- = \int_{S^{d-1}} \int_{\mathbb{R}^d} (f' f'_* - f f_*) B(|v - v_*|, \omega) dv_* d\omega. \tag{4.1.2}$$

This term means that two particles with pre-collisional velocities, one with v and another with v_*, bump into each other, and after colliding, they change their velocities to post-collisional velocities: v' and v'_*. If the collision is elastic, momentum and energy are conserved, i.e.

$$v + v_* = v' + v'_*, \quad |v|^2 + |v_*|^2 = |v'|^2 + |v'_*|^2,$$

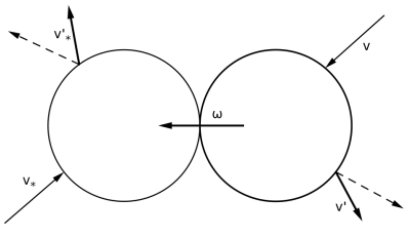


Figure 4.1.1: Velocity change after the collision.

and this leads to:

$$v' = v - \frac{1}{2}(g - |g|\omega), \quad v'_* = v_* + \frac{1}{2}(g - |g|\omega). \tag{4.1.3}$$

where ω is a unit vector on S^{d-1}, the unit sphere defined in R^d space. See figure 4.1.1. We use the shorthands $f' = f(t, x, v')$ and $f'_* = f(t, x, v'_*)$. Q^+ and fQ^- parts are called gaining and losing part respectively, reflecting how many more or fewer particles having the velocity v. B is called collisional cross-section, or collision kernel. It is always positive, and is used to measure the repartition of post-collisional velocities given the pre-collisional velocities. It varies according to the nature of particles considered. For hard sphere particles, $B(z, \omega) \sim z$. For particles governed by power law, $B(z, \omega) = b(\omega)|z|^\gamma$, and specifically when $\gamma = 0$, the particles are called the Maxwell molecule. The Maxwell molecule is not physical but is used very often for the purpose of obtaining explicit calculation in agreement with physical observables. In physics, numerous examples have grazing collision [50, 47], and possess non-integrable $b(\omega)$, so mathematically we have to go through "cut-off" process by replacing it with some bounds that are integrable [87].

There are several assumptions we made: firstly, we assume particles change direction only through colliding with other particles. Since collision is short range, the interaction is very sensitive to the exact value of both position and velocity. If long range potential is considered, the equation would undergo dramatic change, for example one gets Vlasov equation in Coulomb

case where collision term is replaced by effective field [26]. Secondly, we have only considered binary collisions: in rarefied gas, the collisions that involve more than three particles are so rare and are simply ignored. Finally, we assume the collision is elastic, and this does not have to be the real case: energy decay sometimes takes place for granular gases when collisions are inelastic [187, 18, 32, 71].

4.1.1 Properties of the collision term

In this subsection we state some important properties of the collision term [33]. All the derivations are purely formal.

Conservation law

Cross-section may vary, but the first $d+2$ moments of the collision term are always zero. In fact, given arbitrary smooth function ϕ, we have:

$$\begin{aligned}\int Q(f,f)\phi(v)\mathrm{d}v &= \int B(f'f'_* - ff_*)\phi \mathrm{d}v\mathrm{d}v_*\mathrm{d}\omega \\ &= \frac{1}{2}\int B(f'f'_* - ff_*)(\phi+\phi_*)\mathrm{d}v\mathrm{d}v_*\mathrm{d}\omega \\ &= \frac{1}{4}\int B(f'f'_* - ff_*)(\phi+\phi_*-\phi'-\phi'_*)\mathrm{d}v\mathrm{d}v_*\mathrm{d}\omega\end{aligned}$$

In the derivation, we used that the cross-section B is invariant to $|v-v'|$. For elastic collision, immediately, one has that the quadratic functions in v are collisional invariants: let $\phi = \left(1, v, \frac{1}{2}|v|^2\right)^T$, the first $d+2$ moments are zero:

$$\begin{aligned}< Q > &= \int Q(f)dv = 0, \\ < vQ > &= \int vQ(f)dv = 0, \\ < \frac{1}{2}v^2Q > &= \int \frac{1}{2}|v|^2Q(f)dv = 0.\end{aligned} \tag{4.1.4}$$

Based on these formulas, we take moments of the Boltzmann equation and obtain conservation laws. To do that, we firstly define the following macroscopic quantities:

$$\begin{aligned}
\rho &= \int f \mathrm{d}v, & \rho u &= \int v f \mathrm{d}v, \\
E &= \frac{1}{2}\rho u^2 + \rho e = \frac{1}{2}\int |v|^2 f \mathrm{d}v, & e &= \frac{3}{2}T = \frac{1}{2\rho}\int f|v-u|^2 \mathrm{d}v, \\
S &= \int (v-u)\otimes(v-u) f \mathrm{d}v, & q &= \frac{1}{2}\int (v-u)|v-u|^2 f \mathrm{d}v.
\end{aligned} \tag{4.1.5}$$

ρ is called local density, T is local temperature and u is local bulk velocity. All these quantities are "local" in space. With these, we have:

$$\begin{aligned}
&\partial_t \rho + \nabla_x \cdot (\rho u) = < Q >= 0, \\
&\partial_t (\rho u) + \nabla_x \cdot (S + \rho u^2) = \frac{1}{\varepsilon} < vQ >= 0, \\
&\partial_t E + \nabla_x \cdot (Eu + Su + q) = \frac{1}{\varepsilon} < \frac{1}{2}|v|^2 Q >= 0.
\end{aligned} \tag{4.1.6}$$

H theorem

The conservation law does not reflect the built in time-irreversibility of the Boltzmann equation. The irreversibility is seen through the so called "Boltzmann's H-theorem", which basically says that entropy of a system keep decaying in time. We multiply Q with $\log f$. Simple algebra gives:

$$\int Q(f,f)\log f \mathrm{d}v = \frac{1}{4}\int B(f'f'_* - ff_*)\log\frac{ff_*}{f'f'_*}\mathrm{d}v\mathrm{d}v_*\mathrm{d}\omega \leq 0$$

Note again the discussion on integrability and such are all waived, and the computation is purely formal.

From here we have two conclusions.

- Entropy decay. Define entropy as $\int f \log f \mathrm{d}v$, then we have:

$$\partial_t \int f \log f \mathrm{d}v + \nabla_x \cdot \int v f \log f \mathrm{d}v = \int Q(f,f) \log f \mathrm{d}v \leq 0 \tag{4.1.7}$$

- Entropy achieves its minimum when the system gets to the equilibrium state. Mathematically we set $\int Q(f,f) \log f \mathrm{d}v = 0$, and the solution to this is that $\log f$ is a quadratic function, i.e. f is a Gaussian distribution. We call the Gaussian function that are equipped with proper macroscopic quantities Maxwellian (macroscopic quantities are computed from f and thus f appears as the subindex):

$$M_f(t,x,v) = \rho(t,x) \left(\frac{1}{2\pi T(t,x)} \right)^{d/2} \exp\left(-\frac{(v-u(t,x))^2}{2T(t,x)} \right). \tag{4.1.8}$$

4.1.2 Non-dimensionalization and rescaling process

In different regimes, one observes the system with different time and length scales and acquires different behavior of the solution. We firstly perform the non-dimensionalization. Denote l_0 and t_0 as length and time scale, and a reference temperature is marked as T_0. The thermal speed is defined out of it:

$$c \sim \sqrt{\frac{k_b T_0}{m}} \tag{4.1.9}$$

with k_b being the Boltzmann constant and m is the molecular mass. Rescale the length and time as following:

$$\tilde{t} = \frac{t}{t_0}, \quad \tilde{x} = \frac{x}{l_0}, \quad \text{and } \tilde{v} = \frac{v}{c}. \tag{4.1.10}$$

Another time scale associated with the intensity of the collision cross-section is called mean free time τ, a parameter used to reflect the time a particle undergoes between two collisions.

$$\int M_f(v)M_f(v_*)B\mathrm{d}\sigma\mathrm{d}v_*\mathrm{d}v = \frac{N}{l_0^3\tau}, \tag{4.1.11}$$

where M_f is the Maxwellian defined in (4.1.8). Evidently, bigger B means that the system has more frequent collisions, and thus the mean free time is shorter. Mean free path is

$$\lambda = c\tau, \tag{4.1.12}$$

reflecting the distance a particle travels between two collisions. Finally we re-normalize the number density and the collision kernel as well:

$$\tilde{f}(\tilde{t},\tilde{x},\tilde{v}) = \frac{l_0^3c^3}{N}f(t,x,v), \quad \tilde{B}(\tilde{v}-\tilde{v}_*,\sigma) = \rho\tau B(v-v_*,\sigma). \tag{4.1.13}$$

The equation thus becomes (dropping all the tildes):

$$\frac{l_0}{ct_0}\partial_t f + v\cdot\nabla_x f = \frac{l_0}{\lambda}Q(f,f). \tag{4.1.14}$$

Define the Knudsen number as the mean free path over typical domain length $\frac{\lambda}{l_0}$, denoted by Kn. It is divided from the collision term Q. The ratio in front of the $\partial_t f$ term is called the Strouhal number, denoted by St. We tune these two parameters and could observe various phenomena in different regime. In general, as $Kn \to 0$, one gets the Euler equations in the first asymptotic limit, and the Navier-Stokes equation in the second. If St is similarly small, the macroscopic equation are incompressible; and if it is $\mathcal{O}(1)$, one gets compressible limit.

4.2 Asymptotic limit

We only derive the compressible Euler limit in this thesis, i.e. Knudsen number is small, replaced by ε, while the Strouhal number is of $\mathcal{O}(1)$:

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f, f). \tag{4.2.1}$$

While deriving for asymptotic limit of the Boltzmann equation, different ansatz would give different hierarchies of PDEs. In history, there have been two dominating approaches: one is called Hilbert expansion, that dates back to Hilbert's original work in [101], in which he sought for solutions of formal power series in ε:

$$f(t, x, v) = \sum_k \varepsilon^k f_k(t, x, v). \tag{4.2.2}$$

Another approach is called Chapman-Enskog expansion developed independently by Chapmann and Enskog [37]. Their idea is to expand the solution close to the local Maxwellian:

$$f(t, x, v) = M_f \left(1 + \sum_{k>0} \varepsilon^k g_k(t, x, v) \right). \tag{4.2.3}$$

Here M_f is the local Maxwellian function given in (4.1.8), and g is regarded as perturbation that should vanish at infinite time. These two approaches give different expansion processes, but in the leading order of ε, both of them imply that the distribution function should be a Gaussian function, and in particular, Chapmann-Enskog expansion assign this Gaussian function the local Maxwellian:

$$f_0 = M_f = \rho(t, x) \left(\frac{1}{2\pi T(t, x)} \right)^{d/2} \exp\left(-\frac{(v - u(t, x))^2}{2T(t, x)} \right),$$

with ρ, T u defined as in (4.1.5). The idea of Chapmann-Enskog expansion is that since M_f and f has the same moments, and thus one could plug it back into (4.1.6) and close the system. In this leading order, the compressible Euler equation is obtained:

$$\begin{aligned} &\partial_t \rho + \nabla \cdot (\rho u) = 0, \\ &\partial_t (\rho u) + \nabla \cdot (\rho u \otimes u + \rho T \mathbb{I}) = 0, \\ &\partial_t E + \nabla \cdot ((E + \rho T) u) = 0. \end{aligned} \tag{4.2.4}$$

This finishes the leading order asymptotic expansion. In the next order we obtains compressible Navier-Stokes limit, by plugging $f = M_f(1 + \varepsilon g_1)$ back into the original Boltzmann equation, and solving for the kernel of linearized Boltzmann collision term. We omit the process here.

Remark 4.2.1. *A side note on rigorous proofs on the Boltzmann equation. In general there are three ways in connecting the two set of equations. One approach is based on truncated asymptotic expansions. Rigorous justification is done by Caflisch [30] for compressible Euler equation, up to the appearance of the singular time for the Euler equation. Similar result for incompressible Navier-Stokes equations is obtained by De Masi, Esposito and Lebowitz [44], where they also require smoothness of the solutions to the macroscopic equations. Another approach is under the perturbation framework: one expands the distribution function around a renormalized target Maxwellian (in infinite time), and analyze the spectrum of the linearized Boltzmann equation. The spectrum was firstly explored by Ellis and Pinskii [59]. The idea was later used by Nishida for compressible Euler equations [159], and by Bardos and Ukai for incompressible Navier-Stokes equations [8]. As in the first approach, the results obtained rely on the existence of classical solutions to the macroscopic equations, and thus either becomes invalid in finite time or requires smallness on the initial data. However, physically, one should expect that the existence and smoothness of the Boltzmann equation should not rely on that*

of the macroscopic equations. In fact, the global existence is done in a series of work by Lions, initiated in [54] and summarized in [146]. To get rid of these non-physical assumptions, Bardos, Golse and Levermore proposed the third approach [6, 5], which is to link the two sets of equations in the weak sense, and they sought for possibility to obtain Leray [140] solutions to incompressible Navier-Stokes equation. The program was complete by Golse and Saint-Raymond in [78, 77]. None of these theoretical works will be discussed in this thesis, but they are milestones in the development of the whole theory.

Chapter 5

Asymptotic preserving numerical methods

In this chapter we investigate the asymptotic preserving numerical methods. Before going into details of two methods in the following two sections, we firstly briefly introduce this concept.

Usually multi-scale problems contain some parameters that could vary across several scales. As the parameters are manually sent either big or small, one carries asymptotic expansions on these parameters, yielding asymptotic limits that are described by averaged or effective equations. In our case, the Euler equations are the asymptotic limit of the Boltzmann equation as the Knudsen number goes to zero.

Numerically, in these asymptotic regimes, the original equation could be difficult to compute: the small scales need to be resolved in discretization, generating tremendous computational cost. So whenever possible, we tend to simply solve the asymptotic or macroscopic equation as it is more efficient. However, in practise, usually the macroscopic model breaks down in part of the domain, and two or more regimes co-exist. When this happens, a natural idea is to do decompose the domain into parts and compute different sets of equations in different regions. Under this framework, one needs to know specifically where to set up the boundary and how to couple the two systems, but neither is easy. Some kinetic equations can be computed in this line of thinking including [23, 45, 46, 76, 128, 138], but the idea does not

work well for the Boltzmann equation. In fact, understanding the structure of the boundary layer is a challenging problem unsolved for decades by itself [14, 7, 4].

Another approach is to develop asymptotic-preserving (AP) schemes, an approach that we are pursuing in this thesis. It firstly appeared for neutron transport equation in the diffusive regime [134, 135], and was widely used later on in many other models. The idea of this concept is to develop cheap numerical methods only for the microscopic equation, and the methods should be robust in all regimes, so they automatically capture the asymptotic limit in the stiff regime, with fixed discretization. More accurate error analysis could be found in [111]. Specifically, in our case, for the Boltzmann equation, as summarized by Jin [110], a scheme is AP if:

- it preserves the discrete analogy of the Chapman-Enskog expansion, namely, it is a suitable scheme for the kinetic equation, yet, when holding the mesh size and time step fixed and letting the Knudsen number go to zero, the scheme becomes a suitable scheme for the limiting fluid dynamic Euler equations;
- implicit collision terms can be implemented efficiently.

Typically, to relax the time step from the control of the Knudsen number, one needs to use an L-stable numerical scheme, and thus the collision term should be computed implicitly. However, as seen above, the Boltzmann collision term is non-local and nonlinear, and numerically inverting it is impossible. All of these conflicts are making the designing of AP schemes for the Boltzmann equation difficult.

There are several variations of the AP property, with some of them easier to achieve than others. They are weakly-AP, relaxed-AP, and strongly-AP, defined as follows (see [65]):

- *weakly-AP*. If the data are within $O(\varepsilon)$ of the local equilibrium initially, they remain so for all future time steps;

- *relaxed-AP*. For non-equilibrium initial data, the solution will be projected to the local equilibrium beyond an initial layer (after several time steps).

- *strongly-AP*. For non-equilibrium initial data, the solution will be projected to the local equilibrium immediately in one time step.

In general, the strongly-AP property is preferred, and was the designing principle for most classical AP schemes [109, 29]. The relaxed-AP is a concept introduced recently in [65], which was shown numerically to be sufficient to capture the hydrodynamic limit when the Knudsen number goes to zero. The weakly-AP is even weaker: it requires the solutions do not move away from the equilibrium if they are close to it initially. It serves as a necessary condition for the AP property.

We mention here several AP schemes recently developed. One approach was to use the micro-macro decomposition method [139] (see its multi-species extension in [119]), but the issue of designing an efficient implicit collision term, which is necessary for numerical stability independent of the Knudsen number, is still unsolved. An earlier approach introduced by Gabetta, Pareschi and Toscani uses the truncated Wild Sum for uniform numerical stability of the collision term [70]. In the following sections, we introduce in detail two methods. The first one is the BGK penalization introduced in [65] and its stronger version can be found in [199]. The idea was later on extended to hyperbolic systems [68], Fokker-Planck-Landau equation [123], quantum Boltzmann equation [64], quantum Fokker-Planck-Landau equation [104] and multi-species system [112]. Higher order accuracy in time is achieved with IMEX in [53]. The second one is the Exponential Runge-Kutta method, which was developed in [52] for homogeneous equation, and later on generalized by Pareschi and the author for space inhomogeneous case [143].

5.1 BKG-penalization method

This approach was firstly developed in [65]. The idea is to use a BGK operator $\beta(M_f - f)$ [15] to approximate the collision term $Q(f, f)$ and the difficulty in inverting Q is transfered to implicitly computing the BGK operator. The detail is the following: we split the collision operator into the BGK part and the rest. Since the BGK term is a good approximation to the collision term, after the subtraction, the remainder is less stiff, and is treated explicitly, and the BGK term is stiff, and should be computed implicitly. However, although it is also a nonlinear term on f, and is not invertible either, it is discovered in [42, 167] that this term could be computed in an explicit manner. We firstly rewrite the equation in the following:

$$\partial_t f + v \cdot \nabla_x f = \frac{Q(f) - P_b(f)}{\varepsilon} + \frac{P_b(f)}{\varepsilon}. \tag{5.1.1}$$

with $P_b = \beta(M_f - f)$ is a BGK term standing for penalization, and β is constant to be determined. $P_b(f)$ is expected to cancel the leading stiff order in Q, the first term on the right is less stiff, and one could use explicit methods to handle it without worrying about the stability requirements on time discretization. The second term however, needs to be treated implicitly, for example:

$$\frac{f^{l+1} - f^l}{\Delta t} + v \cdot \nabla_x f^l = \frac{Q(f^l) - P_b(f^l)}{\varepsilon} + \frac{P_b(f^{l+1})}{\varepsilon}. \tag{5.1.2}$$

The superscript l stands for the time step. We also use shorthands $Q^l \triangleq Q(f^l)$, $P_b^l \triangleq P_b(f^l)$ for convenience. With simple algebra, one gets:

$$f^{l+1} = \frac{\varepsilon f^l + \Delta t(Q^l - \beta^l(M_f^l - f^l)) - \varepsilon h v \cdot \nabla_x f^l + \beta^{l+1} h M_f^{l+1}}{\varepsilon + \beta^{l+1} h}. \tag{5.1.3}$$

h is time discretization. The computation of $v \cdot \nabla_x f$ is very mature, and the reader could refer to [141] for varies numerical shock-capturing highly accurate methods. The computation of $Q(f)$ is also a topic that have been studied through years [156, 165, 67]. We mainly focus on computing M_f and selection of β in the following.

- **The computation of M_f^{l+1}:** One only needs the evaluation of the macroscopic quantities for the next time step in order to compute M_f. To do that, we take the first $d+2$ moments numerically of the numerical scheme (5.1.2). The moments on the right hand side disappear and one gets:

$$\begin{aligned}
\rho^{l+1} &= \rho^l - h \int v \cdot \nabla_x f^l \, \mathrm{d}v, \\
(\rho u)^{l+1} &= (\rho u)^l - h \int v \otimes v \nabla_x f^l \, \mathrm{d}v, \\
E^{l+1} &= E^l - h \int \frac{1}{2} |v|^2 v \cdot \nabla_x f^l \, \mathrm{d}v,
\end{aligned}$$

  and the temperature T^{l+1} is simply: $T^{l+1} = \frac{2E^{l+1} - (\rho u^2)^{l+1}}{dn^{l+1}}$. With these, M_f is defined by (4.1.8).

- **The selection of β:** The selection of β could be tricky. Generically, it plays a role of Frechet derivative:

$$Q(f) \sim Q(M_f) + \nabla Q(M)[M - f]. \tag{5.1.4}$$

  Considering $Q(M_f) = 0$, β should be chosen to simulate $\nabla Q(f)$. In fact, often in time, one chooses $\beta^l = \sup \left| \frac{Q(f)}{f - M_f} \right|$, or $\beta^l = \sup \left| \frac{Q^l - Q^{l-1}}{f^l - f^{l-1}} \right|$. Later on in [199], the authors discovered that we could tune this parameter to ensure the positivity of the distribution

function. To be more clear, if β satisfies the following:

$$\begin{cases} \beta^l = \sup\, Q^-(f^l) \\ \beta^{l+1} = \beta^l(1+h\kappa^l), \quad \text{with } \kappa^l = \max\left\{\sup\, \frac{M_f^l - M_f^{l+1}}{hM_f^{l+1}}, 0\right\} \end{cases}.$$

then the distribution function in (5.1.3) is positive.

This method was proved to be relaxed-AP in [65], and the strong-AP version is obtained in [199]. The proof will be omitted from here.

5.2 Exponential Runge-Kutta method

This is another AP method proposed in [52] that essentially could achieve arbitrary high order of accuracy. It was designed for the spatial homogeneous Boltzmann equation, namely the transport term $v \cdot \nabla_x f$ is not included. For space inhomogeneous case, one has to use time-splitting, for example, the Strang splitting, which is 2nd order in time, therefore the high accuracy obtained in the collision step is lost. This drawback was later on fixed in [143] where the authors directly manipulate the entire space inhomogeneous Boltzmann equation without splitting. The immediate consequence is the arbitrarily high order of accuracy.

The idea is to reformulate the equation in a way so that the explicit methods, when applied upon them, automatically pushed the distribution function to the right Maxwellian. Some new parameters appear in the reformulation, and numerically one needs to select values for them in a smart way. In fact, up to now two choices are found, both of which are proved to lead to AP. In the following subsections, we describe the reformulation of the equation and present the two numerical methods in 5.2.1, followed with the numerical properties in subsection 5.2.2. The numerical examples will be given in subsection 5.2.3.

5.2.1 Numerical methods

In this subsection we firstly reformulate the Boltzmann equation so that the Exponential RK method could be carried out, and then introduce the two numerical schemes.

Reformulation of the problem and notations

We need to formulate the equation in a way such that generic explicit methods, as long as consistent, could also preserve the right asymptotic limit. Inspired by [52], we have the following:

$$\begin{aligned}
\partial_t \left[(f-\tilde{M})e^{\mu t/\varepsilon}\right] = & \quad \partial_t(f-\tilde{M})e^{\mu t/\varepsilon} + (f-\tilde{M})\tfrac{\mu}{\varepsilon}e^{\mu t/\varepsilon} && (5.2.1)\\
= & \left[\tfrac{1}{\varepsilon}(Q+\mu f-\mu\tilde{M}) - \partial_t\tilde{M} - v\cdot\nabla_x f\right]e^{\mu t/\varepsilon} && (5.2.2)\\
. = & \left[\tfrac{1}{\varepsilon}(P-\mu\tilde{M}) - \partial_t\tilde{M} - v\cdot\nabla_x f\right]e^{\mu t/\varepsilon}.
\end{aligned}$$

In the derivation, μ is a constant, and $\tilde{M}$ is an arbitrary function, called equilibrium function, the scheme for whom will be carefully designed later. Note that $\tilde{M}$ is not the local Maxwellian M_f. P is usually regarded as the gain part of the collision, defined as:

$$P = Q + \mu f. \tag{5.2.3}$$

The equation obtained is always equivalent to the original Boltzmann equation, regardless of the form of $\tilde{M}$ and μ, but numerically, as we will see later, to capture the correct asymptotic limit, they are selected under restrictive criteria. In fact, if one applies the simple forward Euler scheme onto it, clearly one gets f going to $\tilde{M}$ exponentially fast. So to realize asymptotic preserving properties, $\tilde{M}$ has to be chosen close enough to the real limit, the local Maxwellian M_f. To do that, we have two methods, and depending on whether or not $\tilde{M}$ change with respect to time in each time step, we have ExpRK-F (standing for fixed $\tilde{M}$) and ExpRK-V

(varying $\tilde{M}$) respectively.

Exponential RK schemes with fixed equilibrium function

In the equation (5.2.1), it is easy to see that the trickiest term is the time evolution of the equilibrium function $\partial_t \tilde{M}$. So in our first try, we explore possibility to get rid of this term, i.e. we look for an a-priori constant function so that $\partial_t \tilde{M}$ is manually erased from the equation in each time step. If succeed, (5.2.1) is modified accordingly:

$$\partial_t \left[(f - \tilde{M}) e^{\mu t/\varepsilon} \right] = \left[\frac{1}{\varepsilon} (P - \mu \tilde{M}) - v \cdot \nabla_x f \right] e^{\mu t/\varepsilon}. \tag{5.2.4}$$

As argued above, $\tilde{M}$ need to be chosen close enough to the Maxwellian M_f. However, as one could expect, this asymptotic limit itself is a function of time and $\partial_t M_f \neq 0$. So we are going to numerically evolve M_f firstly, and set $\tilde{M}$ as the updated M_f for each time step. The general explicit Runge-Kutta framework to numerically solve (5.2.4) is:

$$\begin{cases} \text{Step } i\text{:} & (f^{(i)} - \tilde{M}) e^{c_i \lambda} = (f^l - \tilde{M}) + \sum_{j=1}^{i-1} a_{ij} \frac{h}{\varepsilon} \left[P^{(j)} - \mu \tilde{M} - \varepsilon v \cdot \nabla_x f^{(j)} \right] e^{c_j \lambda} \\ \text{Final Step:} & (f^{l+1} - \tilde{M}) e^{\lambda} = (f^l - \tilde{M}) + \sum_{i=1}^{\nu} b_i \frac{h}{\varepsilon} \left[P^{(i)} - \mu \tilde{M} - \varepsilon v \cdot \nabla_x f^{(i)} \right] e^{c_i \lambda} \end{cases}, \tag{5.2.5}$$

where $\tilde{M}$ is defined by the macroscopic quantities $(\rho^{l+1}, u^{l+1}, E^{l+1})$ computed from:

$$\begin{cases} \text{Step } i: & \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}^{(i)} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}^{l} - h \sum_{j=1}^{i-1} a_{ij} \nabla_x \cdot \begin{pmatrix} \rho u \\ \rho u \otimes u + \rho T \\ (E + \rho T)\, u \end{pmatrix}^{(j)} \\ \text{Final Step:} & \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}^{l+1} = \begin{pmatrix} \rho \\ \rho u \\ E \end{pmatrix}^{l} - h \sum_{i=1}^{\nu} b_i \nabla_x \cdot \begin{pmatrix} \rho u \\ \rho u \otimes u + \rho T \\ (E + \rho T)\, u \end{pmatrix}^{(i)} . \end{cases} \tag{5.2.6}$$

We explain the notations above, ν is the number of stages of the Runge-Kutta method, with a_{ij} and b_i as its coefficients. $c_i = \sum_j a_{ij}$. Superscript n denote the l-th time step, and (i) denote the sub-stage i in l-th time step. h is the time discretization, and $P^{(j)} = P(f^{(j)})$ for simplicity.

Remark 5.2.1.

- *Note that this method gives us a simple way to couple macro-solver with micro-solver. When ε is considerably big, the accuracy of the method is controlled by the micro-solver. And as ε vanishes, the method pushes f going to M, which is defined by macroscopic quantities computed through the Euler equation while the order of accuracy is given by the macro-solver.*

- *In principle it is possible to adopt other strategies to compute a more accurate time independent equilibrium function in intermediate regions. For example one can use the ES-BGK Maxwellian [66] at time $n+1$, or one can use the Navier-Stokes equation as the macro-counterpart. Here however we do not explore further in these directions.*

In each time step, we update macroscopic quantities (5.2.6) first, and use them to define

$\tilde{M}$. Plug it back into (5.2.5) and update the distribution function f.

Exponential Runge-Kutta schemes with time varying equilibrium function

In this approach we keep the $\partial_t \tilde{M}$ term and try to find an efficient numerical method to compute it. In fact, we consider the most natural choice of equilibrium function, namely the local Maxwellian $\tilde{M} = M_f$. Therefore, $\tilde{M}$ is a Gaussian profile that has the same first $d+2$ moments with f, and the evolution of the moments are governed by:

$$\partial_t \int \phi f dv + \int \phi v \cdot \nabla_x f dv = 0, \tag{5.2.7}$$

with $\phi = \left[1, v, \frac{v^2}{2}\right]^T$. The RK methods are:

Step i:

$$\begin{cases} (f^{(i)} - M^{(i)})e^{c_i\lambda} &= (f^l - M^l) \\ &\quad + \sum_{j=1}^{i-1} a_{ij} \frac{h}{\varepsilon} \left[P^{(j)} - \mu M^{(j)} - \varepsilon v \cdot \nabla_x f^{(j)} - \varepsilon \partial_t M^{(j)}\right] e^{c_j\lambda}, \\ \int \phi f^{(i)} dv &= \int \phi f^l dv + \sum_{j=1}^{i-1} a_{ij} \left(-h \int \phi v \cdot \nabla_x f^{(j)} dv\right); \end{cases} \tag{5.2.8a}$$

Final Step:

$$\begin{cases} (f^{l+1} - M^{l+1})e^{\lambda} &= (f^l - M^l) \\ &\quad + \sum_{i=1}^{\nu} b_i \frac{h}{\varepsilon} \left[P^{(i)} - \mu M^{(i)} - \varepsilon v \cdot \nabla_x f^{(i)} - \varepsilon \partial_t M^{(i)}\right] e^{c_i\lambda}, \\ \int \phi f^{l+1} dv &= \int \phi f^l dv + \sum_{i=1}^{\nu} b_i \left(-h \int \phi v \cdot \nabla_x f^{(i)} dv\right). \end{cases} \tag{5.2.8b}$$

Note that the coupling between the two equation is embedded: $M^{(j)}$ are computed through $\int \phi f^{(j)} \mathrm{d}v$, and will be used to update $f^{(i)}$ for $i > j$. In each sub-stage i, one needs to evaluate

the following: $f^{(j)}$, $M^{(j)}$, $\partial_t M^{(j)}$, $P^{(j)}$, $v\cdot\nabla_x f^{(j)}$ for all $j<i$, and $M^{(i)}$ that is evaluated at the current time sub-stage. $f^{(j)}$ and $M^{(j)}$ are obtained from previous time sub-stages, and $P^{(j)}$ could be computed through spectral method developed in [165], for the $N\log N$ improvement, see [156]. $v\cdot\nabla_x f^{(j)}$ is computed using standard ENO or WENO method to match up the time accuracy designed. We focus on computation of $M^{(i)}$ and the tricky term $\partial_t M$:

Computation of $M^{(i)}$:

solve the second equation of (5.2.8a), to get evaluation of macroscopic quantities at $t^l+c_i h$. Then the Maxwellian $M^{(i)}$ is given by (4.1.8).

Computation of $\partial_t M^{(j)}$:

Write $\partial_t M$ as

$$\partial_t M=\partial_\rho M\partial_t\rho+\nabla_u M\cdot\partial_t u+\partial_T M\partial_t T, \tag{5.2.9}$$

and $\partial_t\rho$, $\partial_t u$ and $\partial_t T$ can be computed from taking moments of the original equation

$$\begin{aligned}\partial_t\begin{pmatrix}\rho\\ \rho u\\ \frac{d\rho T}{2}+\frac{1}{2}\rho u^2\end{pmatrix} &= \partial_t\int\begin{pmatrix}1\\ v\\ \frac{v^2}{2}\end{pmatrix}M dv=\partial_t\int\begin{pmatrix}1\\ v\\ \frac{v^2}{2}\end{pmatrix}f dv\\ &= -\int\begin{pmatrix}1\\ v\\ \frac{v^2}{2}\end{pmatrix}v\cdot\nabla_x f dv.\end{aligned} \tag{5.2.10}$$

To be specific, with data at sub-stage (j) in d-dimensional space, one has

$$\partial_t M^{(j)}=\partial_\rho M^{(j)}\partial_t\rho^{(j)}+\nabla_u M^{(j)}\cdot\partial_t u^{(j)}+\partial_T M^{(j)}\partial_t T^{(j)}, \tag{5.2.11}$$

with

$$\partial_\rho M^{(j)} = \frac{M^{(j)}}{\rho^{(j)}},\ \partial_u M^{(j)} = M^{(j)}\frac{v-u^{(j)}}{T^{(j)}},$$
$$\partial_T M^{(j)} = M^{(j)}\left[\frac{(v-u^{(j)})^2}{2(T^{(j)})^2} - \frac{d}{2T^{(j)}}\right],$$

and all the macroscopic quantities associated with $f^{(j)}$, namely $\rho^{(j)}$, $u^{(j)}$ and $T^{(j)}$ are given by:

$$\begin{aligned}
\partial_t \rho^{(j)} &= -\int v\cdot\nabla_x f^{(j)}\mathrm{d}v,\\
\partial_t u^{(j)} &= \frac{1}{\rho^{(j)}}\left(u^{(j)}\int v\cdot\nabla_x f^{(j)}dv - \int v\otimes v\cdot\nabla_x f^{(j)}\mathrm{d}v\right),\\
\partial_t T^{(j)} &= \frac{1}{d\rho^{(j)}}\left(-\frac{2E^{(j)}}{\rho^{(j)}}\partial_t\rho^{(j)} - 2\rho^{(j)}u^{(j)}\partial_t u^{(j)} - \int v^2 v\cdot\nabla_x f^{(j)}\mathrm{d}v\right).
\end{aligned}$$

5.2.2 Properties of ExpRK schemes

The numerical schemes we designed possess some very nice features. We are especially interested in positivity and asymptotic preserving.

Positivity and monotonicity properties

The distribution function describes the number of particles at certain phase point, and should keep being positive along the evolution. However, it is hard to maintain positivity numerically, especially when the schemes used are high order. Luckily, our ExpRK-F method, under some very mild assumption, shows positivity. To that end, we use a very powerful tool discovered in [191] and called Shu-Osher representation, and the analysis follows [83, 84]. We review Shu-Osher representation first before presenting the proof.

For a large set of ODEs:

$$\partial_t y = F(t,y), \tag{5.2.12}$$

The general Runge-Kutta writes as:

$$\begin{cases} \text{Step } i\text{:} & y^{l,(i)} = y^l + h\sum_{j=1}^{i-1} a_{ij} F(t^l + c_j h, y^{l,(j)}) \\ \text{Final step:} & y^{l+1} = y^l + h\sum_{i}^{\nu} b_i F(t^l + c_i h, y^{l,(i)}) \end{cases}. \tag{5.2.13}$$

We intend to represent it as:

$$\begin{cases} \text{Step } i\text{:} & y^{(i)} = \sum_{j=1}^{i-1} \left[\alpha_{ij} y^{(j)} + h\beta_{ij} F(t^l + c_j h, y^{(j)})\right], \\ \text{Final step:} & y^{l+1} = \sum_{j=1}^{\nu} \left[\alpha_{\nu+1j} y^{(j)} + h\beta_{\nu+1j} F(t^l + c_j h, y^{(j)})\right]. \end{cases} \tag{5.2.14}$$

Note that there are more parameters one could adjust in (5.2.14) than in (5.2.13), and to make these two formulas equivalent, the selection of coefficients in (5.2.14) is not unique. For consistency, we require $\sum_{j=1}^{i-1} \alpha_{ij} = 1$, and for β, we use the most natural choice:

$$\beta_{ij} = \alpha_{ij}\left(c_i - c_j\right). \tag{5.2.15}$$

Through simple algebra, one could prove that this choice is equivalent to the original one proposed in [191] where $\beta_{ij} = a_{ij} - \sum \alpha_{ik} a_{kj}$. In fact, assume one has $y^{(j)} = y^l + h\sum_{k=1}^{j-1} a_{jk} F^{(k)}$, $\forall\ j < i$, where $F^{(k)}$ is a shorthand for $F(t^l + c_k h, y^{(k)})$, then, our choice gives:

$$\begin{aligned} y^{(i)} &= \sum_{j=1}^{i-1} \left[\alpha_{ij} y^{(j)} + \alpha_{ij}(c_i - c_j) h F^{(j)}\right] \\ &= \sum_{j<i} \left[\alpha_{ij}\left(y^l + h\sum_{k<j} a_{jk} F^{(k)}\right) + \alpha_{ij}(c_i - c_j) h F^{(j)}\right] \\ &= y^l + h\sum_{j<i}\left(\sum_{k=j+1}^{i-1} \alpha_{ik} a_{kj} + \alpha_{ij}(c_i - c_j)\right) F^{(j)} \end{aligned} \tag{5.2.16}$$

This clearly requires $a_{ij} = \alpha_{ij}(c_i - c_j) + \sum \alpha_{ik} a_{kj}$, which is exactly the classical Shu-Osher representation: $a_{ij} = \beta_{ij} + \sum \alpha_{ik} a_{kj}$.

With this, our numerical scheme could be transformed into:

$$\begin{cases} f^{(i)} - \tilde{M})e^{c_i\lambda} = \sum_j e^{c_j\lambda}\left\{\alpha_{ij}(f^{(j)} - \tilde{M}) + \beta_{ij}\frac{h}{\varepsilon}\left[P^{(j)} - \mu\tilde{M} - \varepsilon v\cdot\nabla_x f^{(j)}\right]\right\} \\ (f^{l+1} - M)e^{\lambda} = \sum_j e^{c_j\lambda}\left\{\alpha_{\nu+1j}(f^{j} - \tilde{M}) + \beta_{\nu+1j}\frac{h}{\varepsilon}\left[P^{(j)} - \mu\tilde{M} - \varepsilon v\cdot\nabla_x f^{(j)}\right]\right\} \end{cases}.$$

We claim under very mild assumption, the ExpRK-F method gives positive distribution function. The assumption we are going to make is:

Assumption 5.2.1. *For a given $f > 0$ there exists $h^* > 0$ such that*

$$f - h\, v\cdot\nabla_x f > 0, \qquad \forall\ 0 < h \le h^*.$$

Note that this assumption is very mild, and is the minimal requirement on f in order to obtain a non negative scheme. The assumption is naturally valid since $f - v\cdot\nabla_x f$ is simply the forward Euler scheme for a transport equation with constant coefficient (v in this case). Next we can state:

Theorem 5.2.1. *Consider an ExpRK-F method defined by* (5.2.5)*, and $\beta_{ij} \ge 0$ in* (5.2.15)*. There exist $h_* > 0$ and $\mu_* > 0$ such that $f^{l+1} \ge 0$ provided that $f^l \ge 0$, $\mu \ge \mu_*$ and $0 < h \le h_*$.*

Proof. We prove it for sub-stage i, and the analysis for the final step could be carried out in

the same manner. Simple algebra gives, for $i = 1, \cdots, \nu$, $j < i$

$$\begin{aligned} f^{(i)} =& \tilde{M}\left(1 - \sum_j e^{(c_j - c_i)\lambda}\left(\alpha_{ij} + \lambda\beta_{ij}\right)\right) \\ &+ \sum_{j=1}^{i-1} \lambda\beta_{ij} e^{(c_j - c_i)\lambda} \frac{P^{(j)}}{\varepsilon} \\ &+ \sum_{j=1}^{i-1} \alpha_{ij} e^{(c_j - c_i)\lambda}\left(f^{(j)} - \frac{h\beta_{ij}}{\alpha_{ij}} v \cdot \nabla_x f^{(j)}\right). \end{aligned} \tag{5.2.17}$$

To prove positivity, it is enough to show that this is a convex combination, and each element is positive. To show each element is positive, i.e.

$$\tilde{M} > 0, \quad P^{(j)} > 0, \quad f^{(j)} - \frac{h\beta_{ij}}{\alpha_{ij}} v \cdot \nabla_x f^{(j)} > 0. \tag{5.2.18}$$

Positivity of $\tilde{M}$ is obvious, and $P^{(j)}$ is positive if one has big enough μ:

$$\mu \geq \mu_* = \sup |Q^-| \Rightarrow P = Q + \mu f = Q^+ - fQ^- + \mu f > 0.$$

To handle the positivity of the transport term, we adopt **Assumption** 5.2.1. The term is positive as long as h_* is small enough:

$$0 < h \leq h_* = \min_{ij}\left(\frac{\alpha_{ij}}{\beta_{ij}} h^*\right).$$

We also check the convexity of (5.2.17), it should be proved that

$$\sum_j e^{(c_j - c_i)\lambda}\left(\alpha_{ij} + \lambda\beta_{ij}\right) \leq 1. \tag{5.2.19}$$

In fact, it is easy to show that the quantity on the left hand side is a decreasing function for $\lambda \geq 0$, that gets its maximum value $\sum_j \alpha_{ij} = 1$ at $\lambda = 0$. One could simply take the derivative

with respect to λ and will find it is negative:

$$\frac{d}{d\lambda}\left(\sum_j e^{-\Delta_{ij}\lambda}(\alpha_{ij}+\lambda\beta_{ij})\right)$$

$$=\sum_j e^{-\Delta_{ij}\lambda}\left(-\Delta_{ij}(\alpha_{ij}+\lambda\beta_{ij})+\beta_{ij}\right) \tag{5.2.20}$$

$$=\sum_j e^{-\Delta_{ij}\lambda}\left(-\beta_{ij}\Delta_{ij}\lambda+\beta_{ij}-\alpha_{ij}\Delta_{ij}\right)<0 \tag{5.2.21}$$

Here $\Delta_{ij}=c_i-c_j$. In the last step, $\beta_{ij}=\alpha_{ij}(c_i-c_j)$ is used. This finishes the proof on convexity and we conclude that the scheme gives positive distribution function in each time step. □

Since the proof above is based on a convexity argument, we also have monotonicity of the numerical solution or SSP property. Thus the building block of our exponential schemes is naturally given by the optimal SSP schemes which minimize the stability restriction on the time stepping. We refer to [84] for a review on SSP Runge-Kutta schemes.

Remark 5.2.2.

- *Note that the proof above does not rely on the value λ take, i.e. the scheme is positive uniformly in ε.*

- *Optimal second and third order SSP explicit Runge-Kutta methods such that $\beta_{ij}\geq 0$ have been developed in the literature. However the classical third order SSP method by Shu and Osher [191] does not satisfy $c_j\leq c_i$ for $j<i$. Note that standard second order midpoint and third order Heun methods satisfy the assumptions in the Theorem 1(see Table 1.1 page 135 in [98]).*

  In [83] it was proved that all four stage, fourth order RK methods with positive CFL coefficient h_* must have at least one negative β_{ij}. The most popular fourth order method

using five stage with nonnegative β_{ij} has been developed in [176]. In [176] the authors also proved that any method of order greater then four will have negative β_{ij}.

- *Positivity of ExpRK-V schemes is much more difficult to achieve because of the involvement of the $\partial_t M$ term. However, we can prove*
  - *ρ is positive;*
  - *the negative part of T is $O(h\varepsilon)$.*

  We leave both proofs to the appendix B.1.

Contraction and Asymptotic Preservation

In this section, we show ExpRK-F and ExpRK-V are asymptotic preserving numerical method. The whole idea relies on the contraction property of the collision operator P [52].

Formally, if one check the formulas (5.2.5) and (5.2.8b), as $\varepsilon \to 0$, $\lambda \to \infty$, and the exponential functions as a penalty that pushes f approaching to $\tilde{M}$. The two methods above assign $\tilde{M}$ with macroscopic quantities that numerically solve the Euler equation, and thus AP follows directly. The proof in this section is carried out in a rigorous manner: we carefully track in each time step the difference between the distribution function and the correct Maxwellian function and provide accurate estimation on the convergence rate. Especially our proof shows that the argument could also handle the tough case, namely, $c_\nu = 1$. When this happens, the formal argument does not carry through.

The proof relies on the following assumption:

Assumption 5.2.2. *There is a constant C big enough, such that $|P(f,f) - P(g,g)| < C\,|f-g|$ where $|\cdot|$ denotes a proper metric.*

This assumption is on the contraction of the operator P, and is generally true for most physically meaningful metric. In appendix B.2, we show the validity of the proof with the

metric d_2 in space $P_s(\mathbb{R}^d)$ (also see [183]).

Under this assumption, considering $P(M,M) = Q(M,M) + \mu M = \mu M$, one has

$$|P(f,f) - \mu M| < C\,|f - M|\,. \tag{5.2.22}$$

The proofs for both approaches: ExpRK-F and ExpRK-V are exactly the same, and we only show it for ExpRK-F for simplicity, with any given explicit Runge-Kutta scheme.

For AP property, one needs to show that as $\varepsilon \to 0$, the scheme gives correct Euler limit. To do this, basically one needs to prove that f goes to the Maxwellian function whose macroscopic quantities solve the Euler equation (4.2.4).

Let us define

$$\begin{aligned} &d_i = \left| f^{(i)} - \tilde{M} \right|, \quad D_i = \left| v \cdot \nabla_x f^{(i)} \right|, \quad d_0 = \left| f^l - \tilde{M} \right|, \quad \vec{e} = [1, 1, \cdots, 1]^T, \\ &\vec{d} = [d_1, d_2, \cdots, d_\nu], \quad \vec{D} = [D_1, D_2, \cdots, D_\nu]^T. \end{aligned} \tag{5.2.23}$$

Moreover $\mathbb{A}$ is a lower-triangular matrix and $\mathbb{E}$ is a diagonal matrix given by

$$\mathbb{A}_{ij} = \frac{\lambda}{\mu} a_{ij} e^{(c_j - c_i)\lambda}, \quad \mathbb{E} = \mathrm{diag}\{e^{-c_1\lambda}, e^{-c_2\lambda}, \cdots, e^{-c_\nu\lambda}\}.$$

Lemma 5.2.1. *Based on the definitions above, for ExpRK-F one has*

$$\vec{d} \le d_0 \left(\mathbb{I} - C\mathbb{A}\right)^{-1} \cdot \mathbb{E} \cdot \vec{e} + \varepsilon \left(\mathbb{I} - C\mathbb{A}\right)^{-1} \cdot \mathbb{A} \cdot \vec{D}.$$

Proof. It is just direct derivation from (5.2.5):

$$(f^{(i)} - \tilde{M}) e^{c_i\lambda} = (f^n - \tilde{M}) + \sum_{j=1}^{i-1} a_{ij} \frac{\lambda}{\mu} e^{c_j\lambda} (P^{(j)} - \mu\tilde{M} - \varepsilon v \cdot \nabla_x f^{(j)}). \tag{5.2.24}$$

We take the norm of the equation and use the triangle inequality. With the contraction assumption, one gets:

$$\left|f^{(i)}-\tilde{M}\right| \leq \left|f^{l}-\tilde{M}\right| e^{-c_i\lambda} + \sum_j a_{ij}\frac{\lambda}{\mu}e^{(c_j-c_i)\lambda}\left(C\left|f^{(j)}-\tilde{M}\right| + \varepsilon\left|v\cdot\nabla_x f^{(j)}\right|\right). \tag{5.2.25}$$

A condense matrix form is:

$$\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_\nu \end{pmatrix} \leq \mathbb{E}\begin{pmatrix} d_0 \\ d_0 \\ \vdots \\ d_0 \end{pmatrix} + C\mathbb{A}\begin{pmatrix} d_1 \\ d_2 \\ \vdots \\ d_\nu \end{pmatrix} + \varepsilon\mathbb{A}\begin{pmatrix} D_1 \\ D_2 \\ \vdots \\ D_\nu \end{pmatrix}.$$

So

$$\vec{d} \leq d_0\mathbb{E}\cdot\vec{e} + C\mathbb{A}\cdot\vec{d} + \varepsilon\mathbb{A}\cdot\vec{D} \tag{5.2.26}$$

$$\vec{d} \leq d_0\left(\mathbb{I}-C\mathbb{A}\right)^{-1}\cdot\mathbb{E}\cdot\vec{e} + \varepsilon\left(\mathbb{I}-C\mathbb{A}\right)^{-1}\cdot\mathbb{A}\cdot\vec{D} \tag{5.2.27}$$

which completes the proof. □

Lemma 5.2.2.

$$\left|f^{l+1}-\tilde{M}\right| \leq \left|f^{l}-\tilde{M}\right| R_1(\lambda) + \vec{R}_2\cdot\vec{D}, \tag{5.2.28}$$

with R_1 and R_2 defined as:

$$R_1(\lambda) = e^{-\lambda}\left(1+\frac{C\lambda}{\mu}\vec{b}\cdot\mathbb{E}^{-1}\left(\mathbb{I}-C\mathbb{A}\right)^{-1}\mathbb{E}\cdot\vec{e}\right) \tag{5.2.29}$$

$$\vec{R}_2(\lambda) = \frac{\varepsilon\lambda}{\mu}e^{-\lambda}\vec{b}\cdot\mathbb{E}^{-1}\cdot\left(\mathbb{I}-C\mathbb{A}\right)^{-1}\cdot\left(\mathbb{I}+C\mathbb{A}\right) \tag{5.2.30}$$

Proof. It is just a simple derivation. Define

$$k_i = \frac{h}{\varepsilon}(P^{(i)} - \mu\tilde{M} - \varepsilon v \cdot \nabla_x f^{(i)})e^{c_i\lambda}. \tag{5.2.31}$$

Apparently, the previous lemma leads to

$$|\vec{k}| \leq \frac{\lambda}{\mu}\mathbb{E}^{-1} \cdot \left(C\vec{d} + \varepsilon\vec{D}\right). \tag{5.2.32}$$

Back to (5.2.5), one has

$$\left(f^{l+1} - \tilde{M}\right) = \left(f^l - \tilde{M}\right)e^{-\lambda} + \sum_{s=1}^{\nu} b_i k_i e^{-\lambda}, \tag{5.2.33}$$

which implies

$$\begin{aligned}
\left|f^{l+1} - \tilde{M}\right| \leq & d_0 e^{-\lambda} + \frac{\lambda}{\mu}e^{-\lambda}\vec{b}^T \cdot \mathbb{E}^{-1} \cdot \left(C\vec{d} + \varepsilon\vec{D}\right) && \text{(5.2.34a)}\\
\leq & e^{-\lambda}\left(d_0 + \frac{\lambda}{\mu}\vec{b} \cdot \mathbb{E}^{-1} \cdot \left(C\left(\mathbb{I} - C\mathbb{A}\right)^{-1} \cdot \left(d_0\mathbb{E} \cdot \vec{e} + \varepsilon\mathbb{A} \cdot \vec{D}\right) + \varepsilon\vec{D}\right)\right) && \text{(5.2.34b)}\\
\leq & d_0 e^{-\lambda}\left(1 + \frac{C\lambda}{\mu}\vec{b} \cdot \mathbb{E}^{-1} \cdot (\mathbb{I} - C\mathbb{A})^{-1} \cdot \mathbb{E} \cdot \vec{e}\right) && \text{(5.2.34c)}\\
& + \frac{\varepsilon\lambda}{\mu}e^{-\lambda}\vec{b} \cdot \mathbb{E}^{-1} \cdot (\mathbb{I} - C\mathbb{A})^{-1} \cdot (\mathbb{I} + C\mathbb{A}) \cdot \vec{D}. && \text{(5.2.34d)}
\end{aligned}$$

Here $\vec{b} = [b_1, b_2, \cdots, b_\nu]$ is a row vector. The result (5.2.27) is also used. We conclude with the definition of $R_{1/2}$.

$$\left|f^{l+1} - \tilde{M}\right| \leq \left|f^l - \tilde{M}\right| R_1(\lambda) + \vec{R}_2(\lambda) \cdot \vec{D}.$$

□

The two lemmas are pure algebra computation, but they provide the exact formula on how distribution function convergence towards the Maxwellian. Apparently the smaller R_1 is,

the faster the function converges, and R_2 represents the drift from the transportation, and is expected to be small in the limit. Our next task is to prove that they are both small, and the proof is built on the great property the matrix $\mathbb{A}$ possess: it is usually a lower triangular matrix, and a strict lower triangular matrix for explicit Runge-Kutta, and thus it is a nilpotent.

Theorem 5.2.2. *The method ExpRK-F defined by (5.2.5) is AP for general explicit Runge-Kutta method with* $0 \le c_1 \le c_2 \le \cdots \le c_\nu < 1$.

Proof. Obviously if $R_1(\lambda) = O(\varepsilon)$ and $R_2(\lambda) = O(\varepsilon)$ for ε small enough, the theorem holds. In fact, for explicit Runge-Kutta method, $\mathbb{A}$ is a strict lower triangular matrix, and thus a nilpotent, then one has the following

$$\mathbb{E}^{-1}\left(\mathbb{I} - C\mathbb{A}\right)^{-1}\mathbb{E} = \mathbb{E}^{-1}\left(\mathbb{I} + C\mathbb{A} + C^2\mathbb{A}^2 + \cdots + C^{\nu-1}\mathbb{A}^{\nu-1}\right)\mathbb{E} \tag{5.2.35a}$$

$$= \mathbb{I} + \mathbb{B} + \mathbb{B}^2 + \cdots + \mathbb{B}^{\nu-1} \tag{5.2.35b}$$

where $\mathbb{A}^\nu = 0$, definition $\mathbb{B} = C\mathbb{E}^{-1}\mathbb{A}\mathbb{E}$ and $\mathbb{E}^{-1}\mathbb{A}^2\mathbb{E} = \mathbb{E}^{-1}\mathbb{A}\mathbb{E}\mathbb{E}^{-1}\mathbb{A}\mathbb{E}$ are used. According to the definition of $\mathbb{A}$ and $\mathbb{E}$, it can be computed that

$$\mathbb{B}_{ij} = C\mathbb{A}_{ij}e^{c_i\lambda - c_j\lambda} = \frac{C\lambda}{\mu}a_{ij}.$$

Thus $\mathbb{I} + \sum_k \mathbb{B}^k$ is a matrix such that: the element on the kth diagonal is of order $O(\lambda^k)$. This leads to obvious result

$$R_1(\lambda) = e^{-\lambda}\left(1 + \frac{C\lambda}{\mu}\vec{b}\cdot\mathbb{E}^{-1}\left(\mathbb{I} - \mathbb{A}\right)^{-1}\mathbb{E}\cdot\vec{e}\right) = O(e^{-\lambda}\lambda^{\nu-1}) < O(\varepsilon)$$

Similar analysis can be carried to $R_2(\lambda)$ to show that it vanishes to zero as $\varepsilon \to 0$.
So as $\varepsilon \to 0$, $|f^{l+1} - \tilde{M}| \to 0$. By definition, $\tilde{M}$ is defined by macroscopic quantities computed directly from the limit Euler equation, thus the numerical scheme is AP, which finishes the

proof. □

The derivation of the scheme ExpRK-V is essentially the same, and in the end, one still has, in a condense form

$$\left|f^{l+1} - M^{l+1}\right| \leq \left|f^{l} - M^{l}\right| R_1(\lambda) + \vec{R}_2 \cdot \vec{D} \tag{5.2.36}$$

with R_1, $\vec{R}_2$, $\mathbb{E}$, $\mathbb{A}$ defined in the same way as in (6.3.7), but $D_i = |v \cdot \nabla_x f^{(i)} + \partial_t M^{(i)}|$. Following the same computations, one could prove that this method is AP too, but the proof is omitted for brevity.

Theorem 5.2.3. *The method ExpRK-V defined by (5.2.8) is AP for general explicit Runge-Kutta method.*

5.2.3 Numerical example

In this subsection we show three numerical examples. We are especially interested in checking high order of accuracy and asymptotic preserving.

Convergence rate test

In this example, we use smooth data to check the convergence rate of both methods. The problem is adopted from [65]: 1 dimensional in x and 2 dimensional in v. Initial distribution is given by

$$f(t=0,x,v) = \frac{\rho_0(x)}{2}\left(e^{-\frac{|v-u_1(x)|^2}{T_0(x)}} + e^{\frac{|v-u_2(x)|^2}{T_0(x)}}\right) \tag{5.2.37}$$

with

$$\rho_0(x) = \frac{1}{2}\left(2 + \sin\left(2\pi x\right)\right),$$
$$u_1(x) = [0.75, -0.75]^T, \qquad u_2(x) = [-0.75, 0.75]^T,$$
$$T_0(x) = \frac{1}{20}\left(5 + 2\cos\left(2\pi x\right)\right).$$

Domain is chosen as $x \in [0, 1]$ and periodic boundary condition on x is used. Note that the definition of ρ_0, $u_{1/2}$ and T_0 do not represent the real number density, average velocity and temperature.

As one can see, the initial data is summation of two Gaussian functions centered at u_1 and u_2 respectively, and is far away from the Maxwellian. To check the convergence rate, we use $N_x = 128, 256, 512, 1024$ grid points on x space, and $N_v = 32$ points on v space. Time stepping Δt is chosen to satisfy CFL condition with CFL number being 0.5. We measure the L_1 error of ρ and compute the decay rate through the following formula [199]

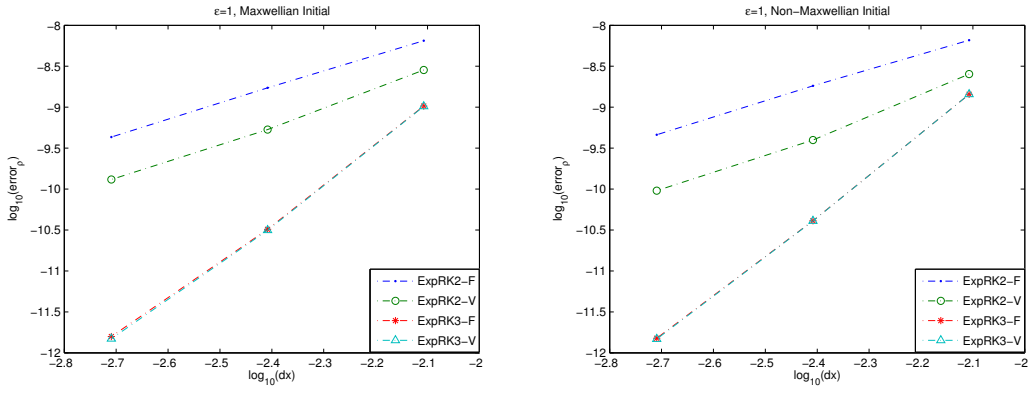
$$\text{error}_{\Delta x} = \max_{t=t^n} \frac{\|\rho_{\Delta x}(t) - \rho_{2\Delta x}(t)\|_1}{\|\rho_{2\Delta x}(t)\|_1}, \tag{5.2.38}$$

with $\Delta x = \frac{1}{N_x}$. The subindex of ρ indicate the discretization on x. Theoretically, a kth order numerical scheme should give $\text{error}_{\Delta x} < C\left(\Delta x\right)^k$ for Δx small enough.

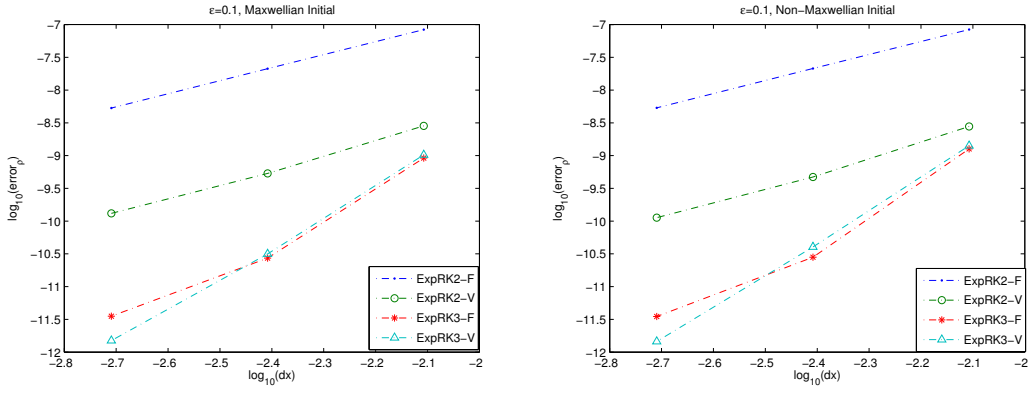
We compute this problem using spectral method [156] in v, WENO of order 3/5 [190] for x. For time discretization, we use the second and third order Runge-Kutta from [98], Table 1 page 135. We denote the four schemes under consideration as ExpRK2-F, ExpRK2-V, ExpRK3-F and ExpRK3-V.

The results for $\epsilon = 1, 0.1, 10^{-3}, 10^{-6}$ with either Maxwellian initial data or non-Maxwellian initial data are shown in Figure 5.2.3-5.2.3. We also give the convergence rate Table 5.2.3. One can see that in kinetic regime, when $\varepsilon = 1$, the two methods are almost the same, but

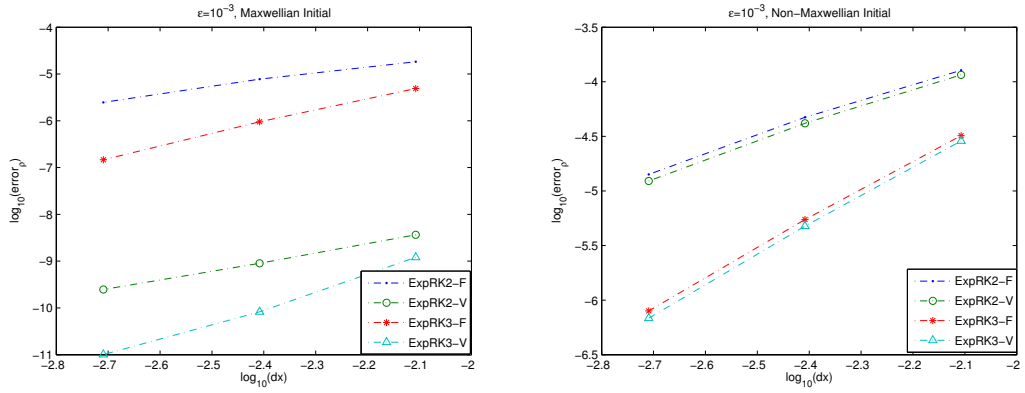
as ε becomes smaller, in the intermediate regime, for example $\varepsilon = 0.1$ for the second order schemes and $\varepsilon = 10^{-3}$ for second and third order schemes with Maxwellian data, ExpRK-V performs better then ExpRK-F. In the hydrodynamic regime, however, the two methods give similar results again shown by the two pictures for $\varepsilon = 10^{-6}$. It is remarkable that the third order methods achieve almost order 5 (the maximum achievable by the WENO solver) in many regimes.



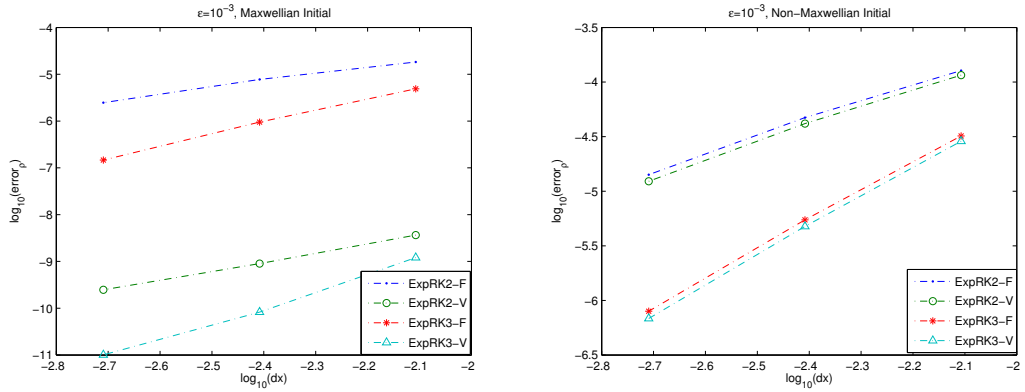
(a) $\varepsilon = 1$



(b) $\varepsilon = 0.1$



(c) $\varepsilon = 10^{-3}$



(d) $\varepsilon = 10^{-6}$

Figure 5.2.1: ExpRK method. Convergence rate test. In each picture, 4 lines are plotted: the lines with dots, circles, stars and triangles on them are given by results of ExpRK2-F, ExpRK2-V, ExpRK3-F and ExpRK3-V respectively. The left column is for Maxwellian initial data, and the right column is for initial data away from Maxwellian (5.2.37). Each row, from the top to the bottom, shows results of $\epsilon = 1, 0.1, 10^{-3}, 10^{-6}$ respectively.

Initial Distribution		Maxwellian Initial		Non-Maxwellian Initial	
N_x		128/256/512	256/512/1024	128/256/512	256/512/1024
$\varepsilon = 1$	ExpRK2-F	1.91327	1.99502	1.84968	1.98504
	ExpRK2-V	2.41608	2.02347	2.67733	2.05436
	ExpRK3-F	4.99725	4.35014	5.12959	4.76788
	ExpRK3-V	5.02508	4.40379	5.13515	4.79080
$\varepsilon = 0.1$	ExpRK2-F	1.98218	1.99539	1.97725	1.99454
	ExpRK2-V	2.41411	2.02293	2.56620	2.05830
	ExpRK3-F	5.07621	2.94707	5.49587	3.00335
	ExpRK3-V	5.02220	4.39651	5.13859	4.79264
$\varepsilon = 10^{-3}$	ExpRK2-F	1.23711	1.64976	1.43331	1.73501
	ExpRK2-V	2.02344	1.85924	1.47466	1.75496
	ExpRK3-F	2.36140	2.69178	2.55225	2.78275
	ExpRK3-V	3.86882	3.03223	2.59114	2.80353
$\varepsilon = 10^{-6}$	ExpRK2-F	2.56137	2.04519	2.56137	2.04519
	ExpRK2-V	2.56137	2.04519	2.56383	2.04859
	ExpRK3-F	5.08829	4.56695	5.08830	4.56699
	ExpRK3-V	5.08830	4.56704	4.91909	3.80638

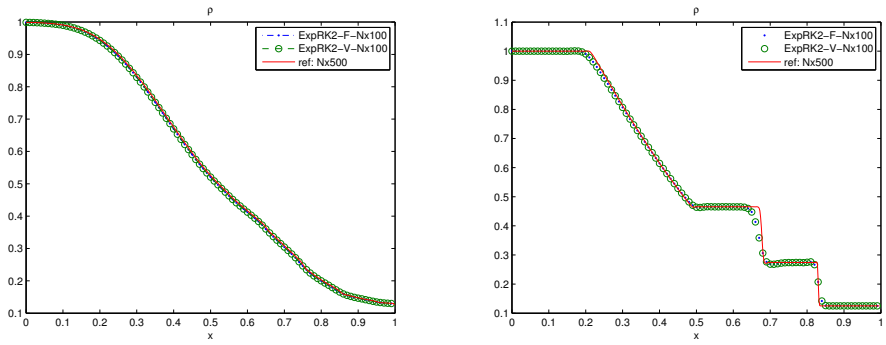
Table 1: Convergence rate for ExpRK methods with different initial data, in different regimes.

A Sod problem

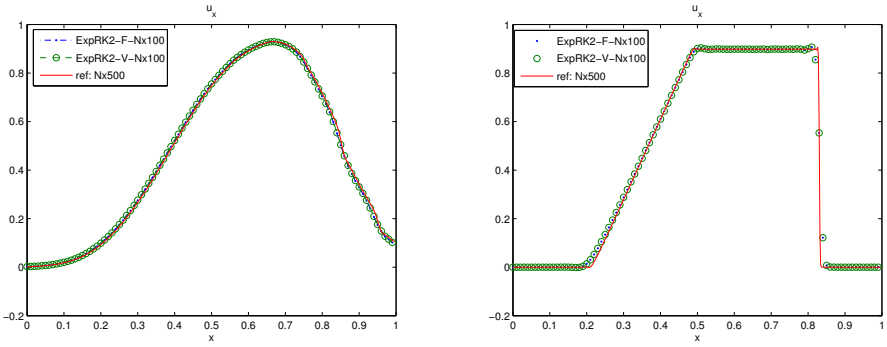
This simple example is adopted from [199] to check accuracy and AP of the numerical methods. It is a Riemann problem, and the solution to the associated Euler limit is a Sod problem.

$$\begin{cases} (\rho, u_x, u_y, T) = (1, 0, 0, 1), & \text{if } x < 0; \\ (\rho, u_x, u_y, T) = (1/8, 0, 0, 1/4), & \text{if } x > 0; \end{cases}$$

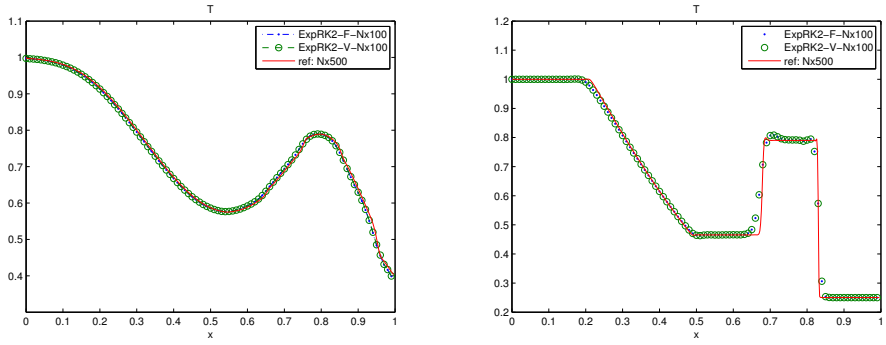
As a reference, we compute the problem using forward Euler with dense mesh and small time step when $\varepsilon = 0.01$: $\Delta x = 1/500$ and $h = 0.0001$. In Figure 5.2.3 (left), we show that when $\varepsilon = 0.01$ is comparably big, both methods give the same numerical results with the reference. In Figure 5.2.3 (right), AP property is shown: it is clear that for $\epsilon = 10^{-6}$, numerical results capture the Euler limit – the Euler limit is computed by kinetic scheme [166]. All plots are given at time $t = 0.2$.



(a) ρ



(b) u_x



(c) T

Figure 5.2.2: Exp-RK method. The Sod problem. Left column: $\varepsilon = 0.01$. The solid line is given by explicit scheme with dense mesh, while dots and circles are given by ExpRK2-F and ExpRK2-V respectively, both with $N_x = 100$. $h = \Delta x/20$ satisfies the CFL condition with CFL number being 0.5. Right column: For $\varepsilon = 10^{-6}$, both methods capture the Euler limit. The solid line is given by the kinetic scheme for the Euler equation, while the dots and circles are given by ExpRK2-F and ExpRK2-V. They perform well in rarefaction, contact line and shock.

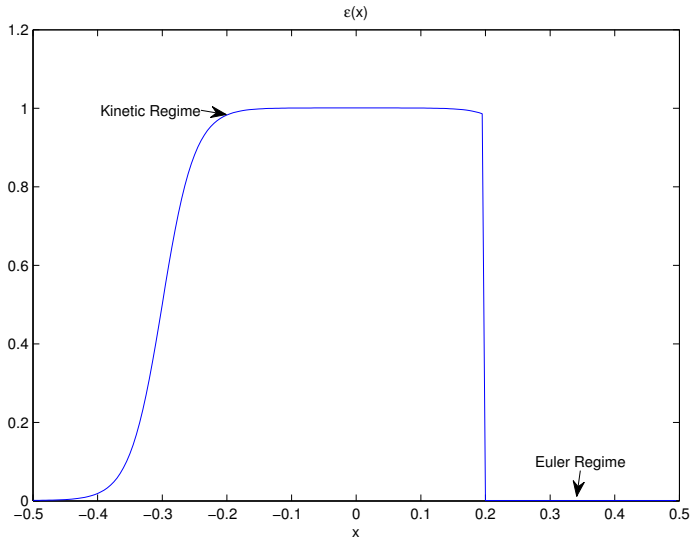


Figure 5.2.3: Mixing Regime: $\varepsilon(x)$

Mixing regime

In this example [199], we show numerical results to a problem with mixing regime. This problem is difficult because ε vary with respect to space. As what we do in the first example, we take identical data along one space direction, so it is $1D$ in space but $2D$ in velocity. An accurate AP scheme should be able to handle all ε with considerably coarse mesh. Domain is chosen to be $x \in [-0.5, 0.5]$, with ε defined by:

$$\epsilon = \begin{cases} \epsilon_0 + 0.5\left(\tanh\left(6 - 20x\right) + \tanh\left(6 + 20x\right)\right)\right) & x < 0.2; \\ \epsilon_0 \quad x > 0.2 \end{cases} \tag{5.2.39}$$

where ϵ_0 is 10^{-3}. So ε raise up from 10^{-3} to $\mathcal{O}(1)$, and suddenly drop back to 10^{-3} as shown in Figure (5.2.3). Initial data is the give as:

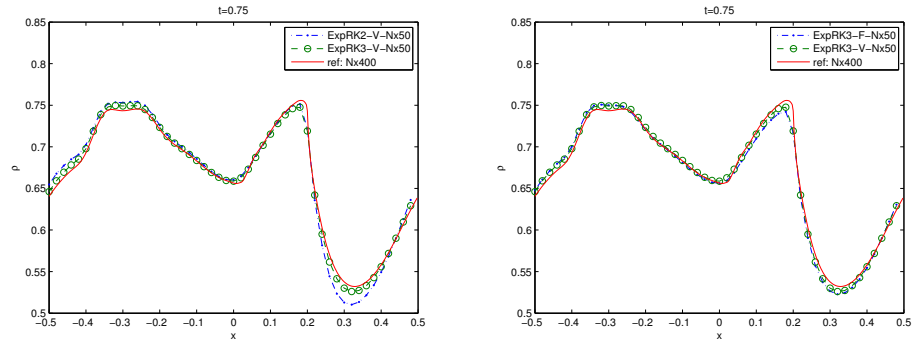
$$f(t=0,x,v) = \frac{\rho_0(x)}{4\pi T_0(x)}\left(e^{-\frac{|v-u_0(x)|^2}{2T_0(x)}} + e^{-\frac{|v+u_0(x)|^2}{2T_0(x)}}\right). \tag{5.2.40}$$

with

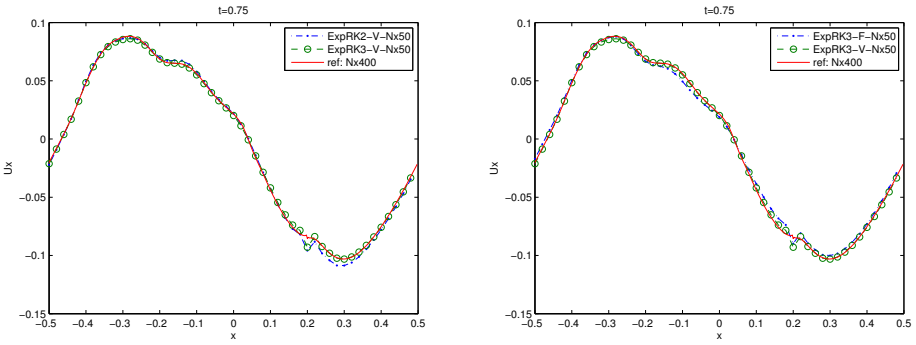
$$\begin{cases} \rho_0(x) = \frac{2+\sin(2\pi x+\pi)}{3} \\ u_0(x) = \frac{1}{5}\begin{pmatrix} \cos(2\pi x+\pi) \\ \\ 0 \end{pmatrix} \\ T_0(x) = \frac{3+\cos(2\pi x+\pi)}{4} \end{cases} . \tag{5.2.41}$$

Periodic boundary condition on x is applied. We compute the problem using ExpRK2/3-F/V. Reference solution is given by Runge-Kutta 2 with fine grids: $\Delta x = 0.0025$.

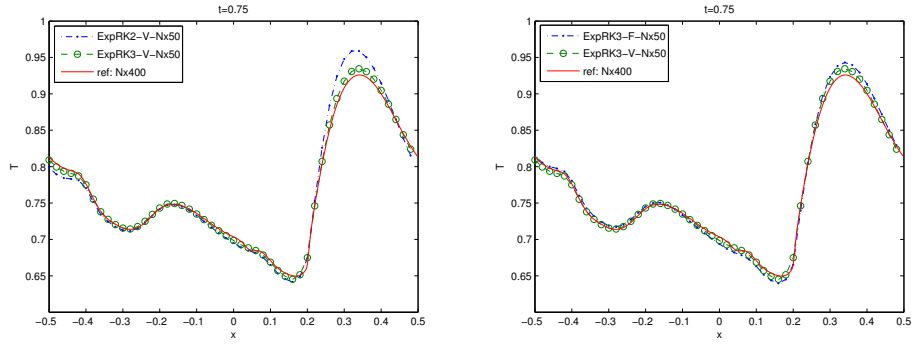
Results are plotted in Figure 5.2.3. Both methods give excellent results simply taking a CFL condition of 0.5 whereas explicit methods are forced to operate on a time scale 1000 times smaller. In particular, ExpRK3-V performs well uniformly on ε by giving a more accurate description of the shock profiles.



(a) ρ



(b) u_x



(c) T

Figure 5.2.4: ExpRK method. The mixing regime problem. The left column shows comparison of RK2 and RK3 using the ExpRK-V. The solid line is the reference solution with a very fine mesh in time and $\Delta x = 0.0025$, the dash line is given by RK3 and the dotted line is given by RK2, both with $N_x = 50$ points. The right column compare two methods, both given by RK3, with the reference. The dash line is given by ExpRK-V, and the dotted line is given by ExpRK-F. $N_x = 50$ for both. h is chosen to satisfy CFL condition, in our case, the CFL number is chosen to be 0.5.

Chapter 6

Multi-species systems

In this chapter we tackle the multi-species problem. To analyze the multi-species system can be very tricky theoretically. The theory carried out for mono-species could not be naïvely applied over: different particles do not use the same collisional cross-section; they exchange momentum and energy; particles with different masses converge to the Maxwellian in different speed, and all these are generating new difficulties on the theoretical level. Despite this, on the numerical level, we could still explore extensions of the developed computational treatments. New difficulties are encountered as well, however, most of the currently available methods mainly only rely on the limits of the systems, which are still known, and thus the difficulties emerged from collision kernels are all avoided. The arrangement of this chapter is the following: we review the general properties of the multi-species Boltzmann in section 6.1, and in the following two sections 6.2 and 6.3, we generalize the two numerical solvers mentioned in the previous chapter.

6.1 The multi-species system

In this section, we show the multi-species Boltzmann equation and its properties. We also discuss a simplified BGK model.

6.1.1 The multi-species Boltzmann equation

Given a N-species system, we use $f_i(t,x,v)$ to represent the distribution function of the i-th species at time t on the phase space (x,v), and $f=(f_1,f_2,\cdots,f_N)^T$ is a vector. The Boltzmann equation for the multi-species system is given by [63]:

$$\partial_t f_i + v\cdot\nabla_x f_i = Q_i(f,f),\ t\geq 0,\ (x,v)\in\mathbb{R}^d\times\mathbb{R}^d, \tag{6.1.1}$$

and the collision operator now is a collection of collisions among all species:

$$Q_i(f,f)=\sum_{k=1}^{N} Q_{ik}(f,f), \tag{6.1.2a}$$

$$Q_{ik}(f,f)(v)=\int_{S_+^{d-1}}\int_{\mathbb{R}^d}(f_i'f_{k^*}'-f_if_{k^*})B_{ik}(|v-v_*|,\Omega)dv_*d\Omega. \tag{6.1.2b}$$

We use the same notation for all the variables, but there are some new changes: the new collision kernel B_{ik} now depends on both two species i and k. $B_{ik}=B_{ki}$ for symmetry, (thus $Q_{ik}=Q_{ki}$); the computation of the post-collisional velocities v' and v_*' also change: the mass difference should also be taken into account:

$$v'=v-\frac{2\mu_{ik}}{m_i}(g\cdot\Omega)\Omega,\quad v_*'=v_*+\frac{2\mu_{ik}}{m_k}(g\cdot\Omega)\Omega, \tag{6.1.3}$$

with $\mu_{ik}=\frac{m_im_k}{m_i+m_k}$ being the reduced mass and m_i, m_k being the mass for species i and k respectively. This deduction is based on momentum and energy conservations:

$$m_iv+m_kv_*=m_iv'+m_kv_*',\quad m_i|v|^2+m_k|v_*|^2=m_i|v'|^2+m_k|v_*'|^2.$$

6.1.2 Properties of the multi-species Boltzmann equation

In d-dimensional space, we define the macroscopic quantities for species i: n_i is the number density; ρ_i is the mass density (note due to mass differences, mass density and number density have different values); u_i is the average velocity; E_i is the total energy; e_i is the specific internal energy; T_i is the temperature; S_i is the stress tensor; and q_i is the heat flux vector, given by:

$$\begin{aligned}
\rho_i &= m_i n_i = m_i \int f_i \mathrm{d}v, & E_i &= \frac{1}{2}\rho_i u_i^2 + n_i e_i = \frac{1}{2} m_i \int |v|^2 f_i dv, \\
\rho_i u_i &= m_i \int v f_i \mathrm{d}v, & e_i &= \frac{d}{2} T_i = \frac{m_i}{2n_i} \int f_i |v-u_i|^2 dv, \\
q_i &= \frac{1}{2} m_i \int (v-u_i)|v-u_i|^2 f_i \mathrm{d}v, & S_i &= \int (v-u_i)\otimes(v-u_i) f_i dv,
\end{aligned} \tag{6.1.4}$$

We also have global quantities for the mixture: the total mass density ρ, the number density n, the mean velocity $\bar{u}$, the total energy E, the internal energy $n\bar{e}$ and the mean temperature $\bar{T} = \frac{2\bar{e}}{d}$ are defined by:

$$\rho = \sum_i \rho_i, \quad n = \sum_i n_i, \quad \rho\bar{u} = \sum_i \rho_i u_i, \tag{6.1.5a}$$

$$E = n\bar{e} + \frac{1}{2}\rho|\bar{u}|^2 = \frac{d}{2} n\bar{T} + \frac{\rho}{2}|\bar{u}|^2 = \sum_i E_i. \tag{6.1.5b}$$

Conservations

In the gas mixture system, for each species, the mass is conserved, but the momentum and energy are not. Species keep exchanging momentum and energy till equilibrium state is achieved. Mathematically, while taking moments, one only gets zero from the first moment of the collision term, and the second and the third moments are nontrivial. Except for very special cases, usually these moments cannot even be explicitly obtained. We list here the

explicit expressions of the moments of the Maxwell molecule.

$$\begin{cases} < m_i Q_i > & = \int m_i Q_i(f) \mathrm{d}v = 0 \\ < m_i v Q_i > & = \int m_i v Q_i(f) \mathrm{d}v = \sum_{k=1} 2\mu_{ik}\chi_{ik} n_i n_k [u_k - u_i] \\ < \frac{1}{2} m_i v^2 Q_i > & = \int \frac{m_i}{2} |v|^2 Q_i(f) \mathrm{d}v \\ & = \sum_{k=1} 2 m_i \chi_{ik} n_i n_k \left[\left(\frac{\mu_{ik}}{m_i} \right)^2 \left(|u_k - u_i|^2 + 2\frac{e_k}{m_k} + 2\frac{e_i}{m_i} \right) + \frac{\mu_{ik}}{m_i} \left((u_k - u_i) \cdot u_i - 2\frac{e_i}{m_i} \right) \right] \end{cases}, \tag{6.1.6}$$

where $\chi_{ik} = \int (\cos\theta)^2 B_{ik}(\theta) \mathrm{d}\theta$, with $\theta = \arccos\left(\frac{g\cdot\Omega}{|g|}\right)$. One can check [75] for power law molecules.

Based on these formulas, when taking moments of the Boltzmann equation, one obtains the corresponding evolution of the macroscopic quantities. Taking the 1D Maxwell molecule for example:

$$\begin{aligned} &\partial_t \rho_i + \partial_x(\rho_i u_i) = < m_i Q_i > = 0, \qquad \text{or} \qquad \partial_t n_i + \partial_x(n_i u_i) = 0, \\ &\partial_t(\rho_i u_i) + \partial_x(S_i + \rho_i u_i^2) = \frac{1}{\varepsilon} < m_i v Q_i > = \frac{1}{\varepsilon} \sum_k 2 B_{ik} n_i n_k \mu_{ik} [u_k - u_i], \\ &\partial_t E_i + \partial_x(E_i u_i + S_i u_i + q_i) = \frac{1}{\varepsilon} < \frac{1}{2} m_i |v|^2 Q_i > = \frac{1}{\varepsilon} \sum_k 2 B_{ik} n_i n_k \left(\frac{\mu_{ik}^2}{m_i m_k} \right) (a + b), \end{aligned} \tag{6.1.7}$$

where $a = (m_k u_k + m_i u_i) \cdot (u_k - u_i)$, $b = 2(e_k - e_i)$. The macroscopic equations are complicated. The $\frac{1}{\varepsilon}$ terms on the right hand side are also making this set of equation very stiff, not to say that in most cases this explicit expressions are not even available. Directly designing macro solver for this set of equation can be very tedious, and is not in our interest. We seek for neat algorithms that aim at the essential physical meaningful limits: in fact, if we look at the system as a whole, it will be seen that the total momentum and total energy are still conserved. We

sum up the momentum and energy equations for all species and get:

$$\partial_t(\rho\bar{u}) + \partial_x\left(\sum_i S_i + \sum_i \rho_i u_i^2\right) = \frac{1}{\varepsilon}\sum_i < m_i v_i Q_i >= 0,$$
$$\partial_t E + \partial_x\left(\sum_i E_i u_i + \sum_i S_i u_i + \sum_i q_i\right) = \frac{1}{\varepsilon}\sum_i < \frac{1}{2} m_i v_i^2 Q_i >= 0. \tag{6.1.8}$$

The conservation hold for all kinds of cross-sections.

The local Maxwellian

As for single species system, we look for the local equilibrium here as well: it is reached when the gaining part and losing part of all collision terms balance out, namely $Q_i(f) = 0$ for each i [1]:

$$f_i = \overline{M}_i = n_i\left(\frac{m_i}{2\pi\bar{T}}\right)^{d/2} e^{-\frac{m_i|v-\bar{u}|^2}{2\bar{T}}}, \tag{6.1.9}$$

where $\bar{T}$ is the mean temperature and $\bar{u}$ is mean velocity defined in (6.1.5a) and (6.1.5b). We call this Maxwellian the "unified Maxwellian" because the velocity $\bar{u}$ and temperature $\bar{T}$ are given by those for the entire system instead of those for each single species.

Evidently, we now have requirement on both microscopic level and macroscopic level: the distribution functions should behave like Maxwellian functions, and the macroscopic physical observables are required to be the same, namely, average velocity u and temperature T. Also note that m_i appears inside the exponential, indicating that different mass leads to difference variance of the Gaussian function. In particular, the heavier the particles are, the more likely they are going to concentrate around $\bar{u}$ on the velocity space.

The Euler limit

The standard Chapman-Enskog expansion is still valid. We expand f_i around the unified Maxwellian (6.1.9), the collision term vanishes, and the system yields its Euler limit [1]:

$$\begin{aligned} &\partial_t \rho_i + \nabla \cdot (\rho_i \bar{u}) = 0, \\ &\partial_t (\rho \bar{u}) + \nabla \cdot (\rho \bar{u} \otimes \bar{u} + n\bar{T}\mathbb{I}) = 0, \\ &\partial_t E + \nabla \cdot ((E + n\bar{T})\bar{u}) = 0. \end{aligned} \tag{6.1.10}$$

Here $\mathbb{I}$ is the identity matrix. Note that in the equation for ρ_i, we have $\bar{u}$ instead of u_i as in (6.1.7). This is because when $\varepsilon \to 0$, $u_i \to \bar{u}$ and $T_i \to \bar{T}$ for all i.

6.1.3 A BGK model

In history, many BGK models have been proposed to simulate the multi-species Boltzmann equation. Unlike the situation in single-species system, here the definition of the BGK operator is very vague. For a period, people were not sure on whether to design one BGK term for each single collision operator, or for their summation. However, several criteria in designing BGK models are clear:

- positivity: the distribution function should be positive;
- indifferentiability: when different species share the same mass, equations of the system should be consistent with the single species Boltzmann equation.

Some models suffer from the loss of positivity [72], while others fail to satisfy the indifferentiability principle [88]. Here we present one model proposed by Andries, Aoki and Perthame [1] that meets both standards.

The model is:

$$\partial_t f_i + v \cdot \nabla_x f_i = \frac{\nu_i}{\varepsilon}(\widetilde{M}_i - f_i), \tag{6.1.11}$$

where ν_i is the collision frequency and $\widetilde{M}_i$ is a Maxwellian function:

$$\nu_i = \sum_k n_k \chi_{ik}, \quad \widetilde{M}_i = n_i \left(\frac{m_i}{2\pi \widetilde{T}_i}\right)^{d/2} e^{-\frac{m_i|v-\widetilde{u}_i|^2}{2\widetilde{T}_i}}. \tag{6.1.12}$$

The way $\widetilde{M}$ is defined is to capture the moments of the collision Q, i.e. $\nu_i\left(\widetilde{M}_i - f_i\right)$ shares the same first $d+2$ moments as Q_i. As stated above, usually explicit expression is hard to obtain for the moments of the collision term, and here we only write down that for the Maxwell molecule (6.1.6):

$$\nu_i \rho_i \widetilde{u}_i - \nu_i \rho_i u_i = < m_i v Q_i > = \sum_{k=1}^{n} 2\mu_{ik}\chi_{ik} n_i n_k [u_k - u_i], \tag{6.1.13a}$$

$$\begin{aligned} &\nu_i \widetilde{E}_i - \nu_i E_i \\ =& < \frac{1}{2} m_i v^2 Q_i > \\ =& \sum_{k=1}^{n} 2 m_i \chi_{ik} n_i n_k \left[\left(\frac{\mu_{ik}}{m_i}\right)^2 \left(|u_k - u_i|^2 + 2\frac{e_k}{m_k} + 2\frac{e_i}{m_i}\right) + \frac{\mu_{ik}}{m_i}\left((u_k - u_i)\cdot u_i - 2\frac{e_i}{m_i}\right)\right], \end{aligned} \tag{6.1.13b}$$

and $\widetilde{T}_i = \left(2\widetilde{E}_i - \rho_i \widetilde{u}_i^2\right)/(n_i d)$. Note that in this case the right hand side of equation (6.1.13a) is just a linear operator of macroscopic velocities. Also, given known u, the right hand side of (6.1.13b) is linear on e. To write it in a condense form, we define a matrix $\mathbb{L}$:

$$(\mathbb{L})_{ij} = \begin{cases} 2\mu_{ij}\chi_{ij} n_i n_j, \ i \neq j, \\ -2\sum_k \mu_{ik}\chi_{ik} n_i n_k, \ i = j. \end{cases} \tag{6.1.14}$$

It is easily seen that under this definition, (6.1.13a) is:

$$\nu_i \rho_i \widetilde{u}_i - \nu_i \rho_i u_i = \mathbb{L} u. \tag{6.1.15}$$

Note that $\mathbb{L}$ is a symmetric matrix with each row summing up to 0, and all non-diagonal entries are positive. Since $\mathbb{L}$ is a symmetric weakly diagonally dominant matrix, it is semi-negative definite, i.e. all its eigenvalues are non-positive.

We also mention another type of Maxwellian, which is defined by macroscopic quantities u_i and T_i for each species. We call it the "species Maxwellian":

$$M_i = n_i \left(\frac{m_i}{2\pi T_i} \right)^{d/2} e^{-\frac{m_i |v-u_i|^2}{2T_i}}. \tag{6.1.16}$$

Remark 6.1.1. *$M_i - f_i$ can not be used as a BGK operator. In the multi-species system, one has to introduce some mechanism into the collision term that captures the interactions between species. $M_i - f_i$ gives no communication between the species, so it cannot be used to express the multi-species collision.*

6.2 BGK-penalization method

We generalize the BGK-penalization method [65] to compute the multi-species system. We show the numerical method in subsection 6.2.1, its properties in 6.2.2, and the numerical examples can be found in 6.2.4. In subsection 6.2.3, we deal with a special case when the two particles have disparate masses.

6.2.1 Numerical method

Here we use the strategy discovered in [65] and write our scheme as:

$$\frac{f_i^{l+1} - f_i^l}{h} + v \cdot \nabla_x f_i^l = \frac{Q_i(f^l) - P_i(f^l)}{\varepsilon} + \frac{P_i(f^{l+1})}{\varepsilon}. \tag{6.2.1}$$

The superscript l stands for the time step and h is time discretization. We also use shorthands $Q_i^l \triangleq Q_i(f^l)$, $P_i^l \triangleq P_i(f^l)$ for convenience. P_i is chosen to be the BGK operator:

$$P_i = \beta(\overline{M}_i - f_i), \tag{6.2.2}$$

where β is a positive constant carefully chosen to take care of stability and positivity, to be explained later. A simple algebraic manipulation on gives:

$$f_i^{l+1} = \frac{\varepsilon f_i^l + h(Q_i^l - \beta^l(\overline{M}_i^l - f_i^l)) - \varepsilon h v \cdot \nabla_x f_i^l + \beta^{l+1}\Delta t \overline{M}_i^{l+1}}{\varepsilon + \beta^{l+1} h}. \tag{6.2.3}$$

We show the computation of $\overline{M}_i$, selection of β, computation of the collision term Q_i and the flux term $v \cdot \nabla_x f$ in order.

The computation of $\overline{M}^{l+1}$

The selection of $\overline{M}_i$ is very crucial. As will be proved later, the distribution function f is forced to $\overline{M}_i$ algebraically. So the best choice for $\overline{M}_i$ would be simply take the asymptotic limit of the distribution function f, which is the unified Maxwellian function defined in (6.1.9). We leave the AP proof to the next subsection. In order to update f, we need to find $\overline{M}_i$ in the next time step, and thus a simple solver for all its associated macroscopic quantities including ρ_i, $\bar{u}$ and $\bar{T}$. To do that, we make use of the fact that these quantities are for the entire system instead of single species, and are governed by the limiting Euler equation (6.1.10). Instead of

directly solving this equation, we use the kinetic scheme by taking the moments of (6.2.1):

$$\begin{aligned} n_i^{l+1} &= n_i^l - h\int v\cdot\nabla_x f_i^l\,dv,\\ (\rho\bar{u})^{l+1} &= (\rho\bar{u})^l - h\sum_i m_i\int v\otimes v\nabla_x f_i^l\,dv,\\ E^{l+1} &= E^l - h\sum_i\int \frac{m_i}{2}|v|^2 v\cdot\nabla_x f_i^l\,dv, \end{aligned}$$

and $\bar{T}$ follows naturally from (6.1.5b):

$$\bar{T}^{l+1} = \frac{2E^{l+1} - (\rho\bar{u}^2)^{l+1}}{dn^{l+1}}.$$

The collision term Q

The computation of collision terms for multi-species systems is much more tricker that in single-species case. We desire spectral accuracy, and adopt the spectral method introduced in [165]. To the our best knowledge, this hard problem has never been addressed in any literature, and the results presented here is only partial results obtained in [112], and though the idea works for many types of particles, we have access to explicit expression only for the simplest case: 1D Maxwell molecule. Here we present the computation of its Q_{ik}.

Use a ball $B(0,S)$ to represent the domain of the compactly supported distribution f. We periodize f on $v\in[-L,L]^d$ with $L\geq(3+\sqrt{2})S$. L is chosen much larger than S to avoid non-physical collision at different periods of the periodized f. Define the Fourier transform as:

$$\hat{f}(x;p) = \int f(x;v)e^{-ip\cdot v}\,dv,\quad f(x;v) = \frac{1}{(2L)^d}\sum_p \hat{f}(x;p)e^{ik\cdot v}. \tag{6.2.4}$$

Plugging into the collision term (6.1.2b):

$$Q_{ik} = \iint B_{ik} \left[f_i' f_{k^*}' - f_i f_{k^*} \right] dv_* d\Omega \equiv Q_{ik}^+ - f_i Q_{ik}^-,$$

where we define the gaining part and losing part as:

$$Q_{ik}^+ = \iint B_{ik} \left(f_i' f_{k^*}' \right) dv_* d\Omega, \quad f_i Q_{ik}^- = \iint B_{ik} \left(f_i f_{k^*} \right) dv_* d\Omega = f_i \iint B_{ik} f_{k^*} dv_* d\Omega.$$

Using the Fourier transform (6.2.4), one gets:

$$Q_{ik}^+ = \iint \frac{B_{ik}}{(2L)^{2d}} \left[\sum_p \sum_q \hat{f}_i(x;p) e^{ip\cdot v'} \hat{f}_k(x;q) e^{iq\cdot v_*'} \right] dv_* d\Omega.$$

For easier computation, one can rewrite equations (6.1.3) to:

$$v' = v - \frac{\mu_{ik}}{m_i}(g - |g|\omega), \qquad v_*' = v_* + \frac{\mu_{ik}}{m_k}(g - |g|\omega) = v - g + \frac{\mu_{ik}}{m_k}(g - |g|\omega).$$

Note the domain for ω is the entire unit sphere S^{d-1} instead of the semi-sphere for Ω. Then:

$$\begin{aligned} Q_{ik}^+ &= \frac{1}{(2L)^{2d}} \sum_{p,q} \hat{f}_i^p \hat{f}_k^q \iint B_{ik} e^{i(p\cdot v' + q\cdot v_*')} dv_* d\omega \\ &= \frac{1}{(2L)^{2d}} \sum_{p,q} \hat{f}_i^p \hat{f}_k^q e^{i(p+q)\cdot v} \iint B_{ik} e^{i\lambda\cdot g + i|g|\gamma\cdot\omega} dv_* d\omega, \end{aligned}$$

where $\lambda = \frac{-m_k}{m_i+m_k} p + \frac{-m_k}{m_i+m_k} q$ and $\gamma = \frac{m_k}{m_i+m_k} p - \frac{m_i}{m_i+m_k} q$. Given a specific B_{ik} one can analytically compute the integration above. The expression, however, can be very tedious, even in one dimensional space, especially when B_{ik} depends on $|g|$. Note that the integration domain for g is not symmetric. For the 1D Maxwell molecule, B_{ik} is a constant, and can be

pulled out of the integral, making the computation much easier. In this case,

$$v' = v - \frac{2m_k}{m_i + m_k}(v - v_*) = \frac{m_i - m_k}{m_i + m_k}v + \frac{2m_k}{m_i + m_k}v_*,$$
$$v'_* = v + \frac{m_i - m_k}{m_i + m_k}(v - v_*) = \frac{2m_i}{m_i + m_k}v - \frac{m_i - m_k}{m_i + m_k}v_*.$$

Plugging in Q_{ik}^+, one gets:

$$Q_{ik}^+ = \frac{B_{ik}}{(2L)^2}\sum_{p,q}\hat{f}_i^p\hat{f}_k^q e^{i\left(\frac{m_i-m_k}{m_i+m_k}p+\frac{2m_i}{m_i+m_k}q\right)v}\int e^{i\left(\frac{2m_k}{m_i+m_k}p-\frac{m_i-m_k}{m_i+m_k}q\right)v_*}dv_*.$$

One can also write Q_{ik}^+ as a summation of its Fourier modes $Q_{ik}^+(v) = \frac{1}{(2L)^2}\sum_l \hat{Q}_{ik}^{l+}e^{ilv}$ where

$$\begin{aligned}\hat{Q}_{ik}^{l+} &= \int Q_{ik}^+ e^{-ilv}dv \\ &= \frac{B_{ik}}{(2L)^2}\sum_{p,q}\hat{f}_i^p\hat{f}_k^q\int e^{i\left(\frac{m_i-m_k}{m_i+m_k}p+\frac{2m_i}{m_i+m_k}q-l\right)v}dv\int e^{i\left(\frac{2m_k}{m_i+m_k}p-\frac{m_i-m_k}{m_i+m_k}q\right)v_*}dv_* \\ &= B_{ik}\sum_{p,q}\hat{f}_i^p\hat{f}_k^q\,\mathrm{sinc(a)}\,\mathrm{sinc(b)},\end{aligned}$$

where $a = (\frac{m_i-m_k}{m_i+m_k}p + \frac{2m_i}{m_i+m_k}q - l)L$, and $b = (\frac{2m_k}{m_i+m_k}p - \frac{m_i-m_k}{m_i+m_k}q)L$. The FFT and the inverse FFT are used to speed up the computation.

The computation for $f_iQ_{ik}^-$ is much simpler in this special case:

$$f_iQ_{ik}^- = f_i\int B_{ik}f_k dv_k = f_i n_k B_{ik}.$$

The choice of the free parameter β

β should be chosen as the maximum value of the Frechet derivative $\nabla Q_i(f)$ [65]. Numerically, for positivity, we could find its minimum value. We split the collision Q into the gaining part and the losing part $Q_i = Q_i^+ - f_iQ_i^-$ by $Q_i^+ = \sum_k Q_{ik}^+$ and $f_iQ_i^- = f_i\sum_k Q_{ik}^-$. Plug back

in the scheme (6.2.3), one can rearrange the scheme:

$$f_i^{l+1} = \frac{\varepsilon\left(f_i^l - hv\cdot\nabla_x f_i^l\right) + hQ_i^+(f^l) + \left[\left(\beta^l - Q_i^-(f^l)\right) f_i^l + \beta^{l+1}\overline{M}_i^{l+1} - \beta^l\overline{M}_i^l\right] h}{\varepsilon + \beta^{l+1}h}.$$

To guarantee positivity, it is sufficient to require the followings for all i [199]:

$$\beta^l > Q_i^-(f^l), \quad \beta^l\overline{M}_i^l > \beta^{l-1}\overline{M}_i^{l-1}.$$

The flux term $v\cdot\nabla_x f_i$

Here we give the numerical flux in 1D. Multi-D could be dealt with in the same fashion. Use $v\partial_x f_{i,j}$ to denote the flux term for species i at the grid point x_j. A shock-capturing finite volume method we use is [141]:

$$v\partial_x f_{i,j} = \nu(f_{i,j_1} - f_{i,j_1-1}) - \frac{\Delta x}{2}\nu(\mathrm{sgn}(\nu) - \nu)(\sigma_{i,j_1} - \sigma_{i,j_1-1}), \tag{6.2.5}$$

where $\nu = \frac{v}{\Delta x}$, Δx is the mesh size. j_1 is chosen to be j for $v > 0$ and $j+1$ for $v < 0$. $\sigma_{i,j} = \frac{f_{i,j+1}-f_{i,j}}{\Delta x}\phi_{i,j}$ where $\phi_{i,j}$ is the slope limiter. For the van Leer limiter, it takes value as $\phi(\theta) = \frac{\theta+|\theta|}{\theta+1}$ and $\theta_{i,j} = \frac{f_{i,j}-f_{i,j-1}}{f_{i,j+1}-f_{i,j}}$ reflects the smoothness around grid point x_j.

6.2.2 The AP property of the time discretization

We prove asymptotic preserving in this subsection. This method is not strong-AP, but we show proof on weak-AP and relaxed-AP. We firstly revisit the time discrete scheme (6.2.1):

$$\frac{f_i^{l+1} - f_i^l}{h} + v\cdot\nabla_x f_i^l = \frac{Q_i^l - \beta(\overline{M}_i^l - f_i^l)}{\varepsilon} + \frac{\beta(\overline{M}_i^{l+1} - f_i^{l+1})}{\varepsilon}. \tag{6.2.6}$$

We will show in order that this method is weakly-AP for the Maxwell molecule, and relaxed-AP for the BGK model given in subsection 6.1.3. AP is trivial for $\mathcal{O}(1)$ ε, and we only prove it when $h \gg \varepsilon$.

In the proof, we are going to use the following notation: $\mathbb{L}$ defined in (6.1.14); $\lambda(\mathbb{M})$ is used to describe the spectral radius of matrix $\mathbb{M}$; δu_i^l and δT_i^l are variation of u_i and T_i from $\bar{u}$ and $\bar{T}$ respectively; $\mathbb{D}$ is a diagonal matrix with its elements being ρ_is.

$$\lambda(\mathbb{M}) = \sup_k (|\lambda_k(\mathbb{M})|), \quad \delta u_i^l = u_i^l - \bar{u}^l, \quad \delta T_i^l = T_i^l - \bar{T}^l,$$
$$\mathbb{D} = \mathrm{diag}\{\rho_1, \rho_2, \cdots, \rho_N\},$$

where $\lambda_k(\mathbb{M})$ are eigenvalues of $\mathbb{M}$.

Weakly-AP

Lemma 6.2.1. *For the Maxwell molecule, if $\delta u_i^l = \mathcal{O}(\varepsilon)$ and $\delta T_i^l = \mathcal{O}(\varepsilon)$ for $\forall i$, then $\delta u_i^{l+1} = \mathcal{O}(\varepsilon)$ and $\delta T_i^{l+1} = \mathcal{O}(\varepsilon)$.*

Proof. Rewrite scheme (6.2.6) as:

$$f_i^{l+1} - \overline{M}_i^{l+1} = \frac{\varepsilon(-\overline{M}_i^{l+1} + \overline{M}_i^l) - \varepsilon h v \cdot \nabla_x f_i^l}{\varepsilon + \beta h} + \frac{h Q_i^l}{\varepsilon + \beta h} - \left(\overline{M}_i^l - f_i^l\right). \tag{6.2.7}$$

Take the first moment on both sides. On the left hand side, one gets $(\rho_i u_i)^{l+1} - (\rho_i \bar{u})^{l+1}$, while on the right hand side, the first term is $O(\varepsilon)$. The second term gives:

$$\begin{aligned} \frac{h}{\varepsilon + \beta h} < m_i v Q_i^l > &= \frac{h}{\varepsilon + \beta h} \sum_k 2\chi_{ik}\mu_{ik} n_i n_k [u_k^l - u_i^l] \\ &= \frac{h}{\varepsilon + \beta h} \sum_k 2\chi_{ik}\mu_{ik} n_i n_k (\delta u_k^l - \delta u_i^l) = \mathcal{O}(\varepsilon). \end{aligned} \tag{6.2.8}$$

The third term gives:

$$< m_i v(\overline{M}_i^l - f_i^l) >= \rho_i \left(\bar{u}^l - u_i^l\right) = \mathcal{O}(\varepsilon).$$

So the entire right hand side is of $\mathcal{O}(\varepsilon)$, thus the term on the left hand side, $(\rho_i u_i)^{l+1} - (\rho_i \bar{u})^{l+1} = \mathcal{O}(\varepsilon)$, i.e. $\delta u_i^{l+1} = \mathcal{O}(\varepsilon)$. Similar analysis can be carried out for δT. □

Remark 6.2.1. *In the proof, we used that the collision kernel is for the Maxwell molecule in (6.2.8). The proof can be extended to other collision kernels as long as one can show the moments of Q is of $\mathcal{O}(\varepsilon)$ whenever $\delta u = \mathcal{O}(\varepsilon)$, which is usually the case.*

Theorem 6.2.1. *The method is weakly-AP; namely, if $\overline{M}_i^l - f_i^l = \mathcal{O}(\varepsilon)$, then $\overline{M}_i^{l+1} - f_i^{l+1} = \mathcal{O}(\varepsilon)$.*

Proof. Since $\overline{M}_i^l - f_i^l = \mathcal{O}(\varepsilon)$, both $P_i(f^l)$ and $Q_i(f^l)$ are of $\mathcal{O}(\varepsilon)$. Plugging back into the scheme (6.2.7), one gets $f_i^{l+1} - \overline{M}_i^{l+1} = \mathcal{O}(\varepsilon)$. □

Relaxed-AP

Lemma 6.2.2. *For the Maxwell molecules, when $h \ll 1$, in the limit of $\varepsilon \to 0$, there $\exists L$, such that $\forall l > L$, $\delta u^l = \mathcal{O}(\varepsilon)$, given big enough β.*

Proof. We prove the result for the 1D case. The proof for higher dimension can be carried out similarly. The proof follows that of [65]. Take moments of numerical scheme (6.2.6), one gets:

$$\frac{(\rho u)^{l+1} - (\rho u)^l}{h} + \partial_x \int v^2 \, m f^l \mathrm{d}v = \frac{1}{\varepsilon}(\mathbb{L}^l u^l + \beta^l \mathbb{D}^l \delta u^l - \beta^{l+1} \mathbb{D}^{l+1} \delta u^{l+1}).$$

Simple algebra gives:

$$\begin{aligned}\left(\varepsilon+\beta^{l+1}h\right)\mathbb{D}^{l+1}\delta u^{l+1} &= \left((\varepsilon+\beta^l h)\mathbb{D}^l + h\mathbb{L}^l\right)\delta u^l \\ &\quad + \varepsilon\left((\rho\bar{u})^l - (\rho\bar{u})^{l+1}\right) - \varepsilon h\partial_x\int v^2\, m f^l \mathrm{d}v, \\ \Rightarrow\quad (\varepsilon+\beta^{l+1}h)\left(\mathbb{D}^l+\mathcal{O}(h)\right)\delta u^{l+1} &= \left[(\varepsilon+\beta^l h)\mathbb{D}^l + h\mathbb{L}^l\right]\delta u^l + \mathcal{O}(\varepsilon).\end{aligned}$$

In the derivation, we used $\mathbb{L}^l\bar{u}=0$ and $\mathbb{D}^{l+1}=\mathbb{D}^l+\mathcal{O}(h)$. Invert the coefficients on the left hand side, one gets:

$$\delta u^{l+1} = \left[\frac{\beta^l}{\beta^{l+1}}\mathbb{I} + \frac{1}{\beta^{l+1}}(\mathbb{D}^l)^{-1}\mathbb{L}^l + \mathcal{O}(h)\right]\delta u^l + \mathcal{O}(\varepsilon).$$

Since the eigenvalues for $\mathbb{L}$ are non-positive, if one chooses $\beta^{l+1}+\beta^l > \lambda((\mathbb{D}^l)^{-1}\mathbb{L}^l)$, given small enough h, the spectrum of the big bracket on the left hand side is controlled by 1, and thus in the limit of $\varepsilon\to 1$, δu would decrease to $\mathcal{O}(\varepsilon)$, and we get our conclusion. □

The same analysis can be carried out for T. We call this property proved above "macro-AP". It is a property only exists in multi-species problems.

Remark 6.2.2. *The proof above is also valid for the BGK model in subsection 6.1.3. It can be extended to other kinds of collision kernel as well, but the corresponding* $\mathbb{L}$ *may not be linear on* u*, and we cannot find explicit expression about the requirement on* β*. In that case,* $\lambda(\mathbb{L})$ *stands for the spectrum of the nonlinear operator* $\mathbb{L}$*.*

Theorem 6.2.2. *If the problem is macro-AP, then,* $\exists L$ *such that* $\forall l > L$, $M_i^l - \overline{M}_i^l = \mathcal{O}(\varepsilon)$, *and* $\widetilde{M}_i^l - \overline{M}_i^l = \mathcal{O}(\varepsilon)$.

Proof. It is a straightforward conclusion from the lemma above, and from the definition for $\widetilde{M}$ in (6.1.12). □

Remark 6.2.3. *Up to now, we have shown that M_i approaches to $\overline{M}_i$ for the Boltzmann collision operator with the Maxwell molecule collision kernel. Rearranging scheme* (6.2.6)*, one gets:*

$$f^{l+1} - \overline{M}^{l+1} = \frac{(\varepsilon + \beta h)(f^l - \overline{M}^l) + hQ^l}{\varepsilon + \beta h} + \frac{\varepsilon(\overline{M}^l - \overline{M}^{l+1}) - \varepsilon h\, v \cdot \nabla_x f^l}{\varepsilon + \beta h}. \tag{6.2.9}$$

The second term on the right is of $\mathcal{O}(\varepsilon)$. So, one can get relaxed-AP only if it can be shown that Q and $f - \overline{M}$ have opposite signs. We can prove this for limited form of Q, say the BGK operator introduced in subsection 6.1.3. However, as will be seen in section 6.2.4, numerically the scheme is relaxed-AP for the general Boltzmann collision.

Theorem 6.2.3. *The scheme is relaxed-AP for the BGK operator $Q = \nu(\widetilde{M} - f)$ defined in section 2.3.*

Proof. Plug in the definition for Q, (6.2.9) writes:

$$\begin{aligned} f^{l+1} - \overline{M}^{l+1} &= \frac{\varepsilon + \beta h - \nu h}{\varepsilon + \beta h}(f^l - \overline{M}^l) + \frac{\nu h(\widetilde{M}^l - \overline{M}^l)}{\varepsilon + \beta h} + \mathcal{O}(\varepsilon) \\ &= \frac{\varepsilon + \beta h - \nu h}{\varepsilon + \beta h}(f^l - \overline{M}^l) + \mathcal{O}(\varepsilon). \end{aligned}$$

The convergence rate to the unified Maxwellian $\overline{M}$ is apparently:

$$\alpha = \frac{\varepsilon + \beta h - \nu h}{\varepsilon + \beta h}, \tag{6.2.10}$$

In the limit of $\varepsilon \to 0$, if one has $\beta > \frac{\nu}{2}$, then $|\alpha| < 1$, thus $|f - \overline{M}|$ keeps diminishing until reaching to $\mathcal{O}(\varepsilon)$, and we get the relaxed-AP. □

6.2.3 Disparate masses

This section is for the system of gas mixture with disparate masses in the homogeneous case, i.e. data given is uniform on physical domain. The mathematical problem was first pointed out by Grad [86], and has attracted great interests since then. The fundamental example is plasma, for which, the basic derivation can be found in [177, 25]. For these systems, it is the different time scalings for different species to reach to the equilibria that makes the problem difficult. Generally speaking, the light species should be able to get to the equilibrium faster, that is to say there is a time period when the light species is in hydrodynamic regime while the heavy species is, on the other hand, in kinetic regime. Analyses of the scalings of the collision operators have been done based on both postulate physical consideration [39, 126] and formal derivation [48, 49].

Theoretical rescaling analysis

In homogeneous space, the disparate masses system should be written as:

$$\begin{cases} \partial_t f_H = Q_H = Q_{HH} + Q_{HL} = \int B_{HH}(f'_H f'_{H^*} - f_H f_{H^*})\, dv_* + \int B_{HL}(f'_H f'_{L^*} - f_H f_{L^*})\, dv_*, \\ \partial_t f_L = Q_L = Q_{LL} + Q_{LH} = \int B_{LL}(f'_L f'_{L^*} - f_L f_{L^*})\, dv_* + \int B_{LH}(f'_{H^*} f'_L - f_{H^*} f_L)\, dv_*. \end{cases}$$

Now the small parameter that makes the collision term stiff is the ratio of mass $\varepsilon = \sqrt{m_L/m_H}$ where the sub-indices H and L stand for heavy and light respectively. While assuming that the two species have densities and temperatures of the same order of magnitude, one could obtain that f_H is much narrower than f_L as shown in Figure 6.2.1. To analyze the magnitude of the collision terms, we define $\tilde{f}_H(v) = f_H(\varepsilon v)$ to stretch f_H to a function that has comparable variance as f_L. As derived in [48, 49], the scaling ratio of the two collision terms is $Q_H/Q_L = \mathcal{O}(\varepsilon)$, which means that the collision Q_L has stronger effect, and that the light species gets to

the hydrodynamical regime much faster. For convenience, we write both Q_H and Q_L as $\mathcal{O}(1)$

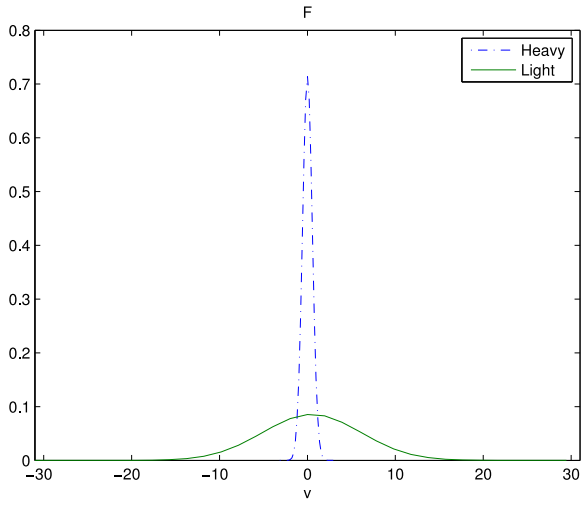


Figure 6.2.1: Distribution function for disparate masses system. The distribution function for heavy species is much narrower than that of the light species.

term, and put ε in front of Q_H to represent its magnitude. The system turns out to be:

$$\partial_t f_H = \varepsilon Q_H, \quad \partial_t f_L = Q_L. \tag{6.2.11}$$

One can also rescale the time by τ and obtain:

$$\partial_t f_H = \frac{\varepsilon}{\tau} Q_H, \quad \partial_t f_L = \frac{1}{\tau} Q_L. \tag{6.2.12}$$

When $\tau = \mathcal{O}(\varepsilon)$, the light species is in hydrodynamic regime but the heavy one is still in kinetic regime; and when $\tau = \mathcal{O}(\varepsilon^2)$, both species should be close to the equilibria.

Remark 6.2.4. *The inhomogeneous problem gets even harder to analyze, especially when the different species have different spatial rescaling coefficients. But numerically it makes very little difference: one simply needs to add the flux term $v \cdot \nabla_x f$ to the homogeneous scheme.*

The numerical scheme

The scheme we adopt for (6.2.11) is:

$$\frac{f_H^{l+1} - f_H^l}{\Delta t} = \frac{\varepsilon}{\tau}\left(Q_H^l - \beta\left(M_H^l - f_H^l\right)\right) + \frac{\beta\varepsilon}{\tau}\left(M_H^{l+1} - f_H^{l+1}\right), \tag{6.2.13a}$$

$$\frac{f_L^{l+1} - f_L^l}{\Delta t} = \frac{1}{\tau}\left(Q_L^l - \beta\left(M_L^l - f_L^l\right)\right) + \frac{\beta}{\tau}\left(M_L^{l+1} - f_L^{l+1}\right), \tag{6.2.13b}$$

where $\beta = \mathcal{O}(1)$.

Theorem 6.2.4. *This scheme yields the following behavior at: $\mathcal{O}(\frac{1}{\varepsilon})$ and $\mathcal{O}(\frac{1}{\varepsilon^2})$.*

- *at $\tau = \mathcal{O}(\frac{1}{\varepsilon})$, the scheme is first order consistent to $\partial_t f_H = Q_H$, and f_L^l is an $\mathcal{O}(\varepsilon)$ approximation of $\overline{M}_L$;*
- *at $\tau = \mathcal{O}(\frac{1}{\varepsilon^2})$, both f_H^l and f_L^l are within $\mathcal{O}(\varepsilon)$ of the unified Maxwellians $\overline{M}_H$ and $\overline{M}_L$ respectively.*

Proof. To show the second statement:

At this time scale, $\tau = \mathcal{O}(\varepsilon^2)$, the system turns out to be:

$$\partial_t f_H = \frac{1}{\varepsilon} Q_H, \quad \partial_t f_L = \frac{1}{\varepsilon^2} Q_L.$$

By the same arguments as in the previous sections, one gets:

$$f_H^l - \overline{M}_H^l = \mathcal{O}(\varepsilon), \quad f_L^l - \overline{M}_L^l = \mathcal{O}(\varepsilon^2).$$

for l large enough.

To show the first statement:

At this time scale, $\tau = \mathcal{O}(\varepsilon)$, system (6.2.12) can be written as:

$$\partial_t f_H = Q_H, \quad \partial_t f_L = \frac{1}{\varepsilon} Q_L.$$

The scheme still gives $f_L^l \to \overline{M}_L^l$. One just needs to show that the scheme gives a correct discretization of the equation for f_H. Write the scheme as (set $\tau = \varepsilon$):

$$\frac{f_H^{l+1} - f_H^l}{h} = Q_H^l - \beta(M_H^l - f_H^l) + \beta(M_H^{l+1} - f_H^{l+1}).$$

By pulling the f_H on the right hand side to the left, one gets:

$$\frac{f_H^{l+1} - f_H^l}{h} = Q_H^l - \frac{\beta h}{1+\beta h} Q_H^l + \frac{\beta}{1+\beta h}\left(M_H^{l+1} - M_H^l\right).$$

The second and the third terms on the right are both $\mathcal{O}(h)$, i.e. the scheme gives a first order discretization to $\partial_t f_H = Q_H$. □

6.2.4 Numerical examples

For comparison, the examples chosen are similar to those in [119]. We also perturb the data on the level of macroscopic quantities. For all the examples below: when ε is not very small so that solving the Boltzmann equation is still possible by using the basic explicit scheme with a resolved mesh, we compute the reference solution using forward Euler with very fine mesh, and compare our results with that; when ε is unbearably small for the forward Euler, we compare our results to the Euler limit. The Euler equations are computed using CLAWPACK Euler solver [142].

A stationary shock

In this example, we show numerical solution to a Riemann problem of two species. The analytical solution to the Euler equations is a stationary shock. Here the subscripts 1 and 2 stand for different species.

$$\begin{cases} m_1 = 1, m_2 = 1.5, n_1 = n_2 = 1, u_1 = 1.8, u_2 = 1.3, T_1 = 0.3, T_2 = 0.35, & \text{if } x < 0; \\ m_1 = 1, m_2 = 1.5, n_1 = n_2 = 1.401869, u_1 = u_2 = 1.07, T_1 = T_2 = 0.8605, & \text{if } x > 0. \end{cases} \tag{6.2.14}$$

The initial distribution for f is given by summation of two Gaussian functions, so it is far away from the unified Maxwellian $\overline{M}$:

$$f_i(t = 0, x, v) = \sum_{\ell=1}^{2} A_{i,\ell} \exp\big(-B_{i,\ell}(v - C_{i,\ell})^2\big), \qquad i = 1, 2, \tag{6.2.15}$$

with

$$B_{i,\ell} = \frac{\rho_i}{4E_i - 2\rho_i u_i^2(1 + \kappa^2)}, \quad A_{i,\ell} = \frac{n_i}{2}\sqrt{\frac{B_{i,\ell}}{\pi}}, \quad C_{i,1} - u_i = u_i - C_{i,2} = \kappa u_i \tag{6.2.16}$$

i.e. the two Gaussian functions have the same height and variation, but their centers are $2\kappa u$ away from each other. In the numerical experiment, we choose $\kappa = 0.2$, $\Delta x = 10^{-2}$ and h is chosen to satisfy the CFL condition: 10^{-3} in our simulation. Numerically we check whether the scheme gives the Euler limit when ε goes to zero; and whether it matches well with the forward Euler method with relatively fine mesh when ε is big. We first show in Figure 6.2.2 that as ε goes to zero, the numerical solution converges to the Euler limit, the stationary shock in this case. In Figure 6.2.3, we show that the AP scheme matches very well with the numerical results given by the forward Euler method for $\varepsilon = 10^{-1}$. Then we show in Figure 6.2.4, that given an initial data far away from the unified Maxwellian $\overline{M}$, f gets close to $\overline{M}$ quickly with

$\varepsilon = 10^{-5}$. This verifies that the scheme is relaxed-AP numerically. Figure 6.2.5 shows that smaller ε gives faster convergence to the equilibria for macroscopic quantities.

A Sod problem

In this example, we solve a Sod problem. The initial data are given by:

$$\begin{cases} m_1 = m_2 = 1, n_1 = 1, n_2 = 1.2, u_1 = 0.6, u_2 = -0.5, T_1 = T_2 = 0.709, & \text{if } x < 0; \\ m_1 = m_2 = 1, n_1 = 0.125, n_2 = 0.2, u_1 = -0.2, u_2 = 0.125, T_1 = T_2 = 0.075, & \text{if } x > 0. \end{cases} \tag{6.2.17}$$

The initial distribution is chosen far away from the Maxwellian as defined in (6.2.15) and (6.2.16) with κ again chosen as 0.2. For all ε, we choose $\Delta x = 10^{-2}$ and $h = 10^{-3}$. In this problem, $m_1 = m_2$, so we first show the numerical indifferentiability in Figure 6.2.6, that is: computing the problem as a multispecies system gives the same result as computing the monospecies Boltzmann equation. In Figure 6.2.7, we show that as ε goes to zero, the numerical solution converges to the Euler limit. For ε as big as 10^{-1} and 10^{-2}, we compare the results with those of the forward Euler with a fine mesh. They match well as shown in Figure 6.2.8. In Figure 6.2.9, we show that for $\varepsilon = 10^{-4}$, although the initial f is far away from the unified Maxwellian $\overline{M}$, as time evolves, it converges to $\overline{M}$. This numerically verifies the relaxed-AP property. In Figure 6.2.10, we show the evolution of u with different ε. Apparently different species gradually share the same velocity, and the smaller ε is, the faster the convergence is.

A disparate masses problem

In this example, we solve the disparate masses system. Define $\varepsilon = \sqrt{\frac{m_L}{m_H}}$, we want to verify that the light species gets close to the unified Maxwellian $\overline{M}_L$ faster than the heavy one. We

solve an inhomogeneous problem with the following initial data:

$$m_H = 8, \quad m_L = 0.08, \quad u_H = 0, \quad u_L = 0.5, \quad T_H = T_L = 2.5, \quad n_H = 1, n_L = 1.2.$$

The initial distribution functions are still given by the summation of two Gaussian functions as in (6.2.15), with parameters $A_1 = A_2$ and $B_1 = B_2$ defined in (6.2.16), and C_1 and C_2 defined by: $C_1 - u = u - C_2 = \kappa$. We choose $\kappa = 0.5$ for the heavy species and $\kappa = 4$ for the light one. In Figure 6.2.11, on the left we show the initial distribution functions for the two species, both of which are given by summation of two Gaussian functions and are far away from the Maxwellian. On the right we show several snapshots of the distribution functions as they evolve. In Figure 6.2.12, we show that as time evolves, the velocities converge toward each other. Note that the heavy species weighted more when computing for the mean velocity $\bar{u}$ as in (6.1.5a), thus its average velocity does not change much.

6.3 ExpRK method for the multi-species Boltzmann equation

In this section we explore how to extend the Exponential Runge-Kutta method to multi-species system [144]. Parts of the computation is the same as in the previous section and will be omitted. We present the method in subsection 6.3.1 and prove its properties in subsection 6.3.2. Numerical examples will be give in subsection 6.3.3.

6.3.1 Exponential Runge-Kutta method

It is a simple extension of the original Exponential Runge-Kutta method [52, 143]. One encounters the same difficulty as in the BGK-penalization case. The algorithm consists two parts in each time step:

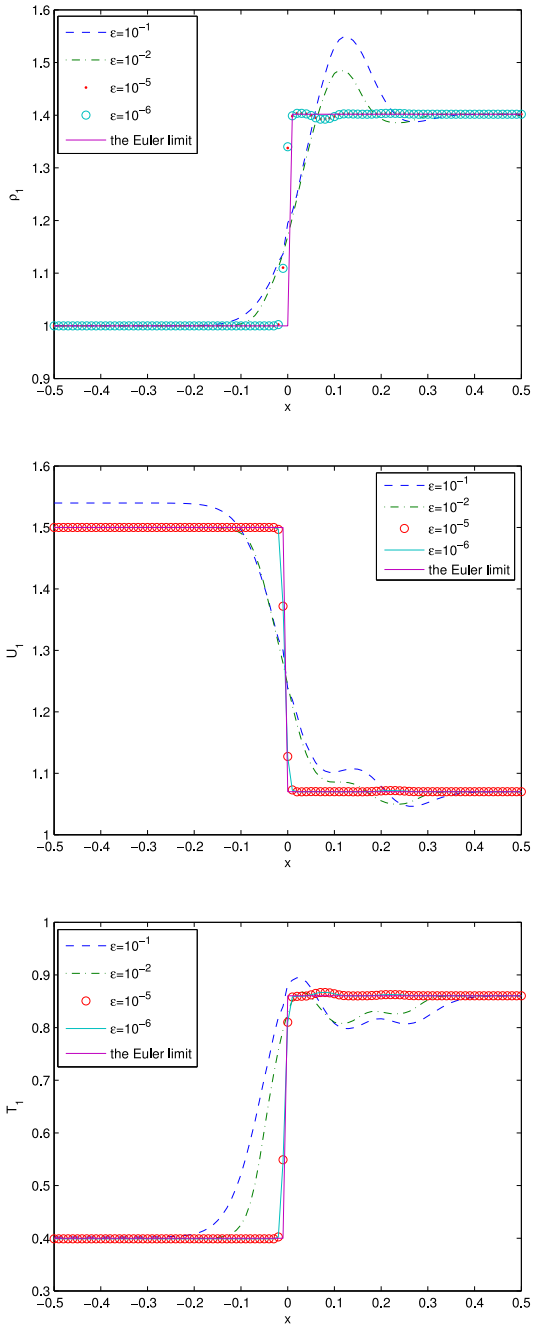


Figure 6.2.2: Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. As $\varepsilon \to 0$, solution of the Boltzmann equation goes to the Euler limit. $t = 0.1$.

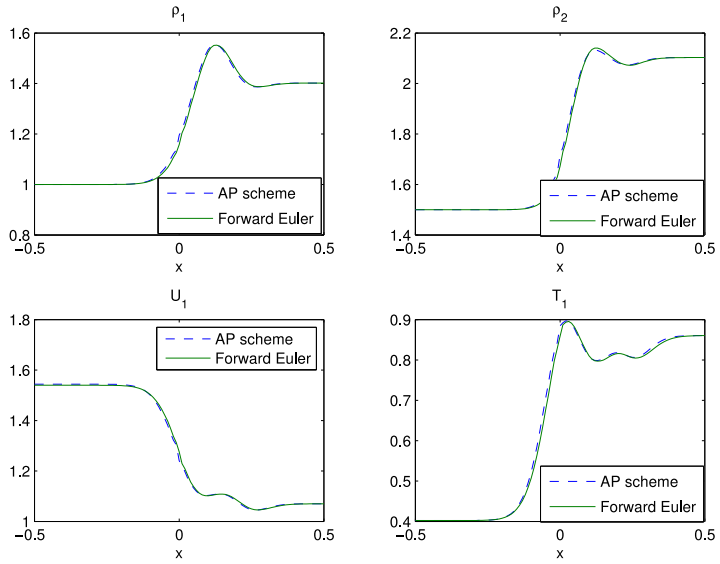


Figure 6.2.3: Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. $\varepsilon = 0.1$, $t = 0.1$. The dashed line is given by the AP scheme, and the solid line is given by the forward Euler with a fine mesh: $\Delta x = 0.01$ and $h = 0.0005$.

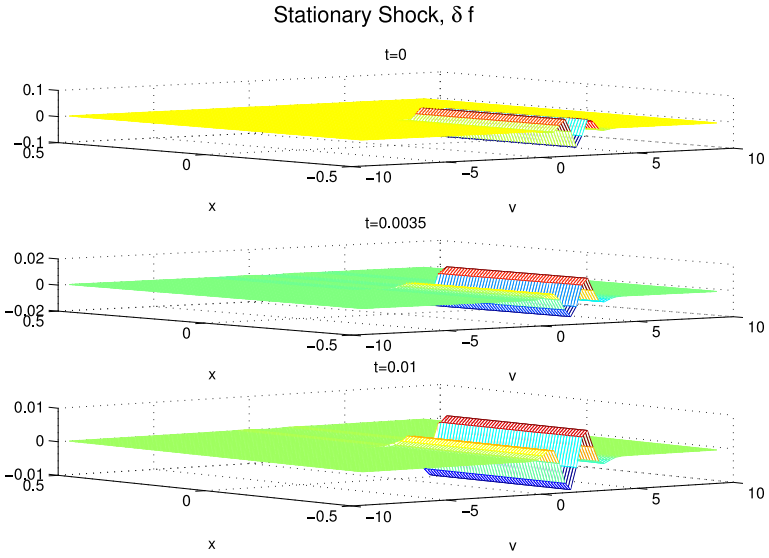


Figure 6.2.4: The stationary shock problem. $\varepsilon = 10^{-5}$, $\delta f = f_1 - \overline{M}_1$ diminishes as time evolves.

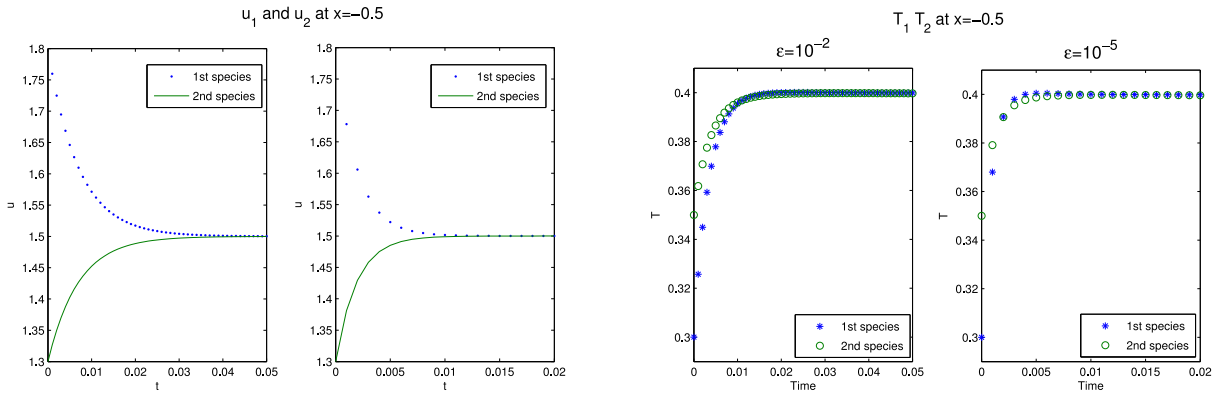


Figure 6.2.5: Multi-species Boltzmann equation. BKG-penalization method. The stationary shock problem. u_1 and u_2 at $x = -0.5$ on the left and T_1, T_2 on the right, as functions of time, for $\varepsilon = 10^{-2}$ and 10^{-5} respectively. Note the different time scales for the two figures.

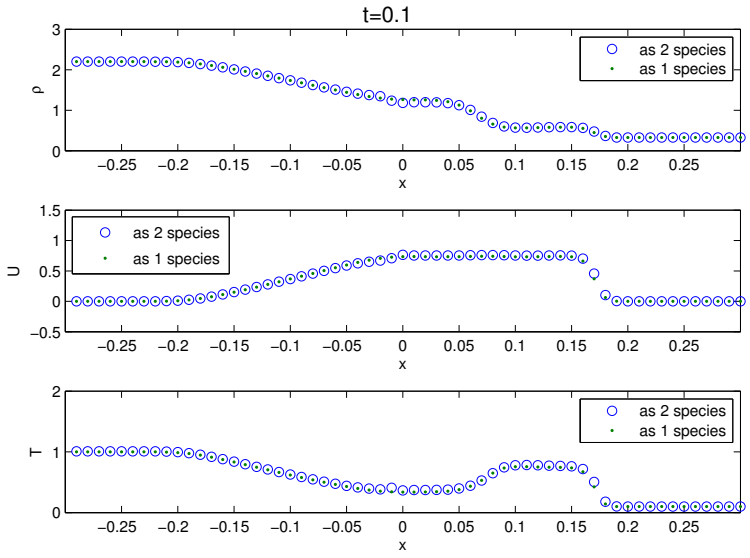


Figure 6.2.6: Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. Indifferentiability. $t = 0.1$. ρ, u and T are computed using two species model ("o") and one species model (".").

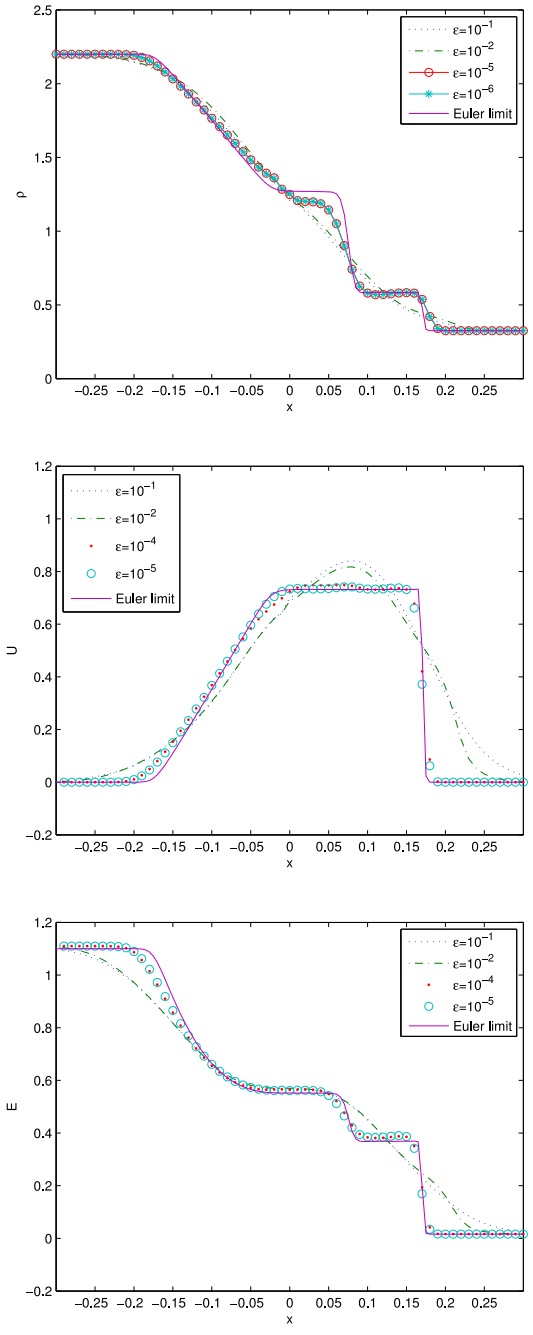


Figure 6.2.7: Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $t = 0.1$. As $\varepsilon \to 0$, the numerical results go to the Euler limit. Solid lines give the Euler limit computed by using [142].

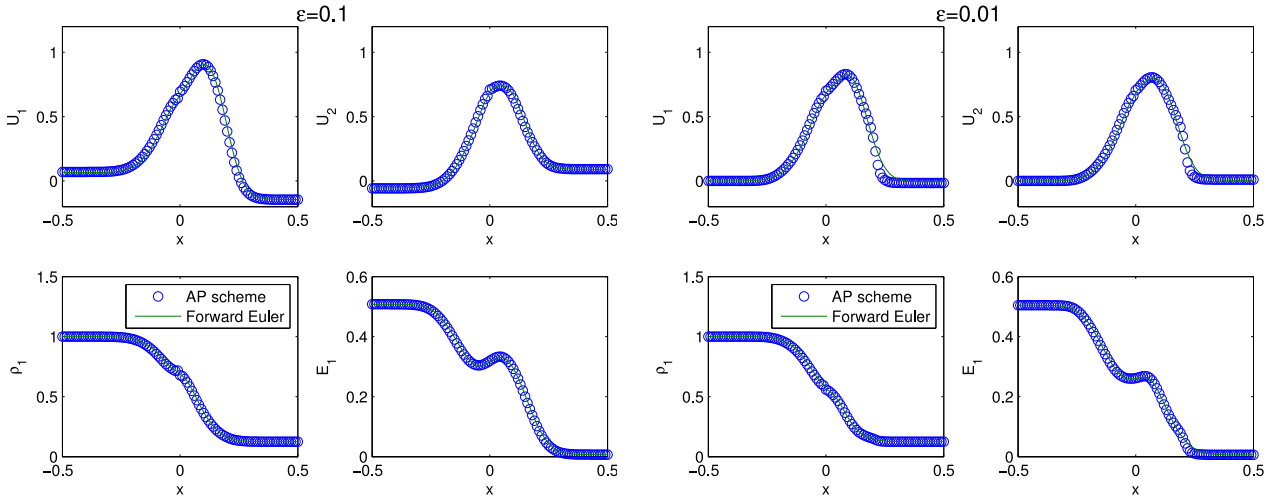


Figure 6.2.8: Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $t = 0.1$. For $\varepsilon = 0.1, 0.01$, we compare the results of the AP scheme, given by the circled lines, and the results of the forward Euler with a fine mesh, given by the solid lines.

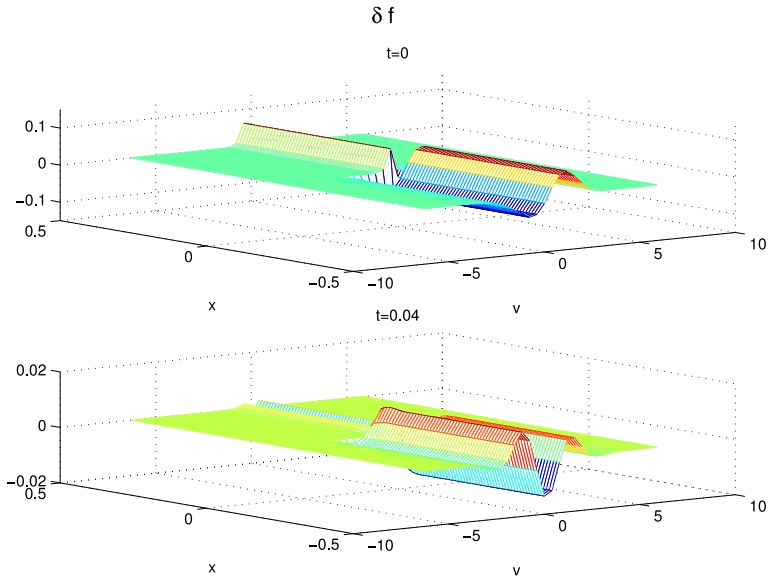


Figure 6.2.9: Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. $\varepsilon = 10^{-4}$. $\delta f = f_1 - \overline{M}_1$.

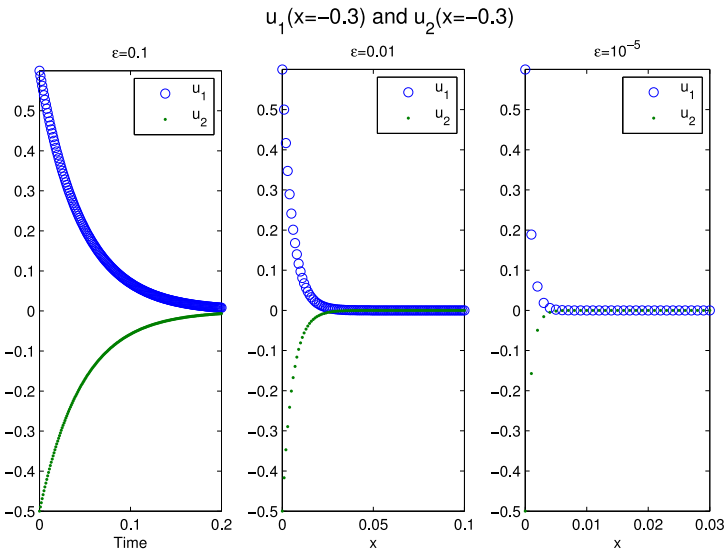


Figure 6.2.10: Multi-species Boltzmann equation. BKG-penalization method. The Sod problem. The three figures show velocities u_1 and u_2 at $x = -0.3$ as functions of time for $\varepsilon = 0.1, 0.01,$ and 10^{-5} respectively. Note different time scales for three figures.

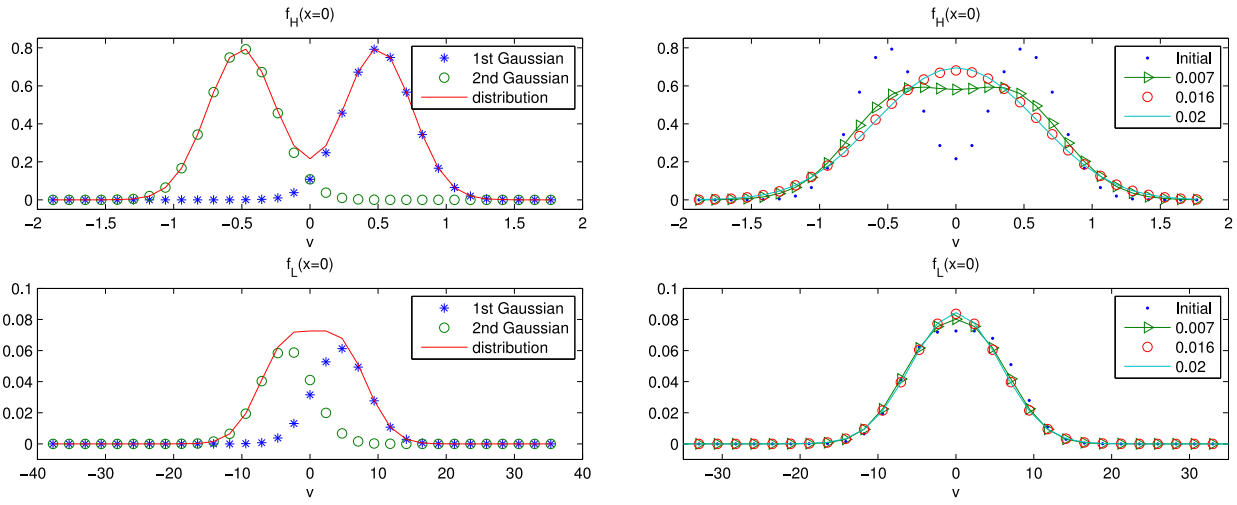


Figure 6.2.11: Multi-species Boltzmann equation. BKG-penalization method. The disparate masses problem. The left figure shows the initial distributions and the right figure show the time evolutions of f_H and f_L. f_H is put at the top and f_L is at the bottom. At $t = 0.007$, f_L is close to $\overline{M}_L$ while f_H is still far away from the equilibrium. Note the different scales for v.

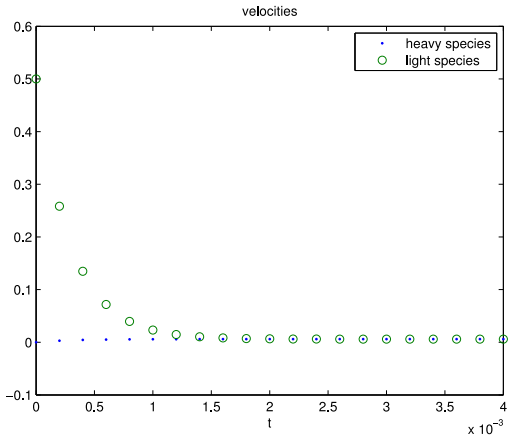


Figure 6.2.12: Multi-species Boltzmann equation. BKG-penalization method. The disparate masses problem. The velocities for the two species converge to each other.

Step 1. Solve the convection part:

$$\partial_t f_i + v \cdot \nabla_x f_i = 0. \tag{6.3.1}$$

Step 2. Solve the collision part:

$$\partial_t f_i = \frac{Q_i}{\varepsilon}. \tag{6.3.2}$$

This is based on the Strang splitting algorithm, and thus the accuracy is second order in time. The convection part is easy. We use the same algorithm presented in section 6.2.1 to deal with the transport term. The collision term, however, is the stiff part: if explicit methods are used, the time step is limited by ε, and if implicit methods are used, one meets the difficulties in numerically inverting Q. Following [52, 143], we reformulate (6.3.2) into an exponential form, and solves the new equation using explicit Runge-Kutta scheme.

Reformulation

Define

$$\mu = \sup_{x,i} |Q_i^-|, \qquad P_i = Q_i + \mu f_i, \tag{6.3.3}$$

and rewrite (6.3.2) as:

$$\begin{aligned}\partial_t\big((f_i-\overline{M}_i)e^{\mu t/\varepsilon}\big) &= \partial_t f_i e^{\mu t/\varepsilon} + (f_i-\overline{M}_i)\frac{\mu}{\varepsilon}e^{\mu t/\varepsilon} = \frac{1}{\varepsilon}(Q_i+\mu f_i-\mu\overline{M}_i)e^{\mu t/\varepsilon}\\ &= \frac{1}{\varepsilon}(P_i-\mu\overline{M}_i)e^{\mu t/\varepsilon},\end{aligned} \tag{6.3.4}$$

in which we have used the fact that ρ_i, $\bar{u}$ and $\bar{T}$ do not change in the collision step, and thus $\partial_t\overline{M}_i = 0$.

Remark 6.3.1.

1. *The equation* (6.3.4) *describes the evolution of the distance between the distribution function f_i and its equilibrium $\overline{M}_i$ multiplied by an integration factor, and thus removes the stiffness in* (6.3.2) *([52]).*

2. *Equation (6.3.4) holds for arbitrary constant μ. The way we choose μ is to guarantee the positivity of P; we refer to [143] for more details.*

Runge-Kutta method

Applying the K-stage Runge-Kutta method to the equation (6.3.4), one gets

$$\begin{cases}\text{Stage } \alpha: & (f^{l,(\alpha)}-\overline{M})e^{\frac{\mu}{\varepsilon}c_\alpha h} = (f^{l+\frac{1}{2}}-\overline{M}) + \dfrac{h}{\varepsilon}\displaystyle\sum_{\beta=1}^{\alpha-1} a_{\alpha\beta}(P^{l,(\beta)}-\mu\overline{M})e^{\frac{\mu}{\varepsilon}c_\beta h},\\ \text{Final stage:} & (f^{l+1}-\overline{M})e^{\frac{\mu}{\varepsilon}h} = (f^{l+\frac{1}{2}}-\overline{M}) + \dfrac{h}{\varepsilon}\displaystyle\sum_{\alpha=1}^{K} b_\alpha(P^{l,(\alpha)}-\mu\overline{M})e^{\frac{\mu}{\varepsilon}c_\alpha h}.\end{cases} \tag{6.3.5}$$

where $\sum_{\beta=1}^{\alpha-1} a_{\alpha\beta} = c_\alpha$, $\sum_\alpha b_\alpha = 1$, and superscripts $l,(\beta)$ stand for the estimate of the quantities at $t = t_l + c_\beta h$. μ is given in (6.3.3) and $f^{l+\frac{1}{2}}$ is obtained after the transport step. The last equation implies:

$$f^{l+1} = \left(1-e^{\lambda}-\sum_\alpha b_\alpha\lambda e^{\lambda(-1+c_\alpha)}\right)\overline{M} + e^{-\lambda}f^{l+\frac{1}{2}} + \sum_\alpha b_\alpha\lambda e^{\lambda(c_\alpha-1)}\frac{P^{l,(\alpha)}}{\mu},\quad \lambda = \frac{\mu h}{\varepsilon}, \tag{6.3.6}$$

In this formulation, one needs to compute $\overline{M}$ and P. The computation is the same as that in section 6.2.1, and will be omitted from here.

6.3.2 Positivity and asymptotic preserving properties

In this section, we discuss the positivity and asymptotic preserving properties of the Exp-RK method introduced above.

Theorem 6.3.1 (Positivity)**.** *The Exp-RK method described by* (6.3.1) *and* (6.3.5) *preserves the positivity property of f_i, i.e. there exist $h_* > 0$ and $\mu_* > 0$ such that $f^{l+1} > 0$ provided $f^l > 0$, if $0 < \Delta x < h_*$ and $\mu > \mu_*$.*

The proof of Theorem 6.3.1 is essentially the same as in the section for single-species system 5.2.2, and we omit the details here.

In order to prove the AP property, we need the following assumption:

Assumption 6.3.1. *The operator P satisfies*

$$\|P(f,f) - P(g,g)\| \lesssim \|f - g\|.$$

Remark that this assumption is true for Maxwell molecule in the d_2 norm defined in $P_s(\mathbb{R}^d)$ space ([183]). This can be easily seen by a similar argument carried in section 5.2.2. For completeness we provide it in appendix B.3.

Assume f^l and g^l are the initial conditions to (6.3.5). Define

$$d_0 = \|f^l - g^l\|, \quad d_\alpha = \|f^{l,(\alpha)} - g^{l,(\alpha)}\|, \quad \alpha = 1, \cdots, K,$$
$$\vec{d} = (d_1, d_2, \cdots, d_K)^T, \quad \vec{b} = (b_1, b_2, \cdots, b_K)^T, \quad \vec{e} = (1, 1, \cdots, 1)^T,$$

and $\mathbb{A}$ is a $K \times K$ strictly lower triangular matrix and $\mathbb{E}$ is a diagonal matrix given by

$$\mathbb{A}_{\alpha\beta} = \frac{\lambda}{\mu} a_{\alpha\beta} e^{(c_\beta - c_\alpha)\lambda}, \quad \beta < \alpha, \quad \text{and} \quad \mathbb{E} = \text{diag}\{e^{-c_1\lambda}, e^{-c_2\lambda}, \cdots, e^{-c_K\lambda}\}, \tag{6.3.7}$$

where $a_{\alpha\beta}$, b_α and c_α are the coefficients of the Runge-Kutta method in (6.3.5), λ is given in (6.3.6) and $\vec{e}$ is a K-dimensional vector. Then we have the following contraction lemma.

Lemma 6.3.1. *After one time step iteration in* (6.3.5)*, the scheme satisfies:*

$$\|f^{l+1} - g^{l+1}\| \le R(\lambda)\|f^l - g^l\| \quad \textit{with} \quad R(\lambda) = e^{-\lambda}\left(1 + \frac{C\lambda}{\mu}\vec{b}^T\mathbb{E}^{-1}(\mathbb{I} - C\mathbb{A})^{-1}\mathbb{E}\vec{e}\right). \tag{6.3.8}$$

where $\mathbb{I}$ is the identity matrix, and $C > 0$ is a constant.

Proof. The equation (6.3.5) implies, for $\alpha = 1, \cdots, K$,

$$(f^{l,(\alpha)} - g^{l,(\alpha)})e^{c_\alpha\lambda} = (f^l - g^l) + \sum_{\beta=1}^{\alpha-1} a_{\alpha\beta}\frac{\lambda}{\mu}e^{c_\beta\lambda}(P_f^{l,(\beta)} - P_g^{l,(\beta)}). \tag{6.3.9}$$

Using the triangle inequality and Assumption 6.3.1 produces

$$\|f^{l,(\alpha)} - g^{l,(\alpha)}\| \le \|f^l - g^l\|e^{-c_\alpha\lambda} + \sum_{\beta=1}^{\alpha-1} a_{\alpha\beta}\frac{\lambda}{\mu}e^{(c_\beta - c_\alpha)\lambda}\left(C\|f^{l,(\beta)} - g^{l,(\beta)}\|\right), \tag{6.3.10}$$

which implies

$$\vec{d} \le d_0\mathbb{E}\vec{e} + C\mathbb{A}\vec{d} \quad \text{and} \quad \vec{d} \le d_0\left(\mathbb{I} - C\mathbb{A}\right)^{-1}\mathbb{E}\vec{e}, \tag{6.3.11}$$

where we have used the fact that $\mathbb{A}$ is a strictly lower triangular matrix.

The final step in (6.3.5) yields

$$\left(f^{l+1} - g^{l+1}\right) = \left(f^l - g^l\right)e^{-\lambda} + \sum_{\alpha=1}^{K}\frac{hb_\alpha}{\varepsilon}\left(P_f^{l,(\alpha)} - P_g^{l,(\alpha)}\right)e^{(c_\alpha - 1)\lambda}, \tag{6.3.12}$$

Then (6.3.11) and Assumption 6.3.1 imply

$$\|f^{l+1} - g^{l+1}\| \leq d_0 e^{-\lambda} + \frac{C\lambda}{\mu} e^{-\lambda} \vec{b}^T \mathbb{E}^{-1} \vec{d} \leq d_0 e^{-\lambda} \left(1 + \frac{C\lambda}{\mu} \vec{b}^T \mathbb{E}^{-1} (\mathbb{I} - C\mathbb{A})^{-1} \mathbb{E} \vec{e}\right), \quad (6.3.13)$$

which completes the proof. □

Lemma 6.3.2. *Under Assumption 6.3.1,* $\|f^l - \overline{M}^l\| = \mathcal{O}(R(\lambda))$ *for each* l.

Proof. Take $g = \overline{M}$ in the previous lemma, one gets

$$\|f^{l+1} - \overline{M}^{l+1}\| \leq \|f^{l+\frac{1}{2}} - \overline{M}^{l+\frac{1}{2}}\| R(\lambda).$$

The convection step yields

$$\|f^{l+\frac{1}{2}} - \overline{M}^{l+\frac{1}{2}}\| < \|f^l - \overline{M}^l\| + \Delta t \|v \cdot \nabla_x \left(f^l - M^l\right)\| + \mathcal{O}(\Delta t) \leq \|f^l - \overline{M}^l\| + \mathcal{O}(\Delta t).$$

Combine these two inequalities, we conclude with the assertion of this lemma. □

Theorem 6.3.2. *The Exp-RK method defined in* (6.3.1) *and* (6.3.5) *is strong AP.*

Proof. By Lemma 6.3.2, it suffices to prove that $R(\lambda) = \mathcal{O}(\varepsilon)$ for small ε. Considering that $\mathbb{A}$ is a strictly lower triangular matrix and is thus a nilpotent, one has

$$\mathbb{E}^{-1} (\mathbb{I} - \mathbb{A})^{-1} \mathbb{E} = \mathbb{E}^{-1} \left(\mathbb{I} + \mathbb{A} + \mathbb{A}^2 + \cdots + \mathbb{A}^{K-1}\right) \mathbb{E} = \mathbb{I} + \mathbb{B} + \mathbb{B}^2 + \cdots + \mathbb{B}^{K-1}, \quad (6.3.14)$$

where $\mathbb{B} = \mathbb{E}^{-1}\mathbb{A}\mathbb{E}$, and we have used $\mathbb{A}^K = 0$. Further by (6.3.7), $\mathbb{B}_{\alpha\beta} = \mathbb{A}_{\alpha\beta} e^{c_\alpha \lambda - c_\beta \lambda} = \frac{\lambda}{\mu} a_{\alpha\beta}$. Thus $\mathbb{I} + \sum_\alpha \mathbb{B}^\alpha$ is a matrix with the elements of at most $\mathcal{O}(\lambda^K)$. Therefore, when $\varepsilon \ll 1$,

$$R(\lambda) = e^{-\lambda} \left(1 + \frac{C\lambda}{\mu} \vec{b} \cdot \mathbb{E}^{-1} (\mathbb{I} - \mathbb{A})^{-1} \mathbb{E} \cdot \vec{e}\right) = \mathcal{O}(e^{-\lambda} \lambda^K) < \mathcal{O}(\varepsilon).$$

This completes the proof. □

Remark 6.3.2. *Note that in the derivation for (6.3.4), the only requirement on $\overline{M}$ is that it does not change with respect to time in the collision part. Analytically, one can replace this $\overline{M}$ by arbitrary function that does not has the time variable. However, the associated numerics simply can not preserve the right asymptotic limit, as indicated in Lemma 6.3.2.*

6.3.3 Numerical example

The examples chosen here is the same as in the previous section 6.2.4 for the convenience of comparisons. We compute the reference solution by the forward Euler method using very small time step and mesh size.

A stationary shock

We compute the same stationary shock problem as shown in 6.2.4. $N = 2$ in (4.1.1), and the initial macroscopic quantities form a shock with zero speed, see (6.2.14). In phase space, the distribution function is given far away from the Maxwellian as in (6.2.15) and (6.2.16) with $\kappa = 0.2$. We choose $h = 0.0005$ and $\Delta x = 0.01$ to compute the equation using forward Euler scheme as a reference result. In Figure 6.3.1, we compare the numerical results given by the Exp-RK method, the BGK penalization method, and the reference solution for $\varepsilon = 1$ and 0.1. In Figure 6.3.2, we verified the AP property of the Exp-RK method, i.e. as ε goes to zero, the numerical results capture the stationary shock. We compare the convergence of the velocities for two species in Figure 6.3.3, where they gradually converge to the mean velocity u, and the smaller ε gives the faster convergence rate.

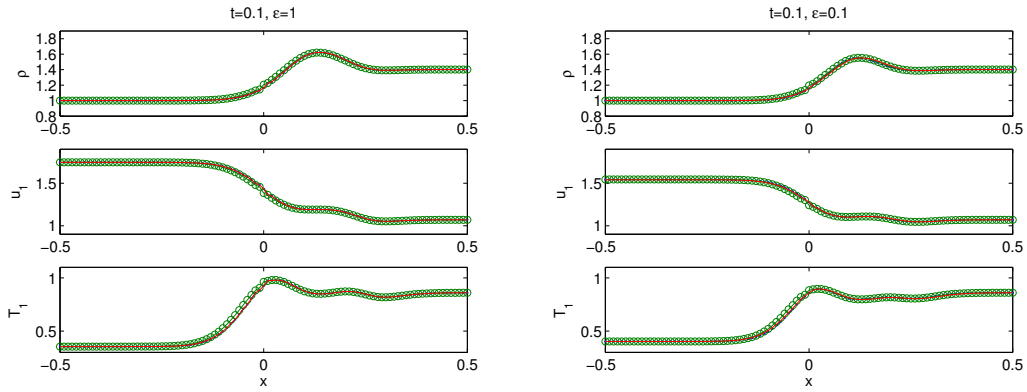


Figure 6.3.1: Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. We compare the numerical results at $t = 0.1$ by the Exp-RK method (the dotted line), the BGK penalization method (the circled line), and the reference solution (the solid line). The left figures are for $\varepsilon = 1$, and the right ones are for $\varepsilon = 0.1$.

A Sod problem

This Sod problem is the same one computed in section 6.2.4. Macroscopic quantities are chosen as in 6.2.17, and the initial distribution function is taken far away from the Maxwellian. We check indifferentialbility in Figure 6.3.4, and compare the numerical results given by this Exponential Runge-Kutta method and the BGK-penalization method, together with reference solution for $\varepsilon = 1$ and 0.1. AP property is verified in Figure 6.3.6 as we take $\varepsilon \to 0$. Figure 6.3.7 shows that in the long time limit, the two species have the same velocities, indicating the Euler limit is achieved. Smaller ε gives faster convergence rate. $f - M$ at $t = 0.1$ for kinetic regime $\varepsilon = 0.1$ and hydrodynamic regime $\varepsilon = 10^{-6}$ are plotted in Figure 6.3.8 reflecting AP on the microscopic level.

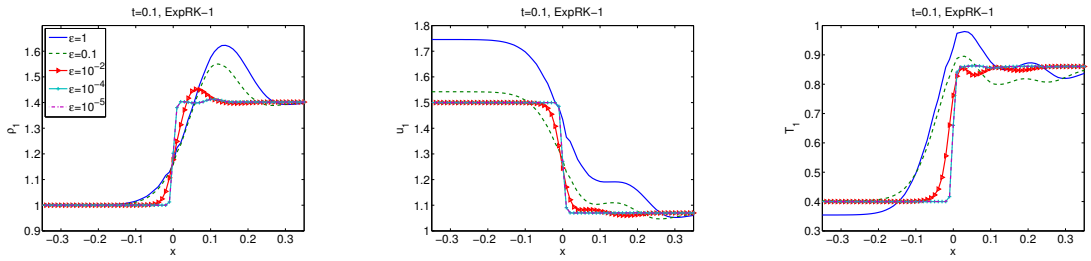


Figure 6.3.2: Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. As ε goes to zero, the results get to the Euler limit, i.e. a stationary shock.

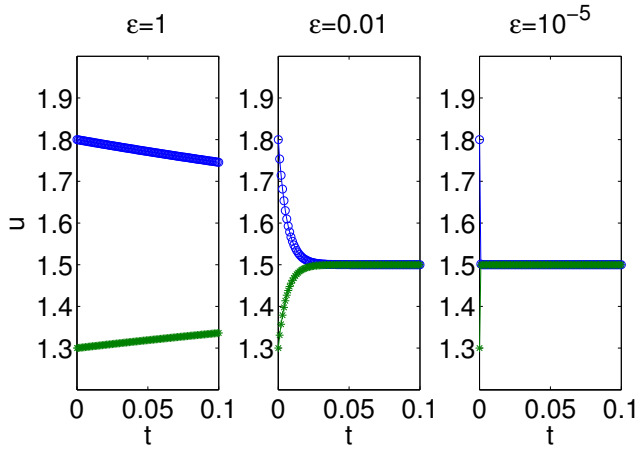


Figure 6.3.3: Multi-species Boltzmann equation. ExpRK method. The stationary shock problem. The velocities converge to the mean velocity at different rates for different ε's. The circles and dots stand for the velocities of the first and second species respectively. The smaller ε gives the faster convergence rate.

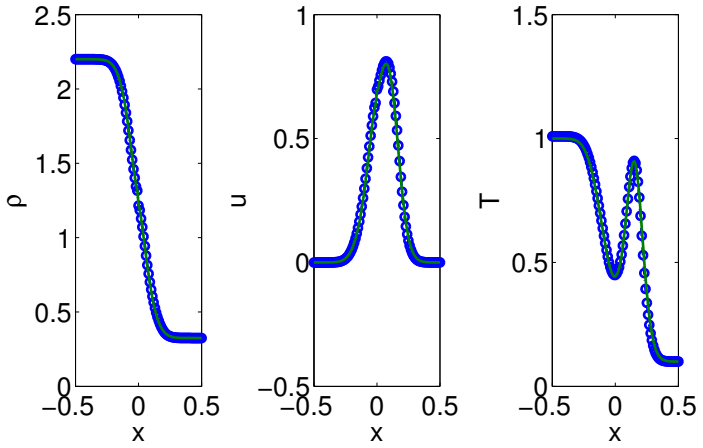


Figure 6.3.4: Multi-species Boltzmann equation. ExpRK method. The Sod problem. Indistinguishability. The solid lines are the results of treating the system as single species, and they agree with the results given by the dashed lines, which are the results of a system that has two species with the same mass.

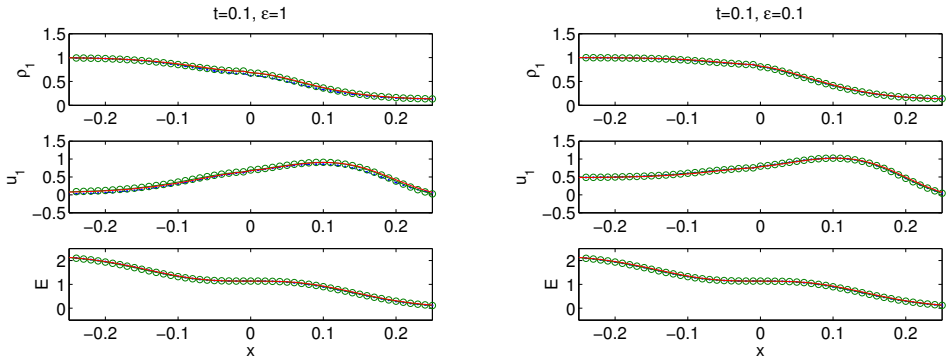


Figure 6.3.5: Multi-species Boltzmann equation. ExpRK method. The Sod problem. At $t = 0.1$, we compare the results give by the Exp-RK method (the dotted line), the BGK penalization method (the circled line) and the reference solution (the solid line). The left figures are for $\varepsilon = 1$, and the right ones are for $\varepsilon = 0.1$.

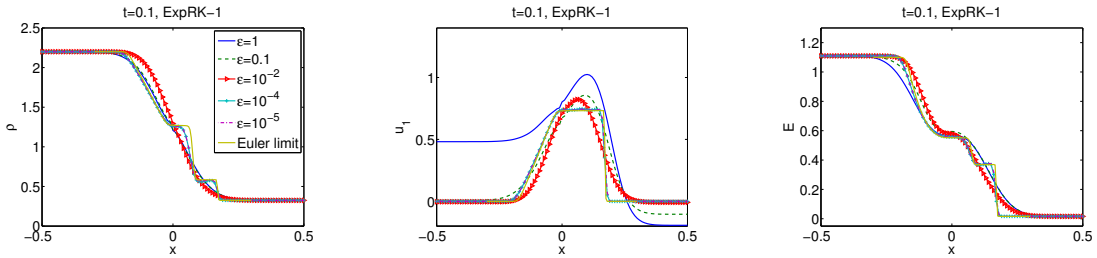


Figure 6.3.6: Multi-species Boltzmann equation. ExpRK method. The Sod problem. As ε goes to zero, the numerical results by the Exp-RK method converge to the Euler limit.

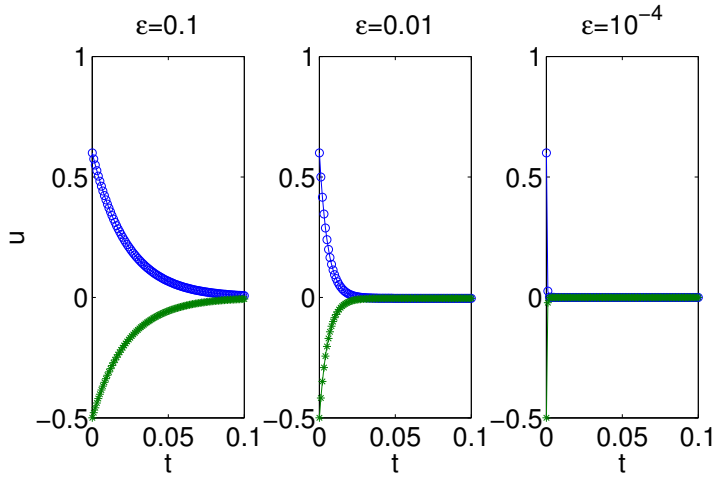


Figure 6.3.7: Multi-species Boltzmann equation. ExpRK method. The Sod problem. The velocities converge to the mean velocity at different rates for different ε's. The circled line stands for species one, and the dotted line stands for species two. The smaller ε gives the faster convergence rate.

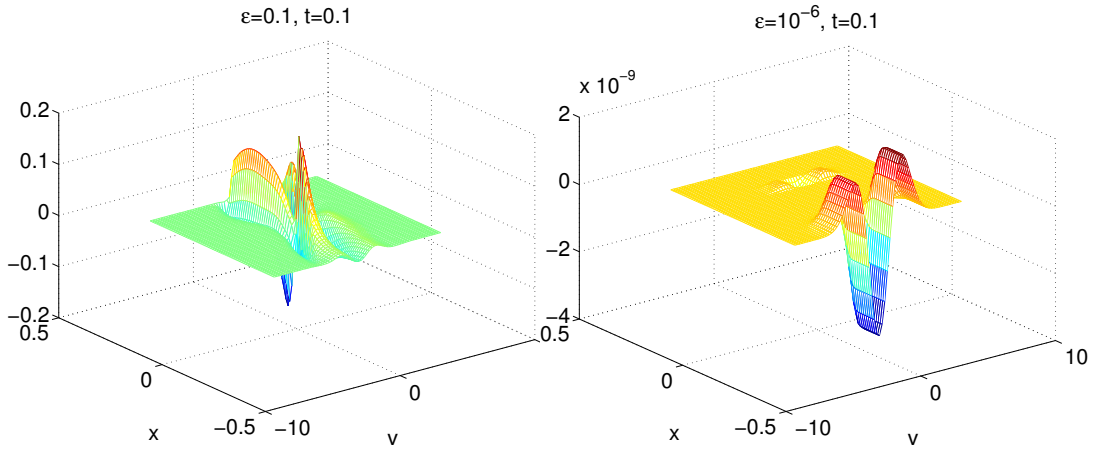


Figure 6.3.8: Multi-species Boltzmann equation. ExpRK method. The Sod problem. $f - M$ for $\varepsilon = 1$ and $\varepsilon = 10^{-6}$ at $t = 0.1$.

Appendices

Appendix A

Quantum system

A.1 Some basic analysis of the semi-classical Liouville systems

To understand the asymptotic behavior of the solution to the Liouville-A system (2.2.19), as mentioned in section 2.2.3, we look at a simpler model system:

$$\begin{cases} \partial_t g + \partial_x f + b(x)\partial_p g = 0\,, \\ \partial_t f + a(p)\partial_x f + \partial_x g + b(x)\partial_p f = \frac{\mathrm{i}}{\varepsilon} c(p) f\,, \\ g(0,x,p) = g_I(x,p), \quad f(0,x,p) = f_I(x,p). \end{cases} \tag{A.1.1}$$

The initial conditions g_I and f_I are bounded smooth functions independent on ε, $b > 0$ and the set of zeros for $c(p)$: $S_c = \{p : c(p) = 0\}$ is measured zero. It is easy to check that (A.1.1) is a linear hyperbolic system, and the solutions g and f are bounded uniformly in ε [43].

A.1.1 Weak convergence

We consider the weak limit of the solution of (A.1.1) in this subsection. To do this, we introduce the inner product $\langle\, \cdot, \cdot \,\rangle$ as

$$\langle\, u, v \,\rangle = \int_0^\infty \int_{\mathbb{R}^2} u(t,x,p)\bar{v}(t,x,p)\,\mathrm{d}x\,\mathrm{d}p\,\mathrm{d}t.$$

Choose an arbitrary test function $h \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^2)$, take the inner product on both side of (A.1.1) w.r.t h, one gets

$$\begin{cases} \langle\, \partial_t g, h \,\rangle - \langle\, f, \partial_x h \,\rangle + \langle\, b\partial_p g, h \,\rangle = 0\,, \\ \langle\, f, \partial_t h \,\rangle + \langle\, af, \partial_x h \,\rangle + \langle\, g, \partial_x h \,\rangle + \langle\, bf, \partial_p h \,\rangle = -\frac{\mathrm{i}}{\varepsilon}\langle\, cf, h \,\rangle\,. \end{cases} \tag{A.1.2}$$

The derivatives in the equation of (A.1.2) are acted on the smooth function h, and the left side is bounded. One gets

$$\langle\, cf, h \,\rangle \to 0 \quad \text{as} \quad \varepsilon \to 0 \quad \text{for all } h \in C_0^\infty(\mathbb{R}^+ \times \mathbb{R}^2).$$

Given that c is almost surely nonzero, and f is bounded, one gets

$$f \rightharpoonup 0 \text{ weakly.}$$

Combined with the first equation in (A.1.2), one gets

$$\partial_t g + b(x)\partial_p g \rightharpoonup 0 \text{ weakly.}$$

A.1.2 Strong convergence: for constant b

In these two subsections, we formally prove that before getting close to the crossing region, $c(p)$ is assumed to be bigger than a constant c_0 that is unrelated to ε. In this region, f is constantly small and controlled by $\mathcal{O}(\varepsilon)$. This subsection is for the case when the speed on p direction is a constant: $b(x) = \beta$. Along the p-characteristic line $p(t) = p_0 + \beta t$, one applies the Fourier transform to the x-variable, and gets:

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{f} = \mathrm{i}R(t)\boldsymbol{f}, \tag{A.1.3}$$

where $\boldsymbol{f}(t,\eta) = (\,\hat{g}(t,\eta,p(t)),\ \hat{f}(t,\eta,p(t))\,)^T$ and

$$R(t) = \begin{pmatrix} 0 & -\eta \\ -\eta & c(p(t))/\varepsilon - \eta a(p(t)) \end{pmatrix}.$$

The two eigenvalues of $R(t)$ are both real, and thus the system above has a bounded solution satisfying

$$|\hat{g}(t,\eta)|^2 + |\hat{f}(t,\eta)|^2 = |\hat{g}_I(\eta)|^2 + |\hat{f}_I(\eta)|^2.$$

The equivalence between norms gives:

$$|\hat{g}(t,\eta)| + |\hat{f}(t,\eta)| < C(|\hat{g}_I(\eta)| + |\hat{f}_I(\eta)|).$$

Adopt it into the solution to (A.1.3), one gets

$$\begin{aligned} |\partial_t g(t,x)| &= \frac{1}{2\pi}\left|\int_{\mathbb{R}} \eta \hat{f}(t,\eta) \mathrm{e}^{\mathrm{i}\eta x} \mathrm{d}\eta\right| \le \frac{1}{2\pi}\int_{\mathbb{R}} |\eta|\,|\hat{f}(t,\eta)| \mathrm{d}\eta \\ &\le C\int_{\mathbb{R}} |\eta|\,\left(|\hat{g}_I(\eta)| + |\hat{f}_I(\eta)|\right) \mathrm{d}\eta \\ |\partial_{tx}^2 g(t,x)| &= \frac{1}{2\pi}\left|\int_{\mathbb{R}} \eta^2 \hat{f}(t,\eta) \mathrm{e}^{\mathrm{i}\eta x} \mathrm{d}\eta\right| \le \frac{1}{2\pi}\int_{\mathbb{R}} |\eta|^2\,|\hat{f}(t,\eta)| \mathrm{d}\eta \\ &\le C\int_{\mathbb{R}} |\eta|^2\,\left(|\hat{g}_I(\eta)| + |\hat{f}_I(\eta)|\right) \mathrm{d}\eta \end{aligned}$$

If the initial conditions $\hat{g}_I$ and $\hat{f}_I$ are smooth enough, and decay fast as $\eta \to \infty$, one could easily get that $\partial_t g(t,x)$ and $\partial_{tx}^2 g(t,x)$ are both bounded in time independent of ε, and thus g and $\partial_x g$ are slowly varying in time.

Remark A.1.1. *In the derivation above, we dropped the $p(t)$-dependence in the functions for simplicity. The partial derivative in the t-variable ∂_t should be understood as taking along the p-characteristic line $p(t)$, i.e. $\partial_t g(t,x) = \partial_t g(t,x,p) + \beta\partial_p g(t,x,p)$.*

Assume that $f_I(x) \equiv 0$, one follows the characteristics of $x(t)$ by solving $\dot{x}(t) = a(t,x)$ and gets:

$$f\big(t, x(t), p(t)\big) = -\int_0^t \exp\left(\frac{\mathrm{i}}{\varepsilon}\int_{t-s}^t c\left(p(\tau)\right)\,\mathrm{d}\tau\right)\,\frac{\partial g}{\partial x}\big(t-s, x(t-s), p(t-s)\big)\,\mathrm{d}s\,.$$

By the assumption, before hitting the crossing region, $c(p(\tau)) > c_0 > 0$, then the stationary phase argument suggests that, given slowly varying $\partial_x g(t, x, p(t))$, $f = \mathcal{O}(\varepsilon)$.

The observations from the above two subsections suggest that $f \rightharpoonup 0$, and before getting close to the crossing region, f is as small as of $\mathcal{O}(\varepsilon)$. Based on these arguments, for the Liouville-A system (2.2.19), and propose the following conjecture: if σ_{12} and σ_{21} are initially zero, then:

Case 1. If $p \ll -\sqrt{\varepsilon}$, then σ_{12} and σ_{21} are of $o(\sqrt{\varepsilon})$;

Case 2. If $p \in [-\sqrt{\varepsilon}, \sqrt{\varepsilon}]$, then σ_{12} and σ_{21} are of $\mathcal{O}(\sqrt{\varepsilon})$, and slowly varying;

Case 3. If $p \gg \sqrt{\varepsilon}$, σ_{12} and σ_{21} are highly oscillatory, and converge to 0 weakly.

A.2 The integration method of a simple model system

In this section, we apply the method in (2.3.6) onto a simple model to show stability.

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{f}(t, p_0 + t) = R(p_0 + t)\,\mathbf{f}(t, p_0 + t),$$

where $\mathbf{f}(t,p) = (g(t,p), f(t,p))^T$, $p = p_0 + t$, and

$$R(p) = \begin{pmatrix} r_{11}(p) & r_{12}(p) \\ -\bar{r}_{12}(p) & r_{22}^{\varepsilon}(p) \end{pmatrix}.$$

with $r_{22}^{\varepsilon}(p) = r_{22}(p) + \frac{\mathrm{i}}{\varepsilon}c(p)$, while r_{11} and r_{22} are purely imaginary, and $c(p)$ real and positive. $r_{11},\ r_{22}$ and r_{12} are independent on ε.

Set up mesh as $t^j = j\Delta t$, and $p^i = -\frac{1}{2} + (i-1)\Delta p$, with Δp and Δt being the mesh size. Denote g_i^j and f_i^j as the numerical result at (t^j, p^i), then (2.3.6) gives

$$f(t,p(t)) \;=\; f_i^j\, \mathrm{e}^{\int_0^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau} - g_i^j\, \bar{r}_{12}(p^i) \int_0^t \mathrm{e}^{\int_s^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}s, \tag{A.2.1a}$$

$$g(t,p(t)) \;=\; g_i^j \;+\; g(t,p(t))\, r_{11}(p(t)) + r_{12}(p^i) \int_0^t \tilde{f}(t^j+s, p^i+s)\,\mathrm{d}s \tag{A.2.1b}$$

Plug (A.2.1a) into (A.2.1b), and evaluate them at (t^{j+1}, p^{i+1}), one obtains:

$$\begin{aligned}\big(1 - r_{11}(p^{i+1})\Delta t\big)\, g_{i+1}^{j+1} \;=\;& g_i^j \left(1 - |r_{12}(p^i)|^2 \int_0^{\Delta t}\int_0^t \mathrm{e}^{\int_s^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}s\,\mathrm{d}t\right)\\ &+ f_i^j\, r_{12}(p^i) \int_0^{\Delta t} \mathrm{e}^{\int_0^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}t\ .\end{aligned} \tag{A.2.2}$$

Written in vector form gives

$$\mathbf{f}_{i+1}^{j+1} \;=\; M_i\, \mathbf{f}_i^j, \tag{A.2.3}$$

with

$$M_i \;=\; \begin{pmatrix} \frac{1-|r_{12}(p^i)|^2 \int_0^{\Delta t}\int_0^t \mathrm{e}^{\int_s^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}s\,\mathrm{d}t}{1-r_{11}(p^{i+1})\Delta t} & \frac{r_{12}(p^i)\int_0^{\Delta t}\mathrm{e}^{\int_0^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}t}{1-r_{11}(p^{i+1})\Delta t} \\ -\bar{r}_{12}(p^i)\int_0^{\Delta t}\mathrm{e}^{\int_s^{\Delta t} r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}s & \mathrm{e}^{\int_0^{\Delta t} r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau} \end{pmatrix}.$$

The following quantities in the matrix M_i should be evaluated very accurately:

$$\begin{aligned} F_0 &= \mathrm{e}^{\int_0^{\Delta t} r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau},\\ F_1 &= \int_0^{\Delta t} \mathrm{e}^{\int_0^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}t,\\ F_2 &= \int_0^{\Delta t}\int_0^t \mathrm{e}^{\int_s^t r_{22}^{\varepsilon}(p^i+\tau)\mathrm{d}\tau}\mathrm{d}s\,\mathrm{d}t. \end{aligned}$$

Remark A.2.1. *These three quantities only depend on the mesh grid point index i but not the*

time steps index j, thus they only need to be computed once at the beginning of the computation.

Given $\varepsilon \ll \Delta t$, the integrands of F_1 and F_2 are highly oscillatory, and one can see that $|F_1| \sim \mathcal{O}(\varepsilon)$ and $|F_2| \sim \mathcal{O}(\varepsilon^2)$. Simple calculation shows that M_i can be written as

$$M_i \,=\, \Omega\, \tilde{M}_i,$$

with

$$\Omega = \mathrm{diag}\left(\frac{1}{1 - r_{11}(p^{j+1})\Delta t},\, 1\right), \qquad \tilde{M}_i = \begin{pmatrix} 1 - |r_{12}(p^i)|^2\, F_2 & r_{12}\, F_1 \\ -\bar{r}_{12}\, \bar{F}_1 & F_0 \end{pmatrix}.$$

With purely imaginary r_{11} and r_{22}^{ε}, it is easy to prove that $\|\Omega\|_\infty \leq 1$, and $\|\tilde{M}_i\|_\infty \leq (1 + \mathcal{O}(\varepsilon\Delta t))$, and thus $\|M_i\|_\infty \leq (1 + \mathcal{O}(\varepsilon\Delta t))$. This implies asymptotic stability of the scheme (A.2.3) independent of $\varepsilon \to 0$.

A.3 The derivations of (3.2.18)

In this section, we give the derivations of the semi-classical Liouville system (3.2.18). We firstly study the Moyal product. From its definition in (3.2.8), one gets:

$$\begin{aligned} A\#B &= A\mathrm{e}^{\frac{\mathrm{i}\varepsilon}{2}\left(\overleftarrow{\nabla_{\mathbf{x}}}\cdot\overrightarrow{\nabla_{\mathbf{p}}} - \overleftarrow{\nabla_{\mathbf{p}}}\cdot\overrightarrow{\nabla_{\mathbf{x}}}\right)} B \\ &= \sum_n \frac{1}{n!}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^n A\left(\overleftarrow{\nabla_{\mathbf{x}}}\cdot\overrightarrow{\nabla_{\mathbf{p}}} - \overleftarrow{\nabla_{\mathbf{p}}}\cdot\overrightarrow{\nabla_{\mathbf{x}}}\right)^n B \\ &= \sum_n \sum_{k=0}^{n} \frac{(-1)^k}{n!}\binom{n}{k}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^n A\left(\overleftarrow{\nabla_{\mathbf{x}}}\cdot\overrightarrow{\nabla_{\mathbf{p}}}\right)^{n-k}\left(\overleftarrow{\nabla_{\mathbf{p}}}\cdot\overrightarrow{\nabla_{\mathbf{x}}}\right)^k B. \end{aligned}$$

In particular, if both symbols depend only on one variable (spatial $\mathbf{x}$ or phase $\mathbf{p}$), the Moyal product becomes the ordinary product:

$$A(\mathbf{x})\#B(\mathbf{x}) = A(\mathbf{x})B(\mathbf{x}), \quad \text{and} \quad A(\mathbf{p})\#B(\mathbf{p}) = A(\mathbf{p})B(\mathbf{p}).$$

With this, one can evaluate the symbol H'.

$$\begin{aligned} H' &= \Theta(\mathbf{x})\#H(\mathbf{x},\mathbf{p})\#\Theta^\dagger(\mathbf{x}) \\ &= \Theta(\mathbf{x})\#\big(U(\mathbf{x})\mathbb{I} + V(\mathbf{x})\big)\#\Theta^\dagger(\mathbf{x}) + \Theta(\mathbf{x})\#\left(\frac{\mathbf{p}^2}{2}\mathbb{I}\right)\#\Theta^\dagger(\mathbf{x}) \\ &= \operatorname{diag}\left\{U + \sqrt{u^2+v^2},\, U - \sqrt{u^2+v^2}\right\} + \Theta(\mathbf{x})\#\left(\frac{\mathbf{p}^2}{2}\mathbb{I}\right)\#\Theta^\dagger(\mathbf{x}) \\ &= \Lambda(\mathbf{x},\mathbf{p}) + \mathrm{i}\varepsilon\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta(\mathbf{x})\Theta^\dagger(\mathbf{x}) + \frac{\varepsilon^2}{2}\nabla_{\mathbf{x}}\Theta(\mathbf{x})\cdot\nabla_{\mathbf{x}}\Theta^\dagger(\mathbf{x}), \end{aligned} \tag{A.3.1}$$

where $\Lambda = U(\mathbf{x}) + \Lambda_V$ with Λ_V defined in (3.1.21). Hence we obtain the equation (3.2.17) in Section 3.2.2. Here the derivation is based on (3.2.15), (3.2.8), and the following identities:

$$\begin{aligned} \Theta(\mathbf{x})\#\left(\frac{\mathbf{p}^2}{2}\mathbb{I}\right) &= \sum_n \frac{1}{n!}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^n \Theta(\mathbf{x})\left(\overleftarrow{\nabla_{\mathbf{x}}}\cdot\overrightarrow{\nabla_{\mathbf{p}}}\right)^n\left(\frac{\mathbf{p}^2}{2}\mathbb{I}\right) \\ &= \frac{\mathbf{p}^2}{2}\Theta + \frac{\mathrm{i}\varepsilon}{2}\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta + \frac{1}{2}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^2\Theta\left(\overleftarrow{\nabla_{\mathbf{x}}}\cdot\overrightarrow{\nabla_{\mathbf{p}}}\right)^2\left(\frac{\mathbf{p}^2}{2}\mathbb{I}\right) \\ &= \frac{\mathbf{p}^2}{2}\Theta + \frac{\mathrm{i}\varepsilon}{2}\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta + \frac{1}{2}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^2\Delta_{\mathbf{x}}\Theta, \\ \left(\frac{\mathbf{p}^2}{2}\Theta(\mathbf{x})\right)\#\Theta^\dagger(\mathbf{x}) &= \frac{\mathbf{p}^2}{2} - \frac{\mathrm{i}\varepsilon}{2}\Theta\,(\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta^\dagger) + \frac{1}{2}\left(\frac{\mathrm{i}\varepsilon}{2}\right)^2\Theta\,\Delta_{\mathbf{x}}\Theta^\dagger, \\ \big(\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta(\mathbf{x})\big)\#\Theta^\dagger(\mathbf{x}) &= \mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta\,\Theta^\dagger - \frac{\mathrm{i}\varepsilon}{2}\nabla_{\mathbf{x}}\Theta\cdot\nabla_{\mathbf{x}}\Theta^\dagger, \\ \Delta_{\mathbf{x}}\Theta(\mathbf{x})\#\Theta^\dagger(\mathbf{x}) &= \Delta_{\mathbf{x}}\Theta\,\Theta^\dagger, \\ \Theta(\mathbf{x})\,(\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta^\dagger(\mathbf{x})) &= -(\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta)\,\Theta^\dagger, \\ -2\nabla_{\mathbf{x}}\Theta(\mathbf{x})\cdot\nabla_{\mathbf{x}}\Theta^\dagger(\mathbf{x}) &= \Theta\Delta_{\mathbf{x}}\Theta^\dagger + \Delta_{\mathbf{x}}\Theta\,\Theta^\dagger. \end{aligned}$$

To finally derive (3.2.18), one needs to go back to the evolution of F' given in (3.2.14):

$$\begin{aligned} \mathrm{i}\varepsilon\frac{\partial F'}{\partial t} &= [\Lambda, F'] - \frac{\mathrm{i}\varepsilon}{2}\{\Lambda, F\} + \frac{\mathrm{i}\varepsilon}{2}\{F, \Lambda\} + \mathrm{i}\varepsilon[\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta\,\Theta^{\dagger}, F'] \;+\; O(\varepsilon^2) \\ &= [\Lambda, F'] - \frac{\mathrm{i}\varepsilon}{2}[\nabla_{\mathbf{p}}\Lambda, \nabla_{\mathbf{x}}F]_+ + \frac{\mathrm{i}\varepsilon}{2}[\nabla_{\mathbf{x}}\Lambda, \nabla_{\mathbf{p}}F]_+ \\ &\qquad + \mathrm{i}\varepsilon\,[\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta\,\Theta^{\dagger}, F'] \;+\; O(\varepsilon^2), \end{aligned} \tag{A.3.2}$$

with $[A, B]_+ = AB + BA$. By definition (3.2.20), it is easy to obtain:

$$[\Lambda, F'] = \begin{pmatrix} 0 & \lambda^+ - \lambda^- \\ \lambda^- - \lambda^+ & 0 \end{pmatrix}, \qquad [\nabla_{\mathbf{p}}\Lambda, \nabla_{\mathbf{x}}F]_+ = 2\begin{pmatrix} \mathbf{p}\cdot\nabla_{\mathbf{x}}f^+ & \mathbf{p}\cdot\nabla_{\mathbf{x}}f^i \\ \mathbf{p}\cdot\nabla_{\mathbf{x}}\overline{f^i} & \mathbf{p}\cdot\nabla_{\mathbf{x}}f^- \end{pmatrix},$$

$$[\nabla_{\mathbf{x}}\Lambda, \nabla_{\mathbf{p}}F]_+ = 2\begin{pmatrix} \nabla_{\mathbf{x}}(U+E)\cdot\nabla_{\mathbf{p}}f^+ & \nabla_{\mathbf{x}}U\cdot\nabla_{\mathbf{p}}f^i \\ \nabla_{\mathbf{x}}U\cdot\nabla_{\mathbf{p}}\overline{f^i} & \nabla_{\mathbf{x}}(U-E)\cdot\nabla_{\mathbf{p}}f^- \end{pmatrix},$$

$$[\mathbf{p}\cdot\nabla_{\mathbf{x}}\Theta\,\Theta^{\dagger}, F'] = \xi\begin{pmatrix} 2\Re\{f^i\} & f^- - f^+ \\ f^- - f^+ & -2\Re\{f^i\} \end{pmatrix},$$

Plug them back in (A.3.2), up to the first order in ε, one gets (3.2.18).

Appendix B

Kinetic theory

B.1 Positivity of the mass density in ExpRK-V

Theorem B.1.1. *The method ExpRK-V defined by (5.2.8) gives positive ρ, and the negative part of T is at most of order $O(h\varepsilon)$.*

To prove this theorem, we firstly check the following lemma.

Lemma B.1.1. *In each sub-stage, the distribution function $f^{(i)}$ and $M^{(i)}$ have the same first $d+2$ moments.*

Proof. We prove this for sub-stage i. Assume for $\forall j < i$, one has

$$\int \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix} (f^{(j)} - M^{(j)}) dv = 0. \tag{B.1.1}$$

Then, one could take moments of the first equation in the scheme (5.2.8a), and gets

$$\int \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix} (f^{(i)} - M^{(i)}) e^{c_i \lambda} dv = \int \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix} (f^n - M^n)\, dv \tag{B.1.2a}$$

$$+ \sum_{j=1}^{i-1} a_{ij} \frac{\lambda}{\mu} e^{c_j \lambda} \int \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix} \left(P^{(j)} - \mu M^{(j)} \right) dv \tag{B.1.2b}$$

$$- \sum_{j=1}^{i-1} a_{ij} \frac{\lambda}{\mu} e^{c_j \lambda} \int \begin{pmatrix} 1 \\ v \\ \frac{v^2}{2} \end{pmatrix} \left(\varepsilon v \cdot \nabla_x f^{(j)} - \varepsilon \partial_t M^{(j)} \right) dv \tag{B.1.2c}$$

(B.1.2a) is zero for sure, (B.1.2b) is zero by definition of P and (B.1.1). (B.1.2c) is zero because of the computation from (5.2.11). Thus it is obvious that $f^{(i)}$ and $M^{(i)}$ share the same moments on each stage. □

With the previous lemma in hand, one could prove Theorem 4.

Proof. As in the previous lemma, we only do the proof for sub-stage i. The final step can be dealt with in the same way. Rewrite the second equation of (5.2.8a) in Shu-Osher representation

$$\int \phi f^{(i)} dv = \sum_{j=1}^{i-1} \left(\alpha_{ij} \int \phi f^{(j)} dv + \beta_{ij} h \int \phi v \cdot \nabla_x f^{(j)} dv \right) \tag{B.1.3}$$

This moment equation is the same as the equation on ρ in the Euler system, and the classical proof for ρ being positive for the Euler equation can just be adopted [83]. To check the

positivity of T, one just need to make use of the last line of the moment equation, i.e.

$$\begin{aligned}\int \frac{v^2}{2} f^{(i)} dv =& \sum_{j=1}^{i-1} \left(\alpha_{ij} \int \frac{v^2}{2} f^{(j)} dv + \beta_{ij} h \int \frac{v^2}{2} v \cdot \nabla_x f^{(j)} dv \right) \\ =& \sum_{j=1}^{i-1} \left(\alpha_{ij} \int \frac{v^2}{2} f^{(j)} dv + \beta_{ij} h \int \frac{v^2}{2} v \cdot \nabla_x M^{(j)} dv \right) \qquad \text{(B.1.4a)} \\ &+ h \sum_{j=1}^{i-1} \beta_{ij} \int \frac{v^2}{2} v \cdot \nabla_x \left(f^{(j)} - M^{(j)} \right) dv \qquad \text{(B.1.4b)}\end{aligned}$$

(B.1.4a) is exactly what one could get when computing for E in the Euler system: the form of M closes it up. So the classical method to prove that $E > \frac{\rho u^2}{2}$ in Runge-Kutta scheme could be used, and the only thing new is from (B.1.4b). However, as proved in the section about AP, the difference between f and M is at most of ε, thus (B.1.4b) is of order $O(h\varepsilon)$. □

B.2 $\|P(f) - P(g)\| \leq \|f - g\|$ for single species

B.2.1 In d_2 norm

We adopt the results from [183]. They denote P_2 the collection of distributions F such that

$$\int_{R^d} |v|^2 dF(v) < \infty$$

A metric d_2 on P_2 is defined by

$$\mathrm{d}_2(F, G) = \sup_\xi \frac{\hat{f}(\xi) - \hat{g}(\xi)}{|\xi|^2} \qquad \text{(B.2.1)}$$

where $\hat{f}$ is the Fourier transform of F

$$\hat{f}(\xi) = \int e^{-i\xi \cdot v} dF(v)$$

One can transform the Boltzmann equation into its Fourier space and obtains[170, 20]

$$\partial_t \hat{f}(t,\xi) = \int_{S^2} B\left(\frac{\xi\cdot n}{|\xi|}\right)\left[\hat{f}(\xi^+)\hat{f}^(\xi^-) - \hat{f}^(\xi)\hat{f}(0)\right] dn \tag{B.2.2}$$

where $\xi^{\pm} = \frac{\xi \pm |\xi| n}{2}$

Theorem B.2.1. $\mathrm{d}_2(P_f, P_g) < \mathrm{d}_2(f,g)$ *for Maxwell molecules with cut-off collision kernel.*

Proof. For Maxwell molecule with cut-off collision kernel $\int B = S$. Thus

$$\sup|Q^-| = \sup\left|\int B f_* d\Omega dv_*\right| = \sup|\rho S| < \infty.$$

Considering $P = Q + \mu f = Q^+ + (\mu - Q^-) f$, it is enough to prove $\mathrm{d}_2(Q_f^+, Q_g^+) < C\mathrm{d}_2(f,g)$ for C big enough. Given

$$\hat{Q}_f^+ = \int_{S^2} B\left(\frac{\xi\cdot n}{|\xi|}\right)\left[\hat{f}(\xi^+)\hat{f}^(\xi^-)\right] dn,$$

one has

$$\frac{\hat{Q}_f^+ - \hat{Q}_g^+}{|\xi|^2} = \int_{S^2} B\left(\frac{\xi\cdot n}{|\xi|}\right)\left[\frac{\hat{f}(\xi^+)\hat{f}(\xi^-) - \hat{g}(\xi^+)\hat{g}(\xi^-)}{|\xi|^2}\right] dn$$

From [183], one gets

$$\left|\frac{\hat{f}(\xi^+)\hat{f}(\xi^-) - \hat{g}(\xi^+)\hat{g}(\xi^-)}{|\xi|^2}\right| \leq \sup\left|\frac{\hat{f}-\hat{g}}{|\xi|^2}\right|$$

Thus, one has:

$$\mathrm{d}_2(Q_f^+, Q_g^+) = \sup_\xi \left|\frac{\hat{Q}_f^+ - \hat{Q}_g^+}{|\xi|^2}\right| \leq S \sup\left|\frac{\hat{f}-\hat{g}}{|\xi|^2}\right| = S\mathrm{d}_2(f,g)$$

□

B.2.2 In L_2 norm

Theorem B.2.2. *Given a system with cut-off collision kernel, if* $\|f\|_\infty < \infty$*, then* $\|P\|_2$ *is controlled by* $\|f\|_2$*.*

Proof. Since $\mu = \sup\|\int Bf_* dv_* d\Omega\| < \infty$, one just need to prove that Q^+ is also a bounded operator. We only consider Maxwell molecule with cut-off collision kernel.

$$\begin{aligned}
\|Q^+(f)\|_2 &= \int \left(\int \int B(\Omega) f' f'_* dv_* d\Omega \right)^2 dv \\
&\leq \int \int \int \left(B(\Omega) f' f'_* \right)^2 dv_* d\Omega dv \\
&\leq \int B^2(\Omega) d\Omega \int (f')^2 dv' \int (f'_*)^2 dv'_* \\
&\leq C\|f\|_2^2 \\
&\leq C\sqrt{\|f\|_\infty \|f\|_1}\|f\|_2
\end{aligned}$$

In this derivation, we used that:

1. the Hessian matrix generated by changing of variable from (v, v_*) to (v', v_*) has determinant one;

2. the collision kernel is a cut-off collision;

3. Hölder Inequality.

Given $\|f\|_1 = \rho$, under the assumption that $\|f\|_\infty$ is bounded, one has that Q^+ is a bounded operator on $\|f\|$. □

B.3 $\|P(f)\| < C\|f\|$ multi-species

B.3.1 In d_2 norm

This is extension from section B.2.1. Here in multipspecies system, the Fourier transform of the collision term is [19]:

$$\partial_t \hat{f}_i(t,\xi) = \sum_j \int_{S^2} B_{ij}\left(\frac{\xi\cdot n}{|n|}\right)\left[\hat{f}(\xi_{ij}^+)\hat{f}(\xi_{ij}^-) - \hat{f}(\xi)\hat{f}(0)\right] dn, \tag{B.3.1}$$

where $\xi_{ij}^+ = \frac{m_i\xi + m_j\xi|\omega|}{m_i+m_j}$ and $\xi_{ij}^- = \frac{m_j}{m_i+m_j}(\xi - |\xi|\omega)$. Unlike in single species case, here, one only has the momentum conservation, which is reflected as:

$$\xi = \xi_{ij}^+ + \xi_{ij}^- \tag{B.3.2a}$$

$$|\xi|^2 \neq |\xi^+|^2 + |\xi^-|^2 = |\xi|^2\left(1 - \frac{4m_im_js}{(m_i+m_j)^2} + \frac{4m_j^2 s}{(m_i+m_j)^2}\right) \tag{B.3.2b}$$

with $s = \frac{1}{2}\left(1 - \frac{\xi\cdot\omega}{|\xi|}\right)$. By definition, $s \in [0,1]$, obviously, the following is bounded, when the system is given and the mass ratio is fixed:

$$0 \leq \frac{|\xi_{ij}^-|^2 + |\xi_{ij}^+|^2}{|\xi|^2} \leq C_{ij}. \tag{B.3.3}$$

Define $C = \max_{i,j} C_{ij}$, then we have the theorem:

Theorem B.3.1. *In multispecies case, there exists a constant $\tilde{C}$ big enough such that*

$$\mathrm{d}_2(P_{\mathbf{f}}, P_{\mathbf{g}}) < \tilde{C}\mathrm{d}_2(\mathbf{f},\mathbf{g})$$

for Maxwell molecule with cut-off collision.

Here d_2 *for vector functions is defined by:* $\mathrm{d}_2(\mathbf{f},\mathbf{g}) = \max_i \mathrm{d}_2(f_i, g_i)$.

Proof. As in the previous theorem, one only need to prove:

$$\mathrm{d}_2(Q^+(\mathbf{f}), Q^+(\mathbf{g})) \leq \tilde{C}\mathrm{d}_2(\mathbf{f}, \mathbf{g}). \tag{B.3.4}$$

This is given by the following:

$$\begin{aligned}
\left|\frac{\hat{Q}_i^+(\mathbf{f}) - \hat{Q}_i^+(\mathbf{g})}{|\xi|^2}\right| &= \left|\sum_j \int \sigma_{ij} \frac{\hat{f}_i(\xi_{ij}^+)\hat{f}_j(\xi_{ij}^-) - \hat{g}_i(\xi_{ij}^+)\hat{g}_j(\xi_{ij}^-)}{|\xi|^2} \mathrm{d}n\right| \\
&\leq \sum_j \int \sigma_{ij} \left(\hat{f}_i^+ \left|\frac{\hat{f}_j^- - \hat{g}_j^-}{|\xi_{ij}^-|^2}\right| \cdot \left|\frac{|\xi_{ij}^-|^2}{|\xi|^2}\right| + \hat{g}_j^- \left|\frac{\hat{f}_i^+ - \hat{g}_i^+}{|\xi_{ij}^+|^2}\right| \cdot \left|\frac{|\xi_{ij}^+|^2}{|\xi|^2}\right| \right) \mathrm{d}n \\
&\leq \sum_j \int \sigma_{ij} \cdot \sup_{\tilde{\xi},j} \left|\frac{\hat{f}_j - \hat{g}_j}{|\tilde{\xi}|^2}\right| \cdot \frac{|\xi_{ij}^+|^2 + |\xi_{ij}^-|^2}{|\tilde{\xi}|^2} \mathrm{d}n \\
&\leq C \cdot S \cdot \sup_{\tilde{\xi},j} \left|\frac{\hat{f}_j - \hat{g}_j}{|\tilde{\xi}|^2}\right|
\end{aligned} \tag{B.3.5}$$

For convenience, denote $\hat{f}_i^+ = \hat{f}_i(\xi_{ij}^+)$ and $\hat{f}_j^- = \hat{f}_i(\xi_{ij}^-)$. In the derivation above, the following are used:

- collision kernel is cut-off, thus $\sum_j \int \sigma_{ij} \mathrm{d}n = S$
- $\frac{|\xi_{ij}^+|^2 + |\xi_{ij}^-|^2}{|\xi|^2}$ is controlled as in (B.3.3)

(B.3.5) is true for all ξ and i, thus one has:

$$\sup_{\xi,i} \left|\frac{\hat{Q}_i^+(\mathbf{f}) - \hat{Q}_i^+(\mathbf{g})}{|\xi|^2}\right| \leq C \cdot S \cdot \sup_{\xi,i} \left|\frac{\hat{f}_i - \hat{g}_i}{|\xi|^2}\right| \tag{B.3.6}$$

Choose $\tilde{C} = C \cdot S$ to finish the proof. □

Remark B.3.1. *The theorem above is based on the assumption the new metrics is well defined. This is true since the triangular inequality still holds. The proof is straightforward and we will skip the details here.*

B.3.2 In L_2 norm

It is an easy extension from section B.2.2. Here $Q_i = \sum_{ik} Q_{ik}$.

Theorem B.3.2. *Given a system of N species, with cut-off collision kernel, if* $\|\mathbf{f}\|_\infty < \infty$, *then* $\|\mathbf{P}\|_2 = sup_i\|P_i\|_2$ *is controlled by* $\|\mathbf{f}\|_2 = sup_i\|f_i\|_2$.

Proof. To prove Q_i^+ is a bounded operator on f, we have the following:

$$\begin{aligned}
\|Q_i^+(\mathbf{f})\|_2^2 &= \int \left(\sum_k \iint B_{ik}(\Omega) f_i' f_{k^*}' dv_* d\Omega \right)^2 dv \\
&\leq \iiint \sum_k \left(B_{ik}(\Omega) f_i' f_{k^*}' \right)^2 dv_* d\Omega dv \\
&\leq \sum_k \left(\int B_{ik}^2(\Omega) d\Omega \int (f_i')^2 dv' \int (f_{k^*}')^2 dv_*' \right) \\
&\leq C \sum_k \|f_i\|_2^2 \|f_k\|_2^2 \\
&\leq \left(C \sum_k \|f_k\|_\infty \|f_k\|_1 \right) \|f_i\|_2^2
\end{aligned}$$

Thus there is a constant $\tilde{C}$ big enough such that $\forall i$, $\|Q_i^-(\mathbf{f})\|_2 < \tilde{C}\|f_i\|_2$. Given Q_i^- is a bounded linear functional on f_i and $\mu < \infty$, one can easily get

$$\|P_i(\mathbf{f})\|_2 < \tilde{C}\|f_i\|_2,$$

thus

$$\sup_i\|P_i(\mathbf{f})\|_2 < \tilde{C}\sup_i\|f_i\|_2.$$

□

Bibliography

[1] P. Andries, K. Aoki, and B. Perthame, *A consistent BGK-type model for gas mixtures*, Journal of Statistical Physics, 106 (2002), pp. 993–1018.

[2] M. Baer and R. Englman, *A study of the diabatic electronic representation within the Born-Oppenheimer approximation*, Molecular Physics, 75 (1992), pp. 293–303.

[3] G. Bal, A. Fannjiang, G. Papanicolaou, and L. Ryzhik, *Radiative transport in a periodic structure*, Journal of Statistical Physics, 95 (1999), pp. 479–494.

[4] C. Bardos, R. Caflisch, and B. Nicolaenko, *The Milne and Kramers problems for the Boltzmann equation of a hard sphere gas*, Communications on Pure and Applied Mathematics, 39 (1986), pp. 323–352.

[5] C. Bardos, F. Golse, and C. Levermore, *Fluid dynamic limits of the Boltzmann equation I*, Journal of Statistical Physics, 63 (1991), pp. 323–344.

[6] C. Bardos, F. Golse, and C. D. Levermore, *Fluid dynamic limits of kinetic equations II: convergence proofs for the Boltzmann equation*, Communications on Pure and Applied Mathematics, 46 (1993), pp. 667–753.

[7] C. Bardos, F. Golse, and Y. Sone, *Half-space problems for the Boltzmann equation: A survey*, Journal of Statistical Physics, 124 (2006), pp. 275–300.

[8] C. Bardos and S. Ukai, *The classical incompressible Navier-Stokes limit of the Boltzmann equation*, Mathematical Models and Methods in Applied Sciences, 01 (1991), pp. 235–257.

[9] P. Bechouche, *Semi-classical limits in a crystal with a coulombian self-consistent potential: Effective mass theorems*, Asymptotic Analysis, 19 (1999), pp. 95–116.

[10] P. Bechouche, N. J. Mauser, and F. Poupaud, *Semiclassical limit for the Schrödinger-Poisson equation in a crystal*, Communications on Pure and Applied Mathematics, 54 (2001), pp. 851–890.

[11] P. Bechouche and F. Poupaud, *Semi-classical limit of a Schrödinger equationfor a stratified material*, Monatshefte für Mathematik, 129 (2000), pp. 281–301.

[12] M. V. Berry, *Quantal phase factors accompanying adiabatic changes*, Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, 392 (1984), pp. 45–57.

[13] ——, *Geometric amplitude factors in adiabatic quantum transitions*, Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences, 430 (1990), pp. 405–411.

[14] C. Besse, S. Borghol, T. Goudon, I. Lacroix-Violet, and J.-P. Dudon, *Hydrodynamic regimes, Knudsen layer, numerical schemes: Definition of boundary fluxes*, Advances in Applied Mathematics and Mechanics, 3 (2011), pp. 519–561.

[15] P. L. Bhatnagar, E. P. Gross, and M. Krook, *A model for collision processes in gases. I. small amplitude processes in charged and neutral one-component systems*, Physical Review, 94 (1954), pp. 511–525.

[16] G. Billing and S. Adhikari, *The time-dependent discrete variable representation method in molecular dynamics*, Chemical Physics Letters, 321 (2000), pp. 197 – 204.

[17] F. Bloch, *Über die quantenmechanik der elektronen in kristallgittern*, Zeitschrift für Physik A Hadrons and Nuclei, 52 (1929), pp. 555–600.

[18] A. Bobylev, J. Carrillo, and I. Gamba, *On some properties of kinetic and hydrodynamic equations for inelastic interactions*, Journal of Statistical Physics, 98 (2000), pp. 743–773.

[19] A. Bobylev and I. Gamba, *Boltzmann equations for mixtures of Maxwell gases: Exact solutions and power like tails*, Journal of Statistical Physics, 124 (2006), pp. 497–516. 10.1007/s10955-006-9044-8.

[20] A. V. Bobylev, *The Fourier transform method in the theory of the Boltzmann equation for Maxwellian molecules*, Akademiia Nauk SSSR Doklady, 225 (1975), pp. 1041–1044.

[21] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems: Foundations, Mathematical Concepts, and Applications in Molecular and Condensed Matter Physics (Theoretical and Mathematical Physics)*, Texts and monographs in physics, Springer, 2003.

[22] M. Born and R. Oppenheimer, *Zur quantentheorie der molekeln*, Annalen der Physik, 389 (1927), pp. 457–484.

[23] J. Bourgat, P. Le Tallec, B. Perthame, and Y. Qiu, *Coupling Boltzmann and Euler equations without overlapping*, in Domain Decomposition Methods in Science and Engineering, A. Quarteroni, J. Périaux, Y. A. Kuznetsov, and O. Widlund, eds., Providence, RI, 1994, AMS, pp. 377–398.

[24] R. Bourquin, V. Gradinaru, and G. Hagedorn, *Non-adiabatic transitions near avoided crossings: theory and numerics*, Journal of Mathematical Chemistry, 50 (2012), pp. 602–619.

[25] S. I. Braginskii, *Transport processes in a plasma*, Reviews of Plasma Physics, 1 (1965), p. 205.

[26] W. Braun and K. Hepp, *The Vlasov dynamics and its fluctuations in the 1/n limit of interacting classical particles*, Communications in Mathematical Physics, 56 (1977), pp. 101–113.

[27] H. Breuer and F. Petruccione, *The Theory of Open Quantum Systems*, Clarendon Press, 2002.

[28] L. Butler, *Chemical reaction dynamics beyond the Born-Oppenheimer approximation*, Annual Review of Physical Chemistry, 49 (1998), pp. 125–171.

[29] R. Caflisch, S. Jin, and G. Russo, *Uniformly accurate schemes for hyperbolic systems with relaxations*, SIAM Journal of Numerical Analysis, 34 (1997), pp. 246–281.

[30] R. E. Caflisch, *The fluid dynamic limit of the nonlinear Boltzmann equation*, Communications on Pure and Applied Mathematics, 33 (1980), pp. 651–666.

[31] R. Carles, P. Markowich, and C. Sparber, *Semiclassical asymptotics for weakly nonlinear Bloch waves*, Journal of Statistical Physics, 117 (2004), pp. 343–375.

[32] J. A. Carrillo, C. Cercignani, and I. M. Gamba, *Steady states of a Boltzmann equation for driven granular media*, Physical Review E, 62 (2000), pp. 7700–7707.

[33] C. Cercignani, *The Boltzmann equation and its applications*, Applied mathematical sciences, Springer-Verlag, 1988.

[34] C. Cercignani, R. Illner, and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, no. v. 106 in Applied Mathematical Sciences Series, Springer-Verlag GmbH, 1994.

[35] L. Chai, S. Jin, and Q. Li, *Semi-classical limit of the Schrödinger equation with periodic potential and band-crossing*, Kinetic and Related Models, (to appear).

[36] L. Chai, S. Jin, Q. Li, and O. Morandi, *Energy surface hopping problem in molecular dynamics: a multiband Wigner approach*, (preprint).

[37] S. Chapman and T. Cowling, *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*, Cambridge Mathematical Library, Cambridge University Press, 1970.

[38] S. Child, *Molecular collision theory*, Dover Books on Chemistry Series, Dover Publications, Incorporated, 1974.

[39] R. M. Chmieleski and J. H. Ferziger, *Transport properties of a nonequilibrium partially ionized gas*, Physics of Fluids, 10 (1967), pp. 364–371.

[40] L. Cohen, *Generalized phase-space distribution functions*, Journal of Mathematical Physics, 7 (1966), pp. 781–786.

[41] J.-M. Combes, P. Duclos, and R. Seller, *The Born-Oppenheimer approximation*, in Rigorous atomic and molecular physics, G. Velo and A. S. Wightman, eds., Plenum Press, 1981, pp. 185–212.

[42] F. Coron and B. Perthame, *Numerical passage from kinetic to fluid equations*, SIAM Journal of Numerical Analysis, 28 (1991), pp. 26–42.

[43] R. Courant and D. Hilbert, *Methods of Mathematical Physics: Volume 2*, Wiley India Pvt. Ltd., 2008.

[44] A. de Masi, R. Esposito, and J. L. Lebowitz, *Incompressible Navier-Stokes and Euler limits of the Boltzmann equation*, Communications on Pure and Applied Mathematics, 42 (1989), pp. 1189–1214.

[45] P. Degond and S. Jin, *A smooth transition model between kinetic and diffusion equations*, SIAM Journal Numerical Analysis, 42 (2005), pp. 2671–2687.

[46] P. Degond, S. Jin, and L. Mieussens, *A smooth transition model between kinetic and hydrodynamic equations*, Journal Computional Physics, 209 (2005), pp. 665–694.

[47] P. Degond and B. Lucquin-Desreux, *The Fokker-Planck asymptotics of the Boltzmann collision operator in the Culomb case*, Mathematical Models and Methods in Applied Sciences, 02 (1992), pp. 167–182.

[48] P. Degond and B. Lucquin-Desreux, *The asymptotics of collision operators for two species of particles of disparate masses*, Mathematical Models and Methods in Applied Sciences, 6 (1996), pp. 405–436.

[49] ——, *Transport coefficients of plasmas and disparate mass binary gases*, Transport Theory and Statistical Physics, 25 (1996), pp. 595–633.

[50] L. Desvillettes, *On asymptotics of the Boltzmann equation when the collisions become grazing*, Transport Theory and Statistical Physics, 21 (1992), pp. 259–276.

[51] A. S. Dickinson and P. R. Certain, *Calculation of matrix elements for one-dimensional quantum-mechanical problems*, The Journal of Chemical Physics, 49 (1968), pp. 4209–4211.

[52] G. Dimarco and L. Pareschi, *Exponential Runge–Kutta methods for stiff kinetic equations*, SIAM Journal Numerical Analysis, 49 (2011), pp. 2057–2077.

[53] ——, *High order asymptotic-preserving schemes for the Boltzmann equation*, Comptes Rendus Mathematique, 350 (2012), pp. 481–486.

[54] R. J. DiPerna and P.-L. Lions, *On the Cauchy problem for Boltzmann equations: Global existence and weak stability*, Annals of Mathematics, 130 (1989), pp. pp. 321–366.

[55] K. Drukker, *Basics of surface hopping in mixed quantum/classical simulations*, Journal of Computational Physics, 153 (1999), pp. 225–272.

[56] J. Duistermaat, *Fourier integral operators*, Progress in Mathematics Series, Birkhauser Verlag GmbH, 1996.

[57] W. E, J. Lu, and X. Yang, *Asymptotic analysis of quantum dynamics in crystals: The Bloch-Wigner transform, Bloch dynamics and Berry phase*, Acta Mathematicae Applicatae Sinica, English Series, (2011), pp. 1–12.

[58] C. Eckart, *The kinetic energy of polyatomic molecules*, Physical Review, 46 (1934), pp. 383–387.

[59] R. Ellis and M. Pinsky, *The first and second fluid approximations to the linearized Boltzmann equation*, Journal de Mathématiques Pures et Appliquées, 54 (1975), pp. 125–156.

[60] C. Fermanian Kammerer, *Wigner measures and molecular propagation through generic energy level crossings*, Reviews in Mathematical Physics, 15 (2003), pp. 1285–1317.

[61] C. Fermanian Kammerer and P. Gérard, *A Landau-Zener formula for non-degenerated involutive codimension 3 crossings*, Annales Henri Poincaré, 4 (2003), pp. 513–552.

[62] C. Fermanian Kammerer and C. Lasser, *Wigner measures and codimension two crossings*, Journal of Mathematical Physics, 44 (2003), pp. 507–527.

[63] J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases*, North-Holland Pub. Co, 1972.

[64] F. Filbet, J. Hu, and S. Jin, *A numerical scheme for the quantum Boltzmann equation efficient in the fluid regime*, Mathematical Modelling and Numerical Analysis, 46 (2012), pp. 443–463.

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

[66] ——, *An asymptotic preserving scheme for the ES-BGK model of the Boltzmann equation*, SIAM Journal Scientifical Computing, 46 (2011), pp. 204–224.

[67] F. Filbet and C. Mouhot, *Analysis of spectral methods for the homogeneous Boltzmann equation*, Transactions of the American Mathematical Society, 363 (2011), pp. 1947–1980.

[68] F. Filbet and A. Rambaud, *Analysis of an asymptotic preserving scheme for relaxation systems*, ESAIM: Mathematical Modelling and Numerical Analysis, 47 (2013), pp. 609–633.

[69] G. Floquet, *Sur les équations différentielles linéaires à coefficients périodiques*, Annales scientifiques de l'École Normale Supérieure, 12 (1883), pp. 47–88.

[70] E. Gabetta, L. Pareschi, and G. Toscani, *Relaxation schemes for nonlinear kinetic equations*, SIAM Journal of Numerical Analysis, 34 (1997), pp. 2168–2194.

[71] I. Gamba, V. Panferov, and C. Villani, *On the Boltzmann equation for diffusively excited granular media*, Communications in Mathematical Physics, 246 (2004), pp. 503–541.

[72] V. Garzò, A. Santos, and J. J. Brey, *A kinetic model for a multicomponent gas*, Physics of Fluids, 1 (1989), pp. 380–384.

[73] P. Gérard, *Mesures semi-classiques et ondes de Bloch*, Séminaire Équations aux dérivées partielles, (1991), pp. 1–19.

[74] P. Gérard, P. Markowich, N. Mauser, and F. Poupaud, *Homogenization limits and Wigner transforms*, Communications on Pure and Applied Mathematics, 50 (1997), pp. 323–379.

[75] E. Goldman and L. Sirovich, *Equations for gas mixtures*, Physics of Fluids, 10 (1967), pp. 1928–1940.

[76] F. Golse, S. Jin, and C. D. Levermore, *A domain decomposition analysis for a two-scale linear transport problem*, Mathematical Modeling and Numerical Analysis, 37 (2002), pp. 869–892.

[77] F. Golse and L. Saint-Raymond, *The Navier-Stokes limit of the Boltzmann equation for bounded collision kernels*, Inventiones mathematicae, 155 (2004), pp. 81–161.

[78] ——, *The Navier-Stokes limit of the Boltzmann equationfor hard potentials*, Journal de Mathématiques Pures et Appliquées, 91 (2009), pp. 508–552.

[79] L. Gosse, *Multiphase semiclassical approximation of an electron in a one-dimensional crystalline lattice II. impurities, confinement and Bloch oscillations*, Journal of Computational Physics, 201 (2004), pp. 344 – 375.

[80] ——, *A case study on the reliability of multiphase WKB approximation for the one-dimensional Schrödinger equation*, in Numerical Methods for Hyperbolic and Kinetic Problems, S. Cordier, T. Goudon, M. Gutnic, and E. Sonnendrucker, eds., European Mathematical Society Publishing House, 2005.

[81] L. Gosse and P. A. Markowich, *Multiphase semiclassical approximation of an electron in a one-dimensional crystalline lattice: I. homogeneous problems*, Journal of Computational Physics, 197 (2004), pp. 387 – 417.

[82] L. Gosse and N. Mauser, *Multiphase semiclassical approximation of an electron in a one-dimensional crystalline lattice-III. from ab initio models to wkb for Schrödinger-Poisson*, Journal of Computational Physics, 211 (2006), pp. 326–346.

[83] S. Gottlieb and C. wang Shu, *Total variation diminishing Runge-Kutta schemes*, Mathematics of Computation, 67 (1998), pp. 73–85.

[84] S. Gottlieb, C. wang Shu, and E. Tadmor, *Strong stability-preserving high-order time discretization methods*, SIAM Review, 43 (2001), pp. 89–112.

[85] H. Grad, *Principles of the Kinetic Theory of Gases*, Handbuch der Physik, 12 (1958), pp. 205–294.

[86] H. Grad, *Theory of rarefied gases*, in Rarefied Gas Dynamics, F. Devienne, ed., Pergamon, London, England, 1960, pp. 100–138.

[87] H. Grad, *Asymptotic Theory of the Boltzmann Equation, II*, in Rarefied Gas Dynamics, Volume 1, J. A. Laurmann, ed., 1963, p. 26.

[88] E. P. Gross and M. Krook, *Model for collision processes in gases: Small-amplitude oscillations of charged two-component systems*, Physical Review, 102 (1956), pp. 593–604.

[89] J. Guillot, J. Ralston, and E. Trubowitz, *Semi-classical asymptotics in solid state physics*, Communications in Mathematical Physics, 116 (1988), pp. 401–415.

[90] G. Hagedorn, *A time dependent Born-Oppenheimer approximation*, Communications in Mathematical Physics, 77 (1980), pp. 1–19.

[91] ——, *Proof of the Landau-Zener formula in an adiabatic limit with small eigenvalue gaps*, Communications in Mathematical Physics, 136 (1991), pp. 433–449.

[92] G. Hagedorn, *Classification and normal forms for avoided crossings of quantum-mechanical energy levels*, Journal of Physics A: Mathematical and General, 31 (1998), p. 369.

[93] G. A. Hagedorn, *High order corrections to the time-dependent Born-Oppenheimer approximation I: Smooth potentials*, Annals of Mathematics, 124 (1986), pp. pp. 571–590.

[94] G. A. Hagedorn, *High order corrections to the time-dependent Born-Oppenheimer approximation. II: Coulomb systems*, Communications in Mathematical Physics, 117 (1988), pp. 387–403.

[95] G. A. Hagedorn and A. Joye, *Landau-Zener transitions through small electronic eigenvalue gaps in the Born-Oppenheimer approximation*, Annales de l'institut Henri Poincaré (A) Physique théorique, 68 (1998), pp. 85–134.

[96] ——, *Molecular propagation through small avoided crossings of electron energy levels*, Reviews in Mathematical Physics, 11 (1999), pp. 41–101.

[97] ——, *A time-dependent BornOppenheimer approximation with exponentially small error estimates*, Communications in Mathematical Physics, 223 (2001), pp. 583–626.

[98] E. Hairer, S. Nørsett, and G. Wanner, *Solving Ordinary Differential Equations I: Nonstiff Problems (Springer Series in Computational Mathematics)*, Springer, 2010.

[99] D. O. Harris, G. G. Engerholm, and W. D. Gwinn, *Calculation of matrix elements for one-dimensional quantum-mechanical problems and the application to anharmonic oscillators*, Journal of Chemical Physics, 43 (1965), pp. 1515–1517.

[100] H. Haug and A. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*, Springer Series in Solid-state Sciences, v. 123, Springer London, Limited, 2008.

[101] D. Hilbert, *Begründung der kinetischen gastheorie*, Mathematische Annalen, 72 (1912), pp. 562–577.

[102] N. Hill, *Gaussian beam migration*, Geophysics, 55 (1990), pp. 1416–1428.

[103] L. Hörmander, *The Weyl calculus of pseudo-differential operators*, Communications on Pure and Applied Mathematics, 32 (1979), pp. 359–443.

[104] J. Hu, S. Jin, and B. Yan, *A numerical scheme for the quantum Fokker-Planck-Landau equation efficient in the fluid regime*, Communications in Computational Physics, (to appear).

[105] Z. Huang, S. Jin, P. Markowich, and C. Sparber, *A Bloch decomposition-based split-step pseudospectral method for quantum dynamics with periodic potentials*, SIAM Journal on Scientific Computing, 29 (2008), pp. 515–538.

[106] Z. Huang, S. Jin, P. Markowich, and C. Sparber, *Numerical simulation of the nonlinear Schrödinger equation with multidimensional periodic potentials*, Multiscale Modeling & Simulation, 7 (2008), pp. 539–564.

[107] Z. Huang, S. Jin, H. Wu, and D. Yin, *Gaussian beam methods for the Dirac equation in the semi-classical regime*, Communications in Mathematical Sciences, 10 (2012), pp. 1301–1315.

[108] V. Jakšić and J. Segert, *Exponential approach to the adiabatic limit and the Landau-Zener formula*, Reviews in Mathematical Physics, 04 (1992), pp. 529–574.

[109] S. Jin, *Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms*, Journal of Computational Physics, 122 (1995), pp. 51–67.

[110] ——, *Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations*, SIAM Journal of Scientific Computing, 21 (1999), pp. 441–454.

[111] ——, *Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review*, in Lecture Notes for Summer School on "Methods and Models of Kinetic Theory" (M&MKT), R. di Matematica della Universita di Parma, ed., Porto Ercole (Grosseto, Italy), June 2010.

[112] S. Jin and Q. Li, *A BGK-penalization-based asymptotic-preserving scheme for the multi-species Boltzmann equation*, Numerical Methods for Partial Differential Equations, 29 (2013), pp. 1056–1080.

[113] S. Jin and X. Li, *Multi-phase computations of the semiclassical limit of the Schrödinger equation and related problems: Whitham vs Wigner*, Physica D: Nonlinear Phenomena, 182 (2003), pp. 46 – 85.

[114] S. Jin, H. Liu, S. Osher, and Y. Tsai, *Computing multivalued physical observables for the semiclassical limit of the Schrödinger equation*, Journal of Computational Physics, 205 (2005), pp. 222–241.

[115] S. Jin, P. Markowich, and C. Sparber, *Mathematical and computational methods for semiclassical Schrödinger equations*, Acta Numerica, 20 (2011), pp. 121–209.

[116] S. Jin and S. Osher, *A level set method for the computation of multivalued solutions to quasi-linear hyperbolic PDEs and Hamilton-Jacobi equations*, Communications in Mathematical Sciences, 1 (2003), pp. 575–591.

[117] S. Jin and P. Qi, *A hybrid Schrödinger/Gaussian beam solver for quantum barriers and surface hopping*, Kinetic and Related Models, 4 (2011), pp. 1097–1120.

[118] S. Jin, P. Qi, and Z. Zhang, *An Eulerian surface hopping method for the Schrödinger equation with conical crossings*, SIAM Multiscale Modeling and Simulation, 9 (2011), pp. 258–281.

[119] S. Jin and Y. Shi, *A micro-macro decomposition-based asymptotic-preserving scheme for the multispecies Boltzmann equation*, SIAM Journal of Scientific Computing, 31 (2010), pp. 4580–4606.

[120] S. Jin, H. Wu, and X. Yang, *Gaussian beam methods for the Schrödinger equation in the semi-classical regime: Lagrangian and Eulerian formulations*, Communications in Mathematical Sciences, 6 (2008), pp. 995–1020.

[121] ——, *A numerical study of the Gaussian beam methods for Schrödinger-Poisson equations*, Journal of Computational Mathematics, 28 (2010), p. 261.

[122] ——, *Semi-Eulerian and high order Gaussian beam methods for the Schrodinger equation in the semiclassical regime*, Communications in Computational Physics, 9 (2011), pp. 668–687.

[123] S. Jin and B. Yan, *A class of asymptotic-preserving schemes for the Fokker-Planck-Landau equation*, Journal of Computational Physics, 230 (2011), pp. 6420–6437.

[124] S. Jin and X. Yang, *Computation of the semiclassical limit of the Schrödinger equation with phase shift by a level set method*, Journal of Scientific Computing, 35 (2008), pp. 144–169.

[125] S. Jin and D. Yin, *Computational high frequency waves through curved interfaces via the Liouville equation and geometric theory of diffraction*, Journal of Computational Physics, 227 (2008), pp. 6106–6139.

[126] E. A. Johnson, *Energy and momentum equations for disparate-mass binary gases*, Physics of Fluids, 16 (1973), pp. 45–49.

[127] A. Joye, *Proof of the Landau-Zener formula*, Asymptotic Analysis, 9 (1994), pp. 209–258.

[128] A. Klar, H. Neunzert, and J. Struckmeier, *Transition from kinetic theory to macroscopic fluid equations: A problem for domain decomposition and a source for new algorithms*, Transport Theory and Statistical Physics, 29 (2000), pp. 93–106.

[129] I. Kôdi Husim, *Some Formal Properties of the Density Matrix*, Nippon Sugaku-Buturigakkwai Kizi Dai 3 Ki, 22 (1940), pp. 264–314.

[130] S. Kube, C. Lasser, and M. Weber, *Monte Carlo sampling of Wigner functions and surface hopping quantum dynamics*, Journal of Computational Physics, 228 (2009), pp. 1947 – 1962.

[131] L. Landau, *Zur Theorie der Energieubertragung. II.*, Physics of the Soviet Union2, (1932), pp. 46–51.

[132] O. E. Lanford, *Time evolution of large classical systems*, in Dynamical Systems, Theory and Applications, J. Moser, ed., vol. 38 of Lecture Notes in Physics, Springer Berlin Heidelberg, 1975, pp. 1–111.

[133] ——, *The hard sphere gas in the Boltzmann-grad limit*, Physica A: Statistical Mechanics and its Applications, 106 (1981), pp. 70 – 76.

[134] E. W. Larsen, J. Morel, and W. F. Miller, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes*, Journal of Computational Physics, 69 (1987), pp. 283 – 324.

[135] E. W. Larsen and J. E. Morel, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes II*, Journal of Computational Physics, 83 (1989), pp. 212–236.

[136] C. Lasser, T. Swart, and S. Teufel, *Construction and validation of a rigorous surface hopping algorithm for conical crossings*, Communications in Mathematical Sciences, 5 (2007), pp. 789–814.

[137] C. Lasser and S. Teufel, *Propagation through conical crossings: an asymptotic semigroup*, Communications on Pure and Applied Mathematics, 58 (2005), pp. 1188–1230.

[138] P. Le Tallec and F. Mallinger, *Coupling Boltzmann and Navier-Stokes equations by half fluxes*, Journal of Computational Physics, 136 (1997), pp. 51–67.

[139] M. Lemou and L. Mieussens, *A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit*, SIAM Journal of Scientific Computing, 31 (2008), pp. 334–368.

[140] J. Leray, *étude de diverses équations intégrales non linéaires et quelques problèmes que pose l'hydrodynamique*, Journal de Mathématiques Pures et Appliquées, 9 (1933), pp. 1–82.

[141] R. J. LeVeque, *Numerical Methods for Conservation Laws*, Lectures in mathematics ETH Zürich, Birkhäuser Verlag, 1992.

[142] R. J. LeVeque and etc., Version 4.3. "http://www.amath.washington.edu/ claw/".

[143] Q. Li and L. Pareschi, *Exponential Runge-Kutta schemes for inhomogeneous Boltzmann equations with high order of accuracy*, arXiv:1208.2622, (preprint).

[144] Q. Li and X. Yang, *Exponential Runge-Kutta methods for the multispecies Boltzmann equation*, (preprint).

[145] J. C. Light and T. Carrington, *Discrete-variable representations and their utilization*, in Advances in Chemical Physics, John Wiley & Sons, Inc., 2007, pp. 263–310.

[146] P.-L. Lions, *Mathematical Topics in Fluid Mechanics, Vol. 1: Incompressible Models*, The Clarendon Press, Oxford University Press, New York, 1996.

[147] P.-L. Lions and T. Paul, *Sur les measures de Wigner*, Revista Matemática Iberoamericana, 9 (1993), pp. 553–618.

[148] F. London, *Über den mechanismus der homoöpolaren bindung*, in Probleme der modernen physik, P. Debye, ed., S. Herzel, 1928.

[149] J. Lu and X. Yang, *Frozen Gaussian approximation for high frequency wave propagation*, Communications in Mathematical Sciences, 9 (2011), pp. 663–683.

[150] J. Lu and X. Yang, *Convergence of frozen Gaussian approximation for high-frequency wave propagation*, Communications on Pure and Applied Mathematics, 65 (2012), pp. 759–789.

[151] P. Markowich, N. Mauser, and F. Poupaud, *A Wigner-function approach to (semi) classical limits: Electrons in a periodic potential*, Journal of Mathematical Physics, 35 (1994), pp. 1066–1094.

[152] C. A. Mead and D. G. Truhlar, *Conditions for the definition of a strictly diabatic electronic basis for molecular systems*, The Journal of Chemical Physics, 77 (1982), pp. 6090–6098.

[153] O. Morandi, *Multiband Wigner-function formalism applied to the Zener band transition in a semiconductor*, Physical Review B, 80 (2009), p. 024301.

[154] O. Morandi and F. Schürrer, *Wigner model for quantum transport in graphene*, Arxiv preprint arXiv:1102.2416, (2011).

[155] H. More, R. Oppenheim, and J. Ross, *The Wigner function and transport theory*,

in Studies in Statistical Mechanics, V. D. Boer and O. E. Uhlenbeck, eds., vol. 1, North-Holland Publishing Company, 1962.

[156] C. Mouhot and L. Pareschi, *Fast algorithms for computing the Boltzmann collision operator*, Mathematics of Computation, (2004).

[157] H. Nakamura, *Nonadiabatic Transition: Concepts, Basic Theories and Applications*, World Scientific Publ., 2002.

[158] H. Nakamura, *Nonadiabatic chemical dynamics: Comprehension and control of dynamics, and manifestation of molecular functions*, in Advances in Chemical Physics, John Wiley & Sons, Inc., 2008, pp. 95–212.

[159] T. Nishida, *Fluid dynamical limit of the nonlinear Boltzmann equation to the level of the compressible Euler equation*, Communications in Mathematical Physics, 61 (1978), pp. 119–148.

[160] F. Odeh and J. Keller, *Partial differential equations with periodic coefficients and Bloch waves in crystals*, Journal of Mathematical Physics, 5 (1964), p. 1499.

[161] S. Osher and J. A. Sethian, *Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations*, Journal of Computational Physics, 79 (1988), pp. 12–49.

[162] G. Panati, H. Spohn, and S. Teufel, *Effective dynamics for Bloch electrons: Peierls substitution and beyond*, Communications in Mathematical Physics, 242 (2003), pp. 547–578.

[163] ——, *Motion of electrons in adiabatically perturbed periodic structures*, Analysis, Modeling and Simulation of Multiscale Problems, (2006), pp. 595–617.

[164] G. Papanicolau, A. Bensoussan, and J. Lions, *Asymptotic Analysis for Periodic Structures*, Studies in Mathematics and its Applications, Elsevier Science, 1978.

[165] L. Pareschi and G. Russo, *Numerical solution of the Boltzmann equation I: Spectrally accurate approximation of the collision operator*, SIAM Journal of Numerical Analysis, 37 (2000), pp. 1217–1245.

[166] B. Perthame, *Boltzmann type schemes for gas dynamics and the entropy property*, SIAM Journal on Numerical Analysis, 27 (1990), pp. pp. 1405–1421.

[167] S. Pieraccini and G. Puppo, *Implicit-Explicit schemes for BGK kinetic equations*, Journal of Scientific Computing, 32 (2007), pp. 1–28.

[168] M. Popov, *A new method of computation of wave fields using Gaussian beams*, Wave Motion, 4 (1982), pp. 85 – 97.

[169] F. Poupaud and C. Ringhofer, *Semi-classical limits in a crystal with exterior potentials and effective mass theorems.*, Communications in Partial Differential Equations, 21 (1996), pp. 1897–1918.

[170] A. Pulvirenti and G. Toscani, *The theory of the nonlinear Boltzmann equation for Maxwell molecules in Fourier representation*, Annali di Matematica Pura ed Applicata, 171 (1996), pp. 181–204.

[171] J. Ralston, *Gaussian beams and the propagation of singularities*, in Studies in Partial Differential Equations, vol. 23, Washington, DC, 1982, Mathematics Association of America, pp. 206–248.

[172] V. Rousse, *Landau-Zener transitions for eigenvalue avoided crossings in the adiabatic and Born-Oppenheimer approsimations*, Asymptotic Analysis, 37 (2004), pp. 293–328.

[173] L. Saint-Raymond, *Hydrodynamic Limits of the Boltzmann Equation*, no. no. 1971 in Lecture Notes in Mathematics, Springer, 2009.

[174] F. Schwabl, *Statistical Mechanics*, Advanced Texts in Physics, Springer-Verlag GmbH, 2002.

[175] M. Sillanpää, T. Lehtinen, A. Paila, Y. Makhlin, and P. Hakonen, *Landau–Zener interferometry in a Cooper-pair box*, Journal of Low Temperature Physics, 146 (2007), pp. 253–262.

[176] R. J. Spiteri and S. J. Ruuth, *A new class of optimal high-order strong-stability-preserving time discretization methods*, SIAM Journal on Numerical Analysis, 40 (2003), pp. pp. 469–491.

[177] L. Spitzer and R. Härm, *Transport phenomena in a completely ionized gas*, Physical Review, 89 (1953), pp. 977–981.

[178] H. Spohn and S. Teufel, *Adiabatic decoupling and time-dependent Born-Oppenheimer theory*, Communications in Mathematical Physics, 224 (2001), pp. 113–132.

[179] P. Stanley and L. Kricka, *Bioluminescence & Chemiluminescence:*, World Scientific, 2002.

[180] G. Sundaram and Q. Niu, *Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and Berry-phase effects*, Physical Review B, (1999), pp. 14915–14925.

[181] N. M. Tanushev, *Superpositions and higher order Gaussian beams*, Communications in Mathematical Sciences, 6 (2008), p. 449475.

[182] N. M. Tanushev, B. Engquist, and R. Tsai, *Gaussian beam decomposition of high frequency wave fields*, Journal of Computational Physics, 228 (2009), pp. 8856–8871.

[183] G. Toscani and C. Villani, *Probability metrics and uniqueness of the solution to the Boltzmann equation for a Maxwell gas*, Journal of Statistical Physics, 94 (1999), pp. 619–637.

[184] J. Tully, *Molecular dynamics with electronic transitions*, The Journal of Chemical Physics, 93 (1990), pp. 1061–1071.

[185] J. Tully and R. Preston, *Trajectory surface hopping approach to nonadiabatic molecular collisions: the reaction of H^+ with D_2*, The Journal of Chemical Physics, 55 (1971), pp. 562–572.

[186] C. Villani, *Chapter 2 a review of mathematical topics in collisional kinetic theory*, in Handbook of Mathematical Fluid Dynamics, S. Friedlander and D. Serre, eds., vol. 1, North-Holland, 2002, pp. 71 – 74.

[187] C. Villani, *Mathematics of granular materials*, Journal of Statistical Physics, 124 (2006), pp. 781–822.

[188] J. von Neumann and E. Wigner, *Über das verhalten von eigenwerten bei adiabatischen prozessen*, Zeitschrift für Physik A Hadrons and Nuclei, 30 (1929), pp. 467–470.

[189] A. I. Voronin, J. M. C. Marques, and A. J. C. Varandas, *Trajectory surface hopping study of the Li + Li2(X1Σ g+) dissociation reaction*, The Journal of Physical Chemistry A, 102 (1998), pp. 6057–6062.

[190] C. wang Shu, *Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws*, tech. rep., Institute for Computer Applications in Science and Engineering (ICASE), 1997.

[191] C. wang Shu and S. Osher, *Efficient implementation of essentially non-oscillatory shock-capturing schemes, II*, Journal of Computational Physics, 83 (1989), pp. 32–78.

[192] H. Weyl, *Quantenmechanik und gruppentheorie*, Zeitschrift für Physik A Hadrons and Nuclei, 46 (1927), pp. 1–46.

[193] E. Wigner, *On the quantum correction for thermodynamic equilibrium*, Physical Review, 40 (1932), pp. 749–759.

[194] C. Wilcox, *Theory of Bloch waves*, Journal d'Analyse Mathématique, 33 (1978), pp. 146–167.

[195] T. Wilson and J. W. Hastings, *Bioluminescence*, Annual Review of Cell and Developmental Biology, 14 (1998), pp. 197–230. PMID: 9891783.

[196] G. Worth and L. Cederbaum, *Beyond Born-Oppenheimer: molecular dynamics through a conical intersection*, Annual Review of Physical Chemistry, 55 (2004), pp. 127–158.

[197] D. Xiao, M. Chang, and Q. Niu, *Berry phase effects on electronic properties*, Reviews of Modern Physics, 82 (2010), p. 1959.

[198] D. Xiao, J. Shi, and Q. Niu, *Berry phase correction to electron density of states in solids*, Physical Review Letter, 95 (2005), p. 137204.

[199] B. Yan and S. Jin, *A successive penalty-based asymptotic-preserving scheme for kinetic equations*, SIAM Journal on Scientific Computing, 35 (2013), pp. A150–A172.

[200] D. Yin, M. Tang, and S. Jin, *The Gaussian beam method for the Wigner equation with discontinuous potentials*, Inverse Problems and Imaging, (to appear).

[201] C. Zachos, D. Fairlie, and T. Curtright, *Quantum Mechanics in Phase Space: An Overview With Selected Papers*, World Scientific Series In 20th Century Physics, World Scientific, 2005.

[202] C. Zener, *Non-adiabatic crossing of energy levels*, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character, 137 (1932), pp. 696–702.

[203] A. Zewail, *Femtochemistry: Ultrafast Dynamics of the Chemical Bond, Volume 1*, World Scientific Series in 20th Century Chemistry, World Scientific, 1994.