On structure preserving numerical schemes for hyperbolic partial differential equations and multiscale kinetic equations

A Thesis
Submitted For the Degree of

Doctor of Philosophy

In

Faculty of Engineering

by

Megala Anandan

Department of Aerospace Engineering
Indian Institute of Science
Bangalore – 560 012 (INDIA)

June, 2024

To the determination of my parents,
Manimegalai and Anandan...

Acknowledgements

I would like to express my deepest gratitude to my supervisor, Prof. S. V. Raghurama Rao, for his unwavering support and guidance throughout this journey. His mentorship has been instrumental in shaping both my research endeavors and personal growth, providing me with the freedom to explore and fostering my development as a researcher. His encouragement and support for research collaborations have been invaluable in enhancing my academic experience.

I am grateful to research collaborators, Dr. Nicolas Crouseilles from Inria Rennes, Dr. Benjamin Boutin from the Univeristy of Rennes, France and Prof. Dr. Mária Lukácová-Medvidová from the Johannes Gutenberg University of Mainz, Germany for their insightful feedback, constructive criticism, and encouragement. Their collective expertise enriched my work and broadened my perspectives. The experiences gained during my research visits provided valuable life lessons.

I extend my gratitude to the chairman of our department Prof. Joseph Mathew and various members of my fellowship review committee for their encouragement and valuable suggestions. I also extend my gratitude to the faculty of the institute, specially Prof. Vittal Rao for offering intriguing courses. I am also thankful to my colleagues and fellow researchers Shubham, Kunal, Balwinder, Aravind, Kedar, Prashant, and Shashi for their support. I also extend my thanks to all the administrative staff involved both directly and indirectly.

I acknowledge the Ministry of Education, Government of India for awarding me the Prime Minister's Research Fellowship. The financial assistance has enabled me to pursue my academic aspirations and contribute to the advancement of knowledge in the field.

Lastly, I would like to express my appreciation to my parents for their love, understanding, and determination. Gratitude is also extended to my friends, with special mention to Vasanth, for the cherished companionship throughout this journey.

Abstract

Natural phenomena are frequently represented through the formulation of differential equations, coupled with specific initial and boundary conditions. Many such models possess inherent structures that are crucial in describing the behaviour of solutions. Unfortunately, numerical discretisations of such models often fail to preserve these structures, leading to inappropriate numerical solutions. The numerical schemes that take special care to preserve the inherent structures of a given differential equation in its discretisation process are known as *structure preserving schemes*. Various structures have been extensively discussed in the existing literature. This thesis focuses on crucial structure preserving strategies such as *entropy stability, asymptotic preservation* and *well-balancing*.

Many hyperbolic systems of partial differential equations (PDEs) have entropy inequalities associated with them. Numerical schemes that are designed to inherently satisfy the entropy inequality are known as *entropy stable schemes*. On the other hand, the hyperbolic system of PDEs can be derived as an approximation of a vector-kinetic model, which also encompasses associated entropy structures. The entropy inequality of the hyperbolic system of PDE can be derived as a moment of the entropy structures of vector-kinetic model. However, this correspondence is not maintained in numerical discretisations. Presented as the first part of this thesis is the development and analysis of a numerical scheme that achieves entropy stability for the vector-kinetic model, along with the proof that it also recovers entropy stability for the given hyperbolic system of PDEs.

Hyperbolic and kinetic equations containing small spatial and temporal scales due to stiff source terms or strong forcing, pose significant challenges for numerical approximation. *Asymptotic preserving (AP) schemes* offer an effective solution for handling these asymptotic regimes, allowing for efficient computations without the need for excessively small mesh sizes and time steps. Unlike traditional domain decomposition methods that involve coupling different models (in different regimes) through interface conditions, AP schemes seamlessly transition between different scales by ensuring automatic adaptation of solvers based on the resolution of scales. Presented as the second part of this thesis is the development and analysis of a high order AP

scheme for diffusive-scaled linear kinetic equations with general initial conditions.
The dimensionless form of barotropic Euler system contains the parameter Mach number which can become small, and this results in the need for an AP scheme. Moreover, this system has an entropy inequality corresponding to a convex entropy function, for all values of the parameter. Hence, this system requires treatment with regard to both the structures: *asymptotic preservation* and *entropy stability*. Presented as a third part of this thesis is the development and analysis of an AP scheme satisfying entropy stability for all values of the parameter in the barotropic Euler system.
In the fourth part of this thesis, the mathematical properties of Lattice Boltzmann Methods (LBMs) derived from vector-kinetic models of hyperbolic PDEs are presented. This LBM framework is extended to hyperbolic PDEs with stiff source terms, where suitable modeling at the vector-kinetic level combined with *well-balancing* is introduced to avoid spurious numerical convection arising from the discretisation of source terms and thereby avoiding wave propagation at incorrect speeds.

List of publications

Published/Accepted for publication in journal

1. Megala Anandan, S. V. Raghurama Rao. Entropy conserving/stable schemes for a vector-kinetic model of hyperbolic systems. *Applied Mathematics and Computation*. Vol. 465, 128410. 15 March 2024.
   DOI: https://doi.org/10.1016/j.amc.2023.128410.
   arXiv url: https://arxiv.org/abs/2302.08014. February 2023.

2. Megala Anandan, Benjamin Boutin, Nicolas Crouseilles. High order asymptotic preserving scheme for diffusive scaled linear kinetic equations with general initial conditions. Accepted for publication in *ESAIM: Mathematical Modelling and Numerical Analysis.*
   arXiv url: https://arxiv.org/abs/2305.13393. May 2023.

Submitted to journal

1. Megala Anandan, S. V. Raghurama Rao. On Lattice Boltzmann Methods based on vector-kinetic models for hyperbolic partial differential equations.
   arXiv url: https://arxiv.org/abs/2401.03952. Jan 2024.
   (Revised manuscript has been submitted towards minor revision suggested by *Computers and Fluids*)

Soon to be submitted to journal

1. Megala Anandan, Mária Lukácová-Medvidová, S. V. Raghurama Rao. An asymptotic preserving scheme satisfying entropy stability for the barotropic Euler system.

Conference proceedings/presentations

1. Megala A, S. V. Raghurama Rao. D2Q9 model of upwind lattice Boltzmann scheme for hyperbolic scalar conservation laws. *8^{th} European Congress on Computational Methods*

in Applied Sciences and Engineering, Scipedia. 05-09 June 2022. Oslo, Norway.
DOI: https://doi.org/10.23967/eccomas.2022.074.

2. Megala A, S. V. Raghurama Rao. A discrete-kinetic entropy conserving and exact discontinuity capturing scheme for scalar conservation laws. Oral presentation at *XVIII International Conference on Hyperbolic Problems: Theory, Numerics, Applications*. Málaga, Spain.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Partial differential equations (PDEs) are used to describe a wide range of physical phenomena like heat conduction, fluid dynamics, and electromagnetism, and are fundamental in various fields such as physics, engineering, and applied mathematics. The general form of PDE consists of an unknown function of several variables and its partial derivatives with respect to those variables. The solution to a PDE is a function that satisfies the PDE along with any given initial or boundary conditions.

PDEs can be classified into several types based on their order, linearity and the number of independent variables. In particular, they are classified into three types based on the nature of equations and behavior of solutions as elliptic, parabolic and hyperbolic PDEs. The reader is referred to the books by Evans [106], Strauss [288], Courant and Hilbert [71], to understand PDEs. In this thesis, we focus on hyperbolic PDE systems and multiscale kinetic equations.

1.1 Hyperbolic systems of PDEs

Hyperbolic PDEs describe wave-like phenomena and the propagation of disturbances through space and time. They are associated with characteristic curves or surfaces that dictate the directions in which disturbances propagate. A system of PDEs is hyperbolic if its flux Jacobian matrix has real eigenvalues and linearly independent eigenvectors. Hyperbolic PDE systems arise in gas dynamics (as Euler's system of gas dynamics), oceanography (as shallow water equations or St. Venant's system), plasma physics (as ideal magnetohydrodynamics (MHD) equations), and so on. Inviscid Burgers' equation is a toy model that eases the understanding of hyperbolic PDE systems.

As a consequence of non-linearity in convection terms, hyperbolic PDE systems allow for the formation of shock waves, expansion waves and contact discontinuities even from smooth initial

data. As derivatives do not exist for the discontinuous solutions, the governing PDE system becomes invalid when discontinuities develop. In this case, weak or integral form that allows piece-wise smooth solutions (known as *weak solutions*) is formulated. Weak solutions that satisfy the weak form are non-unique. Hence, a suitable criterion that allows one to choose physically relevant solutions from the class of weak solutions becomes a necessity. In the context of continuum physics, such admissibility criteria naturally arise from the second law of thermodynamics. Such criteria are incorporated into analytical theory by mandating that the admissible solutions satisfy an entropy inequality corresponding to some convex entropy function.

Hyperbolic PDEs frequently arise in the study of phenomena with disparate scales. For instance, the multiscale nature of barotropic Euler system is revealed by suitable non-dimensionalisation that brings out the square of Mach number in the denominator of a term. In the limit of Mach number approaching zero, this system reduces to incompressible system of equations.

Solving hyperbolic PDEs analytically is often challenging, due to the complex nature of waves. Since hyperbolic PDE systems are systems of balance or conservation laws, numerical methods that allow for discrete conservation (like finite volume method) are usually preferred. The reader is referred to the books by Dafermos [77], Godlewski and Raviart [131], Leray [201], Lax [195], Benzoni-Gavage and Serre [28], to understand hyperbolic PDE systems.

1.1.1 Numerical discretisation of hyperbolic PDE systems

The initial numerical treatment of hyperbolic PDEs began with **finite difference methods**. Researchers like Courant, Friedrichs, and Lewy contributed significantly to the development of finite difference methods in the 1920s and 1930s. Their landmark paper [70] that provides inequality conditions on mesh and time step sizes for convergence, laid the foundation of numerical methods for PDEs. In 1950, von Neumann and Richtmyer introduced artificial viscosity to stabilise the numerical solutions [318]. In another classical paper, Courant, Isaacson and Rees (CIR) [72] introduced upwind schemes, which became immensely popular. In 1960, numerical stability and convergence analyses were discussed by Lax and Wendroff in their seminal paper [193]. These contributions played an important role in the development of numerical methods for PDEs.

Finite volume methods (FVMs) have become popular due to their natural incorporation of conservation principles. These are grouped into two major categories as, *central* and *upwind* discretisation methods. Central schemes are designed to introduce stability while simulating hyperbolic problems, without introducing any specific mechanisms to recognize or deal with hyperbolicity. Some of the popular central schemes were introduced by Lax and Friedrichs

[119, 194], Rusanov [271], Lax and Wendroff [193], MacCormack [210], Richtmyer and Morton [264], Jameson, Schmidt and Turkel [163], Nessyahu and Tadmor [227], Kurganov and Tadmor [187], Liu and Tadmor [209], and Jaisankar and Raghurama Rao [159]. On the other hand, upwind schemes explicitly deal with hyperbolicity by incorporating the information based on wave propagation directions and the characteristic properties of hyperbolic PDE systems. These methods can be categorized into four major groups as: Riemann solvers, flux vector splitting methods, kinetic or Boltzmann schemes, and relaxation schemes.

- *Riemann solvers* are designed such that the Riemann problem formed by the discontinuous representation of solution variable at cell interface is solved either exactly or approximately. These include the developments of Godunov [132], Roe [266], Osher and Solomon [229], Harten, Lax and van Leer (HLL) [148], and Toro, Spruce and Spears [306]. Approximate Riemann solvers are often written in the form of *Flux Difference Splitting (FDS)*, with appropriate split flux vectors.

- *Flux vector splitting (FVS) methods* involve splitting of the flux vector into positive and negative parts based on the eigenvalues of flux Jacobian matrix, to facilitate wave speed splitting for upwinding. Some of the FVS methods specific to Euler's equations of gas dynamics were introduced by Steger and Warming [286] and van Leer [313].
  Some special schemes by Liou and Steffen [208], Jameson [162], Zha and Bilgen [333], and Toro and Vázquez [305] utilize *convection-pressure splitting* of the flux vectors, based on utilizing FVS and/or FDS formats.

- *Kinetic or Boltzmann schemes* utilize the linear nature of transport term in Boltzmann equation for upwinding, together with the fact that the Euler's equations of gas dynamics can be recovered when appropriate moments are taken. Such *moment method strategy* based schemes include the developments of Sanders and Prendergast [275], Reitz [261], Pullin [249], Deshpande and Mandal and Deshpande [85, 211], Perthame [243], Prendergast and Xu [248] and Raghurama Rao and Deshpande [255], apart from others. These schemes are quite robust, but suffer from high numerical diffusion. Some offshoots of the kinetic schemes are the *Discrete Velocity Boltzmann Schemes* and *Variable (or Flexible) Velocity Boltzmann Schemes*. The kinetic schemes of Aregba-Driollet and Natalini [7], Nandagiri and Raghurama Rao [224], Shrinath *et al.* [280], Shashi Shekar Roy and Raghurama Rao [278] are some examples of this sub-category and offer several advantages compared to the traditional kinetic schemes, like simplicity and numerical diffusion control. Lattice Boltzmann methods (LBMs), which involve discretization of discrete ve-

locity Boltzmann equation to solve the continuum equations, are also closely related to the class of kinetic schemes, and are reviewed in chapter 5 (also in our preprint [6]).

- *Relaxation schemes* are numerical schemes based on semi-linear relaxation systems with stiff (relaxation) source terms. The hyperbolic PDE system is recovered from the relaxation system as the relaxation parameter approaches zero. Relaxation schemes were introduced by Jin and Xin [173]. Natalini [226] established the connection between the relaxation systems and discrete velocity Boltzmann systems. The works of Bouchut *et al.* [45], Raghurama Rao and Subba Rao [256], Raghurama Rao and Balakrishna [253], Arun *et al.* [11] Arun and Lukáčová [9], and Coulette *et al.* [69] are some of the schemes belonging to this category.

1.1.2 Higher order discretization strategies

A substantial amount of research has gone into the development of higher order accurate methods for hyperbolic systems, starting from 1970s. Starting from the early development of *Flux Corrected Transport methods* [36] and *flux limited* and *MUSCL* schemes of van Leer [311, 312], significant developments that followed are the shock sensor based central discretization schemes [163], the development of TVD schemes [146], *k-exact reconstruction schemes* [21], ENO schemes [147], WENO schemes [166], ADER schemes [304], spectral volume methods [319], spectral difference schemes [325], discontinuous Galerkin methods [98], Active Flux Schemes [107], and various developments within these categories. For some reviews of these methods, the reader is referred to the following references [315, 268, 161, 191, 304, 321, 105, 320, 15]. All higher order schemes suffer from inherent oscillations (Gibbs phenomena) due the inherent domination of higher order terms in the truncation errors, alternatively explained by the inherent higher order interpolations across large gradients (like shock waves). All the above methods have some mechanism to suppress these oscillations or wiggles, the most popular among them being the *limiter functions*.

1.1.3 Structure preserving schemes

The governing equations (of physical phenomena like fluid flows), which typically represent physical principles in mathematical representation, often contain rich structures associated with them. One typical example is the association of an entropy inequality (or entropy conservation, if shocks are not present) with the fluid flow equations, bridging the connection between fluid dynamics and thermodynamics. It has often been found in the research on development of numerical methods for fluid flows, that mimicking such structures in the discretization process often results in very efficient numerical methods. Thus, *structure preserving methods*, which

ensure that the numerical solutions preserve the inherent structures present in the governing hyperbolic PDE systems, have become popular in the last few years. Focus on some of these structures lead to the concepts of entropy stabilty, asymptotic preservation and well-balancing, among others. These three concepts also form the subject matter of the new algorithms presented in this thesis. A broad overview of the numerical methods based on these concepts is provided here.

- *Entropy stability* has emerged as a non-linear stability criterion for hyperbolic PDE systems, and *entropy stable methods* that inherently satisfy the discrete entropy inequality have become popular. These methods ensure that the numerical solution satisfies a discrete analogue of the entropy inequality satisfied by the continuous solution. Notable contributions in this direction include the developments of Tadmor [294], Ismail and Roe [158], Fjordholm, Mishra and Tadmor [117]. An elaborate review of entropy stable methods along with a novel development in this direction is presented in chapter 2 (also in our article [5]).

- *Asymptotic Preserving (AP)* - The dimensionless form of hyperbolic PDE systems typically contain one or more dimensionless numbers (such as Mach number for Euler's system) that can take different values in different regions of the domain. Numerical schemes utilised for such systems are expected to work efficiently for different values of these dimensionless numbers. Typically, for a small value of such a non-dimensional number, asymptotic analysis provides the useful information of the behaviour of the system. *Asymptotic Preserving (AP) Schemes* ensure that numerical discretization preserves the expected features as the asymptotic limits are reached. Two typical examples of asymptotic preserving schemes are those concerned with (i) the compressible flow solvers at the incompressible limit and (ii) Boltzmann equation (of kinetic theory) based numerical methods at different scalings to yield methods for compressible and incompressible Euler and Navier-Stokes equations.

  - *Low Mach number solvers* In the context of Euler's system, Mach number can be ≥ 1 (compressible regime) or $<< 1$ (incompressible regime). Both these regimes can occur together in a given problem, and a numerical scheme is expected to handle them both efficiently. Different methods such as preconditioning methods, artificial compressibility methods, projection methods, multiple pressure variables methods, and implicit-explicit (IMEX) methods have been developed to extend the usage of compressible solvers to incompressible (or low Mach number) limit.

* *Preconditioning methods* aim to remove the large disparity between different eigenvalues at the low Mach number limit by multiplying the time derivative of compressible flow equations with preconditioning matrix. It is aimed at reducing the stiffness present due to the disparity between different wave speeds at low Mach number limit. An often associated aspect is the reduction of numerical diffusion at low Mach number limit, as typical diffusion is inversely proportional the square of the Mach number. Several researchers such as Choi and Merkle [62], Turkel [307, 308], van Leer, Lee and Roe [316], Hirsch and Hakimi [151] attempted to develop preconditioners for low Mach number limit.
* *Artificial compressibility methods* introduced by Chorin [63] add an artificial compressibility term into the continuity equation, allowing compressible solvers to handle incompressible flows effectively. The method essentially adds a time derivative of pressure to the continuity equation, thus forming a strictly hyperbolic system from the incompressible equations. Based on such a hyperbolicity of the system, it becomes easier to introduce the strategies of compressible flow solvers. Notable contributions in this direction include the developments of Temam [298], Chang and Kwak [58], McHugh and Ramshaw [215], Pappou and Tsangaris [231], and Tamamidis *et al.* [296].
* *Projection methods* introduced by Chorin [64, 65] separate the velocity and pressure updates to improve stability and accuracy in the incompressible limit. Notable contributions in this direction include the developments of Harlow and Welch (Marker-and-Cell method) [145], Strikwerda [289], Bell, Colella and Glaz [22], Brown *et al.* [49], and Weinan and Liu [322].
* *Multiple pressure variables method* involve an operator splitting technique motivated by the asymptotic analysis of Euler's system of equations, in the low Mach number limit. Notable contributions in this direction include the developments of Klein [183], Geratz *et al.* [127], Munz *et al.* [221], and Park and Munz [237].
* *IMEX methods* that involve implicit treatment of stiff terms and explicit treatment of non-stiff terms based on the asymptotic analysis of governing hyperbolic PDE system have lately become popular. Notable contributions in this direction include the developments of Ascher *et al.* [12], Pareschi and Russo [234], and Albi *et al.* [1]. These methods aim to recover the discretisation of incompressible equations as the Mach number (in the context of Euler's system) approaches zero, and hence are *asymptotic preserving*. An elaborate review of these methods

along with a novel development that is both asymptotic preserving and entropy stable is presented in chapter 4.

  - *Kinetic theory based AP methods* - This topic, with the existing of extensive literature over several decades, is intricately linked with other issues of the kinetic theory based numerical methods and thus requires an elaborate description, as presented in the next section.

- *Well-balancing* - Hyperbolic PDE systems often contain source terms due to body forces or chemical reactions. The balance between convection and source terms can support multiple steady solutions that are stable or unstable. A good numerical scheme is expected to maintain the stable steady states with required order of accuracy, and such schemes are known as *well-balanced schemes*. A numerical scheme that is not well-balanced will result in a bad time decay towards steady solutions. Notable contributions in the direction of well-balanced schemes include the developments of van Leer [314], Roe [267], Glaister [130], Greenberg and Leroux [137], Gosse [135], Alouges, Ghidaglia and Tajchman [2], Jin [168], and Perthame and Simeoni [244]. Moreover, if the source term involved is stiff, the imbalance between convection and source terms will result in incorrect wave-speeds and incorrect locations of discontinuities. Several researchers such as, Colella, Majda and Roytburd [67], Ben-Artzi [25], LeVeque and Yee [202], Chorin (Random choice method) [66], Bao and Jin (Random projection method) [17, 18], and Kurganov (Accurate deterministic projection method) [186], attempted to identify the cause and tackle the issue. Lately, modified Godunov [217], implicit-explicit (IMEX) methods [292] and Threshold Values Method (TVM) [330] have been used for stiff source term treatment. A simple and novel way of handling well-balancing in lattice Boltzmann framework is presented in chapter 5.

Further, apart from finite volume methods (FVM), **finite element methods** (FEM) gained significant popularity in solving PDEs during the 1950s and 1960s. However, for hyperbolic systems, finite element methods faced challenges related to oscillations and numerical stability. The basic foundation of the FEM is the Galerkin approximation, which essentially leads to central type discretization and thus needs modifications for application to hyperbolic systems. Various such modifications have been successfully introduced over the past few decades, leading to *Taylor Galerkin methods, Petrov Galerkin methods, Characteristic Galerkin methods and, recently, Discontinuous Galerkin methods*. The books by Hughes [155], Strang and Fix [287], Donea and Huerta [99], Hesthaven and Warburton [149], Kuzmin and Hämäläinen [188], Dolejší and Feistauer [98], Zinkiewicz *et al.* [334] provide insights of some such finite element methods.

1.2 Multiscale kinetic equations

Kinetic equations represent mathematical models for the evolution of particle distribution functions in phase space. They are widely employed to describe the behavior of systems with numerous interacting particles, such as rarefied gases [35, 54, 205, 317], plasmas [299], granular gases [300], semiconductors [213], neutron transport [203], and quantum gases [309]. These equations often exhibit multiscale nature due to disparities in spatial and temporal scales involved. When the scaling parameter approaches a limit value, the kinetic equation asymptotically decomposes into a different equation. For instance, different scalings of the Boltzmann equation lead to different limit equations such as, Euler limit, incompressible Euler limit, and incompressible Navier-Stokes limit. The reader is referred to Bardos, Golse and Levermore [20, 19], Golse and Saint-Raymond [134], and the book by Saint-Raymond [274] for details on different limits of the Boltzmann equation. Traditional numerical methods may struggle to efficiently resolve the behavior of the system, especially when disparities in scales lead to stiffness issues.

1.2.1 Numerical discretisation of multiscale kinetic equations

The Direct Simulation Monte Carlo (DSMC) method by Bird [31] and Nanbu [223] solves the kinetic equation by using statistical sampling to approximate the solution. Although this stochastic method guarantees efficiency and preservation of physical properties, it becomes extremely expensive to avoid statistical fluctuations near continuum regimes. Deterministic numerical methods for kinetic equations are considered here and their broad overview is presented. These consist of semi-Lagrangian methods, discrete velocity methods, spectral methods, domain decomposition methods and asymptotic preserving methods.

- *Semi-Lagrangian methods* consider the Lagrangian nature of transport process, and operate on a fixed computational grid. The kinetic equation is often solved by splitting technique that allows the natural application of Lagrangian method to linear transport terms. Several approaches such as, particle-in-cell methods (Birdsall and Langdon [32]), flux-balance methods (Boris and Book [36]), weighted essentially non-oscillatory (WENO) methods (Carrillo and Vecil [53]), and discontinuous Galerkin (DG) methods (Qiu and Shu [252], Ayuso, Carrillo and Shu [14]) can be used for efficient simulation of transport terms in kinetic equations. Some of the semi-Lagrangian methods applied to Vlasov-Poisson system are due to Cheng and Knorr [60], Crouseilles *et al.* [76], Sonnendrücker *et al.* [284] and Filbet *et al.* [116]. Notable developments of semi-Lagrangian methods for Boltzmann-Gross-Krook (BGK) model of rarefied gas dynamics system are by Russo and Filbet [272], Russo, Santagati and Yun [273], and Dimarco and Loubére [87, 88]. Further,

Lattice Boltzmann methods (LBMs) can be considered as semi-Lagrangian methods with exact transport, and its elaborate review is presented in chapter 5 (also in our preprint [6]).

- *Discrete velocity methods* are known for approximating the Boltzmann equation in velocity space. They have been historically used to ease the mathematical study of rarefied gases (Carleman [51], Broadwell [48], Gatignol [126], Cabannes, Gatignol and Luo [50]). Recently, these methods are related to consistent velocity discretisations of Boltzmann equation (Goldstein *et al.* [133], Rogier and Schneider [269], Mischler [218], Panferov and Heintz [230], Fainsilber *et al.* [108]). However, these methods have limited accuracy and high computational cost when compared to the stochastic methods for evaluation of Boltzmann integral. Yet, these methods are very robust when simple collision operators such as BGK model are used (Mieussens [216]).

- *Spectral methods* use Fast Fourier Transform (FFT) theory for kinetic equations, and have spectral accuracy (*i.e.*, error tends to zero faster than $\mathcal{O}(N^k)$ for any $k < 0$, where N is the number of grid points) if the solution is sufficiently smooth. Several notable contributions were made in this direction for Boltzmann equation (Pareschi and Perthame [232], Pareschi and Russo [233]), Landau equation (Filbet and Pareschi [114]), granular gases (Filbet *et al.* [115]), and quantum gases (Filbet *et al.* [112]). Further, it is possible to speed up the spectral methods by using fast summation algorithms and this makes them competitive with DSMC methods for non-stationary flows (Filbet [111], Wu *et al.* [324], Gamba and Tharkabhushanam [122]).

- *Domain decomposition methods* have been crucial for solving large-scale problems efficiently by dividing the computational domain into subdomains. They involve usage of the kinetic description in domains far from equilibrium, and the numerical discretisation of limit equation in domains near equilibrium. The difficulty, however, is identification, modelling and numerics of the transition zone between different descriptions, and the requirement of moving interfaces to couple different regions. Notable contributions in this direction include the developments of Bourgat *et al.* [47], Bourgat, LeTallec and Tidriri [46], Schneider [276], Tiwari and Klar [302], Tiwari [301], Degond *et al.* [79, 80], Degond and Dimarco [78], and Kolobov *et al.* [185].

- *Asymptotic Preserving (AP) methods* capture the correct limiting behavior as certain scaling parameters approach particular values. This is essential for ensuring that the numerical solution converges to the correct solution in the limiting case, with uniform

accuracy and uniform stability across different values of the scaling parameter. A review of AP schemes for various systems can be found in the article of Jin [169], and technical report of Pareschi and Russo [235]. Some of the AP schemes designed for Boltzmann equation in classical fluid limit are due to Gabetta *et al.* [121], Filbet and Jin [113], and Dimarco and Pareschi [91, 92]. Notable AP schemes for kinetic equation in diffusion limit are due to Jin *et al.* [172], Bennoune *et al.* [27], Lemou and Mieussens [200], Boscarino *et al.* [37], and Dimarco *et al.* [94]. These include exponential Runge-Kutta and implicit-explicit Runge-Kutta (IMEX-RK) strategies. An overview of IMEX-RK technique along with a novel development of AP scheme for kinetic equation in diffusion limit is presented in chapter 3 (also in our preprint [3]).

1.3 Outline of the thesis

The main theme of the thesis is development and analysis of structure preserving numerical methods that yield relevant numerical solutions. Each chapter presents a novel structure preserving numerical scheme for some hyperbolic PDE system or multiscale kinetic equation. The outline is as follows.

The governing systems considered in chapter 2 are hyperbolic PDE system and its vector-kinetic model. Entropy stable schemes for a given hyperbolic system are available in literature. It is also known from literature that the entropy inequality of a hyperbolic system is obtained as a moment of the entropy inequality of vector-kinetic model. However, this correspondence is not necessarily maintained in numerical discretisations. In chapter 2, a novel entropy stable numerical scheme for a vector-kinetic model of hyperbolic PDE systems is developed. It is shown that this scheme also recovers entropy stability of the hyperbolic PDE system.

Chapter 3 is concerned about multiscale kinetic equations with diffusion scaling. In the asymptotic limit, this results in a diffusion equation. A novel high order asymptotic preserving method that recovers a consistent, accurate and stable numerical scheme for the asymptotic limit equation has been developed in this chapter. This scheme can handle general (both well-prepared and non-well-prepared) initial conditions. This chapter also presents the extension of this framework and asymptotic analysis to advection-diffusion asymptotics and inflow boundary problems.

In chapter 4, the governing system considered is the barotropic Euler system. The dimensionless form of this system contains Mach number that can become small. This dimensionless system has an entropy inequality corresponding to a convex entropy function, for all values of Mach number. In the limit of Mach number approaching zero, this system reduces to incompressible system of equations. Thus, it is evident that this requires both asymptotic preservation and

entropy stability. In this chapter, a novel asymptotic preserving scheme that satisfies entropy stability for the barotropic Euler system has been developed.
Chapter 5 is concerned about hyperbolic PDEs and associated vector-kinetic models. Lattice Boltzmann methods (LBMs) have been derived from vector-kinetic models of hyperbolic PDEs in the literature. In this chapter, crucial properties like H-inequality, macroscopic finite difference form, consistency, total variation boundedness and positivity of such LBMs have been analysed, and a novel way to handle well-balancing between convection and source terms of hyperbolic PDEs in the LBM framework has been presented.
Chapter 6 concludes the thesis.

Chapter 2

Entropy conserving/stable scheme for a vector-kinetic model of hyperbolic systems

The moment of entropy equation for vector-BGK model results in the entropy equation for macroscopic model. However, this is usually not the case in numerical methods because the current literature consists mostly of entropy conserving/stable schemes for macroscopic model. In this chapter, we attempt to fill this gap by developing an entropy conserving scheme for vector-kinetic model, and we show that the moment of this results in an entropy conserving scheme for macroscopic model. With the numerical viscosity of entropy conserving scheme as reference, the entropy stable scheme for vector-kinetic model is developed in the spirit of Tadmor [293]. We show that the moment of this scheme results in an entropy stable scheme for macroscopic model. The schemes are validated on several benchmark test problems for scalar and shallow water equations, and conservation/stability of both kinetic and macroscopic entropies are presented.

2.1 Introduction

The connection between entropy functions and symmetrisability of hyperbolic systems was explained in [146, 156], and this led to entropy-based non-linear stability analysis of numerical schemes. In the seminal work in [293, 294], a general condition to conserve/dissipate entropy of a semi-discrete scheme for hyperbolic system was introduced. Following this, many developments on fluxes satisfying entropy conservation/dissipation condition for various hyperbolic systems were made. These include developments specific for shallow water equations

[124, 323, 236], Euler's equations [21, 158, 251, 56, 259, 260, 125, 73, 61, 328], Navier-Stokes equations [326, 212, 257] and magneto hydro-dynamics equations [57]. Recently, several interesting studies such as, entropy stability for conservation laws with non-convex flux functions [197], and characterisation of stability [123] and robustness (for under-resolved flows) [55] of high order entropy stable schemes were carried out.

On the other hand, kinetic entropy formulations were introduced for hyperbolic equations like multi-dimensional scalar conservation laws, isentropic Euler and full Euler equations [245, 206, 207, 86]. Discrete kinetic models with entropy considerations were also proposed for hyperbolic systems [7, 226, 43, 44, 30, 45]. Specifically, in [43] it was shown that the entropy inequalities for a hyperbolic system can be derived as minimisation of entropies of vector-kinetic equation with BGK model. This approach of obtaining entropy inequalities from kinetic-BGK models is a promising strategy to characterise weak solutions of hyperbolic systems [250]. Hence, in this chapter, we attempt to develop entropy stable schemes (in the sense of [293, 294]) for a kinetic model based on [43] and show that they yield entropy stability for the hyperbolic system. This is in contrast to shock capturing schemes [280] based on discrete kinetic models.

A kinetic entropy stable scheme for continuous velocity Boltzmann's equation was recently developed in [160]. Although this scheme is entropy stable in the Euler limit, it employs huge number of velocities (24^3 for one dimensional problems) as the velocity space must be sufficiently resolved to satisfy the collision invariance. In our work, due to the usage of discrete kinetic models instead of continuous velocity Boltzmann's equation, we obtain an entropy stable scheme for the vanishing epsilon limit with very few velocities (as low as 2 for one dimensional problems). Moreover, our formalism is general enough to construct entropy stable scheme for a given hyperbolic system, while the work of [160] is specific to the Euler system.

The chapter is organised as follows. In section 2.2, we briefly describe the entropy framework and entropy conservation/stability conditions required to be satisfied by a semi-discrete scheme for hyperbolic system (or macroscopic model). Then, in section 2.3, we provide a brief description of the vector-BGK model in [43]. In section 2.4, we describe our modification to vector-BGK model, termed as the vector-kinetic model. This modification allows us to obtain entropy flux potentials required for developing entropy preserving scheme for vector-kinetic model. Then, in sections 2.5 and 2.6 we develop entropy conserving and stable schemes for vector-kinetic model, and show that these become entropy conserving and stable schemes for macroscopic model upon taking moments. In section 2.7, we describe the time discretisation strategies employed to complete our scheme. Then, in section 2.8, we verify our schemes on various numerical test problems. Section 2.9 concludes the chapter. The list of symbols used in the chapter are shown in Table 2.1.

Symbol	Description
$\mathbf{U}$	Conserved variable vector in macroscopic model
$\mathbf{G}^{(d)}(\mathbf{U})$	Flux vector (along direction d) in macroscopic model
$\eta(\mathbf{U})$	Entropy function for macroscopic model
$\omega^{(d)}(\mathbf{U})$	Entropy flux function for macroscopic model
$\psi^{(d)}$	Entropy flux potential for macroscopic model
$\mathbf{G}^{(d)^\star}_{i_d \pm \frac{1}{2}}$	Entropy conserving interface flux for macroscopic model
$\mathbf{Q}^{(d)^\star}_{i_d \pm \frac{1}{2}}$	Numerical viscosity corresponding to entropy conserving flux for macroscopic model
$\mathbf{G}^{(d)}_{i_d \pm \frac{1}{2}}$	Entropy stable interface flux for macroscopic model
$\mathbf{F}_m$	Dependent variable vector in vector-kinetic model
$v_m^{(d)}$	Discrete velocities in vector-kinetic model
$v_m^{(d)}\mathbf{F}_m$	Flux (along direction d) of the dependent variable vector in vector-kinetic model
H_m^η	Entropy function for vector-kinetic model
$v_m^{(d)} H_m^\eta$	Entropy flux function for vector-kinetic model
$\chi_m^{(d)}$	Entropy flux potential for vector-kinetic model
$\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d \pm \frac{1}{2}}$	Entropy conserving interface flux for vector-kinetic model
$\mathbf{Q}^{(d)^\star}_{m_{i_d \pm \frac{1}{2}}}$	Numerical viscosity corresponding to entropy conserving flux for vector-kinetic model
$\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d \pm \frac{1}{2}}$	Entropy stable interface flux for vector-kinetic model
$\mathbf{V}$	Entropy variable

Table 2.1: Table of symbols

2.2 Macroscopic model

Consider the hyperbolic system (or macroscopic model),

$$\partial_t \mathbf{U} + \partial_{x_d} \mathbf{G}^{(d)}(\mathbf{U}) = \mathbf{0} \tag{2.1}$$

where $\mathbf{U} : \Omega \times [0, T] \rightarrow \mathbb{R}^p$ and $\mathbf{G}^{(d)}(\mathbf{U}) : \mathbb{R}^p \rightarrow \mathbb{R}^p$, with $d \in \{1, 2, .., D\}$. Here Ω is a convex subset of $\mathbb{R}^D$.

2.2.1 Entropy framework

Here, we briefly recall the underlying theory (presented in [293, 294, 295]) behind development of entropy conserving/stable scheme for eq. (2.1).
If the macroscopic model in eq. (2.1) admits convex entropy-entropy flux pair $\left(\eta(\mathbf{U}), \omega^{(d)}(\mathbf{U})\right)$ that satisfies,

$$\partial_{\mathbf{U}} \omega^{(d)} = \partial_{\mathbf{U}} \eta \cdot \partial_{\mathbf{U}} \mathbf{G}^{(d)} \Leftrightarrow \partial^2_{\mathbf{U}} \eta \cdot \partial_{\mathbf{U}} \mathbf{G}^{(d)} \text{ is symmetric} \tag{2.2}$$

then the following entropy inequality holds.

$$\partial_t \eta(\mathbf{U}) + \partial_{x_d} \omega^{(d)}(\mathbf{U}) \leq 0 \tag{2.3}$$

Equality holds in smooth regions, while strict inequality holds in non-smooth regions.
Due to the convexity of $\eta(\mathbf{U})$, there exists one-one correspondence $\mathbf{U} \rightarrow \mathbf{V} := \partial_{\mathbf{U}} \eta$ such that the following equivalent symmetric form of eq. (2.1) holds true.

$$\partial_{\mathbf{V}} \mathbf{U} \; \partial_t \mathbf{V} + \partial_{\mathbf{U}} \mathbf{G}^{(d)} \; \partial_{\mathbf{V}} \mathbf{U} \; \partial_{x_d} \mathbf{V} = \mathbf{0} \tag{2.4}$$

Here, $\partial_{\mathbf{V}} \mathbf{U} = \left(\partial^2_{\mathbf{U}} \eta(\mathbf{U})\right)^{-1}$ is symmetric positive-definite (due to the convexity of $\eta(\mathbf{U})$) and $\partial_{\mathbf{V}} \mathbf{G}^{(d)} = \partial_{\mathbf{U}} \mathbf{G}^{(d)} \; \partial_{\mathbf{V}} \mathbf{U}$ is symmetric (refer Harten [146] for theorems due to Godunov and Mock).
Further, the compatibility condition in eq. (2.2) can be re-written in terms of entropy variable $\mathbf{V}$, thanks to the convexity of $\eta(\mathbf{U})$ that assures existence of $\left(\partial_{\mathbf{U}} \mathbf{V}\right)^{-1}$.

$$\partial_{\mathbf{V}} \omega^{(d)} = \mathbf{V} \cdot \partial_{\mathbf{V}} \mathbf{G}^{(d)} \tag{2.5}$$

Due to the symmetric nature of $\partial_{\mathbf{V}}\mathbf{G}^{(d)}$, there exist potentials $\psi^{(d)}(\mathbf{V})$ such that $\partial_{\mathbf{V}}\psi^{(d)} = \mathbf{G}^{(d)}(\mathbf{V})$. Therefore, according to eq. (2.5), there exist entropy flux potentials,

$$\psi^{(d)}(\mathbf{V}) = \mathbf{V} \cdot \mathbf{G}^{(d)}(\mathbf{V}) - \omega^{(d)}(\mathbf{V}) \tag{2.6}$$

2.2.2 Entropy conserving scheme

Consider a structured grid with grid size Δx_d along each direction d. Then, a three-point (along each direction d) semi-discrete conservative scheme for eq. (2.1) is,

$$\frac{d}{dt}\mathbf{U}_i + \frac{1}{\Delta x_d}\left(\mathbf{G}^{(d)^\star}_{i_d+\frac{1}{2}} - \mathbf{G}^{(d)^\star}_{i_d-\frac{1}{2}}\right) = \mathbf{0} \tag{2.7}$$

Here i denotes the index for cell centre of each cell/finite volume, and $i_d \pm \frac{1}{2}$ denote indices for right/left interfaces of cell i along direction d. For consistency, the numerical flux $\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} := \mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}}(\mathbf{U}_i, \mathbf{U}_{i_d\pm1})$ is such that $\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}}(\mathbf{U}, \mathbf{U}) = \mathbf{G}^{(d)}(\mathbf{U})$, where $i_d \pm 1$ denote indices for the cell centres of cells to the right/left of cell i along direction d.
The scheme in eq. (2.7) is entropy conserving iff the interface numerical fluxes satisfy the entropy conserving condition (derived in [293, 294]),

$$\left\langle [[\mathbf{V}]]_{i_d+\frac{1}{2}}, \mathbf{G}^{(d)^\star}_{i_d+\frac{1}{2}} \right\rangle = \left[\left[\psi^{(d)}\right]\right]_{i_d+\frac{1}{2}} \tag{2.8}$$

Here, $[[(.)]]_{i_d+\frac{1}{2}}$ denotes the jump $(.)_{i_d+1} - (.)_i$. Then, the following three-point (along each direction d) entropy equality holds true.

$$\frac{d}{dt}\eta\left(\mathbf{V}_i\right) + \frac{1}{\Delta x_d}\left(\omega^{(d)^\star}_{i_d+\frac{1}{2}} - \omega^{(d)^\star}_{i_d-\frac{1}{2}}\right) = 0 \tag{2.9}$$

The interface numerical entropy flux consistent with eq. (2.6) is given by

$$\omega^{(d)^\star}_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\mathbf{V}_i + \mathbf{V}_{i_d\pm1}\right) \cdot \mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\left(\psi^{(d)}_i + \psi^{(d)}_{i_d\pm1}\right) \tag{2.10}$$

Further, the entropy conserving numerical flux $\mathbf{G}^{(d)^\star}_{i_d+\frac{1}{2}}$ satisfying eq. (2.8) can be evaluated along the path $\mathbf{V}_{i_d+\frac{1}{2}}(\xi) = \mathbf{V}_i + \xi\Delta\mathbf{V}_{i_d+\frac{1}{2}}$ as,

$$\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\mathbf{G}^{(d)}_i + \mathbf{G}^{(d)}_{i_d\pm1}\right) - \frac{1}{2}\mathbf{Q}^{(d)^\star}_{i_d\pm\frac{1}{2}}\left[[\mathbf{V}]\right]_{i_d\pm\frac{1}{2}} \tag{2.11}$$

with

$$\mathbf{Q}^{(d)^\star}_{i_d+\frac{1}{2}} = \int_0^1 (2\xi - 1)\, \partial_{\mathbf{V}} \mathbf{G}^{(d)} \left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi) \right) d\xi \tag{2.12}$$

The term $\mathbf{Q}^{(d)^\star}_{i_d\pm\frac{1}{2}}$ which is symmetric (need not be positive-definite) is considered as numerical viscosity coefficient matrix. This counterbalances dispersion from the average flux. Further, the entropy conserving scheme is second order accurate in space (refer [293, 294]). Construction of higher order entropy conserving fluxes as linear combinations of second order accurate entropy conserving fluxes $\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}}$ is discussed in [196].

2.2.3 Entropy stable scheme

The three-point (along each direction d) consistent flux,

$$\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}} = \mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}} \left[[\mathbf{V}]\right]_{i_d\pm\frac{1}{2}} \tag{2.13}$$

with $\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}} = \mathbf{Q}^{(d)}_{i_d\pm\frac{1}{2}} - \mathbf{Q}^{(d)^\star}_{i_d\pm\frac{1}{2}}$ is entropy stable if and only if $\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}}$ is positive-definite. Here $\mathbf{Q}^{(d)}_{i_d\pm\frac{1}{2}}$ is the numerical viscosity coefficient matrix corresponding to entropy stable scheme. The scheme then satisfies the three-point entropy inequality,

$$\begin{aligned}\frac{d}{dt}\eta(\mathbf{V}_i) + \frac{1}{\Delta x_d}\left(\omega^{(d)}_{i_d+\frac{1}{2}} - \omega^{(d)}_{i_d-\frac{1}{2}}\right) = \\ - \frac{1}{4\Delta x_d}\left(\left[[\mathbf{V}]\right]_{i_d+\frac{1}{2}} \cdot \mathbf{D}^{(d)}_{i_d+\frac{1}{2}} \left[[\mathbf{V}]\right]_{i_d+\frac{1}{2}} + \left[[\mathbf{V}]\right]_{i_d-\frac{1}{2}} \cdot \mathbf{D}^{(d)}_{i_d-\frac{1}{2}} \left[[\mathbf{V}]\right]_{i_d-\frac{1}{2}} \right) \leq 0\end{aligned} \tag{2.14}$$

Here, the consistent numerical entropy flux at interface is given by,

$$\omega^{(d)}_{i_d+\frac{1}{2}} = \omega^{(d)^\star}_{i_d+\frac{1}{2}} - \frac{1}{4}\left(\mathbf{V}_i + \mathbf{V}_{i_d+1}\right) \cdot \mathbf{D}^{(d)}_{i_d+\frac{1}{2}} \left[[\mathbf{V}]\right]_{i_d+\frac{1}{2}} \tag{2.15}$$

The entropy stable flux $\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}}$ given by eq. (2.13) is first order accurate in space (refer Tadmor [293, 294]). To achieve higher order accuracy in space, the term $\left[[\mathbf{V}]\right]_{i_d+\frac{1}{2}}$ in eq. (2.13) must be replaced by $\langle\langle \mathbf{V} \rangle\rangle_{i_d+\frac{1}{2}} = \mathbf{V}^-_{i_d+1} - \mathbf{V}^+_i$ where $\mathbf{V}^-_{i_d+1}$ and $\mathbf{V}^+_i$ are higher order reconstructions of $\mathbf{V}$ at interface $i_d + \frac{1}{2}$ (refer [118]).

2.3 Vector-BGK model

In this section, we briefly describe the vector-BGK model presented in [43]. Consider,

$$\partial_t \mathbf{f}_m + \partial_{x_d}\left(v^{(d)}_m \mathbf{f}_m\right) = -\frac{1}{\epsilon}\left(\mathbf{f}_m - \mathbf{F}_m(\mathbf{U})\right) \tag{2.16}$$

where ϵ is the relaxation parameter. Here, $\mathbf{f}_m := \mathbf{f}_m(x_1, .., x_d, .., x_D, v_m^{(1)}, .., v_m^{(d)}, .., v_m^{(D)}, t) \in \mathbb{R}^p$, $\mathbf{F}_m : \mathbb{R}^p \to \mathbb{R}^p$, $m \in \{1, 2, .., M\}$ and M is the number of discrete velocities. Splitting of streaming and relaxation operators in eq. (2.16) gives,

$$\text{Streaming:} \quad \partial_t \mathbf{f}_m + \partial_{x_d} \left(v_m^{(d)} \mathbf{f}_m \right) = \mathbf{0} \tag{2.17}$$

$$\text{Relaxation:} \quad \frac{d}{dt}\mathbf{f}_m = -\frac{1}{\epsilon} \left(\mathbf{f}_m - \mathbf{F}_m(\mathbf{U}) \right) \tag{2.18}$$

Instantaneous relaxation (i.e., $\epsilon = 0$) in the relaxation equation above yields $\mathbf{f}_m = \mathbf{F}_m(\mathbf{U})$. This is inserted into the streaming equation for its evolution. Now, it can be seen that if the following relations are satisfied,

$$\sum_{m=1}^{M} \mathbf{F}_m(\mathbf{U}) = \mathbf{U} \text{ and } \sum_{m=1}^{M} v_m^{(d)} \mathbf{F}_m(\mathbf{U}) = \mathbf{G}^{(d)}(\mathbf{U}) \tag{2.19}$$

then $\sum_{m=1}^{M}$ eq. (2.16) $\to$ eq. (2.1) as $\epsilon \to 0$.

2.3.1 Entropy framework

Following the definition of entropy function for vector-BGK model given by equations (E0)-(E2) in [43], let us define the entropy function $H_m^\eta(\mathbf{f}_m)$ as:

$$H_m^\eta(\mathbf{f}_m) \text{ is a convex function with respect to } \mathbf{f}_m \tag{2.20}$$

$$\sum_{m=1}^{M} H_m^\eta(\mathbf{F}_m(\mathbf{U})) = \eta(\mathbf{U}) \tag{2.21}$$

$$\sum_{m=1}^{M} H_m^\eta(\mathbf{F}_m(\mathbf{U})) \leq \sum_{m=1}^{M} H_m^\eta(\mathbf{f}_m) \tag{2.22}$$

Then, taking inner product of eq. (2.16) with the sub-differential of H_m^η at $\mathbf{F}_m(\mathbf{U})$ and using (2.20), (2.21) and (2.22), the following is obtained.

$$\partial_t H_m^\eta(\mathbf{f}_m) + \partial_{x_d} \left(v_m^{(d)} H_m^\eta(\mathbf{f}_m) \right) \leq \frac{1}{\epsilon} \left(H_m^\eta(\mathbf{F}_m(\mathbf{U})) - H_m^\eta(\mathbf{f}_m) \right)$$

$$\Rightarrow \sum_{m=1}^{M} \left(\partial_t H_m^\eta(\mathbf{f}_m) + \partial_{x_d} \left(v_m^{(d)} H_m^\eta(\mathbf{f}_m) \right) \right) \leq 0$$

$$\Rightarrow \partial_t \eta(\mathbf{U}) + \partial_{x_d} \left(\sum_{m=1}^{M} v_m^{(d)} H_m^{\eta}(\mathbf{F}_m(\mathbf{U})) \right) \leq 0 \text{ in the limit } \epsilon \to 0 \quad (2.23)$$

If $\omega^{(d)}(\mathbf{U}) = \sum_{m=1}^{M} v_m^{(d)} H_m^{\eta}(\mathbf{F}_m(\mathbf{U}))$, then eq. (2.23) is same as eq. (2.3). The reader is referred to [43] for details.

Thus, entropy inequality of the macroscopic model (eq. (2.1)) can be obtained as minimisation of entropies of the vector-BGK model (eq. (2.16)). This inspires one to develop entropy structure preserving numerical schemes for vector-BGK model that recover the entropy inequality of equivalent macroscopic scheme. However, the framework of vector-BGK model does not ensure the existence of $\partial^2_{\mathbf{f}_m} H_m^{\eta}(\mathbf{F}_m(\mathbf{U}))$ which is crucial in obtaining entropy flux potentials that allow for the consistent definition of interface numerical entropy fluxes. Hence, we resort to a much simpler model in the relaxed limit without the stiff relaxation parameter (hereafter referred as *vector-kinetic model*), and make the necessary modification to allow for the definition of entropy flux potentials.

2.4 Vector-kinetic model

In this model, we consider the evolution of relaxed limit ($\epsilon = 0$):

$$\partial_t \mathbf{F}_m + \partial_{x_d} \left(v_m^{(d)} \mathbf{F}_m \right) = \mathbf{0} \quad (2.24)$$

Let us define $\mathbf{F}_m(\mathbf{U})$ as in [43],

$$\mathbf{F}_m(\mathbf{U}) = a_m \mathbf{U} + b_m^{(d)} \mathbf{G}^{(d)}(\mathbf{U}) \quad (2.25)$$

with

$$\sum_{m=1}^{M} a_m = 1, \ \sum_{m=1}^{M} b_m^{(d)} = 0 \quad (2.26)$$

$$\sum_{m=1}^{M} v_m^{(j)} a_m = 0, \ \sum_{m=1}^{M} v_m^{(j)} b_m^{(d)} = \delta_{jd} \quad (2.27)$$

In the light of moment constraints in eqs. (2.26) and (2.27), the definition of $\mathbf{F}_m(\mathbf{U})$ in eq. (2.25) satisfies eq. (2.19).

2.4.1 Entropy framework

Define H_m^η as in [43],

$$H_m^\eta(\mathbf{U}) = a_m \eta(\mathbf{U}) + b_m^{(d)} \omega^{(d)}(\mathbf{U}) \tag{2.28}$$

Due to the constraints in eqs. (2.26) and (2.27), H_m^η satisfies,

$$\sum_{m=1}^{M} H_m^\eta(\mathbf{U}) = \eta(\mathbf{U}) \text{ and } \sum_{m=1}^{M} v_m^{(d)} H_m^\eta(\mathbf{U}) = \omega^{(d)}(\mathbf{U}) \tag{2.29}$$

We assume that the eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$ are positive, unlike in [43] where the eigenvalues are considered to be non-negative. It will be seen that this modification allows the definition of entropy flux potentials required in the construction of entropy preserving numerical scheme. As $\partial_{\mathbf{U}}\mathbf{F}_m$ is now invertible, $\partial_{\mathbf{F}_m} H_m^\eta$ satisfying $\partial_{\mathbf{U}} H_m^\eta = \partial_{\mathbf{F}_m} H_m^\eta \cdot \partial_{\mathbf{U}}\mathbf{F}_m$ exists. Therefore, the inner product of eq. (2.24) with $\partial_{\mathbf{F}_m} H_m^\eta$ gives,

$$\partial_t H_m^\eta + \partial_{x_d}\left(v_m^{(d)} H_m^\eta\right) = 0 \tag{2.30}$$

It can be seen that $\sum_{m=1}^{M}$ (*eq.* (2.30)) becomes eq. (2.3) with equality. Motivated by this, in this chapter, we develop entropy preserving scheme for vector-kinetic model that recovers entropy preservation of equivalent macroscopic scheme.

Lemma 2.1 *If* $\mathbf{F}_m(\mathbf{U})$ *and* $H_m^\eta(\mathbf{U})$ *respectively follow eqs.* (2.25) *and* (2.28) *with constants* a_m, $b_m^{(d)}$ *satisfying the moment constraints in eqs.* (2.26) *and* (2.27) *and rendering the eigenvalues of* $\partial_{\mathbf{U}}\mathbf{F}_m$ *to be positive, then* $\partial_{\mathbf{F}_m} H_m^\eta = \partial_{\mathbf{U}}\eta$.

Proof: Due to the compatibility condition in eq. (2.2), it can be seen from differentiation (with respect to $\mathbf{U}$) of eqs. (2.25) and (2.28) that $\partial_{\mathbf{U}} H_m^\eta = \partial_{\mathbf{U}}\eta \cdot \partial_{\mathbf{U}}\mathbf{F}_m$. Since $\partial_{\mathbf{U}}\mathbf{F}_m$ is invertible, $\partial_{\mathbf{U}}\eta = \partial_{\mathbf{U}} H_m^\eta \cdot (\partial_{\mathbf{U}}\mathbf{F}_m)^{-1}$. We already saw that $\partial_{\mathbf{F}_m} H_m^\eta = \partial_{\mathbf{U}} H_m^\eta \cdot (\partial_{\mathbf{U}}\mathbf{F}_m)^{-1}$. □

This lemma shows that the entropy variables for macroscopic and vector-kinetic models are equal, *i.e.*,

$$\mathbf{V} = \partial_{\mathbf{U}}\eta = \partial_{\mathbf{F}_m} H_m^\eta. \tag{2.31}$$

The choice of constants a_m, $b_m^{(d)}$ satisfying assumptions in the above lemma are discussed in 2.10.

As a consequence of lemma 2.1, we have $\partial^2_{\mathbf{F}_m} H_m^\eta = \partial^2_{\mathbf{U}}\eta \cdot (\partial_{\mathbf{U}}\mathbf{F}_m)^{-1}$. Further, $(\partial^2_{\mathbf{U}}\eta)^{-1} \partial^2_{\mathbf{F}_m} H_m^\eta = (\partial_{\mathbf{U}}\mathbf{F}_m)^{-1}$ can be expressed as

$$\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}} \left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}} \left(\partial^2_{\mathbf{F}_m} H_m^\eta\right) \left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}} \left(\partial^2_{\mathbf{U}}\eta\right)^{\frac{1}{2}} = (\partial_{\mathbf{U}}\mathbf{F}_m)^{-1} \tag{2.32}$$

thanks to the positive-definiteness of $\partial^2_{\mathbf{U}}\eta$. Thus, $\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}$ and $\left(\partial_{\mathbf{U}}\mathbf{F}_m\right)^{-1}$ are similar and therefore their eigenvalues are same.

Lemma 2.2 *If $\partial^2_{\mathbf{U}}\eta$ is positive-definite and eq. (2.32) holds true, then $\partial^2_{\mathbf{F}_m}H^\eta_m$ is positive-definite iff the eigenvalues of $\left(\partial_{\mathbf{U}}\mathbf{F}_m\right)^{-1}$ are positive.*

Proof: $\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}$ is symmetric as $\partial^2_{\mathbf{U}}\eta$ and $\partial^2_{\mathbf{F}_m}H^\eta_m$ are symmetric. Further, we have $\forall \mathbf{y}\neq\mathbf{0}\in\mathbb{R}^p$,

$$\mathbf{y}\cdot\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\mathbf{y}=\mathbf{z}\cdot\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\mathbf{z} \tag{2.33}$$

where $\mathbf{z}=\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\mathbf{y}\neq\mathbf{0}$ (as $\partial^2_{\mathbf{U}}\eta$ is positive-definite).
$\Leftarrow$ If the eigenvalues of $\left(\partial_{\mathbf{U}}\mathbf{F}_m\right)^{-1}$ are positive, then $\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}$ is positive-definite due to eq. (2.32). Then $\partial^2_{\mathbf{F}_m}H^\eta_m$ is rendered positive-definite by eq. (2.33).
$\Rightarrow$ If $\partial^2_{\mathbf{F}_m}H^\eta_m$ is positive-definite, then by eq. (2.33) $\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)\left(\partial^2_{\mathbf{U}}\eta\right)^{-\frac{1}{2}}$ is positive-definite. Then, the eigenvalues of $\left(\partial_{\mathbf{U}}\mathbf{F}_m\right)^{-1}$ are positive due to eq. (2.32). □

Thus, as consequence of lemma 2.1 and lemma 2.2, eq. (2.31) and positive-definiteness of $\partial^2_{\mathbf{F}_m}H^\eta_m$ are guaranteed iff the eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$ are positive. Using the one-to-one correspondence between $\mathbf{U}$ and $\mathbf{V}$, we consider $\mathbf{F}_m\left(\mathbf{U}\right)=\mathbf{F}_m\left(\mathbf{U}\left(\mathbf{V}\right)\right)$. Hence the vector-kinetic model in eq. (2.24) can be expressed in the equivalent symmetric form

$$\partial_{\mathbf{V}}\mathbf{F}_m\partial_t\mathbf{V}+\partial_{\mathbf{V}}\left(v^{(d)}_m\mathbf{F}_m\right)\partial_{x_d}\mathbf{V}=\mathbf{0} \tag{2.34}$$

Here $\partial_{\mathbf{V}}\mathbf{F}_m=\left(\partial^2_{\mathbf{F}_m}H^\eta_m\right)^{-1}$ is symmetric positive-definite. Due to the linearity of vector-kinetic model, $\partial_{\mathbf{V}}\left(v^{(d)}_m\mathbf{F}_m\right)=v^{(d)}_m\partial_{\mathbf{V}}\mathbf{F}_m$ is symmetric. As a result, there exist potentials $\chi^{(d)}_m(\mathbf{V})$ such that

$$\partial_{\mathbf{V}}\chi^{(d)}_m=v^{(d)}_m\mathbf{F}_m \tag{2.35}$$

Further, the compatibility condition

$$\partial_{\mathbf{F}_m}\left(v^{(d)}_m H^\eta_m\right)=\partial_{\mathbf{F}_m}H^\eta_m\cdot\partial_{\mathbf{F}_m}\left(v^{(d)}_m\mathbf{F}_m\right) \tag{2.36}$$

is also satisfied rendering H^η_m as the convex entropy function for vector-kinetic model. Note that this compatibility condition is always true for any convex H^η_m satisfying eq. (2.28) due to the linear nature of vector-kinetic model, unlike the compatibility condition (*eq.* (2.2)) for macroscopic model. In terms of $\mathbf{V}$, the above compatibility condition for vector-kinetic model becomes,

$$\partial_{\mathbf{V}}\left(v^{(d)}_m H^\eta_m\right)=\mathbf{V}\cdot\partial_{\mathbf{V}}\left(v^{(d)}_m\mathbf{F}_m\right) \tag{2.37}$$

thanks to the inverse of $\partial_{\mathbf{F}_m}\mathbf{V}$. Therefore, due to eqs. (2.35) and (2.37), there exist entropy flux potentials

$$\chi_m^{(d)}(\mathbf{V}) = \mathbf{V}\cdot v_m^{(d)}\mathbf{F}_m - v_m^{(d)}H_m^\eta = \partial_{\mathbf{F}_m}H_m^\eta \cdot v_m^{(d)}\mathbf{F}_m - v_m^{(d)}H_m^\eta \tag{2.38}$$

Thus, we have obtained the entropy flux potentials that are crucial in the construction of entropy preserving numerical scheme for vector-kinetic model.

2.5 Entropy conserving scheme for vector-kinetic model

The three-point (along each direction d) semi-discrete conservative scheme for vector-kinetic model in eq. (2.24) on a structured grid is given by,

$$\frac{d}{dt}\mathbf{F}_{m_i} + \frac{1}{\Delta x_d}\left(\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}} - \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d-\frac{1}{2}}\right) = \mathbf{0} \tag{2.39}$$

Here, $\mathbf{F}_{m_i}(t) = \mathbf{F}_m\left(\mathbf{V}_i(t)\right)$ and consistent $\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}} = v_m^{(d)}\mathbf{F}_m(\mathbf{V}_i, \mathbf{V}_{i_d+1})$ is such that $v_m^{(d)}\mathbf{F}_m(\mathbf{V},\mathbf{V}) = v_m^{(d)}\mathbf{F}_m(\mathbf{V})$. Consider the inner product $\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i \cdot \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}}$:

$$\begin{aligned}\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i \cdot \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} = &\frac{1}{2}\left(\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_{i_d\pm1} + \left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i\right)\cdot\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}}\\ &-\frac{1}{2}\left(\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_{i_d\pm1} - \left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i\right)\cdot\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}}\end{aligned}$$

If the interface numerical flux $\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}}$ satisfies the entropy conserving condition,

$$\left\langle \left[\left[\partial_{\mathbf{F}_m}H_m^\eta\right]\right]_{i_d+\frac{1}{2}}, \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}}\right\rangle = \left[\left[\chi_m^{(d)}\right]\right]_{i_d+\frac{1}{2}} \tag{2.40}$$

then,

$$\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i \cdot \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_{i_d\pm1} + \left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i\right)\cdot\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} - \frac{1}{2}\left(\chi^{(d)}_{m_{i_d\pm1}} - \chi^{(d)}_{m_i}\right)$$

Thus, the inner product of eq. (2.39) with $\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i$ gives the three-point entropy equality,

$$\frac{d}{dt}H^\eta_{m_i} + \frac{1}{\Delta x_d}\left(\left(v_m^{(d)}H_m^\eta\right)^\star_{i_d+\frac{1}{2}} - \left(v_m^{(d)}H_m^\eta\right)^\star_{i_d-\frac{1}{2}}\right) = 0 \tag{2.41}$$

iff it satisfies eq. (2.40), and the interface numerical entropy fluxes $\left(v_m^{(d)} H_m^\eta\right)^\star_{i_d \pm \frac{1}{2}}$ consistent with eq. (2.38) are given by,

$$\left(v_m^{(d)} H_m^\eta\right)^\star_{i_d \pm \frac{1}{2}} = \frac{1}{2}\left(\left(\partial_{\mathbf{F}_m} H_m^\eta\right)_i + \left(\partial_{\mathbf{F}_m} H_m^\eta\right)_{i_d \pm 1}\right) \cdot \left(v_m^{(d)} \mathbf{F}_m\right)^\star_{i_d \pm \frac{1}{2}} - \frac{1}{2}\left(\chi_{m_i}^{(d)} + \chi_{m_{i_d \pm 1}}^{(d)}\right) \tag{2.42}$$

It is seen that the entropy flux potentials $\chi_{m_i}^{(d)}$ enable us to consistently relate the two interfacial unknowns, numerical fluxes $\left(v_m^{(d)} \mathbf{F}_m\right)^\star_{i_d \pm \frac{1}{2}}$ and numerical entropy fluxes $\left(v_m^{(d)} H_m^\eta\right)^\star_{i_d \pm \frac{1}{2}}$. Further, let us define the interface numerical fluxes for macroscopic model as the moment of interface numerical fluxes for vector-kinetic model as,

$$\mathbf{G}^{(d)\star}_{i_d \pm \frac{1}{2}} = \sum_{m=1}^{M} \left(v_m^{(d)} \mathbf{F}_m\right)^\star_{i_d \pm \frac{1}{2}} \tag{2.43}$$

Theorem 2.1 *If the three-point semi-discrete conservative scheme (eq. (2.39)) for vector-kinetic model with*

- $\mathbf{F}_{m_i} = a_m \mathbf{U}_i + b_m^{(d)} \mathbf{G}_i^{(d)}, \ \forall i$
- *interface numerical fluxes* $\left(v_m^{(d)} \mathbf{F}_m\right)^\star_{i_d \pm \frac{1}{2}}$ *satisfying the entropy conserving condition in eq. (2.40) and*
- *constants* a_m, $b_m^{(d)}$ *satisfying the moment constraints in eqs. (2.26) and (2.27) while rendering positivity of eigenvalues of* $\partial_{\mathbf{U}} \mathbf{F}_m$

is used, and if the convex entropy function corresponding to it is $H_{m_i}^\eta = a_m \eta_i + b_m^{(d)} \omega_i^{(d)}, \ \forall i$, *then*

1. $\sum_{m=1}^{M}$ *(eq. (2.39)) becomes*

$$\frac{d}{dt}\mathbf{U}_i + \frac{1}{\Delta x_d}\left(\mathbf{G}^{(d)\star}_{i_d + \frac{1}{2}} - \mathbf{G}^{(d)\star}_{i_d - \frac{1}{2}}\right) = \mathbf{0} \tag{2.44}$$

   with $\mathbf{G}^{(d)\star}_{i_d \pm \frac{1}{2}}$ *given by eq. (2.43),*

2. *the interface numerical flux* $\mathbf{G}^{(d)\star}_{i_d \pm \frac{1}{2}}$ *given by eq. (2.43) satisfies the entropy conserving condition for macroscopic model (eq. (2.8)), and*

3. *the three-point entropy equality for macroscopic model (eq. (2.9)) holds true with interface numerical entropy flux* $\omega^{(d)\star}_{i_d \pm \frac{1}{2}}$ *given by eq. (2.10).*

Proof: Due to moment constraint in eq. (2.26), $\sum_{m=1}^{M} \mathbf{F}_{m_i} = \mathbf{U}_i$. Therefore, $\sum_{m=1}^{M}$ (*eq.* (2.39)) becomes eq. (2.44) with $\mathbf{G}^{(d)\star}_{i_d\pm\frac{1}{2}}$ given by eq. (2.43), thus proving 1.
By eq. (2.31), $[[\partial_{\mathbf{F}_m} H^{\eta}_m]]_{i_d\pm\frac{1}{2}} = [[\mathbf{V}]]_{i_d\pm\frac{1}{2}} = [[\partial_{\mathbf{U}}\eta]]_{i_d\pm\frac{1}{2}}$ is not a function of m. Hence, the moment of eq. (2.40) gives,

$$\left\langle [[\mathbf{V}]]_{i_d\pm\frac{1}{2}}, \sum_{m=1}^{M} \left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}} \right\rangle = \left[\left[\sum_{m=1}^{M} \chi_m^{(d)}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.45}$$

From eq. (2.38), it can be seen that $\chi_{m_i}^{(d)} = \mathbf{V}_i.v_m^{(d)}\mathbf{F}_{m_i} - v_m^{(d)}H^{\eta}_{m_i}$, $\forall i$. Hence, $\sum_{m=1}^{M}\chi_{m_i}^{(d)} = \mathbf{V}_i.\sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_{m_i}\right) - \sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_{m_i}\right)$, $\forall i$. We also have $\sum_{m=1}^{M} v_m^{(d)}\mathbf{F}_{m_i} = \mathbf{G}_i^{(d)}$ and $\sum_{m=1}^{M} v_m^{(d)}H^{\eta}_{m_i} = \omega_i^{(d)}$, $\forall i$ due to the action of moment constraint in eq. (2.27) on $\mathbf{F}_{m_i}$ and $H^{\eta}_{m_i}$. Therefore, by eq. (2.6), $\sum_{m=1}^{M}\chi_{m_i}^{(d)} = \psi_i^{(d)}$, $\forall i$. Using this and eq. (2.43) in eq. (2.45), we obtain,

$$\left\langle [[\mathbf{V}]]_{i_d\pm\frac{1}{2}}, \mathbf{G}^{(d)\star}_{i_d\pm\frac{1}{2}} \right\rangle = \left[\left[\psi^{(d)}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.46}$$

This proves 2.
We know that the three-point entropy equality in eq. (2.41) holds true corresponding to the assumptions stated in theorem 2.1. Since $\sum_{m=1}^{M}(H^{\eta}_m)_i = \eta_i$, $\forall i$ (due to the action of moment constraint in eq. (2.26) on $(H^{\eta}_m)_i$), moment of eq. (2.41) gives,

$$\frac{d}{dt}\eta_i + \frac{1}{\Delta x_d}\left(\sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d+\frac{1}{2}} - \sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d-\frac{1}{2}}\right) = 0 \tag{2.47}$$

Since $\left(\partial_{\mathbf{F}_m} H^{\eta}_m\right)_i = \mathbf{V}_i = \left(\partial_{\mathbf{U}}\eta\right)_i$ is not a function of m (by eq. (2.31)), moment of $\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d\pm\frac{1}{2}}$ given by eq. (2.42) yields,

$$\sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\mathbf{V}_i + \mathbf{V}_{i_d\pm1}\right)\cdot\sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\left(\sum_{m=1}^{M}\chi_{m_i}^{(d)} + \sum_{m=1}^{M}\chi_{m_{i_d\pm1}}^{(d)}\right) \tag{2.48}$$

We have already seen that $\sum_{m=1}^{M}\chi_{m_i}^{(d)} = \psi_i^{(d)}$, $\forall i$. Using this and eq. (2.43), we obtain,

$$\sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\mathbf{V}_i + \mathbf{V}_{i_d\pm1}\right)\cdot\mathbf{G}^{(d)\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\left(\psi_i^{(d)} + \psi_{i_d\pm1}^{(d)}\right) \tag{2.49}$$

It can be seen from eq. (2.10) that $\sum_{m=1}^{M}\left(v_m^{(d)}H^{\eta}_m\right)^{\star}_{i_d\pm\frac{1}{2}} = \omega^{(d)\star}_{i_d\pm\frac{1}{2}}$. This proves 3. □

In the light of eq. (2.31) resulting from lemma 2.1, moments involved in the proof of above

theorem become linear since $\partial_{\mathbf{F}_m} H_m^\eta$ is not a function of m. This plays a pivotal role in showing that entropy conserving scheme for vector-kinetic model results in an entropy conserving scheme for macroscopic model.

Remark 2.1 *In the above proof, the three-point entropy equality for macroscopic model (eq. (2.9)) with interface numerical entropy flux $\omega^{(d)^\star}_{i_d\pm\frac{1}{2}}$ given by eq. (2.10) is obtained as moment of three-point entropy equality for vector-kinetic model. Unlike this, we can also obtain eq. (2.9) directly at the macroscopic level as a consequence of $\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M} \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}}$ satisfying the entropy conserving condition for macroscopic model (eq. (2.8)).*

The entropy conserving fluxes satisfying eq. (2.40) can be evaluated using an integral along the path $\mathbf{V}_{i_d+\frac{1}{2}}(\xi) = \mathbf{V}_i + \xi\Delta\mathbf{V}_{i_d+\frac{1}{2}}$ as,

$$\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} = \int_0^1 \left(v_m^{(d)}\mathbf{F}_m\right)\left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) d\xi = \frac{1}{2}\left(v_m^{(d)}\mathbf{F}_{m_i} + v_m^{(d)}\mathbf{F}_{m_{i_d\pm1}}\right) - \frac{1}{2}\mathbf{Q}^{(d)^\star}_{m_{i_d\pm\frac{1}{2}}}\left[\left[\mathbf{V}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.50}$$

where

$$\mathbf{Q}^{(d)^\star}_{m_{i_d+\frac{1}{2}}} = \int_0^1 (2\xi-1)\,\partial_{\mathbf{V}}\left(v_m^{(d)}\mathbf{F}_m\right)\left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) d\xi \tag{2.51}$$

Although $\partial_{\mathbf{V}}\left(v_m^{(d)}\mathbf{F}_m\right)\left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right)$ is symmetric positive-definite, the term $\mathbf{Q}^{(d)^\star}_{m_{i_d+\frac{1}{2}}}$ is only symmetric (need not be positive-definite). This is considered as numerical viscosity coefficient matrix that counterbalances the dispersion from average flux. Integration by parts of $\mathbf{Q}^{(d)^\star}_{m_{i_d+\frac{1}{2}}}$ yields,

$$\mathbf{Q}^{(d)^\star}_{m_{i_d+\frac{1}{2}}} = \int_0^1 \left(\xi-\xi^2\right)\partial_{\mathbf{VV}}\left(v_m^{(d)}\mathbf{F}_m\right)\left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) d\xi\left[\left[\mathbf{V}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.52}$$

Thus,

$$\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(v_m^{(d)}\mathbf{F}_{m_i} + v_m^{(d)}\mathbf{F}_{m_{i_d\pm1}}\right) + O\left(\left|\left[\left[\mathbf{V}\right]\right]_{i_d+\frac{1}{2}}\right|^2\right) \tag{2.53}$$

and hence for smooth functions, we have

$$\frac{1}{\Delta x_d}\left(\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}} - \left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d-\frac{1}{2}}\right) = \frac{1}{2\Delta x_d}\left(\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d+1} - \left(v_m^{(d)}\mathbf{F}_m\right)_{i_d-1}\right) + O\left(\left|\left[\left[x_d\right]\right]_{i_d+\frac{1}{2}}\right|^2\right) \tag{2.54}$$

Therefore, the entropy conserving scheme for vector-kinetic model given by eq. (2.50) is second accurate in space. However, evaluation of a closed form interface flux function using eq. (2.50) is algebraically tedious for a general hyperbolic system.

The closed form expression can be obtained along the same lines as macroscopic model in [294]. Let $\left\{\mathbf{l}^j_{i_d+\frac{1}{2}} \in \mathbb{R}^p\right\}_{j=1}^p$ and $\left\{\mathbf{r}^j_{i_d+\frac{1}{2}} \in \mathbb{R}^p\right\}_{j=1}^p$ be two orthogonal sets of vectors such that

$\left\langle \mathbf{l}^j_{i_d+\frac{1}{2}}, \mathbf{r}^k_{i_d+\frac{1}{2}} \right\rangle = \delta_{jk}$. Let $\mathbf{V}^1_{i_d+\frac{1}{2}} = \mathbf{V}_i$ and

$$\mathbf{V}^{j+1}_{i_d+\frac{1}{2}} = \mathbf{V}^j_{i_d+\frac{1}{2}} + \left\langle \mathbf{l}^j_{i_d+\frac{1}{2}}, [[\mathbf{V}]]_{i_d+\frac{1}{2}} \right\rangle \mathbf{r}^j_{i_d+\frac{1}{2}} \; ; j \in \{1, 2, .., p\} \tag{2.55}$$

Then, we have a path connecting $\mathbf{V}_i$ and $\mathbf{V}_{i_d+1}$ since

$$\mathbf{V}^{p+1}_{i_d+\frac{1}{2}} = \mathbf{V}^1_{i_d+\frac{1}{2}} + \sum_{j=1}^{p} \left\langle \mathbf{l}^j_{i_d+\frac{1}{2}}, [[\mathbf{V}]]_{i_d+\frac{1}{2}} \right\rangle \mathbf{r}^j_{i_d+\frac{1}{2}} = \mathbf{V}_i + [[\mathbf{V}]]_{i_d+\frac{1}{2}} = \mathbf{V}_{i_d+1} \tag{2.56}$$

Now, it can be seen that the numerical flux given by,

$$\left(v^{(d)}_m \mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}} = \sum_{j=1}^{p} \frac{\chi^{(d)}_m \left(\mathbf{V}^{j+1}_{i_d+\frac{1}{2}}\right) - \chi^{(d)}_m \left(\mathbf{V}^{j}_{i_d+\frac{1}{2}}\right)}{\left\langle \mathbf{l}^j_{i_d+\frac{1}{2}}, [[\mathbf{V}]]_{i_d+\frac{1}{2}} \right\rangle} \mathbf{l}^j_{i_d+\frac{1}{2}} \tag{2.57}$$

satisfies the entropy conserving condition in eq. (2.40). However, for the purpose of numerical simulations, we use robust entropy conserving fluxes (satisfying eq. (2.40)) that are derived by defining averages of certain primitive variables and by balancing the coefficients corresponding to jumps in these primitive variables. These fluxes are described in section 2.8.

Remark 2.2 *Higher order entropy conserving (HOEC) fluxes for vector-kinetic model can be constructed as linear combinations of second order entropy conserving fluxes derived in this chapter (along the same lines as in [196] for macroscopic model). Since linear combinations are used, as a consequence of theorem 2.1, the moments of HOEC fluxes for vector-kinetic model will result in HOEC fluxes for macroscopic model.*

Corollary 2.1 *If the assumptions stated in theorem 2.1 hold and entropy conserving flux of the form in eq. (2.50) is used, then*

$$\sum_{m=1}^{M} \mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}} = \mathbf{Q}^{(d)\star}_{i_d\pm\frac{1}{2}} \tag{2.58}$$

Proof: By eqs. (2.43) and (2.50), we obtain

$$\mathbf{G}^{(d)\star}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M} \left(v^{(d)}_m \mathbf{F}_m\right)^\star_{i_d\pm\frac{1}{2}} = \frac{1}{2}\left(\mathbf{G}^{(d)}_i + \mathbf{G}^{(d)}_{i_d\pm1}\right) - \frac{1}{2}\sum_{m=1}^{M} \mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}} [[\mathbf{V}]]_{i_d\pm\frac{1}{2}} \tag{2.59}$$

since $\sum_{m=1}^{M} v^{(d)}_m \mathbf{F}_{m_i} = \mathbf{G}^{(d)}_i, \; \forall i$ due to the action of moment constraint in eq. (2.27) on $\mathbf{F}_{m_i}$.

Further,

$$\sum_{m=1}^{M} \mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}} = \int_0^1 (2\xi - 1) \sum_{m=1}^{M} \partial_{\mathbf{V}} \left(v_m^{(d)} \mathbf{F}_m\right) \left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) d\xi \tag{2.60}$$

and

$$\sum_{m=1}^{M} \partial_{\mathbf{V}} \left(v_m^{(d)} \mathbf{F}_m\right) \left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) = \sum_{m=1}^{M} v_m^{(d)} \partial_{\mathbf{V}} \left(a_m \mathbf{U} + b_m^j \mathbf{G}^j\right) \left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) = \partial_{\mathbf{V}} \mathbf{G}^{(d)} \left(\mathbf{V}_{i_d+\frac{1}{2}}(\xi)\right) \tag{2.61}$$

due to the action of moment constraint in eq. (2.27) on $\partial_{\mathbf{V}} \mathbf{F}_m$. Thus, comparing eqs. (2.12) and (2.60), we obtain $\sum_{m=1}^{M} \mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}} = \mathbf{Q}^{(d)\star}_{i_d\pm\frac{1}{2}}$. □

2.6 Entropy stable scheme for vector-kinetic model

Consider the three-point semi-discrete conservative scheme on structured grid,

$$\frac{d}{dt}\mathbf{F}_{m_i} + \frac{1}{\Delta x_d}\left(\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d+\frac{1}{2}} - \left(v_m^{(d)}\mathbf{F}_m\right)_{i_d-\frac{1}{2}}\right) = \mathbf{0} \tag{2.62}$$

The interface numerical flux $\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}$ is given by,

$$\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}} = \left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}} \left[\left[\partial_{\mathbf{F}_m} H_m^{\eta}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.63}$$

Here, $\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}} = \mathbf{Q}^{(d)}_{m_{i_d\pm\frac{1}{2}}} - \mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}}$. $\mathbf{Q}^{(d)}_{m_{i_d\pm\frac{1}{2}}}$ and $\mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}}$ are the numerical viscosity coefficient matrices corresponding to entropy stable and entropy conserving schemes respectively. $\mathbf{Q}^{(d)\star}_{m_{i_d\pm\frac{1}{2}}}$ is given by eq. (2.51).

Then, the inner product of eq. (2.62) with $\left(\partial_{\mathbf{F}_m} H_m^{\eta}\right)_i$ gives the entropy in-equality,

$$\begin{aligned}
&\frac{d}{dt}H^{\eta}_{m_i} + \frac{1}{\Delta x_d}\left(\left(v_m^{(d)}H_m^{\eta}\right)_{i_d+\frac{1}{2}} - \left(v_m^{(d)}H_m^{\eta}\right)_{i_d-\frac{1}{2}}\right) \\
&= -\frac{1}{4\Delta x_d}\left(\left[\left[\partial_{\mathbf{F}_m}H_m^{\eta}\right]\right]_{i_d+\frac{1}{2}} \cdot \mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} \left[\left[\partial_{\mathbf{F}_m}H_m^{\eta}\right]\right]_{i_d+\frac{1}{2}} + \left[\left[\partial_{\mathbf{F}_m}H_m^{\eta}\right]\right]_{i_d-\frac{1}{2}} \cdot \mathbf{D}^{(d)}_{m_{i_d-\frac{1}{2}}} \left[\left[\partial_{\mathbf{F}_m}H_m^{\eta}\right]\right]_{i_d-\frac{1}{2}}\right) \leq 0
\end{aligned} \tag{2.64}$$

iff $\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}}$ is positive-definite. The interface numerical entropy flux $\left(v_m^{(d)}H_m^{\eta}\right)_{i_d+\frac{1}{2}}$ consistent with eq. (2.38) becomes,

$$\left(v_m^{(d)}H_m^{\eta}\right)_{i_d+\frac{1}{2}} = \left(v_m^{(d)}H_m^{\eta}\right)^{\star}_{i_d+\frac{1}{2}} - \frac{1}{4}\left(\left(\partial_{\mathbf{F}_m}H_m^{\eta}\right)_i + \left(\partial_{\mathbf{F}_m}H_m^{\eta}\right)_{i_d+1}\right) \cdot \mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} \left[\left[\partial_{\mathbf{F}_m}H_m^{\eta}\right]\right]_{i_d+\frac{1}{2}} \tag{2.65}$$

Further, let us define the interface numerical fluxes for macroscopic model as the moment of interface numerical fluxes for vector-kinetic model as,

$$\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M} \left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}} \tag{2.66}$$

Theorem 2.2 *If the three-point semi-discrete conservative scheme (eq. (2.62)) for vector-kinetic model with*

- $\mathbf{F}_{m_i} = a_m\mathbf{U}_i + b_m^{(d)}\mathbf{G}_i^{(d)},\ \forall i$
- *interface numerical fluxes* $\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}$ *satisfying eq. (2.63) and*
- *constants* a_m, $b_m^{(d)}$ *satisfying the moment constraints in eqs. (2.26) and (2.27) while rendering the positivity of eigenvalues of* $\partial_{\mathbf{U}}\mathbf{F}_m$

is used, and if the convex entropy function corresponding to it is $H^{\eta}_{m_i} = a_m\eta_i + b_m^{(d)}\omega_i^{(d)},\ \forall i$, *then*

1. $\sum_{m=1}^{M}$ *eq. (2.62) becomes*

$$\frac{d}{dt}\mathbf{U}_i + \frac{1}{\Delta x_d}\left(\mathbf{G}^{(d)}_{i_d+\frac{1}{2}} - \mathbf{G}^{(d)}_{i_d-\frac{1}{2}}\right) = \mathbf{0} \tag{2.67}$$

   with $\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}}$ *given by eq. (2.66),*

2. *the interface numerical flux* $\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}}$ *given by eq. (2.66) is equal to eq. (2.13), and*
3. *the three-point entropy in-equality for macroscopic model (eq. (2.14)) holds true with interface numerical entropy flux* $\omega^{(d)}_{i_d\pm\frac{1}{2}}$ *given by eq. (2.15).*

Proof: Due to moment constraint in eq. (2.26), $\sum_{m=1}^{M}\mathbf{F}_{m_i} = \mathbf{U}_i$. Therefore, $\sum_{m=1}^{M}$ eq. (2.62) becomes eq. (2.67) with $\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}}$ given by eq. (2.66), thus proving 1.
Since $\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}$ follows eq. (2.63) and $[[\partial_{\mathbf{F}_m}H^{\eta}_m]]_{i_d\pm\frac{1}{2}} = [[\mathbf{V}]]_{i_d\pm\frac{1}{2}} = [[\partial_{\mathbf{U}}\eta]]_{i_d\pm\frac{1}{2}}$ is not a function of m (by eq. (2.31)), eq. (2.66) becomes,

$$\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}}[[\mathbf{V}]]_{i_d\pm\frac{1}{2}} \tag{2.68}$$

By theorem 2.1, $\sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}}$ satisfies entropy conserving condition in eq. (2.8) and hence it is equal to $\mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}}$. We also have $\sum_{m=1}^{M}\mathbf{Q}^{(d)^\star}_{m_{i_d\pm\frac{1}{2}}} = \mathbf{Q}^{(d)^\star}_{i_d\pm\frac{1}{2}}$ by corollary 2.1. Further, $\sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}}$ is positive-definite as $\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}}$ is positive-definite $\forall m$. Therefore, $\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}} = \sum_{m=1}^{M}\mathbf{Q}^{(d)}_{m_{i_d\pm\frac{1}{2}}} - \mathbf{Q}^{(d)^\star}_{i_d\pm\frac{1}{2}}$ is positive-definite, and hence

$$\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}} = \mathbf{G}^{(d)^\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}}\left[\left[\mathbf{V}\right]\right]_{i_d\pm\frac{1}{2}} \tag{2.69}$$

This proves 2.

Corresponding to the assumptions stated in theorem 2.2, the three-point entropy in-equality in eq. (2.64) holds true. Since $\sum_{m=1}^{M}(H_m^\eta)_i = \eta_i,\ \forall i$ (due to the action of moment constraint in eq. (2.26) on $(H_m^\eta)_i$), $\left[\left[\partial_{\mathbf{F}_m}H_m^\eta\right]\right]_{i_d\pm\frac{1}{2}} = \left[\left[\mathbf{V}\right]\right]_{i_d\pm\frac{1}{2}} = \left[\left[\partial_{\mathbf{U}}\eta\right]\right]_{i_d\pm\frac{1}{2}}$ is not a function of m (by eq. (2.31)) and $\sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} = \mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$, moment of eq. (2.64) gives,

$$\begin{aligned}\frac{d}{dt}\eta_i + \frac{1}{\Delta x_d}\left(\sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)_{i_d+\frac{1}{2}} - \sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)_{i_d-\frac{1}{2}}\right) = \\ -\frac{1}{4\Delta x_d}\left(\left[\left[\mathbf{V}\right]\right]_{i_d+\frac{1}{2}}\cdot\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}\left[\left[\mathbf{V}\right]\right]_{i_d+\frac{1}{2}} + \left[\left[\mathbf{V}\right]\right]_{i_d-\frac{1}{2}}\cdot\mathbf{D}^{(d)}_{i_d-\frac{1}{2}}\left[\left[\mathbf{V}\right]\right]_{i_d-\frac{1}{2}}\right)\end{aligned} \tag{2.70}$$

Since $\left[\left[\partial_{\mathbf{F}_m}H_m^\eta\right]\right]_{i_d\pm\frac{1}{2}} = \left[\left[\mathbf{V}\right]\right]_{i_d\pm\frac{1}{2}} = \left[\left[\partial_{\mathbf{U}}\eta\right]\right]_{i_d\pm\frac{1}{2}}$ and $\left(\partial_{\mathbf{F}_m}H_m^\eta\right)_i = \mathbf{V}_i = \left(\partial_{\mathbf{U}}\eta\right)_i$ are not functions of m (by eq. (2.31)), moment of eq. (2.65) yields,

$$\sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)_{i_d+\frac{1}{2}} = \sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)^{\star}_{i_d+\frac{1}{2}} - \frac{1}{4}\left(\mathbf{V}_i + \mathbf{V}_{i_d+1}\right)\cdot\sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}}\left[\left[\mathbf{V}\right]\right]_{i_d+\frac{1}{2}} \tag{2.71}$$

Since $\sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)^{\star}_{i_d+\frac{1}{2}} = \omega^{(d)^\star}_{i_d+\frac{1}{2}}$ (by theorem 2.1) and $\sum_{m=1}^{M}\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} = \mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$, comparison of the above equation with eq. (2.15) yields $\sum_{m=1}^{M}\left(v_m^{(d)}H_m^\eta\right)_{i_d+\frac{1}{2}} = \omega^{(d)}_{i_d+\frac{1}{2}}$. This proves 3. □

Thus, an entropy stable scheme for vector-kinetic model results in an entropy stable scheme for macroscopic model, thanks to eq. (2.31) (resulting from lemma 2.1) that rendered the linearity of moments in the above proof.

Remark 2.3 *In the above proof, the three-point entropy in-equality for macroscopic model (eq. (2.14)) with interface numerical entropy flux* $\omega^{(d)}_{i_d\pm\frac{1}{2}}$ *given by eq. (2.15) is obtained as moment of three-point entropy in-equality for vector-kinetic model. Unlike this, we can also obtain eq. (2.14) directly at the macroscopic level as a consequence of* $\mathbf{G}^{(d)}_{i_d\pm\frac{1}{2}} = \sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}$ *sat-*

isfying the entropy stability condition for macroscopic model $\left(\textit{eq. (2.13) with positive-definite } \mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}}\right)$.

2.6.1 High resolution scheme

Since the interface numerical flux $\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d+\frac{1}{2}}$ contains a term with $[[\mathbf{V}]]_{i_d+\frac{1}{2}}$ which is $O\left(\Delta x_d\right)$, the entropy stable scheme in eq. (2.62) is only first order accurate in space. In order to attain higher order accuracy in space, the interface numerical flux in eq. (2.63) is modified as,

$$\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}} = \left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\mathbf{D}^{(d)}_{m_{i_d\pm\frac{1}{2}}}\left\langle\langle\mathbf{V}\rangle\right\rangle_{i_d\pm\frac{1}{2}} \tag{2.72}$$

where $\left\langle\langle\mathbf{V}\rangle\right\rangle_{i_d+\frac{1}{2}} = \mathbf{V}^-_{i_d+1} - \mathbf{V}^+_i$. Further, $\mathbf{V}^-_{i_d+1} = \mathbf{V}_{i_d+1}\left(x_{d_{i_d+\frac{1}{2}}}\right)$ and $\mathbf{V}^+_i = \mathbf{V}_i\left(x_{d_{i_d+\frac{1}{2}}}\right)$ are higher order reconstructions of $\mathbf{V}$ at interface $i_d+\frac{1}{2}$. We utilise second order reconstructions in obtaining the numerical results, and the details are provided therein section 2.8. The moment of eq. (2.72) becomes,

$$\sum_{m=1}^{M}\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}} = \mathbf{G}^{(d)\star}_{i_d\pm\frac{1}{2}} - \frac{1}{2}\mathbf{D}^{(d)}_{i_d\pm\frac{1}{2}}\left\langle\langle\mathbf{V}\rangle\right\rangle_{i_d\pm\frac{1}{2}} \tag{2.73}$$

It can be easily seen that this is a higher order entropy stable flux for macroscopic model, and it is a consequence of linearity due to eq. (2.31) (resulting from lemma 2.1).

2.7 Time discretisation

Let $\mathcal{F}_{m_i}$ be $-\frac{1}{\Delta x_d}\left(\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d+\frac{1}{2}} - \left(v_m^{(d)}\mathbf{F}_m\right)_{i_d-\frac{1}{2}}\right)$ where $\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}$ is entropy conserving $\left(\left(v_m^{(d)}\mathbf{F}_m\right)^{\star}_{i_d\pm\frac{1}{2}}\right.$ satisfying eq. (2.40)$\left.\right)$ or entropy stable $\left(\left(v_m^{(d)}\mathbf{F}_m\right)_{i_d\pm\frac{1}{2}}\right.$ satisfying eq. (2.63)$\left.\right)$. Then, the semi-discrete entropy conserving/stable schemes in eqs. (2.39) and (2.62) can be represented as,

$$\frac{d}{dt}\mathbf{F}_{m_i} = \mathcal{F}_{m_i} \tag{2.74}$$

Since we utilise second order scheme for entropy conserving/stable spatial discretisations, a third order scheme is required for the temporal derivative so that the entropy production/dissipation due to temporal derivative will not affect the entropy conservation/stability achieved spatially. Hence, the temporal derivative in above equation is discretised using 3-stage third order strong stability preserving Runge-Kutta method (SSPRK$(3,3)$) [281]. After each stage of the RK method, $\mathbf{U}_i$ is evaluated using $\mathbf{U}_i = \sum_{m=1}^{M}\mathbf{F}_{m_i}$, and this is utilised in the evaluation of fluxes required for the next stage.

2.8 Numerical results

In this section, the entropy conserving (EC)/stable (ES) schemes are tested against various physical problems governed by scalar equations and the system of shallow water equations. For each problem, the basic ingredients such as problem description, choice of macroscopic entropy-entropy flux pair, fluxes satisfying entropy conserving/stability conditions in eqs. (2.40) and (2.63), second order reconstructions of entropy stable fluxes and CFL criteria are provided. We use the following error quantifications to study the errors in macroscopic and vector-kinetic entropies at time t.

$$\text{Signed error} = \frac{\sum_i \left((.)_i^t - (.)_i^{t-\Delta t}\right)}{N} \tag{2.75}$$

$$\text{Absolute error} = \frac{\sum_i \left|(.)_i^t - (.)_i^{t-\Delta t}\right|}{N} \tag{2.76}$$

Here, N is the total number of cells or grid points in the computational domain. It can be seen that the signed error allows for cancellations of positive and negative errors present at different spatial locations. An equivalent of this with reference as $t = 0$ instead of $t - \Delta t$ is commonly used in literature in the context of global entropy preservation [258]. However, in order to understand the actual entropy preservation property of a spatially entropy preserving scheme, one needs to use the absolute error that does not allow spatial cancellations. Further, we use the signed error to identify whether the scheme is globally entropy dissipating or not. A positive signed error indicates global entropy production while negative signed error indicates global entropy dissipation. We present the numerical solutions, global entropy vs. time, and error vs. time plots for each problem.

2.8.1 Scalar equations

We consider scalar equations of the form,

$$\partial_t U + \partial_{x_d} G^{(d)}(U) = 0 \tag{2.77}$$

with initial condition $U(x_1, .., x_d, .., x_D, 0) = U_0(x_1, .., x_d, .., x_D)$. We choose suitable convex entropy-entropy flux pair specific to $G^{(d)}(U)$. The constants $a_m, b_m^{(d)}$ in eqs. (2.25) and (2.28) are chosen as described in 2.10. The time step is chosen as

$$\Delta t \leq C \frac{\Delta x}{\lambda} \; ; \; \Delta x = min\left(\Delta x_d\right) \tag{2.78}$$

Here, C is the CFL number. The choice of λ is described in 2.10. The flux

$$\left(v_m^{(d)} F_m\right)^{\star}_{i_d+\frac{1}{2}} = \frac{\chi^{(d)}_{m_{i_d+1}} - \chi^{(d)}_{m_i}}{V_{i_d+1} - V_i} \tag{2.79}$$

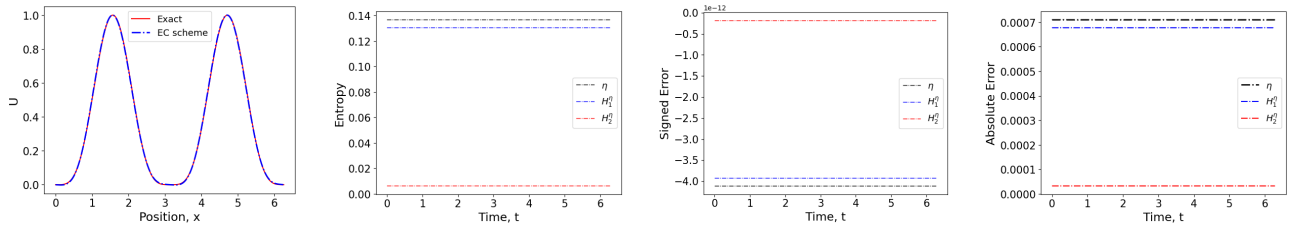
satisfies the entropy conserving condition in eq. (2.40). This is used when $V_{i_d+1} \neq V_i$. When $V_{i_d+1} = V_i$, we do not update the flux, as any value of flux satisfies the entropy conserving condition (*eq.* (2.40)). Here, the entropy variable is $V_i = \left(\partial_U \eta\right)_i$ and the vector-kinetic entropy flux potential is given by $\chi^{(d)}_{m_i} = V_i . \left(v_m^{(d)} F_m\right)_i - \left(v_m^{(d)} H^{\eta}_m\right)_i$.
For entropy stable scheme, we use $\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} \langle\langle \mathbf{V} \rangle\rangle_{i_d+\frac{1}{2}} = \frac{1}{M} \mathbf{R}^{(d)}_{i_d+\frac{1}{2}} \mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}} \left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}}$. For scalar equations, $\mathbf{R}^{(d)}_{i_d+\frac{1}{2}} = 1$ and $\mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}}$ is the absolute wave speed obtained using the average (arithmetic) value of U at cells i and $i_d + 1$. We use the second order reconstruction of $\left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}}$ as explained in section 2.8.2.

2.8.1.1 Linear advection

For the one-dimensional linear advection problem with $G^{(1)}(U) = U$, we choose $\eta(U) = \frac{1}{2}U^2$, and correspondingly $\omega^{(1)}(U) = \frac{1}{2}U^2$ satisfies the compatibility condition in eq. (2.2). The initial condition is $U_0\left(x_1\right) = \left(sin(x_1)\right)^4$. The domain of the problem is $[0, 2\pi)$, and it is discretised using 256 uniform cells. Periodic boundary conditions are used here. Numerical solutions are obtained at $T = 2\pi$.
It can be seen from fig. 2.1a that the numerical solution matches well with the exact solution. Figure 2.1b shows the global entropies over time. It can be seen that the entropies remain nearly constant. The signed and absolute errors in entropies are shown in figs. 2.1c and 2.1d respectively. Since we use second order accurate entropy conserving scheme for vector-kinetic model and Δx is of $O(10^{-2})$, we expect an absolute error of $O(10^{-4})$ in the vector-kinetic entropies. This is observed in fig. 2.1d. The negative signed errors in fig. 2.1c indicate that the $O(\Delta x^2)$ error is globally dissipative in nature. Due to the symmetric nature of the periodic profile, there may be cancellations in errors spatially and we observe a very low signed error of $O(10^{-12})$.
In order to study the convergence of the problem, we use very low CFL of $C = 0.1$. Second order accuracy of the scheme is evident from the results presented in table 2.2. The exact solution is used as reference for the convergence study.

Number of cells, Nx	Δx_1	L_2 norm	$O(L_2)$
32	0.196349541	0.035757668	-
64	0.09817477	0.00781911	2.19
128	0.049087385	0.00140703	2.47
256	0.024543693	0.000249239	2.50

Table 2.2: EOC for linear advection at $T = 2\pi$ using EC scheme with $C = 0.1$



(a) Solutions (b) Entropy functions (c) Signed errors (d) Absolute errors

Figure 2.1: Linear advection at $T = 2\pi$ using EC scheme with $C = 0.1$ and $Nx = 256$

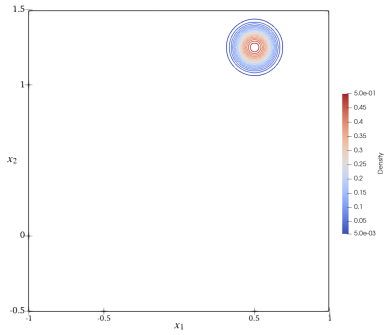
2.8.1.2 Linear rotation

For the two dimensional linear rotation problem, $G^{(1)}(U) = -\left(x_2 - \frac{1}{2}\right) U$ and $G^{(2)}(U) = \left(x_1 - \frac{1}{2}\right) U$. The entropy function is chosen as $\eta(U) = U^2$, and correspondingly the entropy flux functions become $\omega^{(1)}(U) = -\left(x_2 - \frac{1}{2}\right) U^2$ and $\omega^{(2)}(U) = \left(x_1 - \frac{1}{2}\right) U^2$. The initial condition is shown in fig. 2.2a. The domain of the problem is $[-1, 1) \times [-0.5, 1.5)$, and it is discretised using 256×256 uniform cells. The value of U at the boundary is kept fixed throughout the computation, and a CFL of $C = 0.9$ is used.

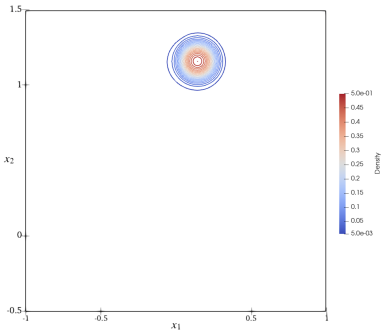
The numerical solution at $T = 0.5$ is shown in fig. 2.2b. Since Δx is of $O(10^{-2})$, one would expect an error of $O(10^{-4})$ in the absolute errors due to the usage of second order accurate entropy conserving scheme. We observe better error of $O(10^{-5})$ in fig. 2.2e. Further, it is interesting to observe the symmetries in errors of H_2^η, H_4^η and H_1^η, H_3^η in fig. 2.2d. However, these symmetries may not be located on the same spatial point. If they were, then the absolute error of macroscopic entropy η would be much smaller than $O(10^{-7})$ (due to cancellations) since it is the sum of vector-kinetic entropies.

2.8.1.3 Non-linear inviscid Burgers' test

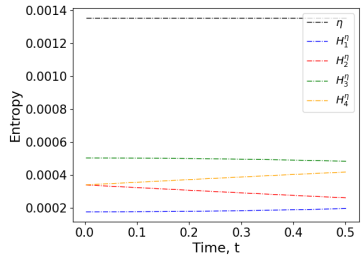
For this non-linear one-dimensional problem with $G^{(1)}(U) = \frac{1}{2}U^2$, we choose $\eta(U) = U^2$, and correspondingly $\omega^{(1)}(U) = \frac{2}{3}U^3$ satisfies the compatibility condition in eq. (2.2). The initial condition is $U_0(x_1) = sin(2\pi x_1)$. The domain of the problem is $[0, 1)$, and it is discretised using



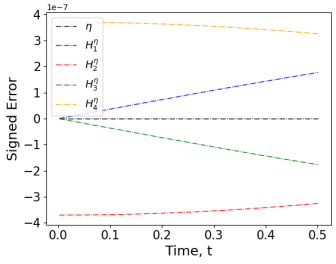
(a) Initial condn. $T = 0$



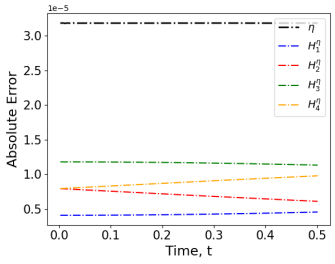
(b) Numerical soln. $T = 0.5$



(c) Entropy functions



(d) Signed errors



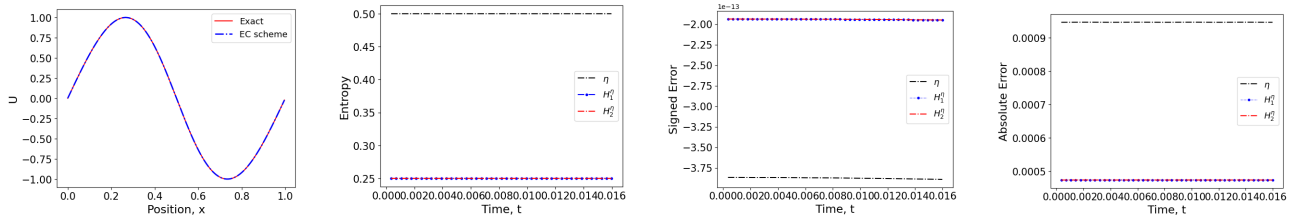
(e) Absolute errors

Figure 2.2: Linear rotation at $T = 0.5$ using EC scheme with $C = 0.9$ and $Nx, Ny = 256$

256 uniform cells. Periodic boundary conditions are used here. We use entropy conserving and entropy stable schemes respectively for obtaining numerical solutions at $T = \frac{0.1}{2\pi}$ and $T = 0.25$ in figs. 2.3 and 2.4.

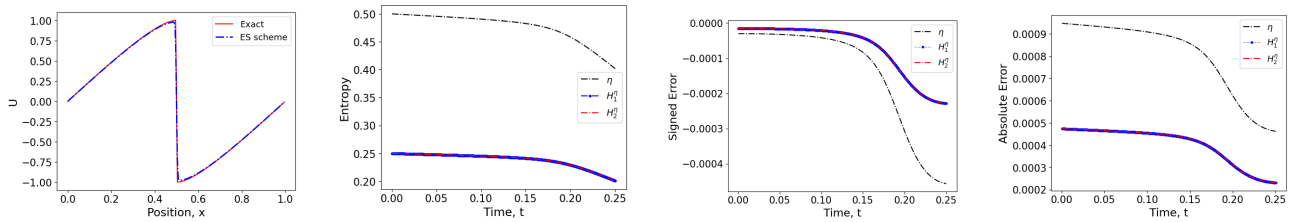
Figures 2.3a and 2.4a show that the numerical solutions match well with the exact solutions. Figures 2.3b and 2.4b show that macroscopic and vector-kinetic entropy functions are conserved and dissipated respectively in the smooth ($T = \frac{0.1}{2\pi}$) and non-smooth ($T = 0.25$) cases. The signed and absolute errors for $T = \frac{0.1}{2\pi}$ are shown in figs. 2.3c and 2.3d. Since we use second order accurate entropy conserving scheme for vector-kinetic model and Δx is of $O(10^{-3})$, we expect an absolute error of $O(10^{-6})$ in the vector-kinetic entropies. However, we observe an absolute error of $O(10^{-4})$ in fig. 2.1d. This might be because the terms multiplying $O(\Delta x^2)$ in the M-PDE of entropy equality are not $O(1)$ due to non-linearities. The negative signed errors in fig. 2.1c indicate that the error is globally dissipative in nature. Due to the symmetric nature of periodic profile, there may be cancellations in errors spatially and we observe a very low signed error of $O(10^{-13})$.

Further, the signed and absolute errors for $T = 0.25$ are shown in figs. 2.4c and 2.4d. Here too, we observe an absolute error of $O(10^{-4})$. Negative signed error of $O(10^{-4})$ indicates entropy dissipation after the formation of discontinuity.



(a) Solutions (b) Entropy functions (c) Signed errors (d) Absolute errors

Figure 2.3: Non-linear inviscid Burgers' test at $T = \frac{0.1}{2\pi}$ using EC scheme with $C = 0.1$ and $Nx = 256$



(a) Solutions (b) Entropy functions (c) Signed errors (d) Absolute errors

Figure 2.4: Non-linear inviscid Burgers' test at $T = 0.25$ using first order ES scheme with $C = 0.1$ and $Nx = 256$

In order to study the convergence of the problem, a very low CFL of $C = 0.1$ is chosen. The reference solution is the exact solution obtained by employing Newton-Raphson iteration with tolerance of 10^{-15}. It is seen from table 2.2 that more than second order accuracy is attained as the grid is refined.

Number of cells, Nx	Δx_1	L_2 norm	$O(L_2)$
64	0.015625	0.000281831	-
128	0.0078125	0.000118395	1.89
256	0.00390625	4.37E-05	3.24

Table 2.3: EOC for non-linear inviscid Burgers' test at $T = \frac{0.1}{2\pi}$ using EC scheme with $C = 0.1$

2.8.2 Shallow water equations

We consider the shallow water equations,

$$\partial_t \begin{bmatrix} \rho \\ \rho u_j \end{bmatrix} + \partial_{x_d} \begin{bmatrix} \rho u_d \\ \rho u_j u_d + p\delta_{dj} \end{bmatrix} = \mathbf{0} \ ; \ p = \kappa \rho^2 \ ; \ j \in \{1, 2, .., D\} \tag{2.80}$$

with initial condition $\mathbf{U}(x_1,..,x_d,..,x_D,0)=\mathbf{U_0}(x_1,..,x_d,..,x_D)$. Here, $\mathbf{U}=\begin{bmatrix}\rho\\ \rho u_j\end{bmatrix}$, $\mathbf{G}^{(d)}(\mathbf{U})=\begin{bmatrix}\rho u_d\\ \rho u_j u_d+p\delta_{dj}\end{bmatrix}$ and $\kappa=\frac{1}{2}$. The notation $h,\ g$ with $h=\rho,\ g=2\kappa=1$ is commonly used in the shallow water community. In this case, $p=\frac{1}{2}gh^2$.

The entropy function is $\eta(\mathbf{U})=\frac{1}{2}\rho u_j u_j+\kappa\rho^2$, and correspondingly the entropy flux functions become $\omega^{(d)}(\mathbf{U})=u_d\left(\frac{1}{2}\rho u_j u_j+2\kappa\rho^2\right)$. $\mathbf{F}_m$ and H_m^η of vector-kinetic model are found using eq. (2.25) and eq. (2.28) respectively. The constants $a_m, b_m^{(d)}$ and λ are chosen as described in 2.10. The time step is chosen as

$$\Delta t\leq C\frac{\Delta x}{\lambda}\ ;\ \Delta x=min\left(\Delta x_d\right) \tag{2.81}$$

Here, C is the CFL number. Let us construct the entropy conserving flux $\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}}$ satisfying eq. (2.40). Consider the arithmetic average $\overline{A}_{i_d+\frac{1}{2}}=\frac{1}{2}\left(A_i+A_{i_d+1}\right)$. This average satisfies $[[AB]]_{i_d+\frac{1}{2}}=\overline{A}_{i_d+\frac{1}{2}}[[B]]_{i_d+\frac{1}{2}}+\overline{B}_{i_d+\frac{1}{2}}[[A]]_{i_d+\frac{1}{2}}$. Hence, the entropy conserving condition in eq. (2.40) can be expressed as,

$$\begin{aligned}\left\langle\begin{bmatrix}2\kappa\left[[\rho]\right]_{i_d+\frac{1}{2}}-\overline{u_k}_{i_d+\frac{1}{2}}\left[[u_k]\right]_{i_d+\frac{1}{2}}\\ \left[[u_j]\right]_{i_d+\frac{1}{2}}\end{bmatrix},\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}}\right\rangle=\\ v_m^{(d)}\left(2\overline{\rho}_{i_d+\frac{1}{2}}\left(a_m\left[[\rho]\right]_{i_d+\frac{1}{2}}+b_m^k\overline{u_k}_{i_d+\frac{1}{2}}\left[[\rho]\right]_{i_d+\frac{1}{2}}\right)+\overline{\rho^2}_{i_d+\frac{1}{2}}\left(b_m^k\left[[u_k]\right]_{i_d+\frac{1}{2}}\right)\right)\end{aligned} \tag{2.82}$$

Equating the terms corresponding to $\left[[\rho]\right]_{i_d+\frac{1}{2}}$ and $\left[[u_j]\right]_{i_d+\frac{1}{2}}$, we obtain

$$\left(v_m^{(d)}\mathbf{F}_m\right)^\star_{i_d+\frac{1}{2}}=\begin{bmatrix}v_m^{(d)}\overline{\rho}_{i_d+\frac{1}{2}}\left(a_m+b_m^k\overline{u_k}_{i_d+\frac{1}{2}}\right)\\ v_m^{(d)}\left(\overline{\rho}_{i_d+\frac{1}{2}}\overline{u_j}_{i_d+\frac{1}{2}}\left(a_m+b_m^k\overline{u_k}_{i_d+\frac{1}{2}}\right)+\kappa b_m^j\overline{\rho^2}_{i_d+\frac{1}{2}}\right)\end{bmatrix} \tag{2.83}$$

This EC flux is second order accurate in space. Let us now derive the entropy stable flux given by eq. (2.63). We know that $\sum_{m=1}^M\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}}=\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$, a positive-definite matrix. We use the robust $\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$ described in [118]. That is,

$$\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}=\mathbf{R}^{(d)}_{i_d+\frac{1}{2}}\mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}}\mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} \tag{2.84}$$

where $\mathbf{R}^{(d)}_{i_d+\frac{1}{2}}$ is a suitably scaled matrix whose columns are eigenvectors of $\partial_{\mathbf{U}}\mathbf{G}^{(d)}$, and $\mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}}$ is the Roe-type diffusion matrix (arithmetic averages are used). The matrices $\mathbf{R}^{(d)}_{i_d+\frac{1}{2}}$ and $\mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}}$

for shallow water equations can be found in [117]. Then, we use $\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} = \frac{1}{M}\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$, $\forall m$, and these are positive-definite.

This results in a first order accurate ES flux. Let us derive the second order accurate ES flux given by eq. (2.72). As in [118], we express $\mathbf{D}^{(d)}_{i_d+\frac{1}{2}} \langle\langle \mathbf{V} \rangle\rangle_{i_d+\frac{1}{2}} = \mathbf{R}^{(d)}_{i_d+\frac{1}{2}} \mathbf{\Lambda}^{(d)}_{i_d+\frac{1}{2}} \left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}}$ where $\left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}} = \mathbf{B}^{(d)}_{i_d+\frac{1}{2}} \mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{1}{2}}$. Here, $\mathbf{B}^{(d)}_{i_d+\frac{1}{2}}$ is a positive diagonal matrix. Now, consider the minmod limiter

$$\mu(A,B) = \begin{cases} \mathrm{s}\ min(|A|,|B|) & \text{if s} = sign(A) = sign(B) \\ 0 & \text{otherwise} \end{cases} \tag{2.85}$$

Then, the reconstruction

$$\left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}} = \mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{1}{2}} - \frac{1}{2}\left(\mu\left(\mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{1}{2}}, \mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{3}{2}}\right) + \mu\left(\mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d-\frac{1}{2}}, \mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{1}{2}}\right)\right) \tag{2.86}$$

results in a second order accurate ES flux. Since $\mathbf{B}^{(d)}_{i_d+\frac{1}{2}}$ is a positive diagonal matrix, the sign property

$$sign\left(\left\langle\left\langle \widetilde{\mathbf{W}} \right\rangle\right\rangle_{i_d+\frac{1}{2}}\right) = sign\left(\mathbf{R}^{(d)^T}_{i_d+\frac{1}{2}} [[\mathbf{V}]]_{i_d+\frac{1}{2}}\right) \tag{2.87}$$

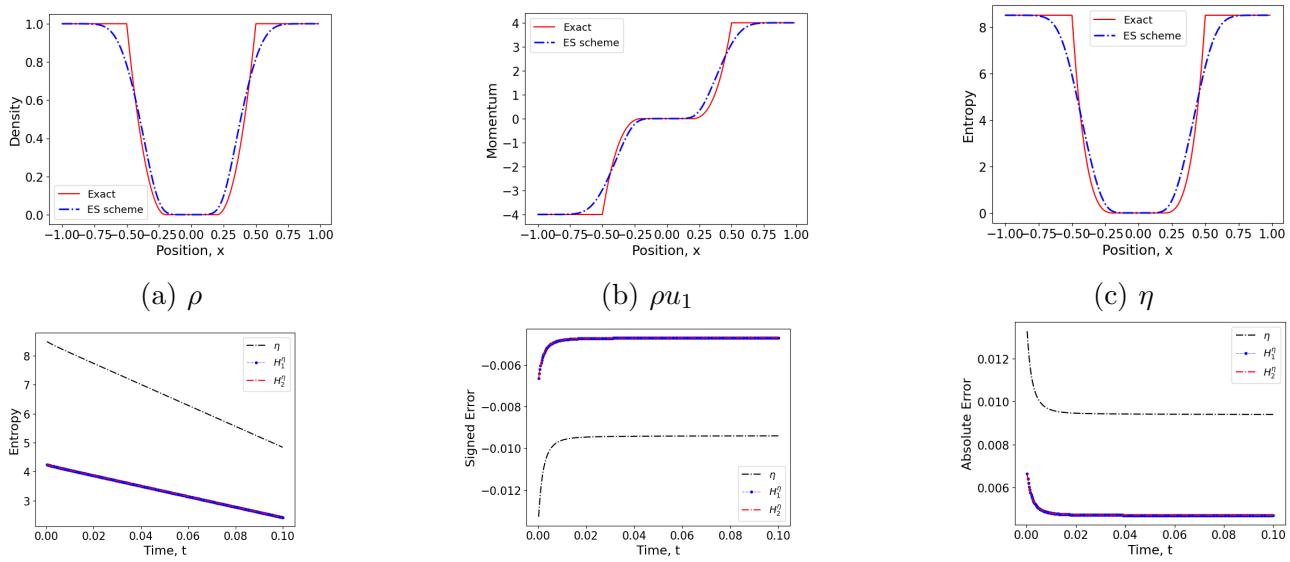
holds true, and the entropy stability is maintained. For vector-kinetic entropy stability, we use $\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}} \langle\langle \mathbf{V} \rangle\rangle_{i_d+\frac{1}{2}} = \frac{1}{M}\mathbf{D}^{(d)}_{i_d+\frac{1}{2}} \langle\langle \mathbf{V} \rangle\rangle_{i_d+\frac{1}{2}}$, $\forall m$.

It may be noted that we have derived the EC fluxes for vector-kinetic model from the vector-kinetic framework. Unlike this, we obtained the ES fluxes for vector-kinetic model based on the diffusion matrices commonly used in literature for macroscopic model. This is because the only requirement for entropy stability is positive-definiteness of $\mathbf{D}^{(d)}_{m_{i_d+\frac{1}{2}}}$, and we achieve this simply by employing the robust $\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$ used for macroscopic model.

2.8.2.1 1D expansion problem

This test case is taken from [117]. The domain of the problem is $[-1, 1)$, and it is discretised using 128 uniform cells. The initial condition is,

$$\rho(x_1, 0) = 1,\ u_1(x_1, 0) = \begin{cases} -4 & \text{if } x_1 < 0 \\ 4 & \text{if } x_1 \geq 0 \end{cases} \tag{2.88}$$



(a) ρ (b) ρu_1 (c) η

(d) Entropy functions (e) Signed error (f) Absolute error

Figure 2.5: SW 1D expansion problem at $T = 0.1$ using first order ES scheme with $C = 0.1$ and $Nx = 128$

Since the density can become very small, non-robust schemes will crash due to the in-ability to maintain positivity of density. Both entropy conserving and second order entropy stable schemes do not maintain the positivity. Hence, we utilise the first order entropy stable flux for vector-kinetic model to obtain the numerical results at $T = 0.1$. The boundary values are kept fixed throughout the computation, and a very low CFL of $C = 0.1$ is used for robustness.

It can be seen from fig. 2.5a that the density remains non-negative. Further, the numerical solutions of density, momentum and entropy match well with the exact solution as shown in figs. 2.5a to 2.5c. Figures 2.5d to 2.5f show entropy functions, their signed and absolute errors over time (for both macroscopic and vector-kinetic entropies). Since Δx is of $O(10^{-2})$, one would expect an absolute error of $O(10^{-2})$ due to the usage of first order entropy stable flux. In fig. 2.5f, we observe a better absolute error of $O(10^{-3})$ in vector-kinetic entropies. Macroscopic entropy which is the sum of vector-kinetic entropies has an absolute error of $O(10^{-2})$. The negative signed errors in fig. 2.5e indicate the global dissipation of macroscopic and vector-kinetic entropies. This can also be seen in fig. 2.5d from the decrease in global macroscopic and vector-kinetic entropies over time. It may be noted that the magnitudes of signed and absolute errors of all entropies in figs. 2.5e and 2.5f are same. This indicates that the first order entropy stable fluxes are dissipating the entropies at almost all spatial points, and not just globally.

2.8.2.2 1D dam break problem

This test case is also from [117]. The domain of the problem is $[-1, 1)$, and it is discretised using 128 uniform cells. The initial condition is,

$$\rho(x_1, 0) = \begin{cases} 15 & \text{if } x_1 < 0 \\ 1 & \text{if } x_1 \geq 0 \end{cases}, \; u_1(x_1, 0) = 0. \tag{2.89}$$

The numerical results obtained using first and second order entropy stable schemes at $T = 0.15$ are shown in figs. 2.6 and 2.7 respectively. The second order entropy stable reconstruction need not produce monotone solutions near discontinuities. Hence, a minmod flux limiter (that combines first and second order entropy stable fluxes) is employed to produce monotone solution near discontinuities. The boundary values are kept fixed throughout the computation, and a CFL of $C = 0.4$ is used.

It can be seen that both first and second order (with minmod limiter) schemes capture the solution profile reasonably well. A positive signed error for H_1^η in figs. 2.6e and 2.7e indicates that the numerical diffusion added for the flux corresponding to H_1^η is not sufficient to account for the entropy dissipation across discontinuities. This is because we have added equal weights of robust $\mathbf{D}^{(d)}_{i_d+\frac{1}{2}}$ to each of the vector-kinetic entropies, irrespective of their entropy dissipation requirements. Nevertheless, the error in macroscopic entropy which is obtained as the sum of vector-kinetic entropies is still negative (indicating entropy dissipation).

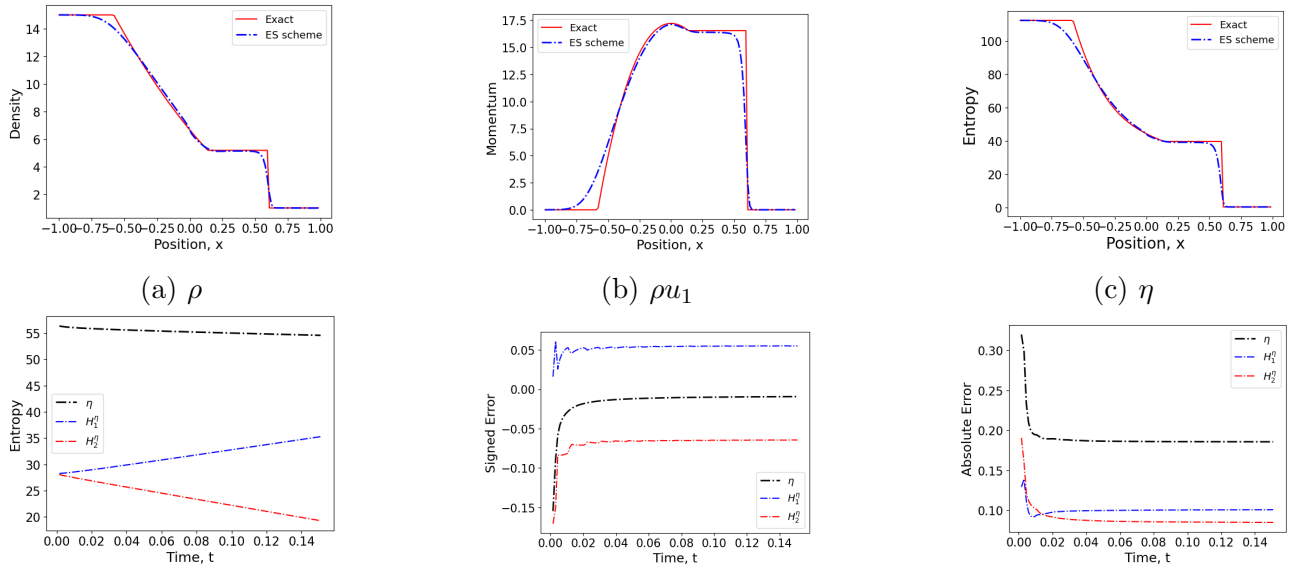
2.8.2.3 2D periodic flow

This test case is taken from the literature on asymptotic preserving schemes [177]. In order to be useful in our context, we have taken the value of asymptotic parameter to be 1. The domain of the problem is $[0, 1) \times [0, 1)$, and it is discretised using 256×256 uniform cells. The initial condition shown in fig. 2.8a is given by,

$$\rho(x_1, x_2, 0) = 1 + sin^2\left(2\pi\left(x_1 + x_2\right)\right) \tag{2.90}$$

$$u_1(x_1, x_2, 0) = u_2(x_1, x_2, 0) = sin\left(2\pi\left(x_1 - x_2\right)\right) \tag{2.91}$$

The numerical results obtained using entropy conserving scheme at $T = 0.1$ are shown in fig. 2.8b. Periodic boundary conditions are employed, and a CFL of C=0.5 is used. It can be seen from fig. 2.8c that the macroscopic and vector-kinetic entropy functions remain almost constant over time. From figs. 2.8d and 2.8e, we observe absolute and signed errors of $O(10^{-3})$ and $O(10^{-10})$ respectively. This huge difference implies that there are spatial cancellations between positive and negative errors. This may be due to the symmetric nature of periodic



(a) ρ (b) ρu_1 (c) η

(d) Entropy functions (e) Signed errors (f) Absolute errors

Figure 2.6: SW 1D dambreak problem at $T = 0.15$ using first order ES scheme with $C = 0.4$ and $Nx = 128$

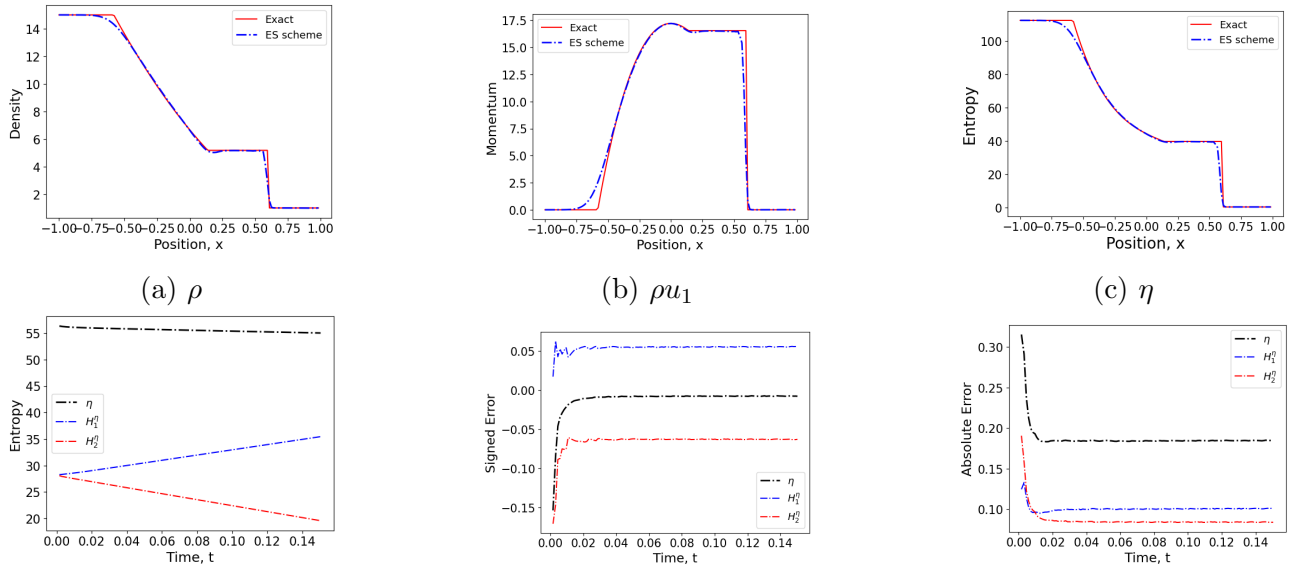
profile. Nevertheless, there is global dissipation of both macroscopic and vector-kinetic entropies as indicated by the negative errors in fig. 2.8d. Order of convergence studies show that the accuracy attained is more than second order, and the results are shown in table 2.4. The reference solution for convergence studies is the numerical solution with refined grid of 512×512.

N	Δx	$\|\|\rho\|\|_{L_2}$	$O(\|\|\rho\|\|)$	$\|\|\rho u_1\|\|_{L_2}$	$O(\|\|\rho u_1\|\|)$	$\|\|\rho u_2\|\|_{L_2}$	$O(\|\|\rho u_2\|\|)$
32	0.03125	0.00162	-	0.00255	-	0.00255	-
64	0.015625	0.000378	2.10	0.000362	2.82	0.000362	2.82
128	0.0078125	5.64×10^{-5}	2.74	5.54×10^{-5}	2.71	5.54×10^{-5}	2.71
256	0.00390625	7.62×10^{-6}	2.89	7.33×10^{-6}	2.92	7.33×10^{-6}	2.92

Table 2.4: EOC for 2D periodic flow at $T = 0.1$ using EC scheme with $C = 0.5$

2.8.2.4 2D Travelling vortex

This test case is also taken from the literature on asymptotic preserving schemes [177]. We have taken the value of asymptotic parameter to be 0.8, so that it will be useful in our context. The domain of the problem is $[0, 1) \times [0, 1)$, and it is discretised using 256×256 uniform cells.



(a) ρ (b) ρu_1 (c) η

(d) Entropy functions (e) Signed errors (f) Absolute errors

Figure 2.7: SW 1D dambreak problem at $T = 0.15$ using second order ES scheme (using minmod limiter) with $C = 0.4$ and $Nx = 128$

The initial condition shown in fig. 2.9a is given by,

$$\rho\left(x_1, x_2, 0\right) = 110 + \left(0.64\left(\frac{1.5}{4\pi}\right)^2\right) Drc\left(x_1, x_2\right)\left(k\left(rc\right) - k\left(\pi\right)\right) \tag{2.92}$$

$$u_1\left(x_1, x_2, 0\right) = 0.6 + 1.5\left(1 + cos\left(rc\left(x_1, x_2\right)\right)\right) Drc\left(x_1, x_2\right)\left(0.5 - x_2\right) \tag{2.93}$$

$$u_2\left(x_1, x_2, 0\right) = 0 + 1.5\left(1 + cos\left(rc\left(x_1, x_2\right)\right)\right) Drc\left(x_1, x_2\right)\left(x_1 - 0.5\right) \tag{2.94}$$

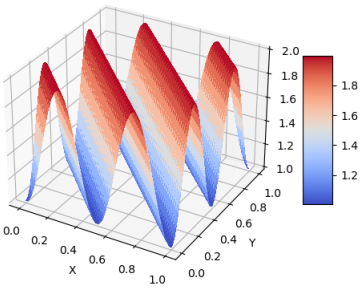
with

$$k\left(q\right) = 2cos\left(q\right) + 2q\ sin\left(q\right) + \frac{1}{8}cos\left(2q\right) + \frac{1}{4}q\ sin\left(2q\right) + \frac{3}{4}q^2 \tag{2.95}$$

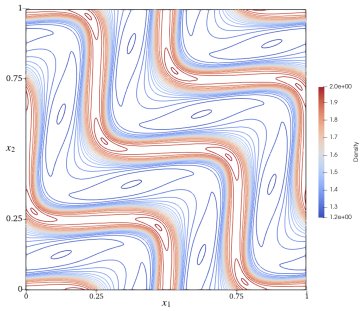
$$rc\left(x_1, x_2\right) = 4\pi\left(\left(x_1 - 0.5\right)^2 + \left(x_2 - 0.5\right)^2\right)^{\frac{1}{2}} \tag{2.96}$$

$$Drc\left(x_1, x_2\right) = \begin{cases} 1 & \text{if } rc\left(x_1, x_2\right) < \pi \\ 0 & \text{otherwise} \end{cases} \tag{2.97}$$

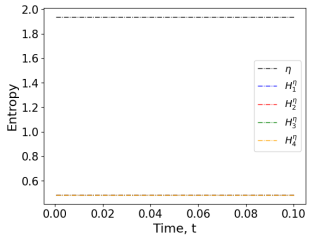
The second order entropy conserving and entropy stable schemes do not distort the structure of vortex, while the first order entropy stable scheme does. We present the numerical results obtained using second order entropy conserving scheme at $T = 0.1$ as shown in fig. 2.9b. Periodic boundary conditions are employed, and a CFL of C=0.5 is used.



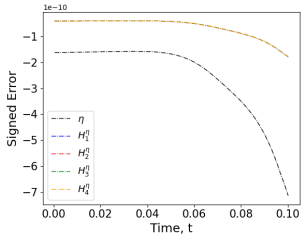
(a) Density at $T = 0$



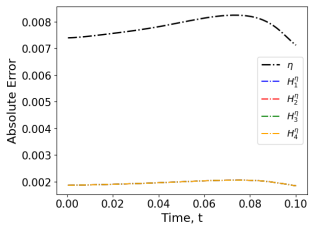
(b) Density contours at $T = 0.1$



(c) Entropy functions



(d) Signed errors



(e) Absolute errors

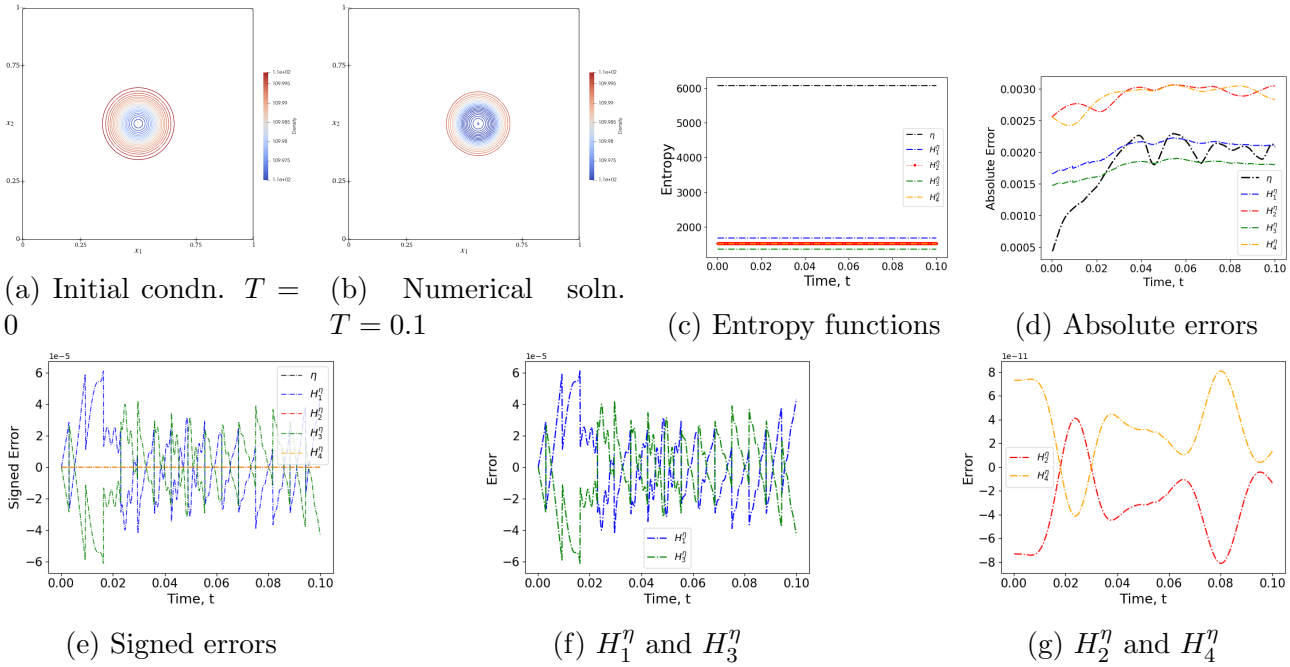
Figure 2.8: SW 2D periodic flow at $T = 0.1$ using EC scheme with $C = 0.5$ and $Nx, Ny = 256$ (Blue, red and green lines are beneath the yellow line)

From fig. 2.9d, we observe that the absolute errors of macroscopic and vector-kinetic entropies are of $O(10^{-3})$. On the other hand, the signed errors in H_2^η and H_4^η are of $O(10^{-11})$ (fig. 2.9g), while those in H_1^η and H_3^η are of $O(10^{-5})$ (fig. 2.9f). Moreover, the signed error profiles of vector-kinetic entropies are symmetric resulting in a much lower signed error of $O(10^{-14})$ for η (not shown in plot). However, these symmetries in signed errors must be located at different spatial points. If they were located at the same spatial points, then we would observe a much lower absolute error in macroscopic entropy, unlike $O(10^{-3})$ in fig. 2.9d.

Order of convergence studies are shown in table 2.5. It is seen that the accuracy attained is more than second order for ρu_1 and ρu_2. For ρ, the required order of accuracy is observed in coarser mesh rather than in fine mesh, and this matches the conclusion made in [263] where the analyses concerning types of vortices (based on their regularity) and their usage for validation of orders of accuracy of numerical methods are discussed.

N	Δx	$\|\|\rho\|\|_{L_2}$	$O(\|\|\rho\|\|)$	$\|\|\rho u_1\|\|_{L_2}$	$O(\|\|\rho u_1\|\|)$	$\|\|\rho u_2\|\|_{L_2}$	$O(\|\|\rho u_2\|\|)$
32	0.03125	0.000156	-	0.00339	-	0.00709	-
64	0.015625	4.39×10^{-5}	1.83	0.000505	2.75	0.00105	2.75
128	0.0078125	2.033×10^{-5}	1.11	0.000105	2.26	0.000174	2.60

Table 2.5: EOC for 2D travelling vortex at $T = 0.1$ using EC scheme with $C = 0.5$



(a) Initial condn. $T = 0$ (b) Numerical soln. $T = 0.1$ (c) Entropy functions (d) Absolute errors

(e) Signed errors (f) H_1^η and H_3^η (g) H_2^η and H_4^η

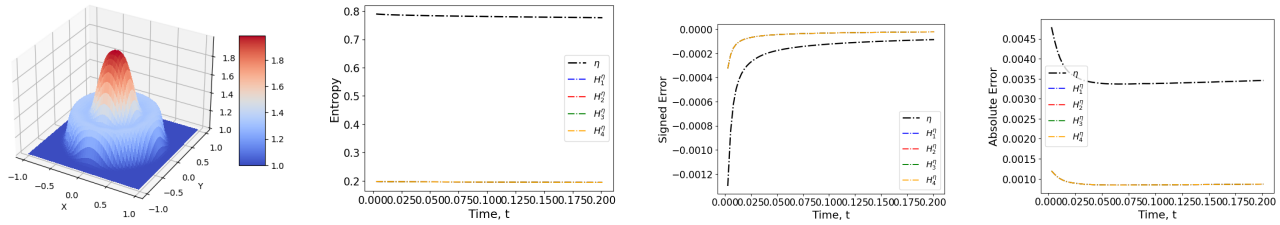
Figure 2.9: SW 2D travelling vortex at $T = 0.1$ using EC scheme with $C = 0.5$ and $Nx, Ny = 256$

2.8.2.5 2D cylindrical dambreak

This test case is taken from [117]. The domain of the problem is $[-1, 1) \times [-1, 1)$, and it is discretised using 100×100 uniform cells. The initial condition is given by,

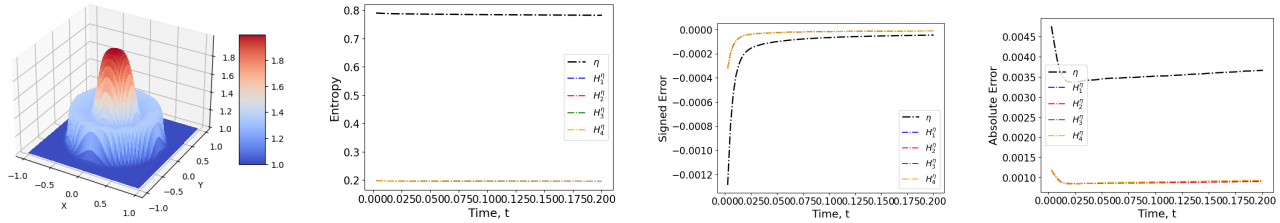
$$\rho\left(x_1, x_2, 0\right) = \begin{cases} 2 & \text{if } \left(x_1^2 + x_2^2\right)^{\frac{1}{2}} < 0.5 \\ 1 & \text{otherwise} \end{cases}, \; u_1\left(x_1, x_2, 0\right) = u_2\left(x_1, x_2, 0\right) = 0 \tag{2.98}$$

The numerical results of first and second order (with minmod limiter) entropy stable schemes at $T = 0.2$ are shown in figs. 2.10a and 2.11a respectively. A CFL of $C = 0.4$ is used, and periodic boundary conditions are employed. From figs. 2.10d and 2.11d, we observe that the



(a) Density at $T = 0.2$ (b) Entropy functions (c) Signed errors (d) Absolute errors

Figure 2.10: SW 2D cylindrical dam-break at $T = 0.2$ using first order ES scheme with $C = 0.4$ and $Nx, Ny = 100$



(a) Density at $T = 0.2$ (b) Entropy functions (c) Signed errors (d) Absolute errors

Figure 2.11: SW 2D cylindrical dambreak at $T = 0.2$ using second order ES scheme (using minmod limiter) with $C = 0.4$ and $Nx, Ny = 100$

absolute errors in entropies are of $O(10^{-3})$. Further, from figs. 2.10c and 2.11c, we observe that the signed errors in entropies are of $O(10^{-4})$. The negative signed errors indicate that there is global dissipation of entropy.

2.9 Summary and Conclusions

The following are the major highlights of the chapter.

- We provided a modification to the vector-BGK model, and this allows us to obtain entropy flux potentials that are required in the consistent definition of interface numerical entropy fluxes. Lemmas 2.1 and 2.2 are essential in obtaining the entropy flux potentials.

- We showed in theorems 2.1 and 2.2 that the moment of entropy conserving/stable schemes for vector-kinetic model results in entropy conserving/stable schemes for macroscopic model. Lemma 2.1 plays a crucial role by rendering the linearities in the involved moments.

- In the numerical tests of scalar smooth problems, we employed our entropy conserving scheme and observed that the macroscopic and all the vector-kinetic entropies involved

are conserved (up to absolute error). We also used signed error to observe global entropy dissipation/production due to higher order terms for which conservation does not apply.

- For shallow water equations, we derived an entropy conserving flux for vector-kinetic model by considering arithmetic averages of primitive variables. We used this entropy conserving scheme on smooth problems such as periodic flow and travelling vortex. In both cases, we observed the conservation of macroscopic and vector-kinetic entropies.

- We considered the 1D expansion problem where non-positivity of density can easily occur in non-robust schemes. For this, we employed the first order entropy stable scheme for vector-kinetic model and observed that the macroscopic and all vector-kinetic entropies involved are dissipative in nature. We also do not encounter non-positivity.

- In the non-smooth category, we considered scalar non-linear inviscid Burgers' test, 1D and 2D cylindrical dam-break problems. The second order entropy stable scheme employed for scalar case dissipates macroscopic and all vector-kinetic entropies. For the shallow water case, we employed the first and second order entropy stable schemes for vector-kinetic model. In 1D dam-break problem, we observed that some of the vector-kinetic entropies are not really dissipative, as their dissipation matrices are not built based on the dissipation requirements near discontinuities. Further research is required on the choice of appropriate robust dissipation matrices for vector-kinetic model.

Thus, the entropy preserving scheme developed in this chapter preserves both vector-kinetic and macroscopic entropy functions. It is interesting to observe that the entropic numerical solutions of macroscopic model do not experience a notable difference when two different routes (via vector-kinetic and macroscopic) are taken.
If the proposed entropy conserving scheme for vector-kinetic model is applied to the Euler's system, the vector-kinetic entropy conserving condition in eq. (2.40) can be satisfied analogous to the ways available in literature to satisfy entropy conserving condition for macroscopic model in eq. (2.8). One can derive the fluxes by utilising an elegant and non-costly route available in literature (for instance, by defining primitive variables, substituting for entropy variables and entropy flux potentials in terms of these primitive variables into eq. (2.40), and equating the coefficients of the jumps in the primitive variables, as introduced in [158] for satisfaction of the condition in eq. (2.8)), and this is a work in progress. It is expected that the moment of such entropy conserving flux functions for vector-kinetic model derived using a particular method (say, [158]) will be an entropy conserving flux function for macroscopic model derived using the same method ([158]).

2.10 Appendix: Choice of constants $a_m, b_m^{(d)}$

We know that the moment of eq. (2.24) becomes the given hyperbolic system in eq. (2.1), if the constants $a_m, b_m^{(d)}$ in eq. (2.25) satisfy the moment constraints in eqs. (2.26) and (2.27). We also know that, if the convex entropy function for vector-kinetic model (*eq.* (2.28)) is used, then the moment of eq. (2.30) becomes eq. (2.3) with equality. Further, positivity of eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$ is an important requirement for obtaining the entropy flux potentials and the results of theorems 2.1 and 2.2. Therefore, in order for the formulation to hold, the constants $a_m, b_m^{(d)}$ are required to satisfy eqs. (2.26) and (2.27) along with the positivity of eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$.
For one dimensional hyperbolic systems, we consider two discrete velocities, *i.e.,* $M = 2$. Let

$$a_1 = \frac{1}{2}, a_2 = \frac{1}{2} \tag{2.99}$$

$$b_1^{(1)} = \frac{1}{2\lambda}, b_2^{(1)} = -\frac{1}{2\lambda} \tag{2.100}$$

If $v_1^{(1)} = \lambda$ and $v_2^{(1)} = -\lambda$, then the moment constraints in eqs. (2.26) and (2.27) are satisfied. Further,

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_1\right) = \text{eig}\left(\frac{1}{2}\mathbf{I} + \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right) \tag{2.101}$$

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_2\right) = \text{eig}\left(\frac{1}{2}\mathbf{I} - \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right) \tag{2.102}$$

Thus, eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$ are $\frac{1}{2} \pm \frac{1}{2\lambda}\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right)$. Therefore, for positivity, we require $\lambda > \sup\left(\left|\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right)\right|\right)$. The supremum is taken over all grid points/cells in the computational domain.
For two dimensional systems, we consider four discrete velocities, *i.e.,* $M = 4$. Let

$$a_1 = \frac{1}{4}, a_2 = \frac{1}{4}, a_3 = \frac{1}{4}, a_4 = \frac{1}{4} \tag{2.103}$$

$$b_1^{(1)} = \frac{1}{2\lambda}, b_2^{(1)} = 0, b_3^{(1)} = -\frac{1}{2\lambda}, b_4^{(1)} = 0 \tag{2.104}$$

$$b_1^{(2)} = 0, b_2^{(2)} = \frac{1}{2\lambda}, b_3^{(2)} = 0, b_4^{(2)} = -\frac{1}{2\lambda} \tag{2.105}$$

If the following holds,

$$v_1^{(1)} = \lambda, v_2^{(1)} = 0, v_3^{(1)} = -\lambda, v_4^{(1)} = 0 \tag{2.106}$$

$$v_1^{(2)} = 0, v_2^{(2)} = \lambda, v_3^{(2)} = 0, v_4^{(2)} = -\lambda \tag{2.107}$$

then the moment constraints in eqs. (2.26) and (2.27) are satisfied. Further,

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_1\right) = \text{eig}\left(\frac{1}{4}\mathbf{I} + \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right) \tag{2.108}$$

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_2\right) = \text{eig}\left(\frac{1}{4}\mathbf{I} + \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(2)}\right) \tag{2.109}$$

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_3\right) = \text{eig}\left(\frac{1}{4}\mathbf{I} - \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right) \tag{2.110}$$

$$\text{eig}\left(\partial_{\mathbf{U}}\mathbf{F}_4\right) = \text{eig}\left(\frac{1}{4}\mathbf{I} - \frac{1}{2\lambda}\partial_{\mathbf{U}}\mathbf{G}^{(2)}\right) \tag{2.111}$$

Thus, eigenvalues of $\partial_{\mathbf{U}}\mathbf{F}_m$ are $\frac{1}{4} \pm \frac{1}{2\lambda}\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right)$ and $\frac{1}{4} \pm \frac{1}{2\lambda}\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(2)}\right)$. Therefore, for positivity, we require $\lambda > 2 \sup\left(\left|\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(1)}\right)\right|, \left|\text{eig}\left(\partial_{\mathbf{U}}\mathbf{G}^{(2)}\right)\right|\right)$. The supremum is taken over all grid points/cells in the domain.

Chapter 3

High order asymptotic preserving scheme for diffusive-scaled kinetic equations

Diffusive scaled linear kinetic equations appear in various applications, and they contain a small parameter ϵ that forces a severe time step restriction for standard explicit schemes. Asymptotic preserving (AP) schemes are those schemes that attain asymptotic consistency and uniform stability for all values of ϵ, with the time step restriction being independent of ϵ. In this chapter, we develop high order AP scheme for such diffusive scaled kinetic equations with both well-prepared and non-well-prepared initial conditions by employing IMEX-RK time integrators such as CK-ARS and A types. This framework is also extended to a different collision model involving advection-diffusion asymptotics, and the AP property is proved formally. A further extension of our framework to inflow boundaries has been made, and the AP property is verified. The temporal and spatial orders of accuracy of our framework are numerically validated in different regimes of ϵ, for all the models. The qualitative results for diffusion asymptotics, and equilibrium and non-equilibrium inflow boundaries are also presented.

3.1 Introduction

This chapter is concerned with the numerical approximation of linear collisional kinetic transport equations in a diffusive scaling. Such models are widely used in applications such as rarefied gas dynamics, neutron transport, and radiative transfer. Due to the presence of a small parameter ϵ (which is the normalized mean free path of the particles), standard explicit schemes suffer from a severe restriction on the numerical parameters so that they experience

extremely high computational cost as $\epsilon \to 0$. In the last decades, the so-called Asymptotic-Preserving (AP) schemes have been proposed to make the numerical passage between the micro and macro scale [180, 181, 169] possible. Indeed, these AP schemes are uniformly stable (ie the numerical parameters can be chosen independent of ϵ and degenerate when $\epsilon \to 0$ to a scheme which is consistent with the asymptotic model. These schemes efficiently deal with multiscale phenomena and are a viable alternative to domain decomposition approaches.

In this chapter, we are concerned with high order in time AP scheme for collisional kinetic equations in the diffusive scaling, possibly involving boundary conditions. Several works can be found in the literature on this topic using splitting method, odd-even or micro-macro decompositions (see [172, 167, 169, 180, 181, 182, 182, 222, 96, 37, 189, 192, 200, 198, 240]). Our work is based on a micro-macro decomposition as introduced in [200] where the unknown f of the stiff kinetic equation is split into an equilibrium part ρ and a remainder g. A micro-macro model (equivalent to the original kinetic one) satisfied by ρ and g can be derived. This micro-macro strategy turns out to be the starting point of several numerical approximation in phase space (using staggered grid [200], particles method [75, 74, 96], Discontinuous Galerkin method [165, 241, 242, 240] or low rank approximation [97, 104]). Regarding the time discretization, a suitable first order semi-implicit time discretization of the micro-macro model has been initially proposed in [200] for which the AP property is ensured for general initial conditions. High order extensions are usually based on IMEX Runge-Kutta methods [52, 12, 234, 40, 38, 95] which turns out to be a useful framework to derive high order AP schemes for stiff kinetic equations under a fluid scaling [91, 92, 93, 1]. but also under a diffusive scaling [164, 165]; in these works however, even if the proposed numerical schemes enjoy the AP property and are high order, the asymptotic diffusion equation is solved explicitly, leading to a stringent parabolic CFL condition for small ϵ. This drawback is overcome using a suitable modification to the semi-implicit time discretization of [200] which results in a first order implicit scheme for the asymptotic diffusion equation that is devoid of the parabolic CFL condition (see [198, 75, 74]). High order versions have been derived and analyzed in [37, 241, 240, 242], leading to a numerical scheme which is asymptotically free from the usual restrictive parabolic condition.

In this chapter, a family of high order IMEX numerical schemes is proposed for linear collisional kinetic equations in the diffusive scaling. According to the collision operator, the asymptotic model can be a pure diffusion equation or an advection-diffusion equation as in [165]. The numerical schemes presented in this chapter are high order, uniformly stable with respect to ϵ and degenerate when $\epsilon \to 0$ to a high order implicit scheme for the pure asymptotic diffusion equation or to a high order IMEX scheme of the asymptotic advection-diffusion equation. From the first order semi-implicit AP numerical scheme [75], the family of high order schemes

proposed in this work is obtained using globally stiffly accurate high order IMEX Runge-Kutta methods, namely type A and type CK [92, 165]. In particular, we discuss the AP property according to the considered class (type A or CK) and according to the initial condition (well-prepared or not). For the two cases (diffusion and advection-diffusion), the AP property is proved with general initial condition (referred as strong AP property in the literature).

This work bears similarities with the series of works [241, 240, 242] in which high order AP schemes are derived and analyzed for linear collisional kinetic equations in the diffusive scaling. However, there are some differences. Indeed, in [241, 240], an artificial weighted diffusion is added and subtracted to get an implicit scheme for the parabolic term, in the spirit of [37]; but, as mentioned in [242], this weighted diffusion term may depend on ϵ and/or the numerical parameters, and has to be chosen according to the considered problem which can affect the performance of the numerical simulations. As in [242], the numerical schemes proposed in this work directly solve the micro-macro system. Another difference lies in the choice of time integrator (type A, in particular) that allows the scheme to be asymptotic preserving when the initial data is not well-prepared, without requiring the reduction of initial time steps (the numerical methods proposed in [241, 242] require the time step to be Δt^p for p^{th} order accurate scheme in the initial few steps). Further, we consider here advection-diffusion problems and problems involving incoming boundary conditions; our family of high order scheme can be easily extended to the half moment micro-macro decomposition introduced in [199] to naturally incorporate incoming boundary conditions, even when non well prepared boundary conditions are considered.

Lastly, we address the space discretization in order to get a fully high order solver of the stiff kinetic equation. Let us mention Discontinuous Galerkin methods developed in [165, 241, 242, 240] for similar purposes. Here we focus on high order classical finite difference methods for the space approximation which only involve a discrete diffusion term to invert. Staggered or non-staggered strategies are discussed.

The chapter is organized as follows. First in Section 3.2, the kinetic and asymptotic diffusion models are introduced. Then in Section 3.3, high order time integrators (using globally stiffly accurate IMEX Runge-Kutta temporal discretization) are proposed, and their AP property in the diffusive limit is addressed in Section 3.4. Section 3.5 is devoted to the space approximation. In Section 3.6, we discuss some extensions to other collision operators and to half moments. In Section 3.7, numerical results are presented, illustrating high order accuracy and the main properties of the schemes.

3.2 Kinetic equation, diffusion limit and micro-macro decomposition

In this section, we introduce the kinetic model in diffusive scaling, and recall the asymptotic limit. Then, the micro-macro decomposition is performed to derive the micro-macro model which serves as a basis for numerical developments.

3.2.1 Linear kinetic equation with diffusive scaling

Let $\Omega \subset \mathbb{R}^d$ be the position space and $V \subseteq \mathbb{R}^d$ be the velocity space with measure $d\mu(v)$. We consider the linear kinetic equation with diffusive scaling,

$$\partial_t f + \frac{1}{\epsilon} v \cdot \nabla_x f = \frac{1}{\epsilon^2} L f, \quad (t, x, v) \in \mathbb{R}^+ \times \Omega \times V \tag{3.1}$$

where $f(t, x, v) \in \mathbb{R}$ is the distribution function (depending on time $t \in \mathbb{R}^+$, space $x \in \Omega \subset \mathbb{R}^d$ and velocity $v \in V \subset \mathbb{R}^d$) and $\epsilon > 0$ measures the dimensionless mean free path of particles or the inverse of relaxation time. We consider the initial condition,

$$f(0, x, v) = f^{\mathsf{init}}(x, v), \quad (x, v) \in \Omega \times V \tag{3.2}$$

and boundary conditions are imposed in space. In this work, we will consider periodic boundary conditions or inflow boundary conditions. The linear collision operator L in eq. (3.1) acts only on the velocity dependence of f, and it relaxes the particles to an equilibrium $M(v)$ which is positive and even. We denote for all velocity dependent distribution functions h,

$$\langle h \rangle_V = \frac{\int_V h(v)\, d\mu}{\int_V M(v)\, d\mu}. \tag{3.3}$$

In particular, we obtain $\langle M \rangle_V = 1$ and $\langle vM \rangle_V = 0$. Further, the operator L is non-positive and self-adjoint in $L^2\left(V, M^{-1} d\mu\right)$, with the following null space and range:

$$\mathcal{N}(L) = \{f : f \in \mathsf{Span}\,(M)\}, \;\; \mathcal{R}(L) = (\mathcal{N}(L))^{\perp} = \{f : \langle f \rangle_V = 0\}. \tag{3.4}$$

Therefore, L is invertible on $\mathcal{R}(L)$ and we denote its pseudo-inverse by L^{-1}.

3.2.2 Diffusion limit

In the limit $\epsilon \to 0$, it is seen from eq. (3.1) that $f \to f_0$ where f_0 belongs to $\mathcal{N}(L)$. Thus, $f_0 = \rho(t, x) M$ where f_0 solves $L f_0 = 0$ and where the limiting density ρ is the solution of the

asymptotic diffusion equation. To derive the diffusion equation, a Chapman-Enskog expansion has to be performed to get $f = f_0 + \epsilon L^{-1}(vM) \cdot \nabla_x \rho + \mathcal{O}(\epsilon^2)$. Integrating with respect to the velocity variable enables to get the diffusion limit

$$\partial_t \rho - \nabla_x \cdot (\kappa \nabla_x \rho) = 0 \text{ with } \kappa = -\left\langle v \otimes L^{-1}(vM) \right\rangle_V > 0. \tag{3.5}$$

3.2.3 Micro-macro decomposition

In this part, we derive a micro-macro model which is equivalent to (3.1), and this is the model that will be discretized in the next sections. First, we consider the standard micro-macro decomposition of the unknown f [200, 198],

$$f = \rho M + g, \quad \text{with} \quad \rho(t,x) = \langle f \rangle_V \text{ and } \langle g \rangle_V = 0. \tag{3.6}$$

We introduce the orthogonal projector Π in $L^2\left(V, M^{-1} d\mu\right)$ onto $\mathcal{N}(L)$: $\Pi h = \langle h \rangle_V M$, which will be useful to derive the micro-macro model. Substituting eq. (3.6) into eq. (3.1) and applying successively Π and $(I - \Pi)$ enables to get the micro-macro model satisfied by (ρ, g)

$$\partial_t \rho + \frac{1}{\epsilon} \nabla_x \cdot \langle vg \rangle_V = 0, \tag{3.7}$$

$$\partial_t g + \frac{1}{\epsilon} (I - \Pi)(v \cdot \nabla_x g) + \frac{1}{\epsilon} vM \cdot \nabla_x \rho = \frac{1}{\epsilon^2} Lg. \tag{3.8}$$

Initial conditions for macro and micro equations become

$$\rho(0,x) = \rho^{\mathsf{init}}(x) = \left\langle f^{\mathsf{init}}(x,\cdot) \right\rangle_V, \tag{3.9}$$

$$g(0,x,v) = g^{\mathsf{init}}(x,v) = f^{\mathsf{init}}(x,v) - \rho^{\mathsf{init}}(x) M(v), \tag{3.10}$$

whereas the boundary conditions for ρ and g become periodic if f is periodic. From the micro part (3.8), a Chapman-Enskog expansion of g can be performed to get

$$g = -\epsilon \left(\epsilon^2 \partial_t - L\right)^{-1} \Big((I - \Pi)(v \cdot \nabla_x g) + vM \cdot \nabla_x \rho \Big) = \epsilon L^{-1}(vM) \cdot \nabla_x \rho + \mathcal{O}(\epsilon^2),$$

under some suitable smoothness assumptions. Inserting this expression in eq. (3.7) leads to eq. (3.5) in the limit $\epsilon \to 0$.

3.3 Time integrators

In this part, we present the family of high order time integrators for the micro-macro model (3.7)-(3.8). We will keep the phase space variables continuous to ease the reading. We first recall the first order temporal scheme which leads to the implicit treatment of the asymptotic diffusion model before introducing the high order version.

3.3.1 First order accurate time integrator

Given ρ^n, g^n that approximate ρ, g at time $t = n\Delta t$, we obtain the solution ρ^{n+1}, g^{n+1} from the following time integration of eqs. (3.7) and (3.8) respectively. We use the following first order implicit-explicit (IMEX) strategy to attain the asymptotic preserving property

$$\rho^{n+1} = \rho^n - \frac{\Delta t}{\epsilon}\nabla_x \cdot \left\langle v g^{n+1} \right\rangle_V , \tag{3.11}$$

$$g^{n+1} = g^n - \frac{\Delta t}{\epsilon}\left(I - \Pi\right)\left(v \cdot \nabla_x g^n\right) - \frac{\Delta t}{\epsilon} v M \cdot \nabla_x \rho^{n+1} + \frac{\Delta t}{\epsilon^2} L g^{n+1}. \tag{3.12}$$

The implicit treatment of density gradient in micro equation (3.12) and fully implicit treatment of the macro equation enables us to get an implicit scheme for diffusion equation in the limit $\epsilon \to 0$.

Although the macro equation is treated in a fully implicit manner, ρ^{n+1} and g^{n+1} can be updated using eqs. (3.11) and (3.12) in an explicit manner. From eq. (3.12), we get

$$g^{n+1} = \left(\epsilon^2 I - \Delta t L\right)^{-1}\left(\epsilon^2 g^n - \epsilon \Delta t \left(I - \Pi\right)\left(v \cdot \nabla_x g^n\right) - \epsilon \Delta t v M \cdot \nabla_x \rho^{n+1}\right). \tag{3.13}$$

Inserting this in eq. (3.11), we obtain the following implicit scheme for the macro unknown

$$\rho^{n+1} = \rho^n - \Delta t \nabla_x \cdot \langle v \left(\epsilon^2 I - \Delta t L\right)^{-1}\left(\epsilon g^n - \Delta t \left(I - \Pi\right)\left(v \cdot \nabla_x g^n\right) - \Delta t v M \cdot \nabla_x \rho^{n+1}\right)\rangle_V,$$

or, expressing ρ^{n+1} as quantities at iteration n

$$\rho^{n+1} = \left(I - \Delta t^2 \nabla_x \cdot \left(\mathcal{D}_{\epsilon,\Delta t}\nabla_x\right)\right)^{-1}\left(\rho^n - \Delta t \nabla_x \cdot \left\langle v \left(\epsilon^2 I - \Delta t L\right)^{-1}\left(\epsilon g^n - \Delta t \left(I - \Pi\right)\left(v \cdot \nabla_x g^n\right)\right)\right\rangle_V\right)$$

with $\mathcal{D}_{\epsilon,\Delta t} = \langle v \otimes \left(\epsilon^2 I - \Delta t L\right)^{-1}\left(vM\right)\rangle_V$. Thanks to this time integrator, ρ^{n+1} can be updated by inverting a diffusion type operator. Following this, g^{n+1} can be found explicitly from the knowledge of ρ^{n+1}. This first order scheme introduced in [198, 75] is the basis of the high order scheme presented below.

3.3.2 High order accurate time integrators

Following previous works [92, 165, 37, 241], we will consider globally stiffly accurate (GSA) IMEX Runge-Kutta (RK) schemes to construct high order time integrators for the micro-macro model eqs. (3.7) and (3.8). An IMEX RK scheme is represented using the double Butcher tableau [52, 12]

$$\begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} \qquad \begin{array}{c|c} c & A \\ \hline & b^T \end{array} \tag{3.14}$$

where $\tilde{A} = (\tilde{a}_{ij})$ and $A = (a_{ij})$ are $s \times s$ matrices which correspond to the explicit and implicit parts of the scheme (A and $\tilde{A}$ respectively are lower triangular and strictly lower triangular matrices). The coefficients $\tilde{c}$ and c are given by $\tilde{c}_i = \sum_{j=1}^{i-1} \tilde{a}_{ij}$, $c_i = \sum_{j=1}^{i} a_{ij}$, and the vectors $\tilde{b} = (\tilde{b}_j)$ and $b = (b_j)$ give quadrature weights that combine the stages. For GSA IMEX RK scheme, we have

$$c_s = \tilde{c}_s = 1 \text{ and } a_{sj} = b_j, \tilde{a}_{sj} = \tilde{b}_j, \quad \forall j \in \{1, 2.., s\}. \tag{3.15}$$

An IMEX RK method is type A if the matrix A is invertible, and it is type CK if the first row of matrix A has zero entries and the square sub-matrix formed by excluding the first column and row of A is invertible. In the special case where the first column of A has zero entries, the scheme is said to be of type CK-ARS. The reader is referred to [92] for more details. In this work, we employ both type A and CK-ARS schemes.

The first order GSA IMEX RK scheme employed in eqs. (3.11) and (3.12) follows the type CK-ARS double Butcher tableau (known as ARS$(1, 1, 1)$),

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1 & 0 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 0 & 1 \\ \hline & 0 & 1 \end{array} \tag{3.16}$$

We now use the general IMEX RK scheme from (3.14) with GSA property eq. (3.15) for obtaining high order accurate time integration of macro and micro eqs. (3.7) and (3.8) respectively. We introduce the following notations in the presentation of our scheme.

$$\mathcal{T} h^{(k)} = (I - \Pi) \left(v \cdot \nabla_x h^{(k)} \right), \tag{3.17}$$

$$\mathcal{D}^{(j)}_{\epsilon, \Delta t} = \left\langle v \otimes \left(\epsilon^2 I - a_{jj} \Delta t L \right)^{-1} (vM) \right\rangle_V, \tag{3.18}$$

$$\mathcal{I}^{(j)}_{\epsilon, \Delta t} = \left(\epsilon^2 I - a_{jj} \Delta t L \right)^{-1}. \tag{3.19}$$

We will construct high order IMEX RK schemes following the first order guidelines (fully implicit treatment of macro equation, implicit treatment of density gradient and relaxation terms and explicit treatment of transport term in micro equation). Given ρ^n, g^n that approximate ρ, g at time $t = n\Delta t$, we obtain the internal RK stage values $\rho^{(j)}$ and $g^{(j)}$, $j = 1, \ldots, s$ as

$$\rho^{(j)} = \rho^n - \sum_{k=1}^{j} a_{jk} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(k)} \right\rangle_V , \tag{3.20}$$

$$g^{(j)} = g^n - \sum_{k=1}^{j-1} \tilde{a}_{jk} \frac{\Delta t}{\epsilon} \mathcal{T} g^{(k)} - \sum_{k=1}^{j} a_{jk} \frac{\Delta t}{\epsilon} v M \cdot \nabla_x \rho^{(k)} + \sum_{k=1}^{j} a_{jk} \frac{\Delta t}{\epsilon^2} L g^{(k)}, \tag{3.21}$$

where, as usual, the summation $\sum_{k=1}^{j-1}$ in the explicit term is zero for $j = 1$.
Although the expressions above are implicit, the stage values $\rho^{(1)}$, $g^{(1)}$ can be found in an explicit manner by using the known quantities ρ^n, g^n, and the stage values $\rho^{(j)}$, $g^{(j)}$, $\forall j \in \{2, 3, \ldots, s\}$ can be found explicitly from ρ^n, g^n and the previous stage values $\rho^{(l)}, g^{(l)}$, $\forall l \in \{1, 2, \ldots, j-1\}$. Indeed, proceeding similarly as for the first order scheme, we get the following expression of $g^{(j)}, j = 1, \ldots, s$ from eq. (3.21),

$$g^{(j)} = \mathcal{I}^{(j)}_{\epsilon,\Delta t} \left(\epsilon^2 g^n - \epsilon \sum_{k=1}^{j-1} \tilde{a}_{jk} \Delta t \mathcal{T} g^{(k)} - \epsilon \sum_{k=1}^{j} a_{jk} \Delta t v M \cdot \nabla_x \rho^{(k)} + \sum_{k=1}^{j-1} a_{jk} \Delta t L g^{(k)} \right). \tag{3.22}$$

Further, we write eq. (3.20) by splitting the summation on k as

$$\rho^{(j)} = \rho^n - \sum_{k=1}^{j-1} a_{jk} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(k)} \right\rangle_V - a_{jj} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(j)} \right\rangle_V ,$$

and inserting eq. (3.22) in the last term leads to the update of $\rho^{(j)}$ for $j = 1, \ldots, s$

$$\begin{aligned} \rho^{(j)} = \left(I - a_{jj}^2 \Delta t^2 \nabla_x \cdot \left(\mathcal{D}^{(j)}_{\epsilon,\Delta t} \nabla_x \right) \right)^{-1} \Bigg(& \rho^n - \sum_{k=1}^{j-1} a_{jk} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(k)} \right\rangle_V \\ & - a_{jj} \Delta t \nabla_x \cdot \Big\langle v \mathcal{I}^{(j)}_{\epsilon,\Delta t} \big(\epsilon g^n - \sum_{k=1}^{j-1} \tilde{a}_{jk} \Delta t \mathcal{T} g^{(k)} \\ & - \sum_{k=1}^{j-1} a_{jk} \Delta t v M \cdot \nabla_x \rho^{(k)} + \frac{1}{\epsilon} \sum_{k=1}^{j-1} a_{jk} \Delta t L g^{(k)} \big) \Big\rangle_V \Bigg), \end{aligned} \tag{3.23}$$

where the definition of $\mathcal{T}, \mathcal{D}^{(j)}_{\epsilon,\Delta t}$ and $\mathcal{I}^{(j)}_{\epsilon,\Delta t}$ are given by eqs. (3.17) to (3.19). After this reformulation, $\rho^{(j)}$ can be computed from (3.23) by inverting a linear elliptic type problem and following

this, $g^{(j)}$ can be found from eq. (3.22). The GSA property in eq. (3.15) guarantees that the solution at $t^{n+1} = (n+1)\Delta t$ is same as the last RK stage values, that is, $\rho^{n+1} = \rho^{(s)}$ and $g^{n+1} = g^{(s)}$.

Remark 3.1 *The IMEX strategy is similar to the one presented in [242] where the Schur complement is employed to make the scheme efficient from a computational point of view. Here, inserting $g^{(j)}$ in the update (3.23) of $\rho^{(j)}$ leads to a similar scheme.*
However, we present here the asymptotic preserving properties of both CK-ARS and type A time integrators and show that the type A time integators require neither well-prepared initial data nor the treatment of reducing initial time steps as in [242].

3.4 Asymptotic preserving property

In this section, we show that the time integrated scheme (3.23)-(3.22) becomes a consistent scheme for the diffusion equation (3.5) in the limit $\epsilon \to 0$. We will discuss the asymptotic preserving property for both CK-ARS and type A time integrators. First, we recall the definition of well-prepared initial data in our context.

Definition 3.1 (Well-prepared initial data) *The initial data $\rho(0,x)$ and $g(0,x,v)$ in eqs. (3.9) and (3.10) are said to be well-prepared if $g(0,x,v) = O(\epsilon)$.*

Lemma 3.1 *Assume that ϵ is sufficiently small. Let $\tilde{a}_{jk}$ and a_{jk} be the coefficients of the RK method (3.14) applied to the scheme (3.20)-(3.21). Then, the following holds:*

1. *CK-ARS case: If $g^n = O(\epsilon)$, then $g^{(1)} = g^n = O(\epsilon)$ and $g^{(j)} = \epsilon L^{-1}(vM)\cdot\nabla_x\rho^{(j)} + O(\epsilon^2)\,,\ \forall j \in \{2,\dots,s\}$.*
2. *Type A case: $g^{(j)} = \epsilon L^{-1}(vM)\cdot\nabla_x\rho^{(j)} + O(\epsilon^2)\,,\ \forall j \in \{1,\dots,s\}$.*

Proof: Let $j \in \{1,\dots,s\}$ such that $a_{jj} \neq 0$. Observe that the operator $\mathfrak{I}^{(j)}_{\epsilon,\Delta t}$ defined in (3.19) admits, for small ϵ, the following expansion:

$$\mathfrak{I}^{(j)}_{\epsilon,\Delta t} = -(a_{jj}\Delta t L)^{-1} + O(\epsilon^2). \tag{3.24}$$

Consider now an A-type time integrator, so with $a_{jj} \neq 0$ for any $j \in \{1,\dots,s\}$, and assume $g^n = O(1)$. From (3.22) and the previous expansion, we obtain

$$g^{(1)} = -(a_{11}\Delta t L)^{-1}\left[-\epsilon a_{11}\Delta t v M\cdot\nabla_x\rho^{(1)}\right] + O(\epsilon^2) = \epsilon L^{-1}(vM)\cdot\nabla_x\rho^{(1)} + O(\epsilon^2).$$

Now, the proof is performed by induction on $j \in \{2, \ldots, s\}$ assuming that for any $k \in \{1, \ldots, j-1\}$, $g^{(k)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(k)} + O(\epsilon^2)$. In particular $g^{(k)} = O(\epsilon)$ and the formula (3.22) has therefore the following expansion:

$$g^{(j)} = -(a_{jj}\Delta t L)^{-1} \left[O(\epsilon^2) - \epsilon \sum_{k=1}^{j} a_{jk} \Delta t v M \cdot \nabla_x \rho^{(k)} + \sum_{k=1}^{j-1} a_{jk} \Delta t L g^{(k)} \right] + O(\epsilon^2).$$

Inserting the induction hypothesis in the last sum, most of the terms in the two sums eliminate so that finally $g^{(j)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(j)} + O(\epsilon^2)$.

The case of a CK-ARS time integrator is slightly different. First $a_{11} = 0$ so that $g^{(1)} = g^n = O(\epsilon)$ by the particular well-prepared assumption. Now $a_{22} \neq 0$ and (3.22) has the following expansion for $j = 2$:

$$g^{(2)} = -(a_{22}\Delta t L)^{-1} \left[O(\epsilon^2) - \epsilon a_{22} \Delta t v M \cdot \nabla_x \rho^{(2)} \right] + O(\epsilon^2) = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(2)} + O(\epsilon^2).$$

Again, the proof is by induction on $j \in \{3, \ldots, s\}$ assuming for any $k \in \{2, \ldots, j-1\}$, $g^{(k)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(k)} + O(\epsilon^2)$. The same computation as above is available since $g^{(1)} = O(\epsilon)$. One has (note that $a_{j1} = 0$ for any j so that the sums start at $k = 2$):

$$\begin{aligned} g^{(j)} &= -(a_{jj}\Delta t L)^{-1} \left[O(\epsilon^2) - \epsilon \sum_{k=2}^{j} a_{jk} \Delta t v M \cdot \nabla_x \rho^{(k)} + \sum_{k=2}^{j-1} a_{jk} \Delta t L g^{(k)} \right] + O(\epsilon^2) \\ &= \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(j)} + O(\epsilon^2). \end{aligned}$$

□

Due to the GSA property of both time integrators, we have $g^{n+1} = g^{(s)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(s)} + O(\epsilon^2) = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{n+1} + O(\epsilon^2)$ for sufficiently small ϵ. Thus, the following are evident from Lemma 3.1:

1. For type CK-ARS, if the initial data is well-prepared (that is, $g^0 = O(\epsilon)$), then $g^n = O(\epsilon)$, $\forall n > 0$.

2. For type A, if the initial data is such that $g^0 = O(1)$, then $g^n = O(\epsilon)$, $\forall n > 0$.

As observed in [92], the initial data does not need to be well-prepared for type A time integrator, unlike type CK-ARS, to ensure AP property.

Theorem 3.1 *Consider the scheme (3.20)-(3.21) approximating the macro-micro model (3.7)-(3.8), with the RK method (3.14) of type A or of type CK-ARS (with well-prepared initial data*

$g^0 = O(\epsilon)$). *Then in the limit* $\epsilon \to 0$, *the scheme* (3.20)-(3.21) *degenerates to the following scheme for the diffusion equation*

$$\rho^{(j)} = \rho^n + \sum_{k=1}^{j} a_{jk}\Delta t \nabla_x \cdot \left(\kappa \nabla_x \rho^{(k)}\right), \ \forall j = 1, \ldots, s, \ \kappa = -\left\langle v \otimes L^{-1}(vM)\right\rangle_V . \tag{3.25}$$

Proof: Corresponding to each case (CK-ARS or type A), we have the following:

Type CK-ARS Assumptions in criterion 1 of Lemma 3.1 are satisfied, and its implications can be utilised. Hence, inserting $g^{(\ell)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(\ell)} + O(\epsilon^2), \ \forall \ell \in \{2, 3, .., s\}$ into eq. (3.20), we get (recall that $a_{j1} = 0$)

$$\begin{aligned}
\rho^{(j)} &= \rho^n - \frac{\Delta t}{\epsilon} \sum_{k=2}^{j} a_{jk} \nabla_x \cdot \left\langle v \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(k)} \right\rangle_V + O(\epsilon), \\
&= \rho^n - \Delta t \sum_{k=2}^{j} a_{jk} \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V \nabla_x \rho^{(k)} \right) + O(\epsilon).
\end{aligned}$$

Type A Assumptions in criterion 2 of Lemma 3.1 are satisfied, and its implications can be utilised. Hence, inserting $g^{(\ell)} = \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(\ell)} + O(\epsilon^2), \ \forall \ell \in \{1, 2, .., s\}$ into eq. (3.20), we get the required result by following the same simplification as before. The only difference is that here $\sum_{k=1}^{j}$ instead of $\sum_{k=2}^{j}$.

□

Remark 3.2 *For type CK-ARS, if the initial data is not well-prepared, computing* $g^{(2)}$ *from* (3.21) *involves* $\epsilon \frac{\tilde{a}_{21}}{a_{22}} L^{-1}(I - \Pi)(v \cdot \nabla_x g^{(1)})$ *which is not of* $O(\epsilon^2)$. *Thus,*

$$g^{(2)} = \epsilon \frac{\tilde{a}_{21}}{a_{22}} L^{-1}(I - \Pi)(v \cdot \nabla_x g^{(1)}) + \epsilon L^{-1}(vM) \cdot \nabla_x \rho^{(2)} + O\left(\epsilon^2\right),$$

and inserting in the macro equation eq. (3.20) *for* $j = 2$ *leads to (since* $a_{21} = 0$*)*

$$\rho^{(2)} = \rho^n - \frac{\tilde{a}_{21}}{a_{22}} \Delta t \left\langle v \otimes L^{-1}\left((I - \Pi) v \nabla_x^2 g^{(1)}\right)\right\rangle_V - a_{22}\Delta t \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V \nabla_x \rho^{(2)} \right) + O(\epsilon),$$

which is not consistent with the diffusion equation. Thus, for CK-ARS, asymptotic consistency cannot be attained if the initial data is not well-prepared.

3.5 Space and velocity discretization

In this section, we present the spatial (for both non-staggered and staggered grids) and velocity discretization strategies that we employ in our numerical scheme.

3.5.1 Discrete velocity method

For the velocity discretization, we will follow the discrete velocity method [170]. Thus, the velocity domain is truncated as $v \in [-v_{\max}, v_{\max}]$, and a uniform mesh is used $v_k = -v_{\max} + k\Delta v$, $k = 1, \dots, N_v (N_v \in \mathbb{N}^\star)$ and $\Delta v = 2v_{\max}/N_v$. Further, $f(t,x,v)$ and $M(v)$ are represented as:

$$f_k(t,x) := f(t,x,v_k), \quad M_k := M(v_k) \text{ for } k = 1, \dots, N_v.$$

Then, according to the definitions (3.3) and (3.6), we have for $j = 1, \dots, N_v$

$$\rho(t,x) \approx \frac{\sum_{k=0}^{N_v-1} f_k \Delta v}{\sum_{k=0}^{N_v-1} M_k \Delta v} \quad \text{and} \quad (\Pi f(t,x,v))_j \approx \frac{\sum_{k=0}^{N_v-1} f_k \Delta v}{\sum_{k=0}^{N_v-1} M_k \Delta v} M_j.$$

For the presentation, we will keep velocity continuous to focus on space discretization.

3.5.2 Space discretization using staggered grid

First, we will consider staggered grid to approximate $g^{(j)}$ and $\rho^{(j)}$ in space following [200]: the two meshes of the space interval $[0,1]$ are $x_i = i\Delta x$ and $x_{i+1/2} = (i+1/2)\Delta x$ for $i = 0, \dots, N_x (N_x \in \mathbb{N}^\star)$, with $\Delta x = L/N_x$. Periodic boundary conditions will be considered in this section.

The expressions for $g^{(j)}$ and $\rho^{(j)}$ in (3.22)-(3.23) are spatially discretised by considering staggered grid: $\rho^{(j)}$ is stored at x_i ($\rho_i^{(j)} \approx \rho^{(j)}(x_i)$), and $g^{(j)}$ is stored at $x_{i+1/2}$ ($g_{i+1/2}^{(j)}(v) \approx g^{(j)}(x_{i+1/2}, v)$). The term $v \cdot \nabla_x g^{(k)}$ in (3.22) and (3.23) is discretised in an upwind fashion as $v \cdot \nabla_x \approx v^+ \cdot \mathbf{G}_{\mathsf{upw}}^- + v^- \cdot \mathbf{G}_{\mathsf{upw}}^+$ where $v^\pm = (v \pm |v|)/2$, $\mathbf{G}_{\mathsf{upw}}^\pm$ denote the $N_x \times N_x$ matrices that approximate ∇_x. For instance, the first order version is

$$\mathbf{G}_{\mathsf{upw}}^- = \frac{1}{\Delta x}\mathsf{circ}([-1, \underline{1}]), \qquad \mathbf{G}_{\mathsf{upw}}^+ = \frac{1}{\Delta x}\mathsf{circ}([\underline{-1}, 1]), \tag{3.26}$$

where the notation circ is defined in section 3.8. With these notations, we get

$$\left(v\partial_x g^{(j)}\right)_{x_{i+1/2}} \approx v^+ \frac{g_{i+\frac{1}{2}}^{(j)} - g_{i-\frac{1}{2}}^{(j)}}{\Delta x} + v^- \frac{g_{i+\frac{3}{2}}^{(j)} - g_{i+\frac{1}{2}}^{(j)}}{\Delta x} = \left(\left(v^+ \mathbf{G}_{\mathsf{upw}}^- + v^- \mathbf{G}_{\mathsf{upw}}^+\right) g^{(j)}\right)_i,$$

where in the last term, the i index has to be understood as the i-th component of the vector.

Instead of first order upwind discretization, one can also use high order upwind discretizations so that the matrices $\mathbf{G}^{\pm}_{\mathsf{upw}}$ will be different. Further, the term $vM\cdot\nabla_x\rho^{(k)}$ in (3.22)-(3.23) and the terms of the form $\nabla_x\cdot\langle(\cdot)\rangle_V$ in (3.23) are discretised using second order central differences as in [200]. In particular, the term $vM\cdot\nabla_x\rho^{(k)}$ is approximated by

$$\left(vM\partial_x\rho^{(k)}\right)_{x_{i+1/2}}\approx vM\frac{\rho^{(k)}_{i+1}-\rho^{(k)}_i}{\Delta x}=\left(vM\mathbf{G}_{\mathsf{cen}_g}\rho^{(k)}\right)_i,\ \mathbf{G}_{\mathsf{cen}_g}=\frac{1}{\Delta x}\mathsf{circ}([\underline{-1},1]). \tag{3.27}$$

Finally, the gradient terms $\nabla_x\cdot\langle(\cdot)\rangle_V$ in (3.23) are approximated as follows

$$\left(\partial_x\langle\cdot\rangle_V\right)_{x_i}=\frac{\left(\langle\cdot\rangle_V\right)_{i+1/2}-\left(\langle\cdot\rangle_V\right)_{i-1/2}}{\Delta x}=\left(\mathbf{G}_{\mathsf{cen}_\rho}\langle\cdot\rangle_V\right)_i,\ \mathbf{G}_{\mathsf{cen}_\rho}=\frac{1}{\Delta x}\mathsf{circ}([-1,\underline{1}]). \tag{3.28}$$

Again, high order centered finite differences methods can be used so that it will give different expressions for $\mathbf{G}_{\mathsf{cen}_\rho}$ and $\mathbf{G}_{\mathsf{cen}_g}$. Let us remark that the term $\nabla_x\cdot\nabla_x=\nabla_x^2$ in (3.23) is approximated by $\mathbf{G}_{\mathsf{cen}_\rho}\mathbf{G}_{\mathsf{cen}_g}$, ie $\mathbf{G}_{\mathsf{cen}_\rho}\mathbf{G}_{\mathsf{cen}_g}=\frac{1}{\Delta x^2}\mathsf{circ}([1,\underline{-2},1])$, which gives the standard second order approximation of the Laplacian.

To ease the reading, we present the fully discrete scheme for first order ARS$(1,1,1)$ but the generalization to high order can be done using the elements of Section 3.3

$$\begin{aligned}
g^{n+1}&=\left(\epsilon^2 I-\Delta t L\right)^{-1}\left(\epsilon^2 g^n-\epsilon\Delta t\left(I-\Pi\right)\left(v^+\mathbf{G}^-_{\mathsf{upw}}+v^-\mathbf{G}^+_{\mathsf{upw}}\right)g^n-\epsilon\Delta t vM\mathbf{G}_{\mathsf{cen}_g}\rho^{n+1}\right)\\
\rho^{n+1}&=\left(I-\Delta t^2\mathbf{G}_{\mathsf{cen}_\rho}\left(\left\langle v\otimes\left(\epsilon^2 I-\Delta t L\right)^{-1}(vM)\right\rangle_V\mathbf{G}_{\mathsf{cen}_g}\right)\right)^{-1}\times\\
&\quad\left(\rho^n-\Delta t\mathbf{G}_{\mathsf{cen}_\rho}\left\langle v\left(\epsilon^2 I-\Delta t L\right)^{-1}\left(\epsilon g^n-\Delta t\left(I-\Pi\right)\left(\left(v^+\mathbf{G}^-_{\mathsf{upw}}+v^-\mathbf{G}^+_{\mathsf{upw}}\right)g^n\right)\right)\right\rangle_V\right).
\end{aligned}$$

3.5.3 Space discretization using non-staggered grid

We also address the case of non-staggered grids which may be more appropriate when high dimensions are considered in space since only one spatial mesh is used: $x_i=i\Delta x$, for $i=0,1,..,N_x$, with $\Delta x=L/N_x$. Let $g^{(j)}$ and $\rho^{(j)}$ in (3.22)-(3.23) $\forall j\in\{1,2,..,s\}$ be approximated in space by $g^{(j)}_i(v)\approx g^{(j)}(x_i,v)$ and $\rho^{(j)}_i\approx\rho^{(j)}(x_i)$. The term $v\cdot\nabla_x g^{(k)}$ in (3.22)-(3.23) is discretised in an upwind fashion as $v\cdot\nabla_x=v^+\mathbf{G}^-_{\mathsf{upw}}+v^-\mathbf{G}^+_{\mathsf{upw}}$, where $v^\pm=(v\pm|v|)/2$. Here, $\mathbf{G}^{\pm}_{\mathsf{upw}}$ denote the matrices that represent an upwind approximation of ∇_x. For instance, the definition (3.26) can be used, but also its third order version

$$\mathbf{G}^-_{\mathsf{upw}}=\frac{1}{6\Delta x}\mathsf{circ}([1,-6,\underline{3},2]),\qquad \mathbf{G}^+_{\mathsf{upw}}=\frac{1}{6\Delta x}\mathsf{circ}([-2,\underline{-3},6,-1]), \tag{3.29}$$

where circ represents the matrix notation described in section 3.8 can be used. The term $vM \cdot \nabla_x \rho^{(k)}$ in (3.22)-(3.23) and the terms of the form $\nabla_x \cdot \langle (\cdot) \rangle_V$ in (3.23) are discretised in central fashion, since these terms act as source in eq. (3.22) and diffusion in (3.23). Here, ∇_x is approximated by central differences as in (3.28) or (3.27) but in the non-staggered case, the same matrix can be used for both terms. As an example, the fourth order central difference produces:

$$\mathbf{G}_{\mathsf{cen}} = \frac{1}{12\Delta x}\mathsf{circ}([1, -8, \underline{0}, 8, -1]). \tag{3.30}$$

The term $\nabla_x \cdot \nabla_x = \nabla_x^2$ in (3.23) is discretised as the matrices product $\mathbf{G}_{\mathsf{cen}}^2 = \mathbf{G}_{\mathsf{cen}}\mathbf{G}_{\mathsf{cen}}$. Like in the staggered grid case, we present the fully discrete scheme for first order ARS$(1,1,1)$ time discretization to ease the reading:

$$\begin{aligned}
g^{n+1} &= \left(\epsilon^2 I - \Delta t L\right)^{-1}\left(\epsilon^2 g^n - \epsilon\Delta t \left(I - \Pi\right)\left(v^+\mathbf{G}_{\mathsf{upw}}^- + v^-\mathbf{G}_{\mathsf{upw}}^+\right) g^n - \epsilon\Delta t v M \mathbf{G}_{\mathsf{cen}}\rho^{n+1}\right)\\
\rho^{n+1} &= \left(I - \Delta t^2 \mathbf{G}_{\mathsf{cen}}\left(\left\langle v \otimes \left(\epsilon^2 I - \Delta t L\right)^{-1}(vM)\right\rangle_V \mathbf{G}_{\mathsf{cen}}\right)\right)^{-1} \times\\
&\quad \left(\rho^n - \Delta t \mathbf{G}_{\mathsf{cen}}\left\langle v\left(\epsilon^2 I - \Delta t L\right)^{-1}\left(\epsilon g^n - \Delta t\left(I - \Pi\right)\left(\left(v^+\mathbf{G}_{\mathsf{upw}}^- + v^-\mathbf{G}_{\mathsf{upw}}^+\right) g^n\right)\right)\right\rangle_V\right)
\end{aligned}$$

Remark 3.3 *We know that the term* $\sum_{k=1}^{j} a_{jk}\frac{\Delta t}{\epsilon}\nabla_x \cdot \left\langle v g^{(k)}\right\rangle_V$ *in* (3.20) *is split into first* $j-1$ *and last* j *contributions, and* $g^{(j)}$ *is substituted for the last* j *contribution, as in* (3.23)*. The gradient in* $\sum_{k=1}^{j-1} a_{jk}\frac{\Delta t}{\epsilon}\nabla_x \cdot \left\langle v g^{(k)}\right\rangle_V$ *of* (3.23) *is discretised using* $\mathbf{G}_{\mathsf{cen}_\rho}$*. Further, the substitution of* $g^{(j)}$ *for the last* j *hints the combination of* $\nabla_x \cdot \nabla_x$ *as* ∇_x^2 *for the terms of* $g^{(j)}$ *involving* $\nabla_x g$ *and* $\nabla_x \rho$*. However, if we choose a spatial discretization for* ∇_x^2 *as* $\mathbf{G}_{\mathsf{diff}}$*, then these terms will experience* $\mathbf{G}_{\mathsf{cen}_\rho}\mathbf{G}_{\mathsf{cen}_g}$ *for the first* $j-1$ *contributions and* $\mathbf{G}_{\mathsf{diff}}$ *for the last* j *contribution of the* $\rho^{(j)}$ *update equation. This disrupts the ODE structure present in RK time discretization, and hence reduction to first order time accuracy was observed numerically. Therefore, in order to retain high order time accuracy, it is important to carry out the space discretization carefully. Hence, we do not introduce a different discretization for* ∇_x^2*, and we retain* $\mathbf{G}_{\mathsf{cen}_\rho}\mathbf{G}_{\mathsf{cen}_g}$ *even for the last* j *contribution of* $\rho^{(j)}$ *equation.*

Remark 3.4 *The matrices introduced for spatial discretization do not change the Chapman-Enskog expansion so that the AP property is still true in the fully discrete form. Thus, we have* $g^{(k)} = \epsilon L^{-1}(vM)\mathbf{G}_{\mathsf{cen}_g}\rho^{(k)} + O(\epsilon^2)$ *for* $k \in \{1, \ldots, s\}$ *by using type A. For CK-ARS with well-prepared data, we have* $g^{(k)} = \epsilon L^{-1}(vM)\mathbf{G}_{\mathsf{cen}_g}\rho^{(k)} + O(\epsilon^2)$ *for* $k \in \{2, \ldots, s\}$*. Inserting this in macro equation, we get the corresponding RK scheme for the diffusion*

$$\rho^{(j)} = \rho^n - \Delta t \sum_{k=1}^{j} a_{jk}\mathbf{G}_{\mathsf{cen}_\rho}\left(\left\langle v \otimes L^{-1}(vM)\right\rangle_V \mathbf{G}_{\mathsf{cen}_g}\rho^{(k)}\right) + O(\epsilon).$$

3.6 Extensions to advection-diffusion collision operator and inflow boundary problems

In this section, we show that our high order AP schemes can be extended to other problems involving advection-diffusion asymptotics and inflow boundaries.

3.6.1 Advection-diffusion asymptotics

In this part, an advection-diffusion collision operator is considered (see [171, 165]),

$$\mathcal{L}f := Lf + \epsilon vM \cdot A \langle f \rangle_V, \quad A \in \mathbb{R}^d, \quad |\epsilon A| < 1, \tag{3.31}$$

where L denotes a collision satisfying the properties listed in Section 3.2. A famous simple example is $Lf = \langle f \rangle_V M - f$.

Using the notations introduced in Section 3.2, we can derive the micro-macro model satisfied by $\rho = \langle f \rangle_V$ and $g = f - \rho M$ by applying Π and $I - \Pi$ to eq. (3.1) with collision $\mathcal{L}$ to get the macro and micro equations in this context

$$\partial_t \rho + \frac{1}{\epsilon} \nabla_x \cdot \langle vg \rangle_V = 0, \tag{3.32}$$

$$\partial_t g + \frac{1}{\epsilon} (I - \Pi)(v \cdot \nabla_x g) + \frac{1}{\epsilon} vM \cdot \nabla_x \rho = \frac{1}{\epsilon^2} Lg + \frac{1}{\epsilon} vM \cdot A\rho. \tag{3.33}$$

A Chapman-Enskog expansion can be performed to get $g = \epsilon L^{-1}(vM) \cdot \nabla_x \rho - \epsilon L^{-1}(vM) \cdot A\rho + \mathcal{O}(\epsilon^2)$. Inserting this in the macro equation (3.32) enables to obtain an advection-diffusion equation in the limit $\epsilon \to 0$:

$$\partial_t \rho + \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V \nabla_x \rho \right) - \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V A\rho \right) = 0. \tag{3.34}$$

The goal is to design a uniformly stable high order time integrators for (3.32)-(3.33) so that they degenerate into a high order time integrator for (3.34) as $\epsilon \to 0$. The extension of the schemes introduced in Section 3.3 will lead to an IMEX discretization of the asymptotic model (3.34), where the advection term is treated explicitly while the diffusion term is implicit.

3.6.1.1 High order time integrator

In this subsection, we present the discretization of macro and micro equations (3.32)-(3.33). As in Section 3.3, in the micro equation, we treat $\frac{1}{\epsilon^2} Lg$ implicitly to ensure uniform stability and the additional term $\frac{1}{\epsilon} vM \cdot A\rho$ explicitly since it will be stabilized by the implicit treatment of the stiffest term. Regarding the macro equation and the remaining terms in micro equation,

we follow the lines from previous Section 3.3. We thus obtain the following high order IMEX RK scheme to approximate (3.32)-(3.33)

$$\rho^{(j)} = \rho^n - \sum_{k=1}^{j} a_{jk} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(k)} \right\rangle_V , \tag{3.35}$$

$$g^{(j)} = g^n - \frac{\Delta t}{\epsilon} \Big[\sum_{k=1}^{j-1} \tilde{a}_{jk} \mathcal{T} g^{(k)} + \sum_{k=1}^{j} a_{jk} v M \cdot \nabla_x \rho^{(k)} - \sum_{k=1}^{j} \frac{a_{jk}}{\epsilon} L g^{(k)} - \sum_{k=1}^{j-1} \tilde{a}_{jk} v M \cdot A \rho^{(k)} \Big], \tag{3.36}$$

where the coefficients $a_{jk}, \tilde{a}_{jk}$ are given by the Butcher tableaux. As in Section 3.3, some calculations are required to make the algorithm explicit. First, we have

$$g^{(j)} = \mathcal{I}^{(j)}_{\epsilon,\Delta t} \left(\epsilon^2 g^n - \epsilon \Delta t \Big[\sum_{k=1}^{j-1} \tilde{a}_{jk} \mathcal{T} g^{(k)} + \sum_{k=1}^{j} a_{jk} v M \cdot \nabla_x \rho^{(k)} - \frac{1}{\epsilon} \sum_{k=1}^{j-1} a_{jk} L g^{(k)} - \sum_{k=1}^{j-1} \tilde{a}_{jk} v M \cdot A \rho^{(k)} \Big] \right), \tag{3.37}$$

with $\mathcal{T} g^{(k)} = (I - \Pi) \left(v \cdot \nabla_x g^{(k)} \right)$ and $\mathcal{I}^{(j)}_{\epsilon,\Delta t} = \left(\epsilon^2 I - a_{jj} \Delta t L \right)^{-1}$. Then, $\rho^{(j)}$ is obtained by inserting $g^{(j)}$ given by (3.37) in the macro equation (3.35) to get

$$\begin{aligned} \rho^{(j)} = \left(I - a_{jj}^2 \Delta t^2 \nabla_x \cdot \left(\mathcal{D}^{(j)}_{\epsilon,\Delta t} \nabla_x \right) \right)^{-1} \Bigg(& \rho^n - \sum_{k=1}^{j-1} a_{jk} \frac{\Delta t}{\epsilon} \nabla_x \cdot \left\langle v g^{(k)} \right\rangle_V \\ & - a_{jj} \Delta t \nabla_x \cdot \Bigg\langle v \mathcal{I}^{(j)}_{\epsilon,\Delta t} \Bigg(\epsilon g^n - \sum_{k=1}^{j-1} \tilde{a}_{jk} \Delta t \mathcal{T} g^{(k)} - \sum_{k=1}^{j-1} a_{jk} \Delta t v M \cdot \nabla_x \rho^{(k)} \\ & + \frac{1}{\epsilon} \sum_{k=1}^{j-1} a_{jk} \Delta t L g^{(k)} + \sum_{k=1}^{j-1} \tilde{a}_{jk} \Delta t v M \cdot A \rho^{(k)} \Bigg) \Bigg\rangle_V \Bigg), \end{aligned} \tag{3.38}$$

where $\mathcal{D}^{(j)}_{\epsilon,\Delta t} = \left\langle v \otimes \left(\epsilon^2 I - a_{jj} \Delta t L \right)^{-1} (vM) \right\rangle_V$. Thus, $\rho^{(j)}$ can be updated by using (3.38) and $g^{(j)}$ can be found explicitly by using (3.37).

3.6.1.2 Asymptotic preserving property

This part is dedicated to the asymptotic preserving property of the scheme (3.38)-(3.37). We first show the AP property of type A time integrator, and we later remark how this property is true for the CK-ARS time integrator with well-prepared initial data. First we have

Lemma 3.2 *If* $g^n = O(1)$ *and* $g^{(k)} = O(\epsilon), \forall k \in \{1, 2, \ldots, j-1\}$, *then* $g^{(j)} = O(\epsilon), \forall j \in$

$\{2,3,..,s\}$ *for small* ϵ. *In particular, we have* $\forall j \in \{2,3,..,s\}$

$$g^{(j)} = \epsilon \sum_{k=1}^{j} \frac{a_{jk}}{a_{jj}} L^{-1}(vM) \cdot \nabla_x \rho^{(k)} - \sum_{k=1}^{j-1} \frac{a_{jk}}{a_{jj}} g^{(k)} - \epsilon \sum_{k=1}^{j-1} \frac{\tilde{a}_{jk}}{a_{jj}} L^{-1}(vM) \cdot A\rho^{(k)} + O(\epsilon^2). \tag{3.39}$$

Proof: Plugging in eq. (3.37) the expansion (3.24) of $\mathcal{I}^{(j)}_{\epsilon,\Delta t}$ given by eq. (3.19), along with the assumptions stated in the Lemma, we obtain (3.39) from which we deduce $g^{(j)} = \mathcal{O}(\epsilon)$ for all $j \in \{2,3,..,s\}$. □

Remark 3.5 *For type A time integrator, if* $g^n = \mathcal{O}(1)$, *we have from* (3.37)*:*

$$g^{(1)} = \epsilon \frac{a_{11}}{a_{11}} vM \cdot \nabla_x \rho^{(1)} + O(\epsilon^2) = O(\epsilon).$$

This satisfies the induction hypothesis in Lemma 3.2. Further, eq. (3.39) *holds by omitting* $\sum_{k=1}^{j-1}$ *terms for* $j = 1$. *Thus, eq.* (3.39) *is true for* $j \in \{1,2,..,s\}$.

Lemma 3.2 enables to get an expansion of $g^{(j)}$ that can be inserted in (3.38) to identify the time discretization of the asymptotic limit. However, this leads to quite involved calculations which requires to introduce some notations.

Definition 3.2 *For* $j \in \{1,2,..,s\}$ *and* $k_1, m \in \{1,2,..,j\}$ *we define*

$$\Pi^m_{j,k_1} = \left\langle v \frac{a_{jk_1}}{a_{k_1 k_1}} \left(\mathfrak{S}^{k_0} \mathfrak{S}^{k_1} \mathfrak{S}^{k_2} \dots \mathfrak{S}^{k_{m-1}} \right) \left(\mathfrak{R}^{k_m} \right) \right\rangle_V, \tag{3.40}$$

with

$$\mathfrak{S}^{k_0} = 1, \qquad \mathfrak{S}^{k_l} = \sum_{k_{l+1}=1}^{k_l - 1} \frac{a_{k_l k_{l+1}}}{a_{k_{l+1} k_{l+1}}} \text{ for } l \in \{1,2,..,m-1\},\ m \geq 2,$$

$$\mathfrak{R}^{k_m} = \sum_{k_{m+1}=1}^{k_m} a_{k_m k_{m+1}} L^{-1}(vM) \cdot \nabla_x \rho^{(k_{m+1})} - \sum_{k_{m+1}=1}^{k_m - 1} \tilde{a}_{k_m k_{m+1}} L^{-1}(vM) \cdot A\rho^{(k_{m+1})}.$$

As usual, we will use the convention $\sum_{j=1}^{q} \equiv 0$ *if* $q \in \mathbb{Z} \backslash \mathbb{N}$.

The term Π^m_{j,k_1} will be useful in the following study and deserves some remarks: the index m denotes the depth of the embedded sums, j corresponds to the current stage and k_1 corresponds to the indexing over previous stages. We continue with the following lemma which gives an induction relation on Π^m_{j,k_1}.

Lemma 3.3 *For $j \geq 2$, we have*

$$\Pi^m_{j,j} = \sum_{k_1=1}^{j-1} \Pi^{m-1}_{j,k_1} \text{ for } m \in \{2,3,..,j\}, \text{ and } \Pi^j_{j,k_1} = 0 \text{ for } k_1 \in \{1,2,..,j-1\}.$$

Proof: For the first relation, considering $k_1 = j$ (with $j \geq 2$) in (3.40) leads to

$$\Pi^m_{j,j} = \left\langle v \left(\mathcal{S}^{k_0}\mathcal{S}^{j}\mathcal{S}^{k_2}\ldots\mathcal{S}^{k_{m-1}}\right)\left(\mathcal{R}^{k_m}\right)\right\rangle_V,$$

since $a_{jj} \neq 0$. Further, since $\mathcal{S}^{k_1=j} = \sum_{k_2=1}^{j-1} \frac{a_{jk_2}}{a_{k_2k_2}}$, we get

$$\Pi^m_{j,j} = \left\langle v \sum_{k_2=1}^{j-1} \frac{a_{jk_2}}{a_{k_2k_2}} \left(\mathcal{S}^{k_0}\mathcal{S}^{k_2}...\mathcal{S}^{k_{m-1}}\right)\left(\mathcal{R}^{k_m}\right)\right\rangle_V$$

By employing the change of variables as $k_\ell \to k_{\ell-1}$ for $\ell \in \{2,3,..,m\}$ in the right hand side of above expression, we get

$$\begin{aligned}\Pi^m_{j,j} &= \left\langle v \sum_{k_1=1}^{j-1} \frac{a_{jk_1}}{a_{k_1k_1}} \left(\mathcal{S}^{k_0}\mathcal{S}^{k_1}\ldots\mathcal{S}^{k_{m-2}}\right)\left(\mathcal{R}^{k_{m-1}}\right)\right\rangle_V \\ &= \sum_{k_1=1}^{j-1} \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\mathcal{S}^{k_0}\mathcal{S}^{k_1}\ldots\mathcal{S}^{k_{m-2}}\right)\left(\mathcal{R}^{k_{m-1}}\right)\right\rangle_V = \sum_{k_1=1}^{j-1} \Pi^{m-1}_{j,k_1},\end{aligned}$$

which proves the first identity.

For the second relation, considering $m = j$ in eq. (3.40) leads to

$$\Pi^j_{j,k_1} = \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\mathcal{S}^{k_0}\mathcal{S}^{k_1}\mathcal{S}^{k_2}...\mathcal{S}^{k_{j-1}}\right)\left(\mathcal{R}^{k_j}\right)\right\rangle_V$$

We first prove the relation for $j = 2$. It is clear from Definition 3.2 that the summation in $\mathcal{S}^{k_1}$ goes from $k_2 = 1$ to $k_2 = k_1 - 1$. For $k_1 = 1$, the summation goes to $k_2 = k_1 - 1 = 0$. Thus, since $\mathcal{S}^{k_1}$ involves $\sum_1^0$ for $k_1 = 1$, it is zero according to the convention. Hence $\Pi^j_{j,k_1} = 0$ for $k_1 = 1$.

We now prove the relation for $j > 2$. From Definition 3.2, it can be seen that the summations in $\mathcal{S}^{k_1}$ and $\mathcal{S}^{k_2}$ go from $k_2 = 1$ to $k_2 = k_1 - 1$ and $k_3 = 1$ to $k_3 = k_2 - 1$ respectively. Thus, the summation in $\mathcal{S}^{k_2}$ can go to atmost $k_3 = k_2 - 1 = (k_1 - 1) - 1 = k_1 - 2$. Proceeding in this manner, we see that the summation in $\mathcal{S}^{k_{j-1}}$ can go to atmost $k_j = k_1 - (j-1)$.

For $k_1 \in \{1,2,..,j-1\}$, $k_j = k_1 - (j-1) \in \mathbb{Z}\backslash\mathbb{N}$ so that $\mathcal{S}^{k_{j-1}} = 0$ and hence $\Pi^j_{j,k_1} = 0$ for $k_1 \in \{1,2,..,j-1\}$ which ends the proof. □

Now, we can use the previous Lemma to identify the asymptotic numerical scheme.

Lemma 3.4 *When $\epsilon \to 0$, the numerical scheme (3.35)-(3.36) degenerates into*

$$\rho^{(j)} = \rho^n + \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left(\sum_{\ell=1}^{j} (-1)^\ell \Pi^\ell_{j,k_1} \right) \quad \text{for } j \in \{1,2,..,s\}, \tag{3.41}$$

where Π^ℓ_{j,k_1} is given by definition 3.2.

Proof: We start with the macro equation in eq. (3.35)

$$\rho^{(j)} = \rho^n - \sum_{k_1=1}^{j} a_{jk_1} \frac{\Delta t}{\epsilon} \nabla_x \cdot \langle v g^{(k_1)} \rangle_V,$$

in which we insert $g^{(k_1)}$ given by eq. (3.39) to get

$$\begin{aligned}
\rho^{(j)} &= \rho^n - \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\sum_{k_2=1}^{k_1} a_{k_1k_2} L^{-1}(vM) \cdot \nabla_x \rho^{(k_2)} \; - \sum_{k_2=1}^{k_1-1} \tilde{a}_{k_1k_2} L^{-1}(vM) \cdot A\rho^{(k_2)} \right) \right\rangle_V \\
&\quad + \frac{\Delta t}{\epsilon} \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\sum_{k_2=1}^{k_1-1} a_{k_1k_2} g^{(k_2)} \right) \right\rangle_V + O(\epsilon) \\
&= \rho^n - \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\mathcal{S}^{k_0} \mathcal{R}^{k_1} \right) \right\rangle_V + \frac{\Delta t}{\epsilon} \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\sum_{k_2=1}^{k_1-1} a_{k_1k_2} g^{(k_2)} \right) \right\rangle_V + O(\epsilon) \\
&= \rho^n - \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \Pi^1_{j,k_1} + \frac{\Delta t}{\epsilon} \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\sum_{k_2=1}^{k_1-1} a_{k_1k_2} g^{(k_2)} \right) \right\rangle_V + O(\epsilon).
\end{aligned}$$

Inserting $g^{(k_2)}$ from eq. (3.39) in the above equation and simplifying as before, we get,

$$\rho^{(j)} = \rho^n - \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left(\Pi^1_{j,k_1} - \Pi^2_{j,k_1} \right) - \frac{\Delta t}{\epsilon} \sum_{k_1=1}^{j} \nabla_x \cdot \left\langle v \frac{a_{jk_1}}{a_{k_1k_1}} \left(\sum_{k_2=1}^{k_1-1} \frac{a_{k_1k_2}}{a_{k_2k_2}} \sum_{k_3=1}^{k_2-1} a_{k_2k_3} g^{(k_3)} \right) \right\rangle_V + O(\epsilon).$$

This procedure can be continued $(j-1)$ times to finally get,

$$\rho^{(j)} = \rho^n + \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left(\sum_{\ell=1}^{j-1} (-1)^\ell \Pi^\ell_{j,k_1} \right)$$

$$
\begin{aligned}
&-(-1)^{j-1}\frac{\Delta t}{\epsilon}\sum_{k_1=1}^{j}\nabla_x\cdot\left\langle v\frac{a_{jk_1}}{a_{k_1k_1}}\left(\sum_{k_2=1}^{k_1-1}\frac{a_{k_1k_2}}{a_{k_2k_2}}\cdots\sum_{k_{j-1}=1}^{k_{j-2}-1}\frac{a_{k_{j-2}k_{j-1}}}{a_{k_{j-1}k_{j-1}}}\sum_{k_j=1}^{k_{j-1}-1}a_{k_{j-1}k_j}g^{(k_j)}\right)\right\rangle_V+O(\epsilon)\\
&=\rho^n+\Delta t\sum_{k_1=1}^{j}\nabla_x\cdot\left(\sum_{\ell=1}^{j-1}(-1)^\ell\Pi^\ell_{j,k_1}\right)\\
&\quad-(-1)^{j-1}\frac{\Delta t}{\epsilon}\sum_{k_1=1}^{j}\nabla_x\cdot\left\langle v\frac{a_{jk_1}}{a_{k_1k_1}}\left(\mathcal{S}^{k_0}\mathcal{S}^{k_1}\ldots\mathcal{S}^{k_{j-2}}\sum_{k_j=1}^{k_{j-1}-1}a_{k_{j-1}k_j}g^{(k_j)}\right)\right\rangle_V+O(\epsilon).
\end{aligned}
$$

We know from definition 3.2 that the summations in $\mathcal{S}^{k_1}$ and $\mathcal{S}^{k_2}$ go from $k_2=1$ to $k_2=k_1-1$ and $k_3=1$ to $k_3=k_2-1$ respectively. Thus, the summation in $\mathcal{S}^{k_2}$ can go to atmost $k_3=k_2-1=(k_1-1)-1=k_1-2$. Proceeding in this manner, we see that the summations in $\mathcal{S}^{k_{j-2}}$ and $\sum_{k_j=1}^{k_{j-1}-1}a_{k_{j-1}k_j}g^{(k_j)}$ go to atmost $k_{j-1}=k_1-(j-2)$ and $k_j=k_1-(j-1)$ respectively. Since the summation in k_1 goes to atmost j in the above equation, k_j in the term $\sum_{k_j=1}^{k_{j-1}-1}a_{k_{j-1}k_j}g^{(k_j)}$ goes to atmost $k_j=k_1-(j-1)=j-(j-1)=1$, and k_{j-1} in $\mathcal{S}^{k_{j-2}}$ goes to atmost $k_{j-1}=k_1-(j-2)=j-(j-2)=2$ and so on. Thus, only $k_j=1$ remains in the last summation so that $\sum_{k_j=1}^{k_{j-1}-1}a_{k_{j-1}k_j}g^{(k_j)}=a_{21}g^{(1)}=\epsilon a_{21}L^{-1}(vM)\cdot\nabla_x\rho^{(1)}+\mathcal{O}(\epsilon^2)=\frac{a_{21}}{a_{11}}\epsilon a_{11}L^{-1}(vM)\cdot\nabla_x\rho^{(1)}+\mathcal{O}(\epsilon^2)=\epsilon\mathcal{S}^{k_{j-1}}\mathcal{R}^{k_j}+\mathcal{O}(\epsilon^2)$. Thus, we have

$$
\begin{aligned}
\rho^{(j)}&=\rho^n+\Delta t\sum_{k_1=1}^{j}\nabla_x\cdot\left(\sum_{\ell=1}^{j-1}(-1)^\ell\Pi^\ell_{j,k_1}\right)\\
&\quad-(-1)^{j-1}\Delta t\sum_{k_1=1}^{j}\nabla_x\cdot\left\langle v\frac{a_{jk_1}}{a_{k_1k_1}}\left(\mathcal{S}^{k_0}\mathcal{S}^{k_1}\ldots\mathcal{S}^{k_{j-1}}\mathcal{R}^{k_j}\right)\right\rangle_V+O(\epsilon)\\
&=\rho^n+\Delta t\sum_{k_1=1}^{j}\left[\nabla_x\cdot\left(\sum_{\ell=1}^{j-1}(-1)^\ell\Pi^\ell_{j,k_1}\right)+\nabla_x\cdot\left((-1)^j\Pi^j_{j,k_1}\right)\right]+O(\epsilon).
\end{aligned}
$$

□

We can now prove the asymptotic property of the scheme (3.35)-(3.36).

Theorem 3.2 *When $\epsilon\to 0$, the scheme (3.35)-(3.36) degenerates into*

$$
\begin{aligned}
\rho^{(j)}=\rho^n&-\Delta t\sum_{k=1}^{j}a_{jk}\nabla_x\cdot\left(\left\langle v\otimes L^{-1}(vM)\right\rangle_V\nabla_x\rho^{(k)}\right)\\
&+\Delta t\sum_{k=1}^{j-1}\tilde{a}_{jk}\nabla_x\cdot\left(\left\langle v\otimes L^{-1}(vM)\right\rangle_V A\rho^{(k)}\right),\ \textit{for } j\in\{1,2,\ldots,s\}. \quad (3.42)
\end{aligned}
$$

Proof: From Lemma 3.4, the asymptotic limit $\epsilon \to 0$ of the macro equation in eq. (3.35) is (for $j \in \{1, 2, .., s\}$)

$$\rho^{(j)} = \rho^n + \Delta t \sum_{k_1=1}^{j} \nabla_x \cdot \left(\sum_{\ell=1}^{j} (-1)^\ell \Pi_{j,k_1}^\ell \right) = \rho^n + \Delta t \nabla_x \cdot \left(\sum_{\ell=1}^{j} (-1)^\ell \left(\Pi_{j,j}^\ell + \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^\ell \right) \right)$$
$$= \rho^n + \Delta t \nabla_x \cdot \left(-\Pi_{j,j}^1 + \sum_{\ell=2}^{j} (-1)^\ell \Pi_{j,j}^\ell + \sum_{\ell=1}^{j} (-1)^\ell \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^\ell \right).$$

Using the recurrence relation given by Lemma 3.3 and a change of indices lead to

$$\rho^{(j)} = \rho^n + \Delta t \nabla_x \cdot \left(-\Pi_{j,j}^1 + \sum_{\ell=2}^{j} (-1)^\ell \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^{\ell-1} + \sum_{\ell=1}^{j} (-1)^\ell \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^\ell \right)$$
$$= \rho^n + \Delta t \nabla_x \cdot \left(-\Pi_{j,j}^1 - \sum_{\ell=1}^{j-1} (-1)^\ell \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^\ell + \sum_{\ell=1}^{j} (-1)^\ell \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^\ell \right)$$
$$= \rho^n + \Delta t \nabla_x \cdot \left(-\Pi_{j,j}^1 + (-1)^j \sum_{k_1=1}^{j-1} \Pi_{j,k_1}^j \right).$$

From Lemma 3.3, we have $\sum_{k_1=1}^{j-1} \Pi_{j,k_1}^j = 0$, so that from Definition 3.2 we get

$$\rho^{(j)} = \rho^n + \Delta t \nabla_x \cdot \left(-\Pi_{j,j}^1 \right) = \rho^n - \Delta t \nabla_x \cdot \left(\left\langle v \frac{a_{jj}}{a_{jj}} \mathcal{S}^{k_0} \mathcal{R}^{k_1=j} \right\rangle_V \right)$$
$$= \rho^n - \Delta t \nabla_x \cdot \left(\left\langle v \left(\sum_{k_2=1}^{k_1} a_{k_1 k_2} L^{-1}(vM) \cdot \nabla_x \rho^{(k_2)} - \sum_{k_2=1}^{k_1-1} \tilde{a}_{k_1 k_2} L^{-1}(vM) \cdot A \rho^{(k_2)} \right) \right\rangle_V \right)_{k_1=j}$$
$$= \rho^n - \Delta t \sum_{k_2=1}^{j} a_{j k_2} \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V \nabla_x \rho^{(k_2)} \right) + \Delta t \sum_{k_2=1}^{j-1} \tilde{a}_{j k_2} \nabla_x \cdot \left(\left\langle v \otimes L^{-1}(vM) \right\rangle_V A \rho^{(k_2)} \right),$$

which ends the proof. □

Remark 3.6 *For CK-ARS schemes with well-prepared initial data, we obtain $g^{(1)} = g^n = O(\epsilon)$ and $\rho^{(1)} = \rho^n$. The presentation in this section will apply for CK-ARS from the second RK stage onwards. For instance, definition 3.2 applies for CK-ARS with the following change in indexes: $j \in \{2, 3, .., s\}$, $k_1, m \in \{2, 3, .., j\}$ and all the summations involved start from 2 instead of 1 since $a_{11} = 0$. The lemmas and theorems that follow also undergo the corresponding change in indexes, and the AP property for CK-ARS can be observed for $j \in \{2, 3, .., s\}$.*

Remark 3.7 *Upon incorporating the spatial matrices corresponding to staggered grid in place*

of the continuous gradient operator, we obtain in the limit $\epsilon \to 0$,

$$\rho^{(j)} = \left(I + a_{jj}\Delta t \mathbf{G}_{cen_\rho}\left(\left\langle v \otimes L^{-1}(vM)\right\rangle_V \mathbf{G}_{cen_g}\right)\right)^{-1} \times \left(\rho^n - \sum_{k=1}^{j-1} a_{jk}\Delta t \mathbf{G}_{cen_\rho}\left(\left\langle v \otimes L^{-1}(vM)\right\rangle_V \mathbf{G}_{cen_g}\rho^{(k)}\right) + \sum_{k=1}^{j-1} \tilde{a}_{jk}\Delta t \mathbf{G}_{cen_\rho}\left(\left\langle v \otimes L^{-1}(vM)\right\rangle_V \mathbf{G}_{avg_g} A\rho^{(k)}\right)\right). \quad (3.43)$$

The matrices $\mathbf{G}_{cen_\rho}, \mathbf{G}_{cen_g}$ *are given in section 3.5.2 and* $\mathbf{G}_{avg_g} = \frac{1}{2}\mathsf{circ}([\underline{1}, 1])$. *Thus,* $\left(\mathbf{G}_{avg_g} A\left(\rho^{(k)}\right)\right)_{i+1/2} = \frac{1}{2}A\left(\rho_{i+1}^{(k)} + \rho_i^{(k)}\right)$. *This results in a central discretization of the advection term in the macro equation. Thus, we obtain a consistent internal RK stage approximation of the advection-diffusion equation in the limit* $\epsilon \to 0$.
To obtain an upwind discretization of the advection term, we use the space operator $\mathbf{G}_{up_g}$ *on* $A\rho^{(k)}$ *instead of* $\mathbf{G}_{avg_g}$. *This is defined as follows:*

$$\left(\mathbf{G}_{up_g} A\left(\rho^{(k)}\right)\right)_{i+1/2} = \begin{cases} A\rho_i^{(k)} & \text{if } A \geq 0 \\ A\rho_{i+1}^{(k)} & \text{if } A < 0 \end{cases}. \quad (3.44)$$

This results in a first order upwind discretization of the advection term. For second order upwind discretization, the following is required:

$$\left(\mathbf{G}_{up_g} A\left(\rho^{(k)}\right)\right)_{i+1/2} = \begin{cases} A\left(\frac{3}{2}\rho_i^{(k)} - \frac{1}{2}\rho_{i-1}^{(k)}\right) & \text{if } A \geq 0 \\ A\left(\frac{3}{2}\rho_{i+1}^{(k)} - \frac{1}{2}\rho_{i+2}^{(k)}\right) & \text{if } A < 0 \end{cases}. \quad (3.45)$$

3.6.2 Inflow Boundaries

So far, periodic boundary conditions were considered. In this part, we consider inflow boundary conditions for f which is solution to (3.1)

$$f(t, x, v) = f_b(t, x, v), \quad (x, v) \in \partial\Omega \times V \text{ such that } v \cdot n(x) < 0, \quad \forall t, \quad (3.46)$$

where f_b is a given function and $n(x)$ denotes the unitary outgoing normal vector to $\partial\Omega$. As mentioned in [200, 199], such boundary conditions cannot be adapted naturally to the standard micro-macro unknowns $\rho(t, x)$ and $g(t, x, v)$ which form a solution to (3.6) and a specific treatment with artificial boundary conditions is required (see [200, 199, 242]). To

overcome this drawback, another micro-macro decomposition is introduced in [199]

$$f = \overline{\rho} M + \overline{g},\ \overline{\rho}(t,x) = \langle f(t,x,\cdot)\rangle_{V_-},\ \langle \overline{g}(t,x,\cdot)\rangle_{V_-} = 0,\ \langle f\rangle_{V_-} = \frac{\int_{V_-} f d\mu}{\int_{V_-} M d\mu}, \tag{3.47}$$

where the velocity domain V_- is defined by

$$V_-(x) = \{v \in V, \omega(x,v) < 0\}, \quad V_+(x) = V \backslash V_-(x). \tag{3.48}$$

The function $\omega(x,v)$ extends $v \cdot n(x)$ in the interior of domain. Some examples of $\omega(x,v)$ for different geometries are provided in [199]. It can be seen that the boundary conditions for $\overline{\rho}(t,x)$ and $\overline{g}(t,x,v)$ can be evaluated from the inflow boundary condition in eq. (3.46). Indeed, for $(x,v) \in \partial\Omega \times V$ such that $v \cdot n(x) < 0$, $\forall t$, we define

$$\overline{\rho}_b(t,x) = \langle f_b(t,x,\cdot)\rangle_{V_-}\ ,\ \ \overline{g}_b(t,x,v) = f_b(t,x,v) - \overline{\rho}_b(t,x) M(v). \tag{3.49}$$

The derivation of the micro-macro model needs to be adapted to this decomposition. The projector Π^- is defined as $\Pi^- h = \langle h\rangle_{V_-} M$. Then, substituting eq. (3.47) into eq. (3.1) and applying Π^- and $I - \Pi^-$ enable to get the macro and micro equations:

$$\partial_t \overline{\rho} + \frac{1}{\epsilon} \langle vM\rangle_{V_-} \cdot \nabla_x \overline{\rho} + \frac{1}{\epsilon} \nabla_x \cdot \langle v\overline{g}\rangle_{V_-} = \frac{1}{\epsilon^2} \langle L\overline{g}\rangle_{V_-}\ , \tag{3.50}$$

$$\partial_t \overline{g} + \frac{1}{\epsilon} \left(I - \Pi^-\right)(v \cdot \nabla_x \overline{g}) + \frac{1}{\epsilon} \left(I - \Pi^-\right) vM \cdot \nabla_x \overline{\rho} = \frac{1}{\epsilon^2} \tilde{L}\overline{g}, \tag{3.51}$$

where $\tilde{L} = (I - \Pi^-) L$. Moreover, it can be seen that $\tilde{L} = (I - \Pi^-) L (I - \Pi^-) = (I - \Pi^-) L (I - \Pi)$ since $\Pi^- h, \Pi h \in \mathcal{N}(L), \forall h$.

The macro equation (3.50) turns out to be more complicated than the one obtained for standard micro-macro decomposition. It can be made simpler by using $\rho = \overline{\rho} + \langle \overline{g}\rangle_V$, $f = \rho M - \langle \overline{g}\rangle_V M + \overline{g}$, obtained from the decompositions (3.6) and (3.47). Applying Π to eq. (3.1) instead of Π^-, we obtain the simpler macro equation,

$$\partial_t \rho + \frac{1}{\epsilon} \nabla_x \cdot \langle v\overline{g}\rangle_V = 0, \tag{3.52}$$

and the micro-macro system that we will consider in the sequel is (3.51)-(3.52).

3.6.2.1 Numerical scheme

In this part, we present the fully discretized scheme to approximate (3.51)-(3.52). The boundary conditions on $\overline{\rho}_b$ and $\overline{g}_b$ in eq. (3.49) will be utilised along with the relation $\rho = \overline{\rho} + \langle \overline{g} \rangle_V$ that allows to link ρ and $\overline{\rho}$ in the interior of the domain. We will use a staggered grid in space following [199] and a high order scheme in time, following the strategy developed previously. To ease the reading, only the first order version will be presented.

We present the space approximation based on a staggered grid. Let us consider the space interval $[0, L]$ with two grids: $x_i = i\Delta x$ and $x_{i+1/2} = (i + 1/2)\Delta x$, $\Delta x = L/(N_x - 1)$. The 'interior' variables such as $\rho, \overline{\rho}$ are stored at grid points x_i with $i = 1, \ldots, N_x - 2$) and $\overline{g}$ is stored at $i+1/2 = 1/2, \cdots, N_x - 3/2$. We also use the variable $\overline{g}_{cl} = \bar{g} \cup \bar{g}_b \in \mathbb{R}^{N_x+1}$. The whole domain including boundary will be considered for the micro unknown $\bar{g}$ so that the components of $\overline{g}_{cl}$ correspond to the grid indices $i+1/2 = -1/2, \cdots, N_x - 1/2$. The matrices corresponding to spatial operators are given by

$$\mathbf{B}^-_{\mathsf{upw}} = \frac{1}{\Delta x}\mathsf{circ}([\underline{-1}, 1])_{(N_x-1)\times(N_x+1)}, \mathbf{B}^+_{\mathsf{upw}} = \frac{1}{\Delta x}\mathsf{circ}([\underline{0}, -1, 1])_{(N_x-1)\times(N_x+1)}, \tag{3.53}$$

$$\mathbf{B}_{\mathsf{cen}_\rho} = \frac{1}{\Delta x}\mathsf{circ}([\underline{-1}, 1])_{(N_x-2)\times(N_x-1)}, \ \mathbf{B}_{\mathsf{avg}} = \frac{1}{2}\mathsf{circ}([\underline{1}, 1])_{(N_x-2)\times(N_x-1)}, \tag{3.54}$$

$$\mathbf{B}_{\mathsf{cen}_g} = \frac{1}{\Delta x}\mathsf{circ}_\mathsf{b}([-1, \underline{1}])_{(N_x-1)\times(N_x-2)}. \tag{3.55}$$

The circ_b definition is presented in section 3.8. Further, we also introduce a vector containing the boundary values of $\overline{\rho}$ as $\overline{\rho}_{bd} = \frac{1}{\Delta x}\left[-\overline{\rho}_{b_{i=0}}, 0, 0, ..., 0, \overline{\rho}_{b_{i=N_x-1}}\right]^T_{(N_x-1)\times 1}$. We now present our scheme by using this matrix notation. For simplicity, we assume that $\overline{\rho}_{bd}$ is time invariant. We also use the following notations:

$$\overline{\mathfrak{T}}h = \left(I - \Pi^-\right)\left(v^+\mathbf{B}^-_{\mathsf{upw}} + v^-\mathbf{B}^+_{\mathsf{upw}}\right)h, \overline{\mathcal{D}}_{\epsilon,\Delta t} = \left\langle v\left(\epsilon^2 I - \Delta t\tilde{L}\right)^{-1}\Delta t\left(I - \Pi^-\right)(vM)\right\rangle_V,$$
$$\overline{\mathcal{E}}_{\epsilon,\Delta t} = \left\langle\left(\epsilon^2 I - \Delta t\tilde{L}\right)^{-1}\Delta t\left(I - \Pi^-\right)(vM)\right\rangle_V, \quad \overline{\mathcal{I}}_{\epsilon,\Delta t} = \left(\epsilon^2 I - \Delta t\tilde{L}\right)^{-1}, \quad \overline{\mathcal{J}} = \left(I - \Pi^-\right)(vM).$$

The micro equation (3.51) is discretised in time as in the previous (periodic) case

$$\overline{g}^{n+1} = \overline{\mathcal{I}}_{\epsilon,\Delta t}\left(\epsilon^2\overline{g}^n - \epsilon\Delta t\overline{\mathfrak{T}}\overline{g}^n_{cl} - \epsilon\Delta t\overline{\mathcal{J}}\mathbf{B}_{\mathsf{cen}_g}\overline{\rho}^{n+1} - \epsilon\Delta t\overline{\mathcal{J}}\overline{\rho}_{bd}\right), \tag{3.56}$$

and for the macro equation (3.52), we obtain

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \frac{1}{\epsilon}\left\langle v\mathbf{B}_{\mathsf{cen}_\rho}\overline{g}^{n+1}\right\rangle_V = 0$$

Substituting $\overline{g}^{n+1}$ in the above equation, we get

$$\rho^{n+1} = \rho^n - \Delta t \mathbf{B}_{\mathsf{cen}_\rho} \left\langle v \overline{\mathcal{I}}_{\epsilon,\Delta t} \left(\epsilon \overline{g}^n - \Delta t \overline{\mathcal{T}} \overline{g}^n_{cl} - \Delta t \overline{\mathcal{J}} \mathbf{B}_{\mathsf{cen}_g} \overline{\rho}^{n+1} - \Delta t \overline{\mathcal{J}} \overline{\rho}_{bd} \right) \right\rangle_V . \tag{3.57}$$

In index notation, we use $\rho_i^{n+1} = \overline{\rho}_i^{n+1} + \frac{1}{2}\langle \overline{g}^{n+1}_{i-1/2} + \overline{g}^{n+1}_{i+1/2} \rangle_V$ (since $\rho = \overline{\rho} + \langle \overline{g} \rangle_V$) to match the two grids. In matrix notation, this becomes $\rho^{n+1} = \overline{\rho}^{n+1} + \mathbf{B}_{\mathsf{avg}} \langle \overline{g}^{n+1} \rangle_V$ with $\mathbf{B}_{\mathsf{avg}}$ given by (3.54). Substituting this into the above equation and inserting the expression for $\overline{g}^{n+1}$ into $\mathbf{B}_{\mathsf{avg}} \langle \overline{g}^{n+1} \rangle_V$ enable to update the interior macro unknown

$$\begin{aligned} \overline{\rho}^{n+1} = \left(I - \epsilon \mathbf{B}_{\mathsf{avg}} \left(\overline{\mathcal{E}}_{\epsilon,\Delta t} \mathbf{B}_{\mathsf{cen}_g} \right) - \Delta t \mathbf{B}_{\mathsf{cen}_\rho} \left(\overline{\mathcal{D}}_{\epsilon,\Delta t} \mathbf{B}_{\mathsf{cen}_g} \right) \right)^{-1} \times \\ \left(\rho^n - \mathbf{B}_{\mathsf{avg}} \left\langle \overline{\mathcal{I}}_{\epsilon,\Delta t} \left(\epsilon^2 \overline{g}^n - \epsilon \Delta t \overline{\mathcal{T}} \overline{g}^n_{cl} - \epsilon \Delta t \overline{\mathcal{J}} \overline{\rho}_{bd} \right) \right\rangle_V \right. \\ \left. - \Delta t \mathbf{B}_{\mathsf{cen}_\rho} \left\langle v \overline{\mathcal{I}}_{\epsilon,\Delta t} \left(\epsilon \overline{g}^n - \Delta t \overline{\mathcal{T}} \overline{g}^n_{cl} - \Delta t \overline{\mathcal{J}} \overline{\rho}_{bd} \right) \right\rangle_V \right) . \end{aligned} \tag{3.58}$$

The right hand side of above expression involves only known quantities so that $\overline{\rho}^{n+1}$ can be updated from (3.58) which can then be used to update $\overline{g}^{n+1}$ in (3.56). Then, we update $\overline{g}^{n+1}_{cl}$ thanks to the boundary conditions (3.49), and finally ρ^{n+1} can be computed from $\rho^{n+1} = \overline{\rho}^{n+1} + \mathbf{B}_{\mathsf{avg}} \langle \overline{g}^{n+1} \rangle_V$. In the limit $\epsilon \to 0$, the above equation becomes,

$$\overline{\rho}^{n+1} = \left(I + \Delta t \mathbf{B}_{\mathsf{cen}_\rho} \left(\left\langle v \otimes \tilde{L}^{-1} \overline{\mathcal{J}} \right\rangle_V \mathbf{B}_{\mathsf{cen}_g} \right) \right)^{-1} \left(\rho^n - \Delta t \mathbf{B}_{\mathsf{cen}_\rho} \left(\left\langle v \otimes \tilde{L}^{-1} \overline{\mathcal{J}} \right\rangle_V \overline{\rho}_{bd} \right) \right)$$

This is a consistent discretization of the diffusion equation in eq. (3.5) since $\langle v \otimes \tilde{L}^{-1} \overline{\mathcal{J}} \rangle_V = \langle v \otimes L^{-1}(vM) \rangle_V = -\kappa$. Further, the high order scheme in time can be constructed in a similar manner as before.

We now present the evaluation of boundary condition on ρ. For $i = 1/2$ and $i = 3/2$, the micro equation in (3.56) simplifies as:

$$\overline{g}^{n+1}_{\frac{1}{2}} = (-\Delta t \tilde{L})^{-1} \left(-\epsilon \Delta t \left(I - \Pi^- \right) v M \frac{\overline{\rho}_1^{n+1} - \overline{\rho}_0^{n+1}}{\Delta x} + 2\epsilon \Delta t \left(I - \Pi^- \right) v^+ \frac{\overline{g}_0^n}{\Delta x} \right) + \mathcal{O}(\epsilon^2) \tag{3.59}$$

$$\overline{g}^{n+1}_{\frac{3}{2}} = (-\Delta t \tilde{L})^{-1} \left(-\epsilon \Delta t \left(I - \Pi^- \right) v M \frac{\overline{\rho}_2^{n+1} - \overline{\rho}_1^{n+1}}{\Delta x} \right) + \mathcal{O}(\epsilon^2) \tag{3.60}$$

since $\overline{g}^n_{-\frac{1}{2}} = 2\overline{g}^n_0 - \overline{g}^n_{\frac{1}{2}}$, and $\overline{g}^n_{i+\frac{1}{2}} = \mathcal{O}(\epsilon)$ for all $n \geq 1$ and $i + \frac{1}{2} = 1/2, 3/2, 5/2$. Inserting these into the macro equation corresponding to $i = 1$ and simplifying, we obtain

$$\rho_1^{n+1} = \rho_1^n - \frac{\Delta t}{\epsilon} \left\langle v \frac{\overline{g}^{n+1}_{\frac{3}{2}} - \overline{g}^{n+1}_{\frac{1}{2}}}{\Delta x} \right\rangle_V \tag{3.61}$$

$$= \rho_1^n - \frac{\Delta t}{\Delta x^2} \left\langle v\tilde{L}^{-1}\left(I-\Pi^-\right)(vM)\right\rangle_V \left(\overline{\rho}_2^{n+1} - 2\overline{\rho}_1^{n+1} + \overline{\rho}_0^{n+1}\right) \tag{3.62}$$

$$-\frac{\Delta t}{\Delta x^2} 2\left\langle v\tilde{L}^{-1}\left(I-\Pi^-\right)(v^+\overline{g}_0^n)\right\rangle_V \tag{3.63}$$

up to $\mathcal{O}(\epsilon)$. Further, observing that $\left\langle v\tilde{L}^{-1}\left(I-\Pi^-\right)(vM)\right\rangle = \langle vL^{-1}(vM)\rangle_V$ and $\left\langle v\tilde{L}^{-1}\left(I-\Pi^-\right)(v^+\overline{g}_0^n)\right\rangle_V = \langle vL^{-1}\left(I-\Pi\right)(v^+\overline{g}_0^n)\rangle_V$, and inserting $\overline{g}_0^n = f_0^n - \overline{\rho}_0^n M$ into the above equation lead to:

$$\rho_1^{n+1} = \rho_1^n - \frac{\Delta t}{\Delta x^2}\langle vL^{-1}(vM)\rangle_V \left(\overline{\rho}_2^{n+1} - 2\overline{\rho}_1^{n+1} + \overline{\rho}_0^{n+1} - 2\frac{\langle vL^{-1}\left(I-\Pi\right)(v^+M)\rangle_V}{\langle vL^{-1}(vM)\rangle_V}\overline{\rho}_0^n \right.$$
$$\left. +2\frac{\langle vL^{-1}\left(I-\Pi\right)(v^+f_0^n)\rangle_V}{\langle vL^{-1}(vM)\rangle_V}\right) \tag{3.64}$$

For both choices $M(v) = 1$ on $V = [-1,1]$ and $M(v) = \frac{1}{\sqrt{2\pi}}e^{-v^2/2}$ on $V = \mathbb{R}$, we observe that $2\frac{\left\langle vL^{-1}(I-\Pi)(v^+M)\right\rangle_V}{\langle vL^{-1}vM\rangle_V} = 1$. Hence, for these cases with constant in time boundary conditions, the above equation simplifies as:

$$\rho_1^{n+1} = \rho_1^n - \frac{\Delta t}{\Delta x^2}\langle vL^{-1}(vM)\rangle_V \left(\overline{\rho}_2^{n+1} - 2\overline{\rho}_1^{n+1} + 2\frac{\langle vL^{-1}\left(I-\Pi\right)(v^+f_0^n)\rangle_V}{\langle vL^{-1}(vM)\rangle_V}\right) \tag{3.65}$$

where $\tilde{\rho}_0 = 2\frac{\left\langle vL^{-1}(I-\Pi)(v^+f_0^n)\right\rangle_V}{\langle vL^{-1}(vM)\rangle_V}$ is the boundary value of the micro-macro numerical scheme in the diffusion limit. For $M(v) = \frac{1}{\sqrt{2\pi}}e^{-v^2/2}$ on $V = \mathbb{R}$, this evaluates to $\tilde{\rho}_0 = 2\sqrt{\frac{2}{\pi}} \simeq 1.59$. In the work by [199], this value is evaluated to be $\tilde{\rho}_0 = 0.75$ for $M(v) = 1$ on $V = [-1,1]$.

3.7 Numerical results

In this section, we present the numerical validation of our high order asymptotic preserving schemes in different configurations.

3.7.1 Diffusion asymptotics

First, we check time and space accuracy for the micro-macro scheme in the diffusion limit.

3.7.1.1 Time order of accuracy

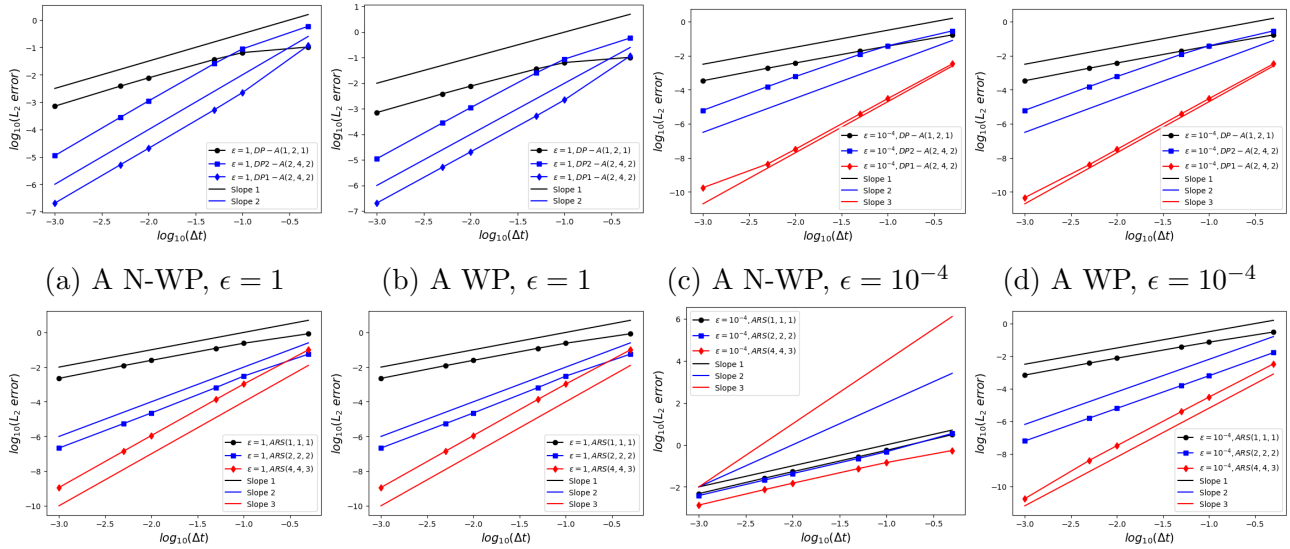
The spatial domain $L = [0, 2\pi]$ of the problem is discretized using $N_x = 50$ grid points. The velocity domain is truncated to $[-v_{\max}, v_{\max}]$ with $v_{\max} = 5$ and we take $\Delta v = 1$. The initial

condition is:

$$\rho(0,x) = 1 + \cos(x)$$
$$\text{Well-prepared data (WP): } g(0,x,v) = \epsilon^2 (I - \Pi)\left(v^2 M\right)\rho(0,x)$$
$$\text{Non-well-prepared data (N-WP): } g(0,x,v) = (I - \Pi)\left(v^2 M\right)\rho(0,x),$$

with $M(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$. Periodic boundary conditions are used on both ρ and g. The spatial terms are discretised by using the atmost-third order accurate matrices on non-staggered grid presented in section 3.5.3. The final time is $T = 0.5$, and the following Δt are considered to validate the different high order time integrators: $\Delta t = 0.5, 0.1, 0.05, 0.01, 0.005, 0.001$. The type A micro-macro schemes constructed using the Butcher tableau corresponding to DP-A$(1,2,1)$, DP2-A$(2,4,2)$ and DP1-A$(2,4,2)$ are considered. Although DP1-A$(2,4,2)$ is second order accurate, the implicit part of it when used separately is third order accurate. Further, we also consider the type CK-ARS micro-macro schemes constructed using Butcher tableau corresponding to ARS$(1,1,1)$, ARS$(2,2,2)$ and ARS$(4,4,3)$. The Butcher tableau of different time integrators utilised are presented in section 3.9.

In fig. 3.1, we plot the time error for the different time integrators in both WP and N-WP cases and for different values of ϵ. Note that the reference solution for each curve is obtained by using the same micro-macro scheme corresponding to that curve with $\Delta t = 10^{-4}$. For $\epsilon = 1$, the required orders of accuracy are recovered for type A schemes with both N-WP and WP initial data, as observed in figs. 3.1a and 3.1b. For $\epsilon = 10^{-4}$, due to the asymptotic degeneracy of our scheme into a fully-implicit scheme for diffusion equation, only the implicit part of the Butcher tableau plays a role. Hence DP1-A$(2,4,2)$ becomes third order accurate in time, while DP-A$(1,2,1)$ and DP2-A$(2,4,2)$ are first and second order accurate respectively. This is shown in figs. 3.1c and 3.1d. On the other hand, CK-ARS schemes with both N-WP and WP initial data for $\epsilon = 1$ recover the required orders of accuracy as shown in figs. 3.1e and 3.1f. However, for $\epsilon = 10^{-4}$, required orders of accuracy are observed only when WP initial data are used (fig. 3.1h). As shown in the analyses presented in previous sections, usage of N-WP initial data for CK-ARS time integrators does not allow the asymptotic accuracy (fig. 3.1g). The required order of accuracy for N-WP initial data with CK-ARS time integrator can be obtained by modifying the initial few time steps as Δt^p for p^{th} order accurate scheme as discussed in [241, 242]. On the other hand, the type A time integrators DP1-A$(2,4,2)$ and DP2-A$(2,4,2)$ that we have used do not require such initial time step reduction for maintaining the required order of accuracy.



(a) A N-WP, $\epsilon = 1$ (b) A WP, $\epsilon = 1$ (c) A N-WP, $\epsilon = 10^{-4}$ (d) A WP, $\epsilon = 10^{-4}$

(e) CK N-WP, $\epsilon = 1$ (f) CK WP, $\epsilon = 1$ (g) CK N-WP, $\epsilon = 10^{-4}$ (h) CK WP, $\epsilon = 10^{-4}$

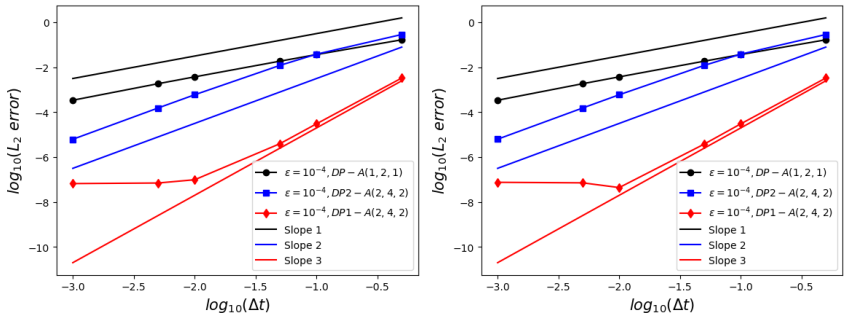
Figure 3.1: Accuracy in time for different type A and CK-ARS time integrators (both WP and N-WP initial data). The reference solution is obtained from the micro-macro with $\Delta t = 10^{-4}$.

Since we proved the asymptotic preserving property, the diffusion solution is used as reference solution in the asymptotic regime ($\epsilon = 10^{-4}$) with $\Delta t = 10^{-4}$ (in fig. 3.2) to check the orders of accuracy of high order integrators. The results are similar to the ones obtained for $\epsilon = 10^{-4}$ in fig. 3.1, except that here we observe a plateau for third order scheme and small Δt. This is due to the $\mathcal{O}(\epsilon^2)$ difference between the schemes based on micro-macro and diffusion models. This error dominates $\mathcal{O}(\Delta t^3)$ error, and hence it is observed.

3.7.1.2 Space order of accuracy

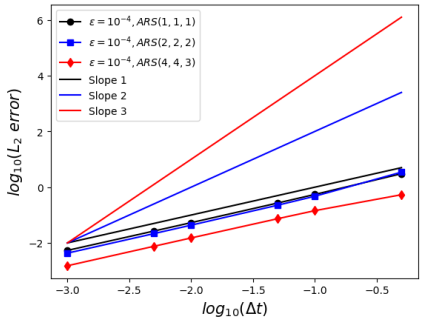
The problem set-up is the same as described in the previous subsection, except for the following changes. Here, we consider the final time to be $T = 0.01$ and $\Delta t = 0.001$ so that the error in time is small enough to study the spatial accuracy. To do so, we consider the following number of points in space: $N_x = 20, 24, 30, 40$ and 60. The reference solution is obtained with $N_x = 120$.

Since the spatial accuracy plots obtained from different time integrators are quite similar, we present only the plots obtained by using DP1-A$(2, 4, 2)$ and ARS$(4, 4, 3)$ for different values of ϵ ($\epsilon = 10^{-4}, 0.2, 1$) in figs. 3.3a and 3.3b. For the spatial discretization, we only show the results obtained by the third order spatial matrices on non-staggered grid presented in section 3.5.3 so that the scheme is expected to be third order accurate in space. In figs. 3.3a and 3.3b, the expected order is observed for the two time integrators and for the three considered values of ϵ.

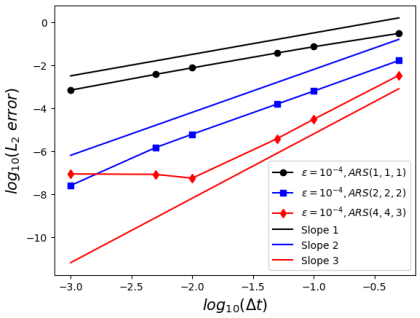


(a) A N-WP

(b) A WP

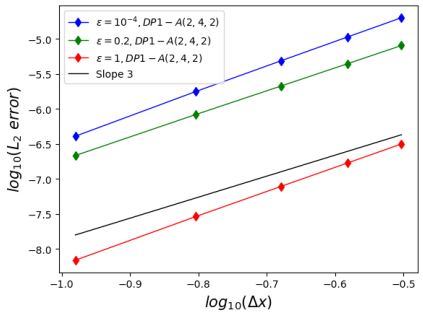


(c) CK N-WP

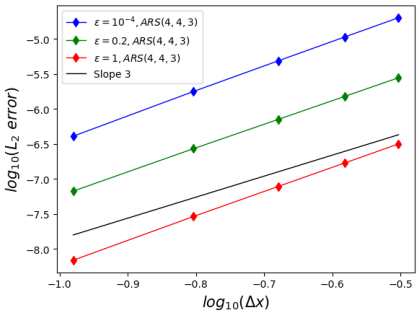


(d) CK WP

Figure 3.2: Accuracy in time for different type A and CK-ARS time integrators (both WP and N-WP initial data). The reference solution is obtained from the diffusion equation with $\Delta t = 10^{-4}$.

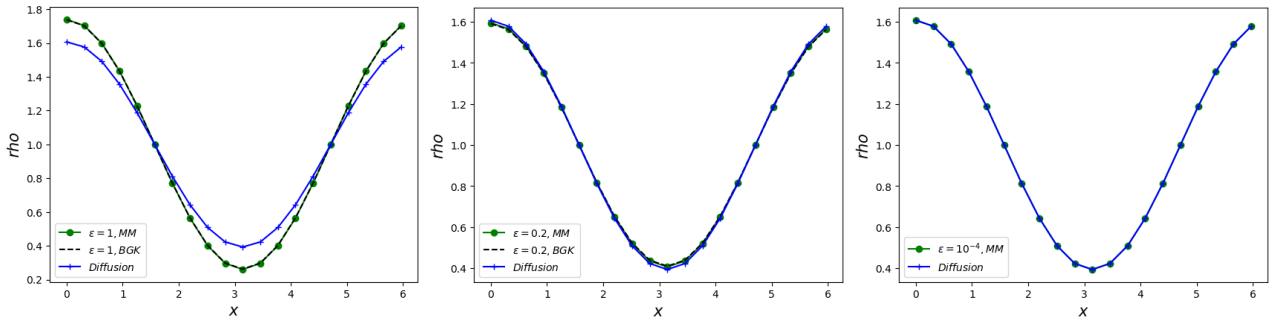


(a) A N-WP



(b) CK-ARS WP

Figure 3.3: Accuracy in space for the third order spatial scheme coupled with DP1-A$(2,4,2)$ (left) and ARS$(4,4,3)$ (right) for the time approximation.



(a) $\epsilon = 1$ (b) $\epsilon = 0.4$ (c) $\epsilon = 10^{-4}$

Figure 3.4: Qualitative results for diffusion asymptotics

3.7.1.3 Qualitative results

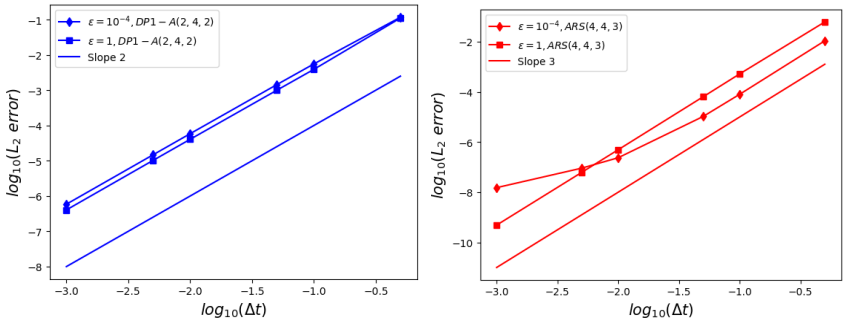
In this part, we compare the density obtained by the micro-macro equation (MM), the linear kinetic equation with BGK collision operator (BGK) and the asymptotic diffusion equation, for different values of ϵ. The MM scheme described in previous sections is utilised, the BGK is discretized using an IMEX (implicit treatment of collision term and explicit treatment of transport term) scheme whereas for the diffusion model, an implicit scheme is used. For all three models, the Butcher tableau corresponding to DP1-A$(2, 4, 2)$ time integrator is used. For the spatial discretization, we use third order scheme on non-staggered grid.

The problem domain $L = [0, 2\pi]$ is discretised using $N_x = 20$ grid points for all the three models. The final time is $T = 0.5$, and $\Delta t = 0.005$. We use the same N-WP initial and boundary conditions described in the previous subsection. Further, we also consider the same velocity discretization as before for both MM and BGK models.

In fig. 3.4a for rarefied regime ($\epsilon = 1$), the MM and BGK models compare very well, while the diffusion model is different as expected. In the intermediate regime ($\epsilon = 0.2$), the BGK and MM models match very well while the diffusion model is slightly different. For $\epsilon = 10^{-4}$, we only compare MM and the diffusion in fig. 3.4c and illustrate the AP property of the time integrators used for MM.

3.7.2 Advection-diffusion asymptotics

In this subsection, we present the time accuracy of our high order micro-macro scheme for the advection-diffusion case. As in the diffusion case, the spatial domain $L = [0, 2\pi]$ is discretised using $N_x = 20$ grid points whereas the velocity domain is $[-v_{\max}, v_{\max}]$ with $v_{\max} = 5$ and



(a) A (b) CK-ARS

Figure 3.5: Accuracy in time. Left: DP1-A$(2,4,2)$ (N-WP initial data). Right: ARS$(4,4,3)$ (WP initial data). The reference solution is obtained from the micro-macro scheme with $\Delta t = 10^{-4}$.

$\Delta v = 1$. The initial condition for the problem is:

$$\rho(0,x) = \sin(x) \tag{3.66}$$

$$\text{Well-prepared data (WP): } g(0,x,v) = \epsilon^2(I-\Pi)\left(v^2 M\right)\rho(0,x) \tag{3.67}$$

$$\text{Non-well-prepared data (N-WP): } g(0,x,v) = (I-\Pi)\left(v^2 M\right)\rho(0,x), \tag{3.68}$$

with $M(v) = \frac{1}{\sqrt{2\pi}}e^{-v^2/2}$. Periodic boundary conditions are used on both ρ and g. The spatial terms are discretised by using the atmost-first order accurate matrices on staggered grid presented in section 3.5.2. The final time is $T = 0.5$, and the following time steps are considered: $\Delta t = 0.5, 0.1, 0.05, 0.01, 0.005, 0.001$. We observe the time order of accuracy for both $\epsilon = 1$ and $\epsilon = 10^{-4}$. We choose the highest order time integrator in both type A and CK-ARS schemes for studying the time accuracy. Hence, we consider DP1-A$(2,4,2)$ and ARS$(4,4,3)$ with N-WP and WP data respectively.

Asymptotically, our micro-macro scheme degenerates to a consistent scheme for the advection-diffusion equation with advection and diffusion terms being treated explicitly and implicitly respectively. Hence, unlike the case of diffusion asymptotics for which an extra order is observed asymptotically, DP1-A$(2,4,2)$ remains second order accurate for $\epsilon = 10^{-4}$ since both explicit and implicit matrices of the Butcher tableau are involved here (fig. 3.5a). For $\epsilon = 1$, the required second order accuracy is observed. Further, the required third order accuracy of ARS$(4,4,3)$ is observed for both $\epsilon = 10^{-4}, 1$ in fig. 3.5b, since well-prepared initial data is considered.

3.7.3 Inflow boundary condition

In this subsection, the high order numerical scheme for micro-macro model that allows inflow boundary conditions is validated numerically. We first present the time accuracy results for high order schemes. Then, some qualitative plots are shown for two tests with zero inflow at the right boundary, and equilibrium and non-equilibrium inflows respectively at the left boundary.

3.7.3.1 Time order of accuracy

If the domain of the problem is a half-plane, $\omega(x,v) = \left[-v, 0, 0, \cdots\right]$ can be chosen $\forall x$ as described in [199]. Here, for numerical purposes, we consider a domain of $L = [0,2]$ and assume that the right boundary does not influence the dynamics.

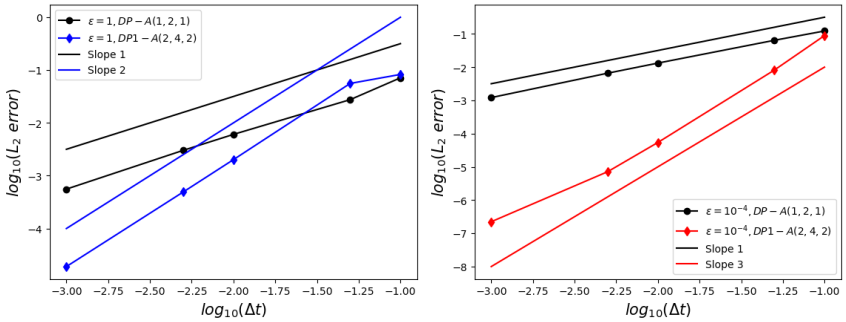
The spatial domain is discretised using $N_x = 20$ grid points and the velocity domain is $[-v_{\max}, v_{\max}]$ with $v_{\max} = 5$ and $\Delta v = 1$. The initial conditions at all interior points and right boundary conditions for the variables $\rho, \overline{\rho}$ and $\overline{g}$ are considered to be 0. The left boundary conditions (for $v_k > 0$) are:

$$f\left(t, x_i = 0, v_k\right) = M(v_k), \quad \overline{\rho}\left(t, x_i = 0\right) = 1, \quad \overline{g}(t, x_{i+1/2} = -\Delta x/2, v_k) = 0, \tag{3.69}$$

with $M(v) = \frac{1}{\sqrt{2\pi}} e^{-v^2/2}$. The final time is $T = 0.1$, and the following time steps are considered to check the accuracy in time: $\Delta t = 0.1, 0.05, 0.01, 0.005, 0.001$. Like in the previous problems, we observe the time order of accuracy for both $\epsilon = 1$ and $\epsilon = 10^{-4}$. The time integrators considered are DP-A$(1,2,1)$ and DP1-A$(2,4,2)$. The reference solution for each curve in fig. 3.6 is obtained by using the same micro-macro scheme corresponding to that curve with $\Delta t = 10^{-4}$. For type A time integrators with $\epsilon = 1$ in fig. 3.6a, first and second order accuracies of DP-A$(1,2,1)$ and DP1-A$(2,4,2)$ are observed. In fig. 3.6b for $\epsilon = 10^{-4}$, first and third order accuracies of DP-A$(1,2,1)$ and DP1-A$(2,4,2)$ respectively are observed. As for the (periodic) diffusion case, DP1-A$(2,4,2)$ turns out to be third order accurate since only the implicit part of Butcher tableau is involved asymptotically. For ARS$(2,2,2)$ and ARS$(4,4,3)$ time integrators (not shown here), order reduction to first order for $\epsilon = 1$ (due to the initial condition). However, for $\epsilon = 10^{-4}$, the required second and third orders respectively are observed.

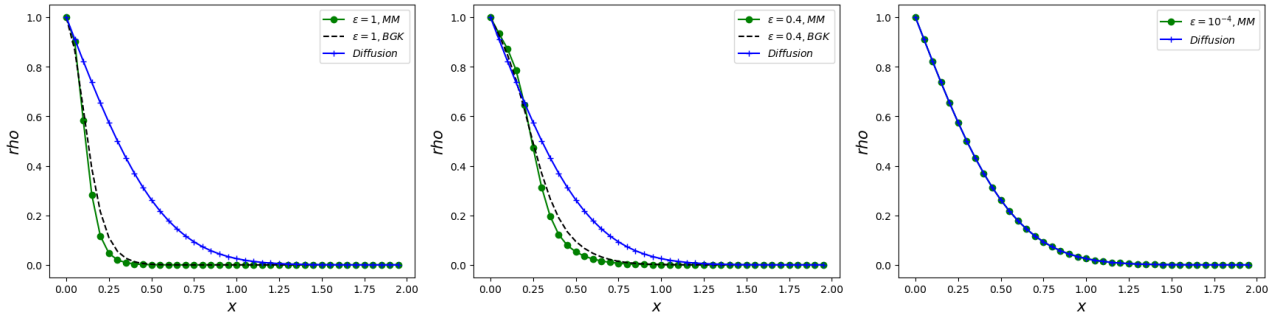
3.7.3.2 Qualitative results for equilibrium inflow

In this part, we consider the same problem as before and present a comparison of density plots obtained by using schemes based on micro-macro (MM), full-kinetic (BGK) and diffusion models, for different regimes of ϵ. The boundary conditions for diffusion model $\rho(t, x = 0) = 1$ and $\rho(t, x = 2) = 0$. The final time is $T = 0.1$, $N_x = 40$ and $\Delta t = 0.001$. Further, we consider



(a) Type A, $\epsilon = 1$ (b) Type A, $\epsilon = 10^{-4}$

Figure 3.6: Accuracy in time with type A schemes for $\epsilon = 1$ (left) and $\epsilon = 10^{-4}$ (right). The reference solution is obtained from the micro-macro for inflow boundaries scheme with $\Delta t = 10^{-4}$.



(a) $\epsilon = 1$ (b) $\epsilon = 0.4$ (c) $\epsilon = 10^{-4}$

Figure 3.7: Qualitative results for equilibrium inflow at the left boundary.

the same velocity discretization as before for both MM and BGK models. The results for MM are obtained by DP1-A$(2, 4, 2)$ time integrator.

In fig. 3.7a for rarefied regime ($\epsilon = 1$), the MM and BGK results are in good agreement. In the intermediate regime ($\epsilon = 0.4$) in fig. 3.7b, the MM and BGK results are still close, and still different from the diffusion one. For $\epsilon = 10^{-4}$, only MM and the diffusion are plotted and are found to be in very good agreement, thereby illustrating the AP property of the numerical scheme for MM.

3.7.3.3 Qualitative results for non-equilibrium inflow

In this part, we consider the same problem as before, but the left boundary condition is chosen as (for $v_k > 0$)

$$f\left(t, x_i{=}0, v_k\right) = v_k M_k, \quad \overline{\rho}\left(t, x_i{=}0\right) = \left\langle f\left(t, x_i{=}0, v_k\right)\right\rangle_{V_-} \tag{3.70}$$

$$\overline{g}\left(t, x_{i+1/2}=-\tfrac{\Delta x}{2}, v_k\right)=2\left(f\left(t, x_i=0, v_k\right)-\overline{\rho}\left(t, x_i=0\right) M_k\right)-\overline{g}\left(t, x_{i+1/2}=\tfrac{\Delta x}{2}, v_k\right). \tag{3.71}$$

The final time and time step are the same as in the previous (equilibrium inflow) case. The time integrator used is DP1-A$(2,4,2)$. Here, we present a comparison of plots obtained by using schemes based on MM, BGK and diffusion models, for different regimes of ϵ. The scheme described in subsection 3.6.2.1 is used for the micro-macro model and a standard BGK approximation where only inflow boundary condition is needed serves as a reference. For diffusion, the diffusion term is treated implicitly and the left boundary condition for diffusion model is obtained from [180] which translates in our context as:

$$\begin{aligned}\rho(t, x_i=0) = & \frac{\sum_{v_k>0} v_k f\left(t, x_i=0, v_k\right) \Delta v}{\sum_{v_k>0} v_k M_k \Delta v} \\ & + \frac{1}{\kappa \sum_{v_k} M_k \Delta v} \sum_{v_k>0} v_k^2 \left(f\left(t, x_i=0, v_k\right) - M_k \frac{\sum_{v_k>0} v_k f\left(t, x_i=0, v_k\right) \Delta v}{\sum_{v_k>0} v_k M_k \Delta v} \right) \Delta v. \end{aligned} \tag{3.72}$$

For the continuous in velocity form, the above boundary value is evaluated to be $\rho_0 = \frac{\pi+4}{2\sqrt{2\pi}} \simeq 1.42$.

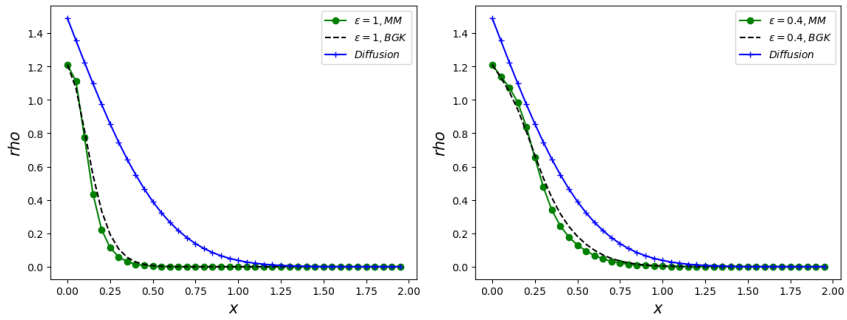
In fig. 3.8a for rarefied regime ($\epsilon = 1$), the MM and BGK models compare very well, while the diffusion model is driven by the macro boundary condition. In the intermediate regime ($\epsilon = 0.4$) in fig. 3.8b, in the MM and BGK results (which are in a good agreement), a boundary layer starts to be created whereas it is not the case for the diffusion model. For $\epsilon = 10^{-4}$, one can see from fig. 3.9 that MM model develops a boundary layer at the left boundary before aligning with the diffusion model in the interior of the domain. This is consistent with the results observed using first order schemes in literature [180, 199, 200, 75].

Further in fig. 3.10, we present the numerical results of MM and diffusion models for $\epsilon = 10^{-4}$ with refined velocity grid to demonstrate the difference between boundary values of 1.42 from (3.72) and 1.59 from micro-macro numerical scheme (derived in subsection 3.6.2.1).

3.8 Appendix: Matrix notation

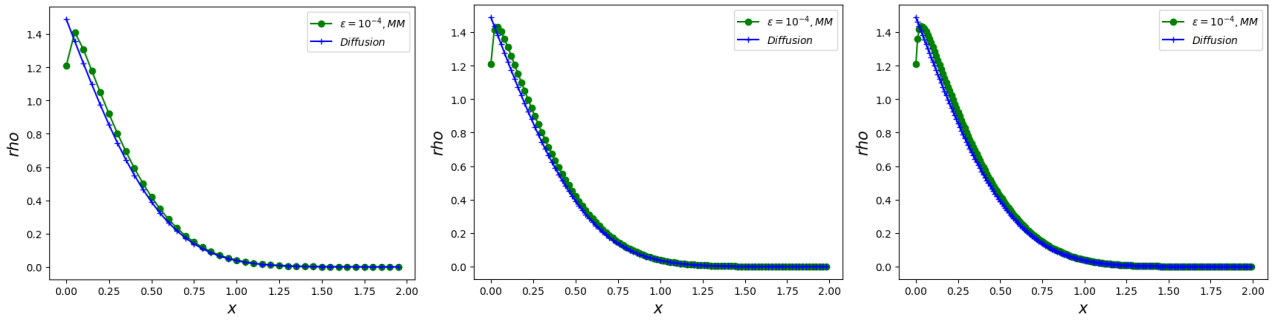
The circ function is given by:

$$\mathsf{circ}([a_1, a_2, .., \underline{a_m}, .., a_M]) = \begin{bmatrix} a_m & a_{m+1} & .. & a_M & 0 & .. & 0 & a_1 & .. & a_{m-1} \\ a_{m-1} & a_m & a_{m+1} & .. & a_M & 0 & .. & 0 & a_1 & .. \\ & & & & & & & & & \\ a_{m+2} & .. & a_M & 0 & .. & 0 & a_1 & .. & a_m & a_{m+1} \\ a_{m+1} & .. & a_M & 0 & .. & 0 & a_1 & .. & a_{m-1} & a_m \end{bmatrix} \tag{3.73}$$



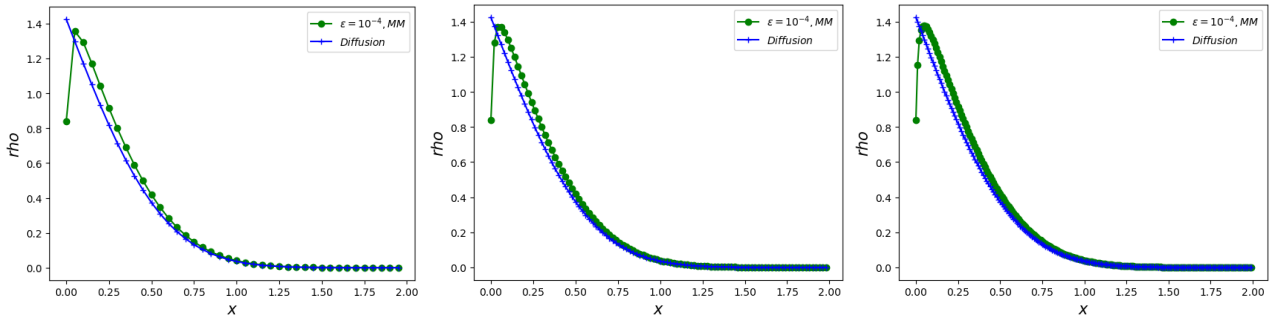
(a) $\epsilon = 1$ (b) $\epsilon = 0.4$

Figure 3.8: Qualitative results for non-equilibrium inflow at the left boundary. $\epsilon = 1, 0.4$. $v_{\max} = 5$ and $\Delta v = 1$. $N_x = 40$



(a) $\epsilon = 10^{-4}, Nx = 40$ (b) $\epsilon = 10^{-4}, Nx = 100$ (c) $\epsilon = 10^{-4}, Nx = 200$

Figure 3.9: Qualitative results for non-equilibrium inflow at the left boundary. $\epsilon = 10^{-4}$. $v_{\max} = 5$ and $\Delta v = 1$



(a) $\epsilon = 10^{-4}, Nx = 40$ (b) $\epsilon = 10^{-4}, Nx = 100$ (c) $\epsilon = 10^{-4}, Nx = 200$

Figure 3.10: Qualitative results for non-equilibrium inflow at the left boundary. $\epsilon = 10^{-4}$. $v_{\max} = 10$ and $\Delta v = 0.125$

The $\mathsf{circ_b}([-1,\underline{1}])_{(N_x-1)\times(N_x-2)}$ function is given by:

$$\mathsf{circ_b}([-1,\underline{1}])_{(N_x-1)\times(N_x-2)} = \begin{bmatrix} 1 & 0 & \ddots & 0 \\ -1 & 1 & 0 & \ddots \\ & & & \\ \ddots & \ddots & -1 & 1 \\ \ddots & \ddots & \ddots & -1 \end{bmatrix}_{(N_x-1)\times(N_x-2)} \tag{3.74}$$

3.9 Appendix: Butcher tableau

The following is the 2-stage second order accurate Butcher tableau ARS$(2,2,2)$:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & \delta & 1-\delta & 0 \\ \hline & \delta & 1-\delta & 0 \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}$$

Here, $\gamma = 1 - \frac{1}{\sqrt{2}}$ and $\delta = 1 - \frac{1}{2\gamma}$.
The following is the 4-stage third order accurate Butcher tableau ARS$(4,4,3)$:

$$\begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 2/3 & 11/18 & 1/18 & 0 & 0 & 0 \\ 1/2 & 5/6 & -5/6 & 1/2 & 0 & 0 \\ 1 & 1/4 & 7/4 & 3/4 & -7/4 & 0 \\ \hline & 1/4 & 7/4 & 3/4 & -7/4 & 0 \end{array} \qquad \begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 & 0 \\ 2/3 & 0 & 1/6 & 1/2 & 0 & 0 \\ 1/2 & 0 & -1/2 & 1/2 & 1/2 & 0 \\ 1 & 0 & 3/2 & -3/2 & 1/2 & 1/2 \\ \hline & 0 & 3/2 & -3/2 & 1/2 & 1/2 \end{array}$$

For type A, we use 2-stage first order accurate Butcher tableau DP-A$(1,2,1)$ ($\gamma \geq \frac{1}{2}$)

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1 & 0 \end{array} \qquad \begin{array}{c|cc} \gamma & \gamma & 0 \\ 1 & 1-\gamma & \gamma \\ \hline & 1-\gamma & \gamma \end{array}$$

The following is the 4-stage second order accurate Butcher tableau DP2-A$(2,4,2)$:

$$\begin{array}{c|cccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1/2 & 1/2 & 0 \\ \hline & 0 & 1/2 & 1/2 & 0 \end{array} \qquad \begin{array}{c|cccc} \gamma & \gamma & 0 & 0 & 0 \\ 0 & -\gamma & \gamma & 0 & 0 \\ 1 & 0 & 1-\gamma & \gamma & 0 \\ 1 & 0 & 1/2 & 1/2-\gamma & \gamma \\ \hline & 0 & 1/2 & 1/2-\gamma & \gamma \end{array}$$

The following is the 4-stage second order accurate Butcher tableau DP1-A$(2, 4, 2)$ which achieves third order accuracy on the DIRK part:

0	0	0	0	0
1/3	1/3	0	0	0
1	1	0	0	0
1	1/2	0	1/2	0
	1/2	0	1/2	0

1/2	1/2	0	0	0
2/3	1/6	1/2	0	0
1/2	−1/2	1/2	1/2	0
1	3/2	−3/2	1/2	1/2
	3/2	−3/2	1/2	1/2

Chapter 4

An asymptotic preserving scheme satisfying entropy stability for the barotropic Euler system

In this chapter, we study structure-preserving numerical methods for low Mach number barotropic Euler equations. Besides their asymptotic preserving properties that are crucial in order to obtain uniformly consistent and stable approximations of the Euler equations in their singular limit as the Mach number approaches 0, our aim is to preserve discrete entropy stability. Suitable acoustic/advection splitting approach combined with time implicit-explicit approximations are used to achieve the asymptotic preserving property. The entropy stability of different space discretisation strategies is studied for different values of Mach number and is validated by the numerical experiments.

4.1 Introduction

Many problems arising in science and engineering often contain dimensionless parameters that appear when suitable non-dimensionalisation is employed. For barotropic/full Euler systems, the parameter ϵ (Mach number) dictates whether the flow is compressible ($\epsilon > 1$) or incompressible ($\epsilon << 1$). It has been rigorously proved in [178, 179, 277] that the solutions of hyperbolic system converge to those of mixed hyperbolic-elliptic incompressible system when ϵ approaches zero. Explicit numerical methods for these systems require restrictive $\epsilon-$dependent stability condition on time step, and hence they become computationally very expensive when ϵ becomes small. Further, Godunov-type compressible flow solvers suffer from loss of accuracy as numerical dissipation is inversely proportional to ϵ [82]. Fully implicit numerical methods, on the other

hand, are very complicated to implement due to non-linearity of the Euler systems. Hence, attempts were made to efficiently extend the compressible flow solvers to incompressible/low Mach number limit consisting of divergence-free constraint on velocity field. In particular, semi-implicit time stepping techniques allow for the compressible solver to transform into an incompressible solver as ϵ becomes small, and its stability requirements are independent of ϵ. Such schemes are called asymptotic preserving (AP) schemes, as firstly introduced by Jin [167] for kinetic equations and later extended to hyperbolic systems (see [169] for review). Semi-implicit time stepping is often achieved by implicit-explicit (IMEX) approach involving implicit treatment of stiff terms and explicit treatment of non-stiff terms. Several IMEX-AP schemes have been formulated by using different strategies to split the flux into stiff and non-stiff parts, and we refer the interested reader to [81, 297, 68, 143, 33, 228, 34, 332, 41, 39, 89, 90, 331, 10]. In addition to asymptotic preserving properties, another crucial property of a numerical method is its stability. For the Euler equations, this means non-linear stability dictated by the second law of thermodynamics, the entropy inequality. Consequently, entropy stability has emerged as a non-linear stability criterion for numerical schemes since the seminal work of Tadmor [293, 294, 295]. Several entropy stable numerical methods for different hyperbolic systems have been developed. These include developments specific to shallow water equations [124, 323, 236, 5], the Euler equations [21, 158, 251, 56, 259, 260, 125, 73, 61, 328], and magnetohydrodynamics equations [57]. However, these entropy stable schemes were proposed for fixed Mach number ϵ being order one. On the other hand, the governing system exhibits entropy inequality for all non-zero values of ϵ. Hence, our aim in this chapter is to develop a numerical scheme that is entropy stable for different values of ϵ, and AP as ϵ approaches zero. As fas as we are aware, this question has not yet been studied in literature. The present chapter makes the first step in this research direction, discusses possible discretisation strategies, and validates them by a series of numerical experiments.

The chapter is organised as follows: Section 4.2 presents the barotropic Euler system, its entropy stability property for different values of Mach number ϵ, and its asymptotic limit as ϵ approaches zero. Section 4.3 presents the numerical method that employs an IMEX-AP time discretisation in the spirit of [89, 39], and three different space discretisation strategies. The asymptotic preserving property of fully discrete scheme is also presented. Section 4.4 presents the numerical validation of our scheme by depicting the AP and entropy stability properties. Section 4.5 concludes the chapter.

4.2 Mathematical model

In this section, we present the barotropic Euler system, its entropy stability property, and its asymptotic limit as Mach number approaches 0.

4.2.1 The barotropic Euler system

Consider the barotropic Euler system,

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{4.1}$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \nabla p(\rho) = \mathbf{0}, \tag{4.2}$$

where $\mathbf{x} \in \Omega \subset \mathbb{R}^d$, $t \in \mathbb{R}^+ \cup \{0\}$, $\rho(\mathbf{x}, t) : \Omega \times \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^+$ is the fluid density, $\mathbf{u}(\mathbf{x}, t) : \Omega \times \mathbb{R}^+ \cup \{0\} \to \mathbb{R}^d$ is the fluid velocity, and $p(\rho(\mathbf{x}, t)) = \kappa \rho^\gamma \in \mathbb{R}^+$ is the pressure. Here, d is the dimension in space, and κ, $\gamma > 1$ are constants. This system is hyperbolic with eigenvalues (in direction $\mathbf{n}$) $\mathbf{u} \cdot \mathbf{n} - c$ and $\mathbf{u} \cdot \mathbf{n} + c$, where $c = \sqrt{\gamma p / \rho}$ is the sound speed, and the conserved quantities are density, ρ and momentum, $\rho \mathbf{u}$. The initial conditions required for the system are $\rho(\mathbf{x}, 0) = \rho^0(\mathbf{x})$ and $\mathbf{u}(\mathbf{x}, 0) = \mathbf{u}^0(\mathbf{x})$, and the boundary is considered to have periodic or zero flux conditions.

We perform non-dimensionalization of the above barotropic Euler system in (4.1) and (4.2) by using the reference values $x_r, t_r, \rho_r, u_r, p_r$. The dimensionless variables are given as,

$$\hat{\mathbf{x}} = \frac{\mathbf{x}}{x_r}, \hat{t} = \frac{t}{t_r}, \hat{\rho} = \frac{\rho}{\rho_r}, \hat{\mathbf{u}} = \frac{\mathbf{u}}{u_r}, \hat{p} = \frac{p}{p_r}. \tag{4.3}$$

Inserting these into (4.1) and (4.2) and omitting the hat, we obtain the dimensionless barotropic Euler system,

$$\partial_t \rho + \nabla \cdot (\rho \mathbf{u}) = 0 \tag{4.4}$$

$$\partial_t (\rho \mathbf{u}) + \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u}) + \frac{1}{\epsilon^2} \nabla p(\rho) = \mathbf{0}, \tag{4.5}$$

where $\epsilon = u_r \sqrt{\rho_r / p_r}$ is proportional to the Mach number. This system is also hyperbolic, and its eigenvalues (in direction $\mathbf{n}$) are $\mathbf{u} \cdot \mathbf{n} - c/\epsilon$ and $\mathbf{u} \cdot \mathbf{n} + c/\epsilon$. Hereafter, we consider the dimensionless form of barotropic Euler system in (4.4) and (4.5) for the presentation of analysis and numerical methods.

4.2.2 Entropy stability property

Most hyperbolic systems in general have entropy inequality associated with them. In this section, we present the entropy inequality corresponding to the system in (4.4) and (4.5). As we will see in what follows, the physical energy plays the role of (mathematical) entropy. Consequently, the entropy inequality reduces to the energy dissipation property.

Let $\mathbf{U} = [\rho, \rho u_1, \dots, \rho u_d]^T$ be the vector of conserved variables and $\mathbf{G}^k(\mathbf{U}) = [\rho u_k, p\delta_{k1}/\epsilon^2 + \rho u_1 u_k, \dots, p\delta_{kd}/\epsilon^2 + \rho u_d u_k]^T$ be its flux vector in k^{th} direction. Here u_i is the i^{th} component of fluid velocity $\mathbf{u}$. In this notation, the barotropic Euler system in (4.4) and (4.5) can be recast as:

$$\partial_t \mathbf{U} + \partial_{x_k} \mathbf{G}^k(\mathbf{U}) = \mathbf{0}. \tag{4.6}$$

The convex function,

$$\eta(\mathbf{U}) = \frac{1}{2}\rho \|\mathbf{u}\|_2^2 + \frac{1}{\epsilon^2}\frac{p(\rho)}{\gamma - 1} \tag{4.7}$$

is an entropy for the system (4.6) as it satisfies,

$$\eta''(\mathbf{U}) \cdot \left(\mathbf{G}^k\right)'(\mathbf{U}) \text{ is symmetric } \iff (\omega_k)'(\mathbf{U}) = \eta'(\mathbf{U}) \cdot \left(\mathbf{G}^k\right)'(\mathbf{U}). \tag{4.8}$$

Here ω_k is the k^{th} component of the entropy flux function $\boldsymbol{\omega}(\mathbf{U}) = \mathbf{u}\left(\eta(\mathbf{U}) + p(\rho)/\epsilon^2\right)$ corresponding to $\eta(\mathbf{U})$. For sufficiently smooth solutions, the inner product of (4.6) with $\eta'(\mathbf{U})$ gives entropy equality

$$\partial_t \eta(\mathbf{U}) + \partial_{x_k} \omega_k(\mathbf{U}) = 0. \tag{4.9}$$

For weak (non-smooth) solutions, we only get

$$\partial_t \eta(\mathbf{U}) + \partial_{x_k} \omega_k(\mathbf{U}) \leq 0 \tag{4.10}$$

due to the convexity of $\eta(\mathbf{U})$. Note that (4.10) is understood in the distributional sense.

4.2.3 Asymptotic limit

Our aim in this section is to define a limiting system of (4.4), (4.5) as $\epsilon \to 0$. We point out that all calculations presented below are formal assuming enough regularity of the corresponding solutions. We assume that solutions can be expanded with respect to $\epsilon-$powers as follows:

$$\rho = \rho_0 + \epsilon\rho_1 + \epsilon^2\rho_2 + \dots, \tag{4.11}$$

$$\mathbf{u} = \mathbf{u}_0 + \epsilon\mathbf{u}_1 + \epsilon^2\mathbf{u}_2 + \dots, \tag{4.12}$$

$$p = p_0 + \epsilon p_1 + \epsilon^2 p_2 + \dots \tag{4.13}$$

The asymptotic behavior as $\epsilon \to 0$ is determined by inserting (4.11), (4.12) and (4.13) into the system in (4.4) and (4.5). Balancing $\mathcal{O}(\epsilon^{-2})$ terms in the momentum conservation equation, we obtain,

$$\nabla p_0 = 0.$$

Hence, p_0 is spatially constant and is function of time alone. Since $p_0 = \kappa \rho_0^\gamma$, ρ_0 is also spatially constant and is function of time alone. Similarly balancing $\mathcal{O}(\epsilon^{-1})$ terms in the momentum conservation equation, we infer that p_1 and ρ_1 are also spatial constants and are functions of only time. Now, balancing $\mathcal{O}(1)$ terms in both mass and momentum conservation equations, we get,

$$\partial_t \rho_0 + \rho_0 \nabla \cdot \mathbf{u}_0 = 0, \tag{4.14}$$

$$\partial_t (\rho_0 \mathbf{u}_0) + \rho_0 \nabla \cdot (\mathbf{u}_0 \otimes \mathbf{u}_0) + \nabla p_2 = \mathbf{0}. \tag{4.15}$$

Here p_2 is interpreted as the hydrostatic pressure. Integrating the $\mathcal{O}(1)$ mass balance in (4.14) on Ω, we get,

$$|\Omega| \partial_t \rho_0 = -\rho_0 \int_\Omega \nabla \cdot \mathbf{u}_0 d\Omega = -\rho_0 \int_{\partial\Omega} \mathbf{u}_0 \cdot \mathbf{n} ds. \tag{4.16}$$

Taking $\mathbf{u} \cdot \mathbf{n} = 0$ on $\partial\Omega$ or considering periodic boundary conditions, we get $\int_{\partial\Omega} \mathbf{u}_0 \cdot \mathbf{n} ds = 0$. Thus, $\partial_t \rho_0 = 0$ and ρ_0 is constant in both space and time, resulting in $\nabla \cdot \mathbf{u}_0 = 0$ according to (4.14). The $\mathcal{O}(1)$ momentum balance in (4.15) therefore becomes,

$$\partial_t \mathbf{u}_0 + \nabla \cdot (\mathbf{u}_0 \otimes \mathbf{u}_0) + \frac{\nabla p_2}{\rho_0} = \mathbf{0}. \tag{4.17}$$

Similarly, integration of $\mathcal{O}(\epsilon)$ mass balance equation and usage of $\mathbf{u} \cdot \mathbf{n} = 0$ on $\partial\Omega$ or periodic boundary conditions result in $\partial_t \rho_1 = 0$, and hence ρ_1 is constant in both space and time.
Further, the initial conditions are assumed to be compatible with the equations of different orders of ϵ (such as, ϵ^{-2}, ϵ^{-1}, ϵ^0). In this chapter, we consider the well-prepared initial conditions, *i.e.*

$$\rho(\mathbf{x}, 0) = \rho^0(\mathbf{x}) = \rho_0^0 + \epsilon^2 \rho_2^0(\mathbf{x}) \tag{4.18}$$

$$u(\mathbf{x}, 0) = u^0(\mathbf{x}) = u_0^0(\mathbf{x}) + \epsilon u_1^0(\mathbf{x}) \tag{4.19}$$

such that ρ_0^0 is constant and $\nabla \cdot \mathbf{u}_0^0 = 0$.

4.3 Numerical method

In this section, we want to construct a numerical method that is both asymptotic preserving and entropy stable. That is, we expect the method to satisfy the asymptotic limits of dimensionless barotropic Euler system in (4.4) and (4.5) as $\epsilon \to 0$, and also satisfy discrete entropy inequality in different regimes of ϵ. To achieve this goal, we use implicit-explicit (IMEX) time discretisation required for attaining asymptotic consistency, and compare the entropy stability property of three different types of space discretisation in different regimes of ϵ. We also present the asymptotic preserving property of considered numerical methods.

4.3.1 Semi-discrete IMEX time discretisation

We begin with the presentation of first order IMEX time discretisation of the barotropic Euler system in (4.4) and (4.5) for clarity.

$$\rho^{n+1} = \rho^n - \Delta t_n \nabla \cdot (\rho \mathbf{u})^{n+1} \tag{4.20}$$

$$(\rho \mathbf{u})^{n+1} = (\rho \mathbf{u})^n - \Delta t_n \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n - \frac{\Delta t_n}{\epsilon^2} \nabla p(\rho)^{n+1} \tag{4.21}$$

Here, $\Delta t_n = t_{n+1} - t_n$. The mass flux $\nabla \cdot (\rho \mathbf{u})$ and the pressure term $\frac{1}{\epsilon^2} \nabla p(\rho)$ are treated implicitly, while $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})$ in the momentum flux is treated explicitly. It is important to treat the mass flux implicitly in order to get $\nabla \cdot \mathbf{u}^{n+1} = 0$ as $\mathcal{O}(1)$ constraint. Indeed, if the mass flux is treated explicitly, then the whole method would become explicit and require severe ϵ dependent time step restriction enforced by stability.
Substituting the momentum equation (4.21) in $\nabla \cdot (\rho \mathbf{u})^{n+1}$ of (4.20), we get,

$$\rho^{n+1} = \rho^n - \Delta t_n \nabla \cdot (\rho \mathbf{u})^n + \Delta t_n^2 \nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^n + \frac{\Delta t_n^2}{\epsilon^2} \Delta p(\rho)^{n+1}. \tag{4.22}$$

Since $p(\rho) = \kappa \rho^\gamma$, the presence of $\Delta p(\rho)^{n+1}$ in the above equation calls for a need to use the non-linear iterative solver to find ρ^{n+1}. To avoid the computational effort, we perform linearisation of $p(\rho)^{n+1}$ around the incompressible constant density ρ_0 as:

$$p(\rho)^{n+1} = p(\rho_0) + (\rho^{n+1} - \rho_0) p'(\rho)|_{\rho=\rho_0} + \mathcal{O}(\epsilon^4). \tag{4.23}$$

The above linearisation is true if the higher derivatives of p are $\mathcal{O}(1)$ and the method is asymptotic preserving (that is, $(\rho^{n+1} - \rho_0) \simeq \mathcal{O}(\epsilon^2)$). We intend to construct our method such that it is asymptotic preserving, and we have used this information *apriori* in the linearisation of

$p(\rho)^{n+1}$. Using this linearisation in (4.22), we get,

$$\rho^{n+1} = \rho^n - \Delta t_n \nabla \cdot (\rho \mathbf{u})^n + \Delta t_n^2 \nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^n + \frac{\Delta t_n^2}{\epsilon^2} p'(\rho)|_{\rho=\rho_0} \Delta \rho^{n+1} + \mathcal{O}(\Delta t_n^2 \epsilon^2). \tag{4.24}$$

In a crude sense, the modified or equivalent partial differential equation of the above time discrete equation is,

$$\partial_t \rho = -\nabla \cdot (\rho \mathbf{u})^{n+1} + \mathcal{O}(\Delta t_n) + \mathcal{O}(\Delta t_n \epsilon^2).$$

Thus, the first order temporal accuracy of the method remains unaffected due to the linearisation as long as $\mathcal{O}(\epsilon^2) \leq \mathcal{O}(1)$. The higher derivatives are considered to be $\mathcal{O}(1)$ in this argument. As indicated by (4.24), (4.21), we have split the part governed by the acoustic waves from the rest non-stiff part. The later models the nonlinear advection waves.
From the algorithmic viewpoint, (4.24) can be solved easily by inversion of a matrix as follows,

$$\rho^{n+1} = \left(I - \left(\frac{\Delta t_n}{\epsilon} \right)^2 p'(\rho)|_{\rho=\rho_0} \Delta \right)^{-1} \left(\rho^n - \Delta t_n \nabla \cdot (\rho \mathbf{u})^n + \Delta t_n^2 \nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^n \right). \tag{4.25}$$

Then, ρ^{n+1} evaluated as above is used to find $p(\rho)^{n+1}$. Inserting this into (4.21), we get $(\rho \mathbf{u})^{n+1}$ and thus the algorithm is complete. (4.25) and (4.21) together form the update equations for first order time semi-discrete scheme.

Next, we present the higher order IMEX Runge Kutta (IMEX-RK) time discretisation of the barotropic Euler system in (4.4) and (4.5). An IMEX-RK time discretisation is represented by the following double Butcher tableau:

$$\begin{array}{c|c} \tilde{c} & \tilde{A} \\ \hline & \tilde{b}^T \end{array} \qquad \begin{array}{c|c} c & A \\ \hline & b^T \end{array} \tag{4.26}$$

where $\tilde{A} = (\tilde{a}_{ij}), A = (a_{ij}) \in \mathbb{R}^{s \times s}$; $c, \tilde{c}, b, \tilde{b} \in \mathbb{R}^s$. The matrices $\tilde{A}, A$ correspond to explicit (strictly lower triangular matrix with diagonal elements as 0) and implicit (lower triangular with non-zero diagonal elements) parts of the scheme. Such A are known as *diagonally implicit* matrices. The coefficients $\tilde{c}$ and c are given by

$$\tilde{c}_i = \sum_{j=1}^{i-1} \tilde{a}_{ij}, c_i = \sum_{j=1}^{i} a_{ij}, \tag{4.27}$$

and the vectors $\tilde{b} = (\tilde{b}_j)$ and $b = (b_j)$ give quadrature weights that combine the stages. For

AP schemes, it turns to be important to work with globally stiffly accurate (GSA) IMEX-RK scheme that satisfies the following property:

$$c_s = \tilde{c}_s = 1 \text{ and } a_{sj} = b_j, \tilde{a}_{sj} = \tilde{b}_j, \quad \forall j \in \{1, 2, \ldots, s\}. \tag{4.28}$$

The GSA property ensures that the update at t_{n+1} is same as the update at s^{th} stage.
The i^{th} stage update (for $i \in \{1, 2, \ldots, s\}$) of the barotropic Euler system in (4.4) and (4.5) is given by,

$$\rho^i = \rho^n - \Delta t_n \sum_{j=1}^{i} a_{ij} \nabla \cdot (\rho \mathbf{u})^j \tag{4.29}$$

$$(\rho \mathbf{u})^i = (\rho \mathbf{u})^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^j - \frac{\Delta t_n}{\epsilon^2} \sum_{j=1}^{i} a_{ij} \nabla p(\rho)^j \tag{4.30}$$

where $\Delta t_n = t_{n+1} - t_n$. Substituting the momentum equation (4.30) in $\nabla \cdot (\rho \mathbf{u})^i$ of (4.29), we get,

$$\rho^i = \rho^n - \Delta t_n \sum_{j=1}^{i-1} a_{ij} \nabla \cdot (\rho \mathbf{u})^j - \Delta t_n a_{ii} \nabla \cdot (\rho \mathbf{u})^n + \Delta t_n^2 a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^j + \frac{\Delta t_n^2}{\epsilon^2} a_{ii} \sum_{j=1}^{i} a_{ij} \Delta p(\rho)^j. \tag{4.31}$$

The above equation requires a nonlinear solver to find ρ^i. Similar to first order method, we perform linearisation around incompressible constant density ρ_0:

$$p(\rho)^i = p(\rho_0) + (\rho^i - \rho_0) p'(\rho)|_{\rho=\rho_0} + \mathcal{O}(\epsilon^4). \tag{4.32}$$

The asymptotic preserving property ($(\rho^i - \rho_0) \simeq \mathcal{O}(\epsilon^2)$) of the method is used a priori in the linearisation. Plugging in (4.31) yields,

$$\rho^i = \left(I - \left(\frac{\Delta t_n}{\epsilon} \right)^2 a_{ii}^2 p'(\rho)|_{\rho=\rho_0} \Delta \right)^{-1} \left(\rho^n - \Delta t_n \sum_{j=1}^{i-1} a_{ij} \nabla \cdot (\rho \mathbf{u})^j - \Delta t_n a_{ii} \nabla \cdot (\rho \mathbf{u})^n \right.$$
$$\left. + \Delta t_n^2 a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^j + \frac{\Delta t_n^2}{\epsilon^2} a_{ii} \sum_{j=1}^{i-1} a_{ij} p'(\rho)|_{\rho=\rho_0} \Delta \rho^j \right). \tag{4.33}$$

Then, ρ^i evaluated as above is used to find $p(\rho)^i$. Inserting this into (4.30), we get $(\rho \mathbf{u})^i$ and thus the evaluation of stage values is complete. (4.33) and (4.30) together form the stage update equations for higher order IMEX-RK time semi-discrete scheme. Further, $\rho^{n+1} = \rho^s$

and $(\rho\mathbf{u})^{n+1} = (\rho\mathbf{u})^s$ due to the GSA property and therefore the algorithm is complete.

4.3.2 Asymptotic preserving property of the time semi-discrete scheme

In this section, we show that the higher order GSA IMEX-RK time semi-discrete scheme (4.31) and (4.30) is asymptotic preserving.

Theorem 4.1 *Assume well-prepared initial conditions in (4.18) and (4.19), the asymptotic expansion ansatz in (4.11)-(4.13), and periodic boundary conditions on ρ and $\mathbf{u}$. Then the time semi-discrete GSA IMEX-RK scheme given by (4.31) and (4.30) satisfies for $\epsilon \to 0$*

$$\rho_0^i \equiv \text{ constant }, \ \rho_1^i \equiv \text{ constant }, \ \rho_0^i + \epsilon\rho_1^i = \rho_0, \tag{4.34}$$

$$\nabla \cdot \mathbf{u}_0^i = 0, \tag{4.35}$$

$$\mathbf{u}_0^i = \mathbf{u}_0^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla \cdot (\mathbf{u}_0 \otimes \mathbf{u}_0)^j - \frac{\Delta t_n}{\rho_0} \sum_{j=1}^{i} a_{ij} \nabla p_2^j, \tag{4.36}$$

for all $i \in \{1, 2, \ldots, s\}$, which is a consistent approximation of the incompressible Euler system (4.14), (4.15).

Proof: Inserting the asymptotic ansatz (4.11)-(4.13) into the momentum update equation (4.30) and equating $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$ terms, we obtain

$$\Delta t_n \sum_{j=1}^{i} a_{ij} \nabla p_0^j = 0, \quad \text{for all } i \in \{1, 2, \ldots, s\} \implies \nabla p_0^i = 0, \quad \text{for all } i \in \{1, 2, \ldots, s\}.$$

Since $p_0^i = \kappa \rho_0^{i\gamma}$, ρ_0^i is spatially constant for all $i \in \{1, 2, \ldots, s\}$. Similarly equating $\mathcal{O}\left(\frac{1}{\epsilon}\right)$ terms in the momentum update equation (4.30), we infer that ρ_1^i is spatially constant for all $i \in \{1, 2, \ldots, s\}$.
Inserting the asymptotic ansatz (4.11)-(4.13) into the mass update equation (4.31) and equating $\mathcal{O}(1)$ terms, we obtain

$$\rho_0^i = \rho_0^n - \Delta t_n \sum_{j=1}^{i-1} a_{ij} \rho_0^j \nabla \cdot (\mathbf{u}_0)^j - \Delta t_n a_{ii} \rho_0^n \nabla \cdot (\mathbf{u}_0)^n + \Delta t_n^2 a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} \rho_0^j \nabla^2 : (\mathbf{u}_0 \otimes \mathbf{u}_0)^j + \Delta t_n^2 a_{ii} \sum_{j=1}^{i} a_{ij} \Delta p_2^j. \tag{4.37}$$

Integrating the above equation on Ω and using periodic boundary conditions on ρ_2 and $\mathbf{u}_0$, we obtain

$$\rho_0^i = \rho_0^n, \quad \text{for all } i \in \{1, 2, \ldots, s\}. \tag{4.38}$$

Repeating the similar procedure for $\mathcal{O}(\epsilon)$ terms of the mass update equation (4.31), we obtain,

$$\rho_1^i = \rho_1^n, \quad \text{for all } i \in \{1, 2, \ldots, s\}. \tag{4.39}$$

Since $\rho_{0,1}^{n+1} = \rho_{0,1}^s$ due to the GSA property of IMEX-RK time discretisation, we have $\rho_{0,1}^{n+1} = \rho_{0,1}^s = \rho_{0,1}^n$. Therefore, $\rho_{0,1}^n = \rho_{0,1}^0 \equiv$ constant, for all $n = 1, 2, \ldots$.
Inserting this into the $\mathcal{O}(1)$ mass and momentum update equations, we get for all $i \in \{1, 2, \ldots, s\}$

$$\sum_{j=1}^{i-1} a_{ij} \nabla \cdot (\mathbf{u}_0)^j + a_{ii} \nabla \cdot (\mathbf{u}_0)^n - \Delta t_n a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla^2 : (\mathbf{u}_0 \otimes \mathbf{u}_0)^j - \frac{\Delta t_n}{\rho_0} a_{ii} \sum_{j=1}^{i} a_{ij} \Delta p_2^j = 0, \tag{4.40}$$

$$\mathbf{u}_0^i = \mathbf{u}_0^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla \cdot (\mathbf{u}_0 \otimes \mathbf{u}_0)^j - \frac{\Delta t_n}{\rho_0} \sum_{j=1}^{i} a_{ij} \nabla p_2^j, \tag{4.41}$$

where $\rho_0 = \rho_0^0 + \epsilon \rho_1^0$. Taking divergence of (4.41) and inserting it into (4.40), we obtain

$$\sum_{j=1}^{i} a_{ij} \nabla \cdot (\mathbf{u}_0)^j = 0, \quad \text{for all } i \in \{1, 2, \ldots, s\} \implies \nabla \cdot (\mathbf{u}_0)^i = 0, \quad \text{for all } i \in \{1, 2, \ldots, s\}. \tag{4.42}$$

□

The above theorem shows the asymptotic consistency of the IMEX-RK time semi-discrete scheme. Due to the GSA property, the expressions for $\rho_0^s, \rho_1^s, u_0^s$ follow for $\rho_0^{n+1}, \rho_1^{n+1}, u_0^{n+1}$. In the next section, we explain the discretisation techniques for spatial derivatives present in the scheme.

4.3.3 Space discretisation

In this section, we discuss various consistent spatial discretisations for the time semi-discrete scheme proposed above. It is important to keep the numerical diffusion coefficients free of the small parameter ϵ, in-order to avoid an uncontrollable growth in numerical diffusion term as $\epsilon \to 0$. In what follows, we present three different types of space discretisation.
We consider the first order time semi-discrete scheme given by (4.25) and (4.21) for presentation of the spatial discretisation. The discretisation of all the spatial derivatives present in this scheme will be explained. The corresponding spatial derivatives in higher order time semi-discrete scheme given by (4.33) and (4.30) will be approximated analogously. The additional terms $\sum_{j=1}^{i-1} a_{ij} \nabla \cdot (\rho \mathbf{u})^j$ and $\sum_{j=1}^{i-1} a_{ij} \Delta p(\rho)^j$ present in the mass update equation (given by (4.33)) will also follow the same discretisation as $\nabla \cdot (\rho \mathbf{u})^n$ and $\Delta p(\rho)^{n+1}$ respectively. For convenience of presentation, we explain the ideas for one-dimensional setting.

4.3.3.1 Type 1

We apply an upwind discretisation for $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in the momentum equation (4.21), while all other first and second derivatives present in the scheme (4.25) and (4.21) are treated in a central fashion. Since our goal is to achieve entropy stability, we do not add numerical diffusion to implicit terms as they are entropy stable with central discretisation.

1. Spatial discretisation of $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21) is given by upwind discretisation

$$D_{upw}(\rho \mathbf{u} \otimes \mathbf{u})_k = \frac{1}{\Delta x}\left((\rho \mathbf{u}\mathbf{u})_{k+\frac{1}{2}} - (\rho \mathbf{u}\mathbf{u})_{k-\frac{1}{2}}\right) \tag{4.43}$$

$$(\rho \mathbf{u}\mathbf{u})_{k+\frac{1}{2}} = \begin{cases} (\rho \mathbf{u})_k \mathbf{u}_{k+\frac{1}{2}} & \text{if } \mathbf{u}_{k+\frac{1}{2}} > 0 \\ (\rho \mathbf{u})_{k+1} \mathbf{u}_{k+\frac{1}{2}} & \text{if } \mathbf{u}_{k+\frac{1}{2}} < 0 \end{cases} \tag{4.44}$$

   where $\mathbf{u}_{k+\frac{1}{2}} = \frac{1}{2}(\mathbf{u}_k + \mathbf{u}_{k+1})$, $k = 1, 2, \ldots$ is an index of spatial discretisation.

2. Spatial discretisation of $\nabla \cdot (\rho \mathbf{u})^n$ in mass update equation (4.25) and $\nabla p(\rho)^{n+1}$ in momentum equation (4.21) are given by central finite difference discretisations

$$D_{cen}(\rho \mathbf{u})_k = \frac{1}{2\Delta x}\left((\rho \mathbf{u})_{k+1} - (\rho \mathbf{u})_{k-1}\right) \tag{4.45}$$

$$D_{cen}(p)_k = \frac{1}{2\Delta x}\left(p_{k+1} - p_{k-1}\right). \tag{4.46}$$

3. Spatial discretisation of $\nabla^2 : (\rho \mathbf{u} \otimes \mathbf{u})^n$ and Δp^{n+1} in mass update equation (4.25) are given by central finite difference discretisations

$$D^2(\rho \mathbf{u} \otimes \mathbf{u})_k = \frac{1}{\Delta x^2}\left((\rho \mathbf{u}\mathbf{u})_{k+1} - 2(\rho \mathbf{u}\mathbf{u})_k + (\rho \mathbf{u}\mathbf{u})_{k-1}\right) \tag{4.47}$$

$$D^2(p)_k = \frac{1}{\Delta x^2}\left(p_{k+1} - 2p_k + p_{k-1}\right), k = 1, 2, \ldots. \tag{4.48}$$

The corresponding terms in higher order time semi-discrete scheme given by (4.33) and (4.30) also follow the same discretisation as described above.

4.3.3.2 Type 2

All the terms in (4.25) and (4.21) follow the discretisation in type 1, except that the term $\nabla \cdot (\rho \mathbf{u})^n$ in the mass update equation (4.25) is treated in an upwind fashion. Spatial discretisation

for $\nabla \cdot (\rho \mathbf{u})^n$ is given by,

$$D_{upw}(\rho \mathbf{u})_k = \frac{1}{\Delta x}\left((\rho \mathbf{u})_{k+\frac{1}{2}} - (\rho \mathbf{u})_{k-\frac{1}{2}}\right) \tag{4.49}$$

$$(\rho \mathbf{u})_{k+\frac{1}{2}} = \begin{cases} \rho_k \mathbf{u}_{k+\frac{1}{2}} & \text{if } \mathbf{u}_{k+\frac{1}{2}} > 0 \\ \rho_{k+1} \mathbf{u}_{k+\frac{1}{2}} & \text{if } \mathbf{u}_{k+\frac{1}{2}} < 0 \end{cases} \tag{4.50}$$

where $\mathbf{u}_{k+\frac{1}{2}} = \frac{1}{2}(\mathbf{u}_k + \mathbf{u}_{k+1})$. The corresponding terms in higher order time semi-discrete scheme given by (4.33) and (4.30) also follow the same discretisation as described above.

4.3.3.3 Type 3

Here, all the terms in (4.25) and (4.21) follow the discretisation in type 1, except that the term $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21) is discretised by using an entropy stable flux. We apply the entropy conservative flux discretisation, and add numerical diffusion for attaining entropy stability. We first derive the entropy conserving and stable fluxes for the full barotropic Euler system, and use only the part of these fluxes corresponding to $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21). All other first and second derivatives in the scheme (4.25) and (4.21) are treated in central fashion as in type 1.

In one space dimension, the entropy variable corresponding to the convex entropy function (4.7) is:

$$\mathbf{V} = \frac{\partial \eta}{\partial \mathbf{U}} = \begin{bmatrix} -\frac{1}{2}u_1^2 + \frac{1}{\epsilon^2}\frac{\kappa\gamma}{\gamma-1}\rho^{\gamma-1} & u_1 \end{bmatrix}^T \tag{4.51}$$

The one-dimensional flux function is, $\mathbf{G} = \begin{bmatrix} \rho u_1 & \frac{1}{\epsilon^2}p + \rho u_1^2 \end{bmatrix}^T$. For entropy conservation, we require (cf. [293, 294, 295])

$$[\mathbf{V} \cdot \mathbf{G}]_{k+\frac{1}{2}} - [\mathbf{V}]_{k+\frac{1}{2}} \cdot \mathbf{G}^*_{k+\frac{1}{2}} = [\omega]_{k+\frac{1}{2}} \tag{4.52}$$

where $[\cdot]_{k+\frac{1}{2}} = (\cdot)_{k+1} - (\cdot)_k$. Thus, we get

$$[\mathbf{V} \cdot \mathbf{G}]_{k+\frac{1}{2}} - [\omega]_{k+\frac{1}{2}} = \left[\frac{1}{\epsilon^2}\frac{2\gamma-1}{\gamma-1}p(\rho)u_1 + \frac{1}{2}\rho u_1^3\right]_{k+\frac{1}{2}} - \left[\frac{1}{2}\rho u_1^3 + \frac{1}{\epsilon^2}\frac{\gamma}{\gamma-1}p(\rho)u_1\right]_{k+\frac{1}{2}} \tag{4.53}$$

$$= \left[\frac{1}{\epsilon^2}p(\rho)u_1\right]_{k+\frac{1}{2}}, \tag{4.54}$$

$$[\mathbf{V}]_{k+\frac{1}{2}} \cdot \mathbf{G}^*_{k+\frac{1}{2}} = -\frac{1}{2}\left[u_1^2\right]_{k+\frac{1}{2}}(\rho u_1)^*_{k+\frac{1}{2}} + \frac{1}{\epsilon^2}\frac{\kappa\gamma}{\gamma-1}\left[\rho^{\gamma-1}\right]_{k+\frac{1}{2}}(\rho u_1)^*_{k+\frac{1}{2}} + \frac{1}{\epsilon^2}\left[u_1\right]_{k+\frac{1}{2}}p^*_{k+\frac{1}{2}}$$

$$+ \left[u_1\right]_{k+\frac{1}{2}} \left(\rho u_1^2\right)^*_{k+\frac{1}{2}}. \quad (4.55)$$

The following interface flux function

$$\mathbf{G}^*_{k+\frac{1}{2}} = \begin{bmatrix} (\rho u_1)^*_{k+\frac{1}{2}} \\ \left(\frac{1}{\epsilon^2}p + \rho u_1^2\right)^*_{k+\frac{1}{2}} \end{bmatrix} = \begin{bmatrix} \overline{\rho}_{\gamma_{k+\frac{1}{2}}} \overline{u_1}_{k+\frac{1}{2}} \\ \frac{1}{\epsilon^2}\overline{p}_{k+\frac{1}{2}} + \overline{\rho}_{\gamma_{k+\frac{1}{2}}} \overline{u_1}^2_{k+\frac{1}{2}} \end{bmatrix} \quad (4.56)$$

with

$$\overline{\rho}_{\gamma_{k+\frac{1}{2}}} = \frac{\gamma-1}{\gamma}\frac{[\rho^\gamma]_{k+\frac{1}{2}}}{[\rho^{\gamma-1}]_{k+\frac{1}{2}}}, \; \overline{u_1}_{k+\frac{1}{2}} = \frac{1}{2}\left(u_{1_{k+1}} + u_{1_k}\right), \overline{p}_{k+\frac{1}{2}} = \frac{1}{2}\left(p_{k+1} + p_k\right) \quad (4.57)$$

satisfies the entropy conserving condition in (4.52). Hence, the entropy conserving spatial discretisation of $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21) is given by

$$D_{EC}(\rho \mathbf{u} \otimes \mathbf{u})_k = \frac{1}{\Delta x}\left((\rho u_1^2)^*_{k+\frac{1}{2}} - (\rho u_1^2)^*_{k-\frac{1}{2}}\right) \quad (4.58)$$

$$(\rho u_1^2)^*_{k+\frac{1}{2}} = \overline{\rho}_{\gamma_{k+\frac{1}{2}}} \overline{u_1}^2_{k+\frac{1}{2}}. \quad (4.59)$$

Note that (4.58), (4.59) yield second order accurate approximation.
To achieve entropy stability, we consider a dissipation matrix that is independent of ϵ:

$$\Lambda = \begin{bmatrix} |\overline{u_1}_{k+\frac{1}{2}}| & 0 \\ 0 & |\overline{u_1}_{k+\frac{1}{2}}| \end{bmatrix}. \quad (4.60)$$

Thus, the entropy stable flux for full barotropic Euler system becomes,

$$\mathbf{G}_{k+\frac{1}{2}} = \mathbf{G}^*_{k+\frac{1}{2}} - \frac{q}{2}\Lambda \left[\mathbf{V}\right]_{k+\frac{1}{2}} \quad (4.61)$$

where $q > 0$ is a suitable constant. This entropy stable flux results in first order spatial accuracy. For second order accuracy, we use (cf. [118])

$$\mathbf{G}_{k+\frac{1}{2}} = \mathbf{G}^*_{k+\frac{1}{2}} - \frac{q}{2}\Lambda \left\langle\langle \mathbf{V} \rangle\right\rangle_{k+\frac{1}{2}} \quad (4.62)$$

$$\left\langle\langle \mathbf{V} \rangle\right\rangle_{k+\frac{1}{2}} = \left[\mathbf{V}\right]_{k+\frac{1}{2}} - \frac{1}{2}\left(\mu\left(\left[\mathbf{V}\right]_{k+\frac{1}{2}}, \left[\mathbf{V}\right]_{k+\frac{3}{2}}\right) + \mu\left(\left[\mathbf{V}\right]_{k-\frac{1}{2}}, \left[\mathbf{V}\right]_{k+\frac{1}{2}}\right)\right) \quad (4.63)$$

$$\mu\left(A, B\right) = \begin{cases} s \min\left(|A|, |B|\right) & \text{if } s = sign(A) = sign(B) \\ 0 & \text{otherwise} \end{cases}. \quad (4.64)$$

Thus, the entropy stable spatial discretisation of $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21) is given by,

$$D_{ES}(\rho \mathbf{u} \otimes \mathbf{u})_k = \frac{1}{\Delta x}\left((\rho u_1^2)_{k+\frac{1}{2}} - (\rho u_1^2)_{k-\frac{1}{2}}\right) \tag{4.65}$$

$$(\rho u_1^2)_{k+\frac{1}{2}} = (\rho u_1^2)^*_{k+\frac{1}{2}} - \frac{q}{2}|\overline{u_1}_{k+\frac{1}{2}}| \begin{cases} [u_1]_{k+\frac{1}{2}} \\ \langle\langle u_1 \rangle\rangle_{k+\frac{1}{2}} \end{cases} = \begin{cases} |\overline{u_1}_{k+\frac{1}{2}}| \left(\overline{\rho}_{\gamma_{k+\frac{1}{2}}} |\overline{u_1}_{k+\frac{1}{2}}| - \frac{q}{2}[u_1]_{k+\frac{1}{2}}\right), & 1^{st} \text{ order} \\ |\overline{u_1}_{k+\frac{1}{2}}| \left(\overline{\rho}_{\gamma_{k+\frac{1}{2}}} |\overline{u_1}_{k+\frac{1}{2}}| - \frac{q}{2}\langle\langle u_1 \rangle\rangle_{k+\frac{1}{2}}\right), & 2^{nd} \text{ order} \end{cases} \tag{4.66}$$

Hence, in this type, $\nabla \cdot (\rho \mathbf{u} \otimes \mathbf{u})^n$ in momentum equation (4.21) is discretised as shown in (4.65)-(4.66). All the other first and second derivatives present in (4.25) and (4.21) are discretised in central fashion as shown in type 1.
The corresponding terms in higher order time semi-discrete scheme given by (4.33) and (4.30) also follow the same discretisation as described above.

4.3.4 Asymptotic preserving property of the fully discrete scheme

In this section, we show the asymptotic consistency of our fully discrete scheme as $\epsilon \to 0$. For this, we present a general theorem that considers all the three types of spatial discretisation.

Theorem 4.2 *Assume well-prepared initial conditions in (4.18) and (4.19), the asymptotic expansion ansatz in (4.11)-(4.13), and periodic boundary conditions on ρ and u_1. Consider the fully discrete scheme*

$$\rho_k^i = \rho_k^n - \Delta t_n \sum_{j=1}^{i-1} a_{ij} D(\rho u_1)_k^j - \Delta t_n a_{ii} D(\rho u_1)_k^n + \Delta t_n^2 a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} D^2(\rho u_1^2)_k^j + \frac{\Delta t_n^2}{\epsilon^2} a_{ii} \sum_{j=1}^{i} a_{ij} D^2(p)_k^j \tag{4.67}$$

$$(\rho u_1)_k^i = (\rho u_1)_k^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} D_{upw/ES} \cdot (\rho u_1^2)_k^j - \frac{\Delta t_n}{\epsilon^2} \sum_{j=1}^{i} a_{ij} D_{cen}(p)_k^j \tag{4.68}$$

with $D = D_{cen/upw}$. Then for $\epsilon \to 0$, a solution of (4.67), (4.68) satisfies

$$\rho_{0_k}^i \equiv \text{ constant }, \ \rho_{1_k}^i \equiv \text{ constant }, \ \rho_{0_k}^i + \epsilon \rho_{1_k}^i = \rho_0, \tag{4.69}$$

$$(u_{1_0})_k^i = (u_{1_0})_k^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} D_{upw/ES} (u_{1_0}^2)_k^j - \frac{\Delta t_n}{\rho_0} \sum_{j=1}^{i} a_{ij} D_{cen}(p_2)_k^j, \tag{4.70}$$

for all $k = 1, 2, \ldots,$ and all $i \in \{1, 2, \ldots, s\}$.

Proof: Substituting the asymptotic *ansatz* into the momentum update equation in (4.68) and equating $\mathcal{O}\left(\frac{1}{\epsilon^2}\right)$ terms, we get

$$\sum_{j=1}^{i} a_{ij} D_{cen}(p_0)_k^j = 0, \text{ for all } i \in \{1,2,\ldots,s\} \implies D_{cen}(p_0)_k^i = 0, \text{ for all } i \in \{1,2,\ldots,s\}. \tag{4.71}$$

Note that this property does not allow us to conclude that $(p_0)_k^i$ is spatially constant. Depending on boundary conditions, checkerboard modes could occur. In order to conclude that $(p_0)_k^i$ is spatially constant, we consider a ghost point on the left and impose $p_{0_{ghost}} = (p_0)_{k=0}$. Then, $(p_0)_k^i$ is spatially constant and hence $(\rho_0)_k^i$ is also spatially constant since $(p_0)_k^i = \kappa\left((\rho_0)_k^i\right)^\gamma$. Similarly equating $\mathcal{O}\left(\frac{1}{\epsilon}\right)$ terms in the momentum balance (4.68), we infer that $(\rho_1)_k^i$ is spatially constant.

Inserting the asymptotic *ansatz* into the mass update equation in (4.67) and equating $\mathcal{O}(1)$ terms, we obtain,

$$\rho_{0_k}^i = \rho_{0_k}^n - \Delta t_n \sum_{j=1}^{i-1} a_{ij} \rho_{0_k}^j D(u_{1_0})_k^j - \Delta t_n a_{ii} \rho_{0_k}^n D(u_{1_0})_k^n + \Delta t_n^2 a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} \rho_{0_k}^j D^2(u_{1_0}^2)_k^j + \Delta t_n^2 a_{ii} \sum_{j=1}^{i} a_{ij} D^2(p_2)_k^j. \tag{4.72}$$

Summing over all the points in the domain and using periodic boundary conditions on ρ_2 and u_{1_0}, we obtain,

$$\rho_{0_k}^i = \rho_{0_k}^n, \quad \text{for all } k, \quad \text{for all } i \in \{1,2,\ldots,s\}. \tag{4.73}$$

Repeating the similar procedure for $\mathcal{O}(\epsilon)$ terms of the mass update equation (4.67), we obtain,

$$\rho_{1_k}^i = \rho_{1_k}^n, \quad \text{for all } k, \quad \text{for all } i \in \{1,2,\ldots,s\}. \tag{4.74}$$

Since $\rho_{0,1_k}^{n+1} = \rho_{0,1_k}^s$ due to the GSA property of IMEX-RK time discretisation, we have $\rho_{0,1_k}^{n+1} = \rho_{0,1_k}^s = \rho_{0,1_k}^n$. Therefore, $\rho_{0,1_k}^n = \rho_{0,1_k}^0 \equiv$ constant, for all $n = 1,2,\ldots$.

Inserting this into the $\mathcal{O}(1)$ mass and momentum update equations, we get for all $i \in \{1,2,\ldots,s\}$,

$$\sum_{j=1}^{i-1} a_{ij} D(u_{1_0})_k^j + a_{ii} D(u_{1_0})_k^n - \Delta t_n a_{ii} \sum_{j=1}^{i-1} \tilde{a}_{ij} D^2(u_{1_0}^2)_k^j - \frac{\Delta t_n}{\rho_0} a_{ii} \sum_{j=1}^{i} a_{ij} D^2(p_2)_k^j = 0, \tag{4.75}$$

$$(u_{1_0})_k^i = (u_{1_0})_k^n - \Delta t_n \sum_{j=1}^{i-1} \tilde{a}_{ij} D_{upw/ES}(u_{1_0}^2)_k^j - \frac{\Delta t_n}{\rho_0} \sum_{j=1}^{i} a_{ij} D_{cen}(p_2)_k^j, \tag{4.76}$$

where $\rho_0 = \rho_{0_k}^0 + \epsilon \rho_{1_k}^0$, for any $k = 1,2,\ldots$. □

Due to the GSA property, the expressions for $\rho^s_{0_k}, \rho^s_{1_k}, (u_{1_0})^s_k$ follow for $\rho^{n+1}_{0_k}, \rho^{n+1}_{1_k}, (u_{1_0})^{n+1}_k$, for all $k = 1, 2, \ldots$. Thus, we have devised an asymptotic preserving IMEX-RK scheme with three different types of space discretisation techniques.

4.4 Numerical results and discussion

In this section, we present the numerical results obtained from our asymptotic preserving IMEX-RK scheme with three different types of spatial discretisation. The numerical results include: entropy, potential energy (PE), kinetic energy (KE) plots and accuracy tables of a standard periodic problem for different values of ϵ; entropy, density and momentum plots for colliding acoustic waves and Riemann problems; entropy, PE, KE and Mach ratio plots for Gresho vortex problems.

4.4.1 Standard periodic problem

The domain of the problem is $\Omega := [0, 1]$, and the initial conditions are:

$$\rho_0(x) = 1 + \epsilon^2 \sin(2\pi x) \tag{4.77}$$

$$u_{1_0}(x) = 1 + \epsilon \sin(2\pi x) \tag{4.78}$$

The parametric values are: $\kappa = 1$ and $\gamma = 2$. The entropy plots and accuracy tables of this problem will be presented for different values of ϵ.

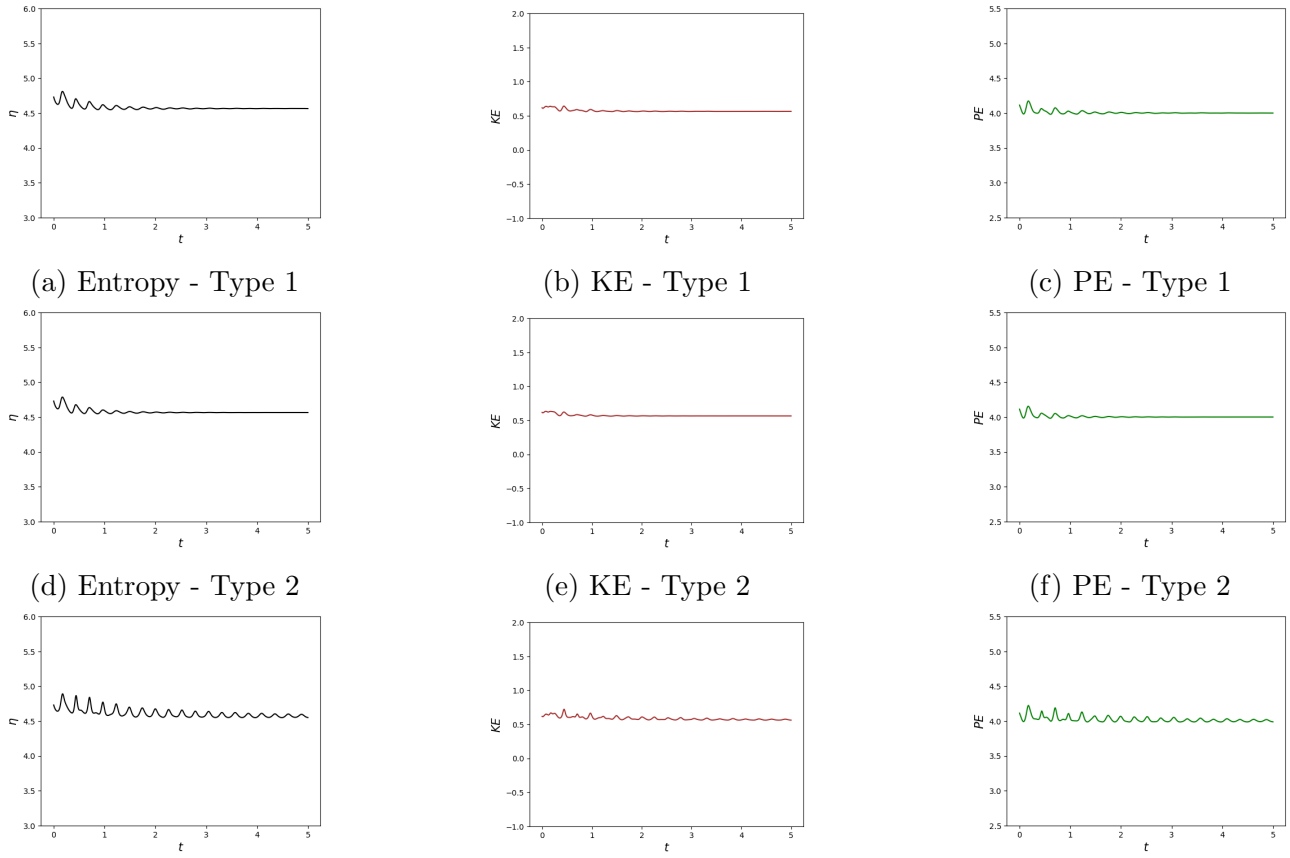
4.4.1.1 Entropy, kinetic energy (KE) and potential energy (PE)

The domain Ω is discretised into $N = 20$ grid points. The first order IMEX scheme $ARS(1, 1, 1)$ is used along with three types of spatial discretisation techniques to obtain the plots on entropy, KE and PE. The time step is chosen as:

$$\Delta t_n = C \frac{\Delta x}{2 \max\limits_{i \in \Omega_N} (u^n_{1_i})} \tag{4.79}$$

where the CFL number, $C = 0.4$ and Ω_N is the discretised set of the domain Ω (that is, $\Omega_N = \{1, 2, \ldots, N\}$).

Figures 4.1, 4.2 and 4.3 show the entropy, KE and PE plots obtained for $\epsilon = 0.5, 0.1$ and 10^{-4} respectively by using type 1, type 2 and type 3 (entropy conserving) spatial discretisation



(a) Entropy - Type 1 (b) KE - Type 1 (c) PE - Type 1

(d) Entropy - Type 2 (e) KE - Type 2 (f) PE - Type 2

(g) Entropy - Type 3 (h) KE - Type 3 (i) PE - Type 3

Figure 4.1: Entropy, KE and PE plots for $\epsilon = 0.5$ using space discretisation types 1,2 and 3

techniques. The global value of the convex entropy function at time t_n is given by,

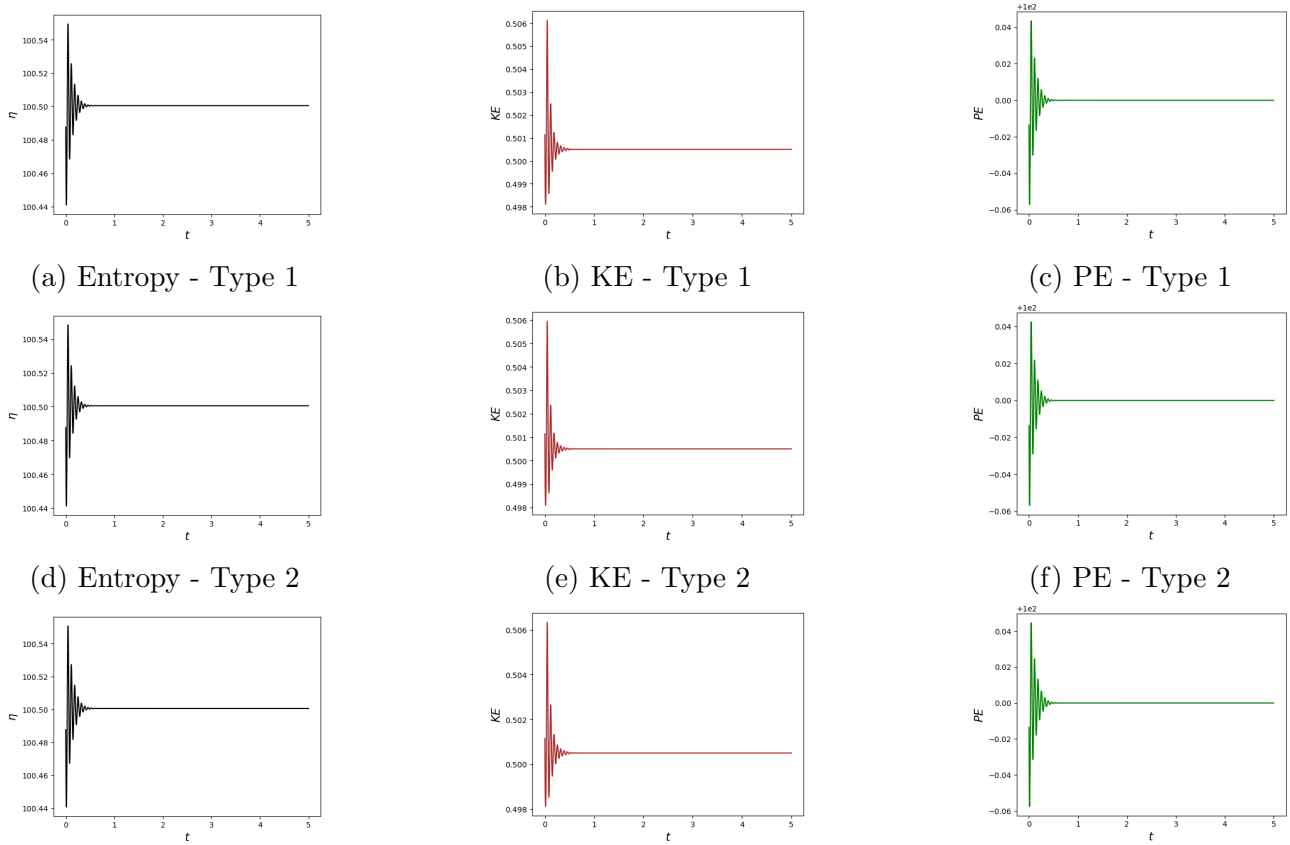
$$\eta^n = KE^n + PE^n, \text{ where } KE^n = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}\rho_k^n\left(u_{1_k}^n\right)^2, \; PE^n = \frac{1}{N}\sum_{k=1}^{N}\frac{1}{\epsilon^2}\frac{p_k^n}{\gamma-1} \tag{4.80}$$

The plots are obtained at time $T = 5$ to depict the long time behaviours of entropy, KE and PE.

$\epsilon = \mathbf{0.5}$: Decaying oscillations are observed in both KE and PE for all the three types of spatial discretisations.

$\epsilon = \mathbf{0.1}$: Both KE and PE are oscillating during the initial transience before becoming constant, for all the three types of spatial discretisations.

$\epsilon = \mathbf{0.0001}$: PE decreases while KE remains almost constant for all the three types of spatial discretisations.



(a) Entropy - Type 1 (b) KE - Type 1 (c) PE - Type 1

(d) Entropy - Type 2 (e) KE - Type 2 (f) PE - Type 2

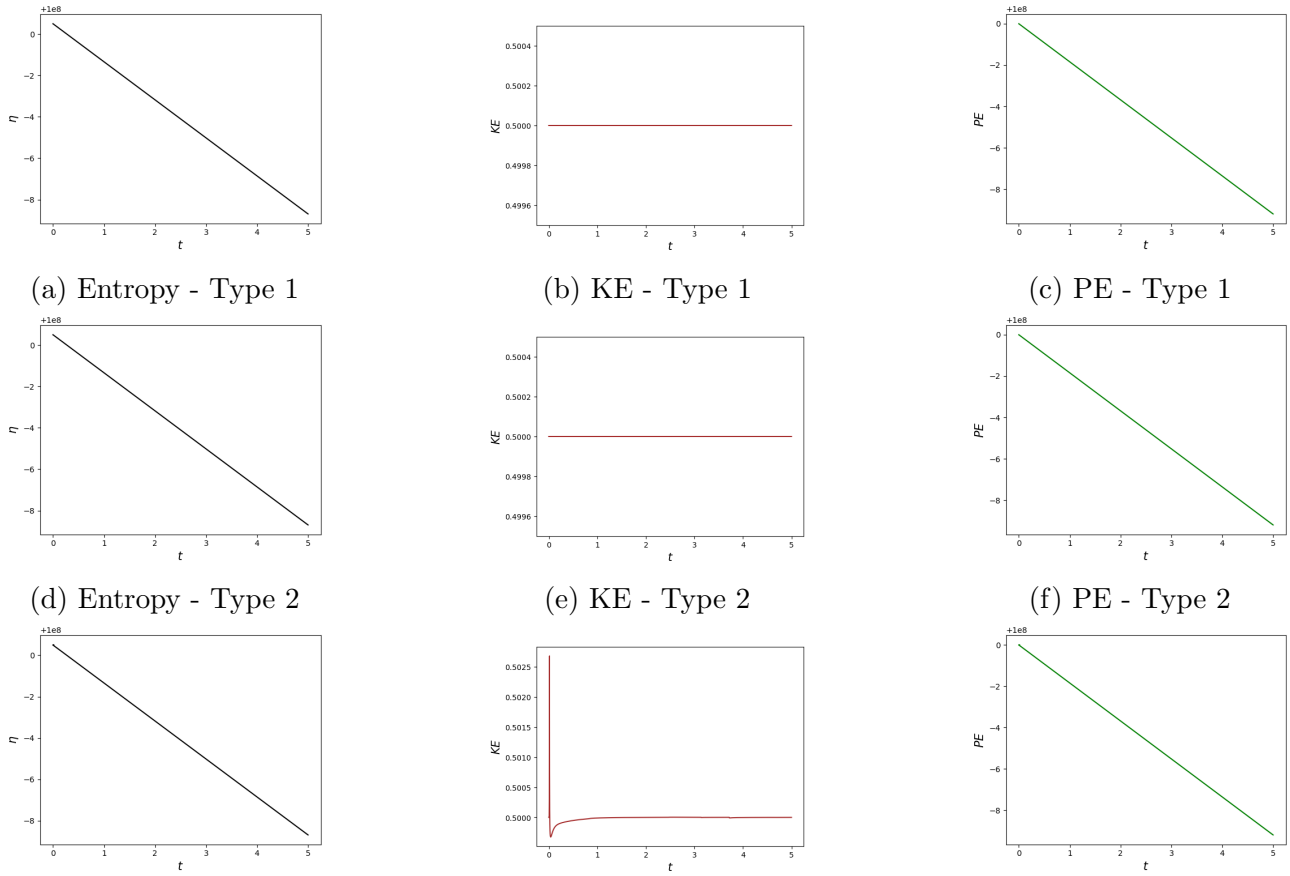
(g) Entropy - Type 3 (h) KE - Type 3 (i) PE - Type 3

Figure 4.2: Entropy, KE and PE plots for $\epsilon = 0.1$ using space discretisation types 1,2 and 3

4.4.1.2 Order of accuracy

In this subsection, we show the order of accuracy of type 3 entropy conserving spatial discretisation (that is second order accurate in space) paired with $ARS(1,1,1)$ (in table 4.1) and $ARS(2,2,2)$ (in table 4.2) IMEX time discretisations that are respectively first and second order accurate in time. We observe the order of accuracy of density ρ by using different number of grid points: $N = 20, 24, 30, 40, 60$. The reference solution is obtained with $N = 120$. The time step is chosen according to (4.79) with CFL number, $C = 0.4$.

For $\epsilon = 0.5$, both $ARS(1,1,1)$ and $ARS(2,2,2)$ give the required order of accuracy. For $\epsilon = 10^{-4}$, both $ARS(1,1,1)$ and $ARS(2,2,2)$ show very small errors of $\mathcal{O}(10^{-10})$ and $\mathcal{O}(10^{-9})$ respectively for all tested values of N. This indicates that the error in ρ is of $\mathcal{O}(\epsilon^2)$, and thus our scheme is asymptotic preserving. For $\epsilon = 0.1$, $ARS(1,1,1)$ gives required order of accuracy whereas degeneracy is observed for large N while using $ARS(2,2,2)$. Such degeneracy in the intermediate regime is commonly observed in literature due to the lack of uniform stability of



(a) Entropy - Type 1 (b) KE - Type 1 (c) PE - Type 1

(d) Entropy - Type 2 (e) KE - Type 2 (f) PE - Type 2

(g) Entropy - Type 3 (h) KE - Type 3 (i) PE - Type 3

Figure 4.3: Entropy, KE and PE plots for $\epsilon = 0.0001$ using space discretisation types 1,2 and 3

IMEX time discretisations involved.

4.4.2 Colliding acoustic waves problem

The domain of the problem is $\Omega = [-1, 1]$, and the initial conditions are:

$$\rho_0(x) = 0.955 + 0.5\epsilon\,(1 - \cos(2\pi x)) \tag{4.81}$$

$$u_{1_0}(x) = -\text{sign}(x)\sqrt{\gamma}\,(1 - \cos(2\pi x)) \tag{4.82}$$

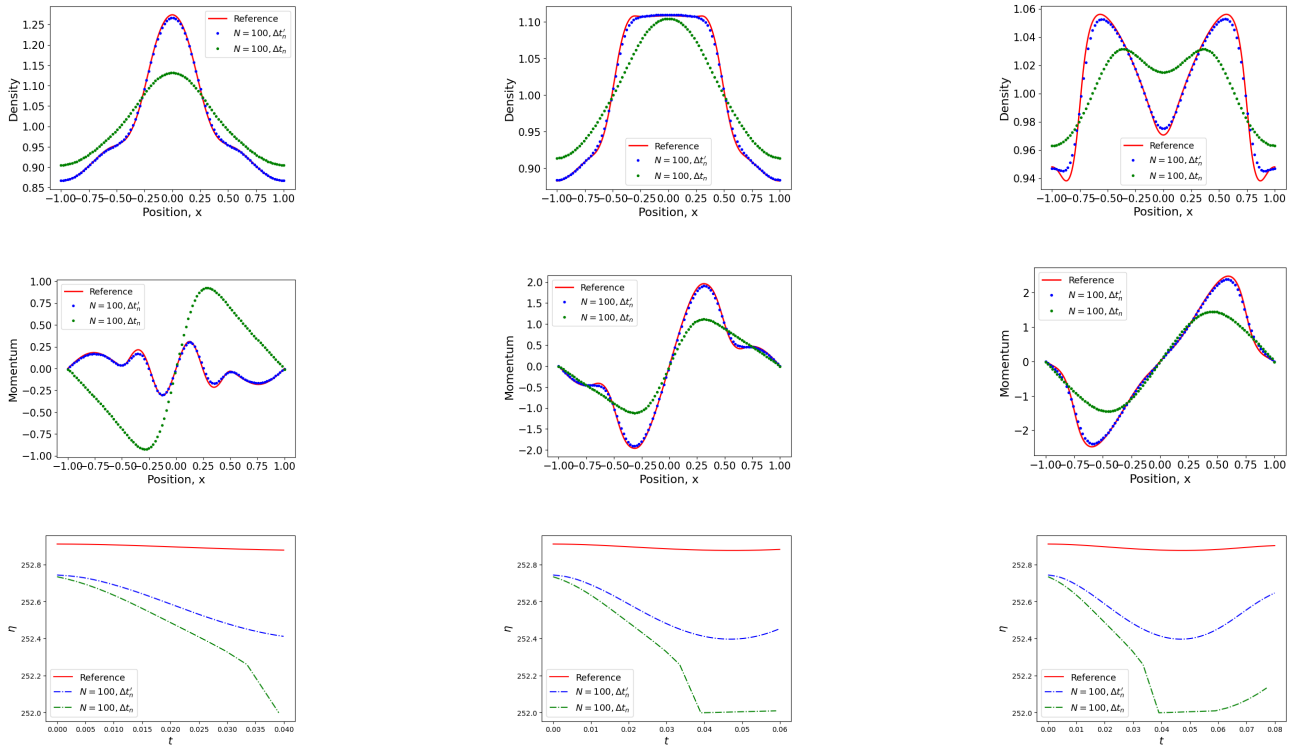
The parametric values are: $\kappa = 1$, $\gamma = 1.4$ and $\epsilon = 0.1$. It is to be noted that the initial condition is not well-prepared. Periodic boundary conditions are used for this problem. Density, momentum and global entropy (vs. time) plots are presented for different values of final time, $T = 0.04, 0.06, 0.08$. $ARS(1,1,1)$ IMEX time discretisation is used. Figures 4.4, 4.5 and 4.6 show the plots obtained by using type 1, type 2 and type 3 space discretisations respectively.

N	Δx	$\epsilon = 0.5, \|\|\rho\|\|_{L_2}$	$\epsilon = 0.5, EOC$	$\epsilon = 0.1, \|\|\rho\|\|_{L_2}$	$\epsilon = 0.1, EOC$	$\epsilon = 10^{-4}, \|\|\rho\|\|_{L_2}$	$\epsilon = 10^{-4}, EOC$
20	0.0526	0.00546	-	0.000185	-	3.9×10^{-11}	-
24	0.0435	0.00370	2.025	0.000149	1.136	3.6×10^{-11}	0.426
30	0.0435	0.00257	1.568	9.3×10^{-5}	2.015	3.1×10^{-10}	-9.28
40	0.0256	0.00143	1.992	6×10^{-5}	1.525	8.0×10^{-11}	4.549
60	0.0169	0.000521	2.432	3.6×10^{-5}	1.200	3.8×10^{-11}	1.809

Table 4.1: EOC for $ARS(1,1,1)$ coupled with spatially 2^{nd} order accurate type 3 discretisation

N	Δx	$\epsilon = 0.5, \|\|\rho\|\|_{L_2}$	$\epsilon = 0.5, EOC$	$\epsilon = 0.1, \|\|\rho\|\|_{L_2}$	$\epsilon = 0.1, EOC$	$\epsilon = 10^{-4}, \|\|\rho\|\|_{L_2}$	$\epsilon = 10^{-4}, EOC$
20	0.0526	0.00597	-	0.000201	-	2.1×10^{-9}	-
24	0.0435	0.00416	1.886	0.000111	3.078	1.9×10^{-9}	0.516
30	0.0435	0.00272	1.839	6.3×10^{-5}	2.447	1.7×10^{-9}	0.488
40	0.0256	0.00160	1.788	4.2×10^{-5}	1.413	1.5×10^{-9}	0.494
60	0.0169	0.00070	2.008	2.5×10^{-5}	1.214	1.3×10^{-9}	0.383

Table 4.2: EOC for $ARS(2,2,2)$ coupled with spatially 2^{nd} order accurate type 3 discretisation



(a) $T = 0.04$ (b) $T = 0.06$ (c) $T = 0.08$

Figure 4.4: Colliding acoustic waves problem with $\epsilon = 0.1$ using type 1 space discretisation. The reference is 2^{nd} order type 3 entropy stable space discretisation with $q = 7$, $N = 1000$ and classical time-step $\Delta t'_n$.

The reference solution for all the cases is obtained by using type 3 space discretisation which is second order accurate in space, with $N = 1000$ and the classical time step

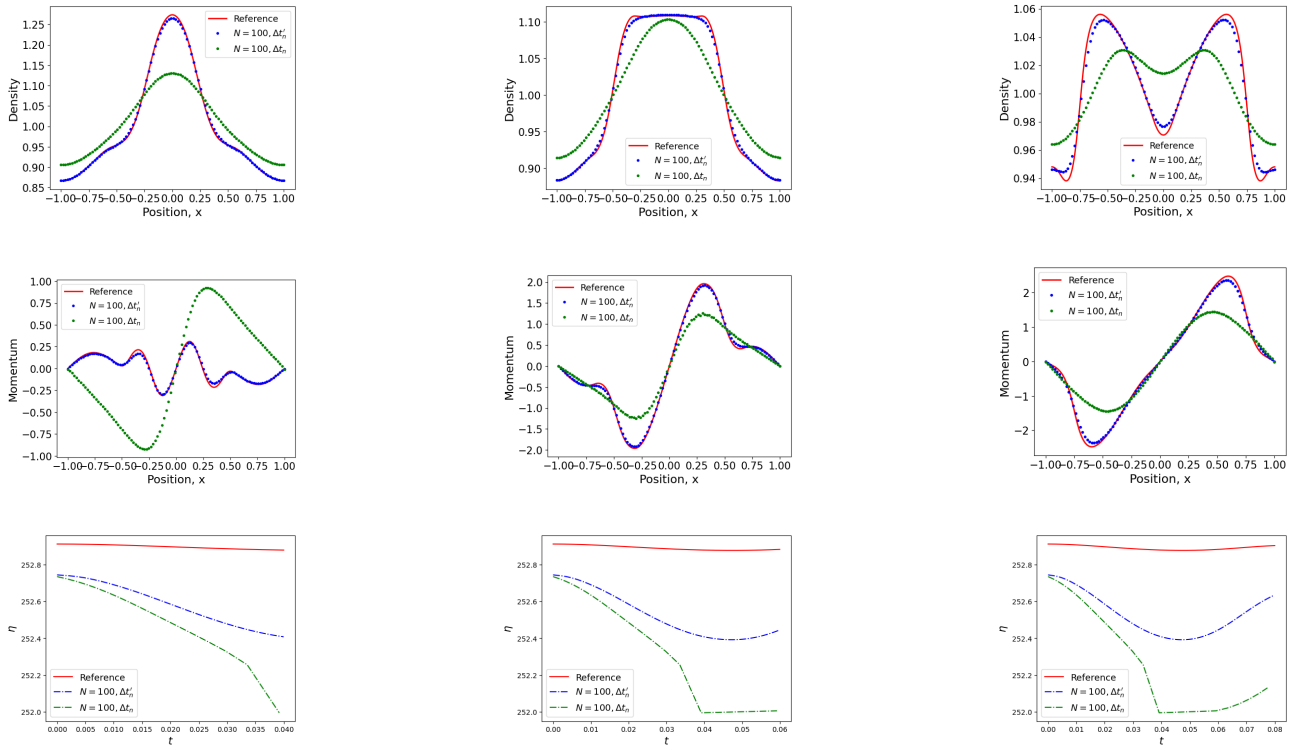
$$\Delta t'_n = C \frac{\Delta x}{2 \max\limits_{i \in \Omega_N} \left(|u^n_{1_i}| + \frac{c^n_i}{\epsilon} \right)}. \tag{4.83}$$

No significant difference is observed between the three types of space discretisation, except that the type 3 discretisation is slightly more accurate compared to the other two types owing to its second order accuracy in space.

4.4.3 Riemann problem

This problem is from [81]. The domain is $\Omega = [0, 1]$ and the initial conditions are:

$$\rho_0(x) = 1, \, (\rho u_1)_0 (x) = 1 - \epsilon^2/2, \text{ if } x \in [0, 0.2] \cup [0.8, 1] \tag{4.84}$$



(a) $T = 0.04$ (b) $T = 0.06$ (c) $T = 0.08$

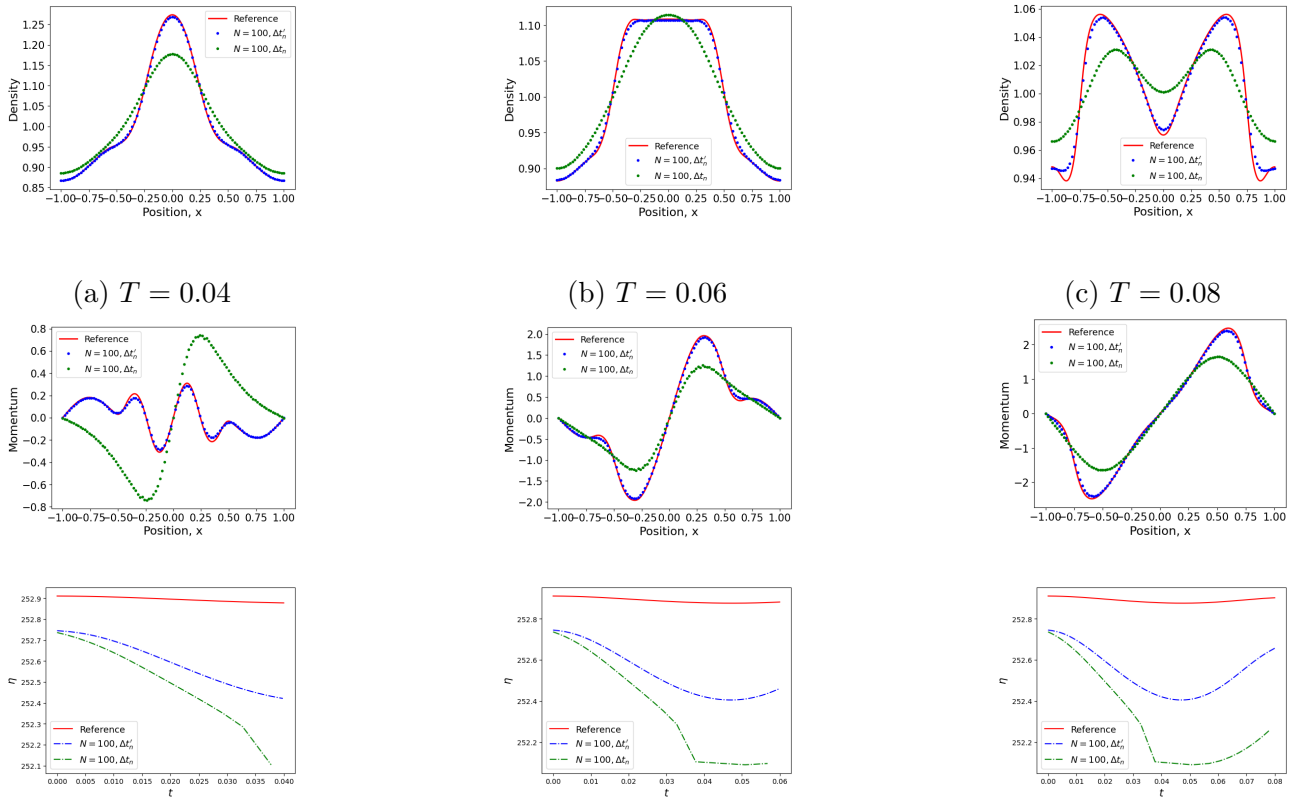
Figure 4.5: Colliding acoustic waves problem with $\epsilon = 0.1$ using type 2 space discretisation. The reference is 2^{nd} order type 3 entropy stable space discretisation with $q = 7$, $N = 1000$ and classical time-step $\Delta t'_n$.

$$\rho_0(x) = 1 + \epsilon^2, \qquad (\rho u_1)_0(x) = 1, \qquad \text{if } x \in (0.2, 0.3] \tag{4.85}$$

$$\rho_0(x) = 1, \ (\rho u_1)_0(x) = 1 + \epsilon^2/2, \ \text{if } x \in (0.3, 0.7] \tag{4.86}$$

$$\rho_0(x) = 1 - \epsilon^2, \qquad (\rho u_1)_0(x) = 1, \qquad \text{if } x \in (0.7, 0.8] \tag{4.87}$$

The parametric values are: $\kappa = 1$ and $\gamma = 2$. Periodic boundary conditions are used for this problem. Density, momentum and global entropy (vs. time) plots are presented at $T = 0.05$ for different values of ϵ, such as $\epsilon = 0.8, 0.3, 0.05$. $ARS(1,1,1)$ IMEX time discretisation is used. Figures 4.7, 4.8 and 4.9 show the plots obtained by using type-2, type-3 1^{st} and 2^{nd} order entropy stable space discretisations respectively. The reference solution for all the cases is obtained by using type-3 first order accurate entropy stable space discretisation with $q = 7$, $N = 1000$ and the classical time step $\Delta t'_n$ in (4.83).



(a) $T = 0.04$ (b) $T = 0.06$ (c) $T = 0.08$

(d) $T = 0.04$ (e) $T = 0.06$ (f) $T = 0.08$

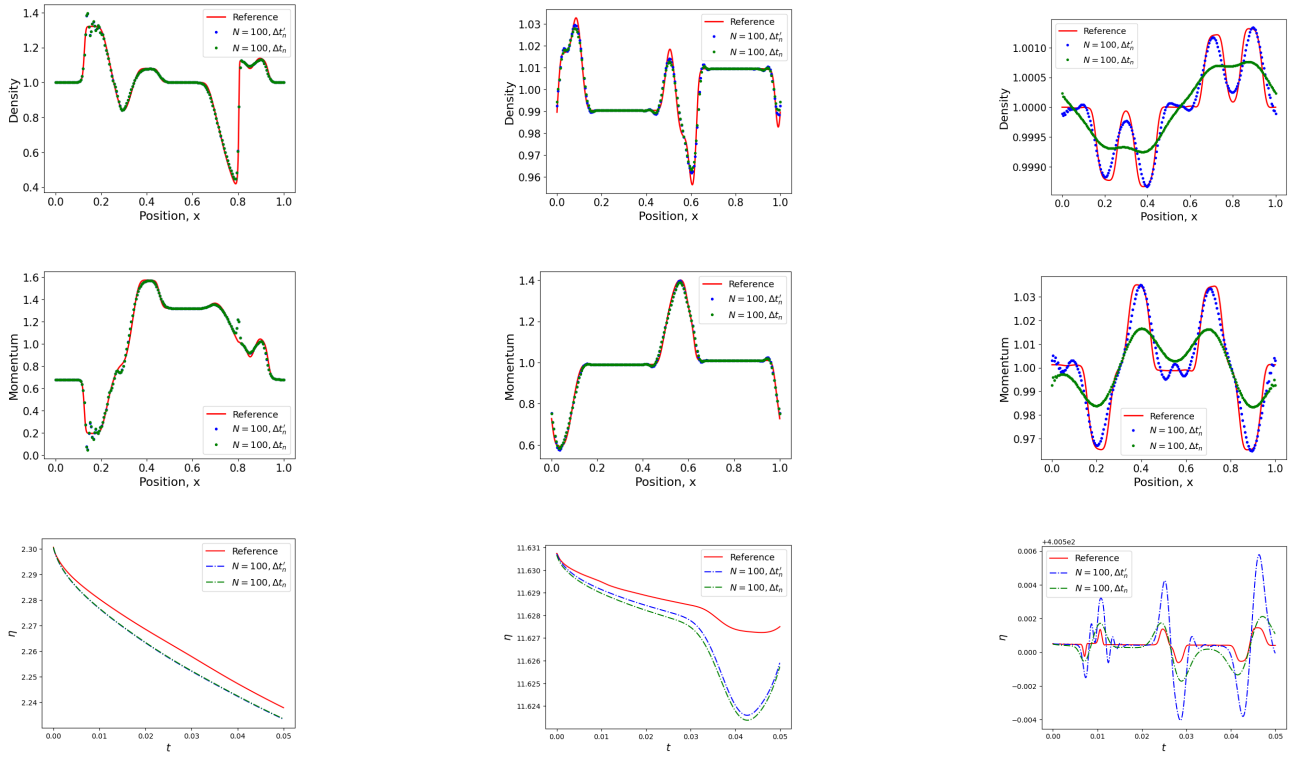
Figure 4.6: Colliding acoustic waves problem with $\epsilon = 0.1$ using 2^{nd} order type 3 entropy stable space discretisation with $q = 7$. The reference is the same scheme with $N = 1000$ and classical time-step $\Delta t'_n$.

4.4.4 Gresho vortex problem

This problem is from [138, 265]. A vortex of radius $R = 0.4$ centered at $(x_{1_0}, x_{2_0}) = (0.5, 0.5)$ is considered at initial time $t = 0$. The initial background state is considered as: $\rho_0 = 1$, $\mathbf{u}_0 = (u_{1_0}, 0)^T$, $p_0 = 1$ and hence $a_0 = \sqrt{\frac{\gamma p_0}{\rho_0}} = \sqrt{\gamma}$. The flow velocity is then given by $u_{1_0} = \epsilon_0 a_0$, where ϵ_0 is the global Mach number.

The radial velocity of the vortex is

$$u_r(r) = u_{1_0} \begin{cases} 2\frac{r}{R} & \text{if } 0 \le r < \frac{R}{2} \\ 2\left(1 - \frac{r}{R}\right) & \text{if } \frac{R}{2} \le r < R \\ 0 & \text{if } r \ge R \end{cases},$$



(a) $\epsilon = 0.8$ (b) $\epsilon = 0.3$ (c) $\epsilon = 0.05$

Figure 4.7: Riemann problem at $T = 0.05$ using type 2 space discretisation. The reference is 1^{st} order type 3 space discretisation with $q = 7$, $N = 1000$ and classical time-step $\Delta t'_n$.

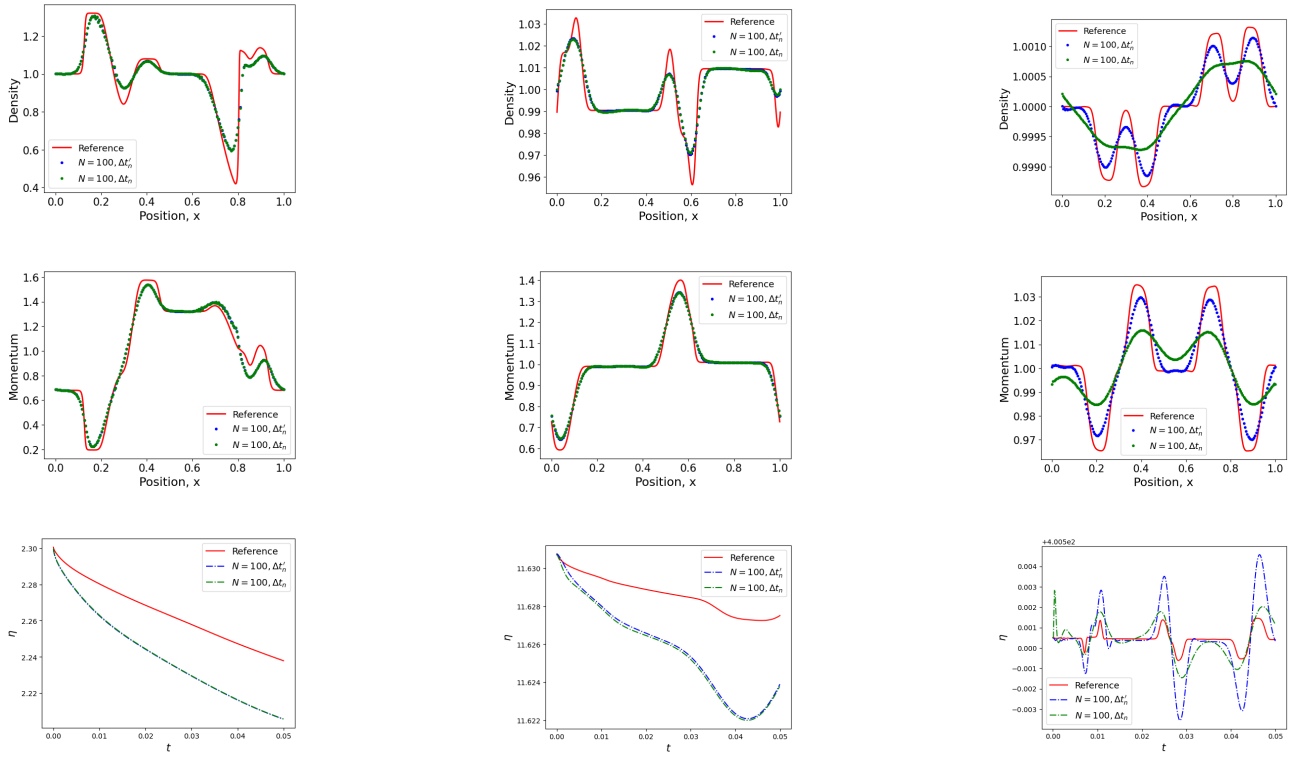
and the velocity components in Cartesian coordinates are

$$u_1(x_1, x_2) = u_{1_0} - \frac{x_2 - x_{2_0}}{r} u_r, \;\; u_2(x_1, x_2) = \frac{x_1 - x_{1_0}}{r} u_r.$$

Here, $r = \sqrt{(x_1 - 0.5)^2 + (x_2 - 0.5)^2}$. Upon balancing the pressure gradient and centrifugal force $\left(i.e., \; \rho_0 \frac{u_r^2}{r} = \frac{\partial p}{\partial r} \right)$, pressure is derived as:

$$p(r) = p_0 + u_0^2 \begin{cases} 2\frac{r^2}{R^2} + 2 - \log 16 & \text{if } 0 \le r < \frac{R}{2} \\ 2\frac{r^2}{R^2} - 8\frac{r}{R} + 4\log\left(\frac{r}{R}\right) + 6 & \text{if } \frac{R}{2} \le r < R \\ 0 & \text{if } r \ge R \end{cases}.$$

We assume adiabatic compression $p = \rho^\gamma$ with $\gamma = 1.4$, and use the asymptotic ansatz: $p = p_0 + u_{1_0}^2 p_2$, $\rho = \rho_0 + u_{1_0}^2 \rho_2$. Comparing $\frac{\partial p}{\partial u_{1_0}} = 2u_{1_0} p_2$ and $\frac{\partial p}{\partial u_{1_0}} = \frac{\partial p}{\partial \rho} \frac{\partial \rho}{\partial u_{1_0}} = \frac{\gamma p}{\rho} 2u_{1_0} \rho_2$, we obtain $\rho_2 = \frac{p_2}{\gamma}$ by noting that $p = 1, \rho = 1$ up-to leading order. This ρ_2 is used in the initial



(a) $\epsilon = 0.8$ (b) $\epsilon = 0.3$ (c) $\epsilon = 0.05$

Figure 4.8: Riemann problem at $T = 0.05$ using 1^{st} order type 3 space discretisation with $q = 7$. The reference is the same scheme with $N = 1000$ and classical time-step $\Delta t'_n$.

condition.

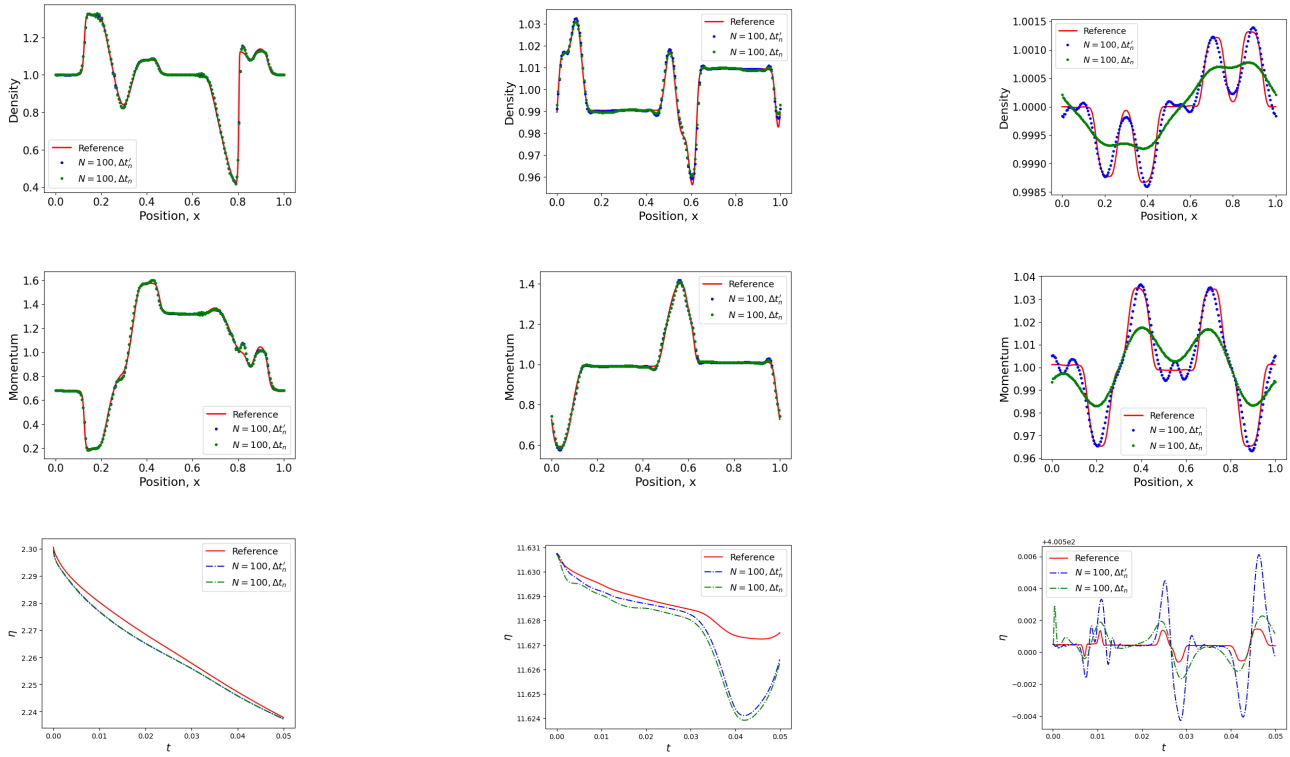
Periodic boundary conditions are imposed in both directions, and the mesh size used is $Nx \times Ny = 100 \times 100$. The problem is simulated using space discretisation type 2 for $\epsilon_0 = 0.1, 0.01, 0.001$. The following quantities are observed:

$$\eta = 1/2\ \rho(u_x^2 + u_y^2) + (1/\epsilon^2)p/(\gamma - 1),\ \text{PE}\ = (1/\epsilon^2)p/(\gamma - 1), \tag{4.88}$$

$$\text{KE}\ = (u_x - u_{x_{\text{inf}}})^2 + u_y^2 \text{ where } u_{x_{\text{inf}}} \text{ is the the background velocity} \tag{4.89}$$

$$\text{Ma ratio}\ = \frac{1}{\epsilon_0}\left(\sqrt{\frac{(u_x - u_{x_{\text{inf}}})^2 + u_y^2}{\gamma p/\rho}}\right) \tag{4.90}$$

Figure 4.10 shows the evolution of entropy (η), KE, PE over time upto $T = R\pi$ and Ma ratio contours at $T = R\pi$.

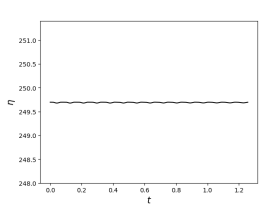


(a) $\epsilon = 0.8$ (b) $\epsilon = 0.3$ (c) $\epsilon = 0.05$

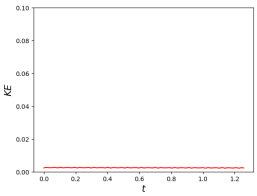
Figure 4.9: Riemann problem at $T = 0.05$ using 2^{nd} order type 3 space discretisation with $q = 7$. The reference is 1^{st} order type 3 space discretisation with $q = 7$, $N = 1000$ and classical time-step $\Delta t'_n$.

4.5 Summary and Conclusions

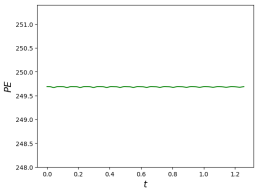
In this chapter, the entropy inequality corresponding to convex entropy function depending on Mach number ϵ has been derived for the baratropic Euler system. Further, numerical schemes satisfying such an entropy stability for different values of ϵ have been developed by using IMEX-AP time discretisation and three space discretisation strategies. The entropy stability and asymptotic preserving properties of the methods have been validated numerically.



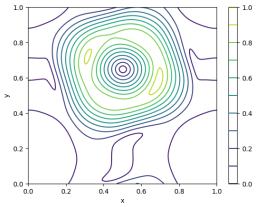
(a) Entropy, $\epsilon = 0.1$



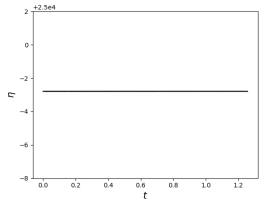
(b) KE, $\epsilon = 0.1$



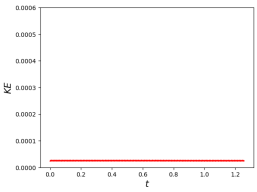
(c) PE, $\epsilon = 0.1$



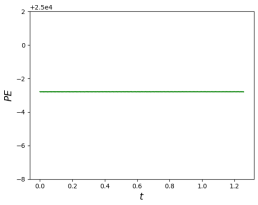
(d) Ma ratio, $\epsilon = 0.1$



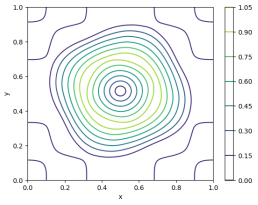
(e) Entropy, $\epsilon = 0.01$



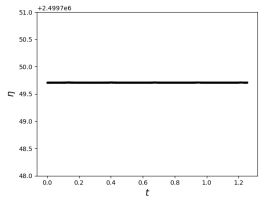
(f) KE, $\epsilon = 0.01$



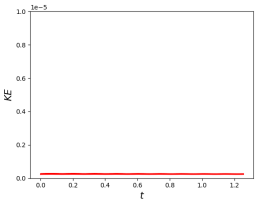
(g) PE, $\epsilon = 0.01$



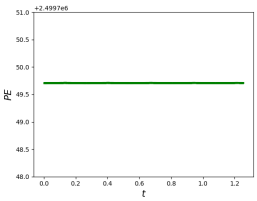
(h) Ma ratio, $\epsilon = 0.01$



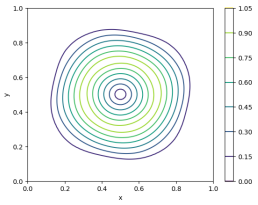
(i) Entropy, $\epsilon = 0.001$



(j) KE, $\epsilon = 0.001$



(k) PE, $\epsilon = 0.001$



(l) Ma ratio, $\epsilon = 0.001$

Figure 4.10: Entropy, KE, PE and Ma ratio plots using space discretisation type 2 for $\epsilon = 0.1, 0.01, 0.001$ on 100×100 grid

4.A Appendix: Butcher tableau

The first order type CK-ARS double Butcher tableau (known as ARS$(1, 1, 1)$) is:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1 & 0 \end{array} \qquad \begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 0 & 1 \\ \hline & 0 & 1 \end{array} \tag{4.91}$$

The following is the 2-stage second order accurate Butcher tableau ARS$(2, 2, 2)$:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & \delta & 1-\delta & 0 \\ \hline & \delta & 1-\delta & 0 \end{array} \qquad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}$$

Here, $\gamma = 1 - \frac{1}{\sqrt{2}}$ and $\delta = 1 - \frac{1}{2\gamma}$.

Chapter 5

On Lattice Boltzmann Methods based on vector-kinetic models for hyperbolic partial differential equations

In this chapter, we are concerned about the lattice Boltzmann methods (LBMs) based on vector-kinetic models for hyperbolic partial differential equations. In addition to usual lattice Boltzmann equation (LBE) derived by explicit discretisation of vector-kinetic equation (VKE), we also consider LBE derived by semi-implicit discretisation of VKE and compare the relaxation factors of both. We study the properties such as H-inequality, total variation boundedness and positivity of both the LBEs, and infer that the LBE due to semi-implicit discretisation naturally satisfies all the properties while the LBE due to explicit discretisation requires more restrictive condition on relaxation factor compared to the usual condition obtained from Chapman-Enskog expansion. We also derive the macroscopic finite difference form of the LBEs, and utilise it to establish the consistency of LBEs with the hyperbolic system. Further, we extend this LBM framework to hyperbolic conservation laws with source terms, such that there is no spurious numerical convection due to imbalance between convection and source terms. We also present a D2Q9 model that allows upwinding even along diagonal directions in addition to the usual upwinding along coordinate directions. The different aspects of the results are validated numerically on standard benchmark problems.

5.1 Introduction

Lattice Boltzmann methods (LBMs) have emerged as a powerful and versatile class of computational techniques for simulating fluid flow and related phenomena. Over the years, they have

gained significant popularity due to their ability to handle a wide range of fluid flow scenarios, from incompressible flows ([327, 246, 154]) to complex multiphase ([139, 247, 42, 110, 204]) and multiscale ([310]) systems. LBMs have been employed for modelling and simulating problems in magnetohydrodynamics ([214, 238, 279, 150]), porous media ([29, 141, 142, 109]), heat transfer ([219, 225, 303]) and turbulence ([184, 144]). The reader is referred to the books [220, 290, 140] for extensive study of LBMs, [59] for review of LBMs for fluid flows, [8] for review of LBMs for heat transfer, and [153] for review of entropic LBMs.

The Lattice Boltzmann equation (LBE) has been shown to approximate the Euler and the Navier-Stokes equations through different approaches such as Chapman-Enskog expansion ([190, 129, 335]), asymptotic expansion ([157, 175, 176]), Maxwellian iteration ([13, 26, 329]), equivalent equation ([100]), and recursive representation ([152]). Some notions of stability, including non-negativity of particle distribution function, of algorithms based on LBE were studied ([128, 101, 102]. Various attempts have been made in which the LBE is shown to be equivalent to mutistep finite difference equation ([291, 83, 103, 120, 24]), and the consistency with macroscopic equations has been shown in [23]. Further, the linkage between LBM and relaxation systems of [173] has been explored in [16, 262] and is being investigated frequently in [136, 282, 283].

While the discussions above correspond to the LBE derived from discretisation of the Boltzmann equation (essentially scalar-kinetic equation) with discrete velocities, we consider the class of LBEs derived from discretisation of vector-kinetic equations introduced in [43, 44, 7]. The vector-kinetic models have been utilised to develop various numerical schemes in the areas of porous media [174], entropy stable methods for hyperbolic systems [5], implicit kinetic relaxation schemes [69], and lattice Boltzmann relaxation schemes [254, 84, 270]. In particular, [69] and [254] present the lattice Boltzmann methods with different equilibrium functions and their resulting Chapman-Enskog expansions. *In this chapter, we present some important properties (such as macroscopic multi-step finite difference form and consistency) of the LBE derived from vector-kinetic equations. We also present a novel way to handle well-balancing of convection and source terms in this framework. Further, we also present an LBM model that allows upwinding along diagonal directions in addition to the usual upwinding along coordinate directions* (presented first in the proceedings of a conference [4]).

The chapter is organised as follows: Section 5.2 presents the mathematical model of hyperbolic conservation law and its vector-kinetic equation. Section 5.3 presents two different ways of deriving LBE from vector-kinetic equation, their Chapman-Enskog expansion and different equilibrium functions. Different properties such as H-inequality, macroscopic multi-step finite difference form, consistency, total variation boundedness and positivity are discussed in section

5.4. The well-balancing technique for hyperbolic partial differential equations with source terms is explained in section 5.5. The D2Q9 model of LBM that allows upwinding along diagonal directions is explained in section 5.6. The numerical validation of the methods is presented in section 5.7. Section 5.8 concludes the chapter.

5.2 Mathematical model

In this section, we describe the hyperbolic conservation law and the vector-kinetic equation that approximates it.

5.2.1 Hyperbolic conservation law

Consider the hyperbolic conservation law

$$\partial_t U + \partial_{x_d} G^d(U) = 0, \tag{5.1}$$

where $U(x_1, x_2, \ldots, x_D, t) : \Omega \times [0, T] \to \mathbb{R}^p$ is the conserved variable and $G^d(U) : \mathbb{R}^p \to \mathbb{R}^p$ is the flux in direction d, for $d \in \{1, 2, \ldots, D\}$. Here $\Omega \subset \mathbb{R}^D$, D and p indicate the number of dimensions and number of equations in the system respectively. $\eta(U)$ is the convex entropy function for (5.1).

5.2.2 Vector-kinetic equation

The hyperbolic conservation law in (5.1) can be approximated by the vector-kinetic equation (VKE) (see [43, 7]),

$$\partial_t f_q + \partial_{x_d} \left(v_q^d f_q \right) = -\frac{1}{\epsilon} \left(f_q - f_q^{eq}(U) \right). \tag{5.2}$$

Here $f_q : \Omega \times [0, T] \to \mathbb{R}^p$, $f_q^{eq} : \mathbb{R}^p \to \mathbb{R}^p$ and $q \in \{1, 2, \ldots, Q\}$ with Q being the number of discrete velocities. ϵ is a positive small parameter. v_q^d is the d^{th} component of the q^{th} discrete velocity. Summing (5.2) over all q, we get

$$\partial_t \sum_{q=1}^{Q} f_q + \partial_{x_d} \sum_{q=1}^{Q} \left(v_q^d f_q \right) = -\frac{1}{\epsilon} \sum_{q=1}^{Q} \left(f_q - f_q^{eq}(U) \right). \tag{5.3}$$

If $\sum_{q=1}^{Q} f_q = \sum_{q=1}^{Q} f_q^{eq} = U$, then

$$\partial_t U + \partial_{x_d} \sum_{q=1}^{Q} \left(v_q^d f_q \right) = 0. \tag{5.4}$$

In the limit $\epsilon \to 0$, we infer from (5.2) that $f_q \to f_q^{eq}(U)$. Thus, we can write f_q as perturbation (in ϵ) of f_q^{eq}:

$$f_q = f_q^{eq} + \epsilon f_q^{neq}, \tag{5.5}$$

where f_q^{neq} consists of the non-equilibrium perturbations.
If $\sum_{q=1}^{Q} v_q^d f_q^{eq} = G^d(U)$, then (5.4) becomes the hyperbolic conservation law (5.1) in the limit $\epsilon \to 0$.

5.3 Lattice Boltzmann equation

In this section, we present explicit and semi-implicit lattice Boltzmann discretisations of the VKE (5.2), their comparison, and their Chapman-Enskog expansions.
Let us use the vector notations: $\mathbf{x} = \begin{bmatrix} x_1, & x_2, & \ldots, & x_D \end{bmatrix}$ and $\mathbf{v_q} = \begin{bmatrix} v_q^1, & v_q^2, & \ldots, & v_q^D \end{bmatrix}$. An explicit Euler discretisation of the VKE (5.2) along $\frac{dx_d}{dt} = v_q^d$ (the characteristic equation) gives

$$f_q\left(\mathbf{x}, t + \Delta t\right) = f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) - \frac{\Delta t}{\epsilon}\left(f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) - f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right)\right). \tag{5.6}$$

Using $\omega = \frac{\Delta t}{\epsilon}$ and rewriting the above equation, we obtain the lattice Boltzmann equation (LBE)

$$f_q\left(\mathbf{x}, t + \Delta t\right) = (1 - \omega) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \omega f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right). \tag{5.7}$$

On the other hand, a semi-implicit discretisation of the VKE (5.2) with implicit treatment of f_q in the collision term gives

$$f_q\left(\mathbf{x}, t + \Delta t\right) = f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) - \frac{\Delta t}{\epsilon}\left(f_q\left(\mathbf{x}, t + \Delta t\right) - f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right)\right). \tag{5.8}$$

Rewriting the above equation as

$$f_q\left(\mathbf{x}, t + \Delta t\right) = \left(\frac{1}{1+\omega}\right) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \left(\frac{\omega}{1+\omega}\right) f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right), \tag{5.9}$$

$$\text{or } f_q\left(\mathbf{x}, t + \Delta t\right) = \left(1 - \tilde{\omega}\right) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \tilde{\omega} f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right), \tag{5.10}$$

an LBE with $\tilde{\omega} = \frac{\omega}{1+\omega}$ is obtained.
If the grid is uniform with spacing Δx_d along direction d and if the velocities are chosen such that $v_q^d = m\frac{\Delta x_d}{\Delta t}$ with $m \in \mathbb{Z}$, $\forall d, q$, then the collision-streaming algorithm

$$\text{Collision: } f_q^*\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) = (1 - \hat{\omega}) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \hat{\omega} f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right) \tag{5.11}$$

$$\text{Streaming: } f_q\left(\mathbf{x}, t + \Delta t\right) = f_q^*\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) \tag{5.12}$$

can be used to numerically implement the LBEs in (5.7) (with $\hat{\omega} = \omega$) and (5.10) (with $\hat{\omega} = \tilde{\omega}$). It is to be noted that the streaming in (5.12) is exact. After evaluating $f_q(\mathbf{x}, t + \Delta t)$, we find U by using $U(\mathbf{x}, t + \Delta t) = \sum_{q=1}^{Q} f_q(\mathbf{x}, t + \Delta t)$. Then, we evaluate $f_q^{eq}(U(\mathbf{x}, t + \Delta t))$ and then proceed with the next time step. *Hereafter, we use $\hat{\omega}$ in the presentation of our theory to commonly represent ω in (5.7) and $\tilde{\omega}$ in (5.10).*

5.3.1 Chapman-Enskog expansion

Taylor expanding the LBEs in (5.7) (with $\hat{\omega} = \omega$) and (5.10) (with $\hat{\omega} = \tilde{\omega}$) and simplifying, we get

$$(\partial_t + \mathbf{v_q} \cdot \nabla) f_q = -\frac{\hat{\omega}}{\Delta t}\left(f_q - f_q^{eq}\right) + \frac{\hat{\omega}}{2}(\partial_t + \mathbf{v_q} \cdot \nabla)\left(f_q - f_q^{eq}\right) + \mathcal{O}(\Delta t^2). \tag{5.13}$$

Consider the perturbation expansion of f_q:

$$f_q = f_q^{eq} + \epsilon f_q^{(1)} + \epsilon^2 f_q^{(2)} + \dots \tag{5.14}$$

Using the above expression, since $\sum_{q=1}^{Q} f_q = \sum_{q=1}^{Q} f_q^{eq} = U$, we infer that the moment of non-equilibrium function leads to $\sum_{q=1}^{Q}\left(\epsilon f_q^{(1)} + \epsilon^2 f_q^{(2)} + \dots\right) = 0$. Each term corresponding to different order of ϵ in this moment expression must individually be zero. Hence $\sum_{q=1}^{Q} f_q^{(i)} = 0, \forall i \in \mathbb{N}$. Multiple scale expansion of derivatives of f_q gives $\partial_t f_q = \left(\epsilon\partial_t^{(1)} + \epsilon^2\partial_t^{(2)} + ...\right) f_q$ and $\mathbf{v_q} \cdot \nabla f_q = \epsilon \mathbf{v_q} \cdot \nabla^{(1)} f_q$.

Using perturbation expansion of f_q and multiple scale expansion of derivatives of f_q in (5.13) and separating out $O(\epsilon)$ and $O(\epsilon^2)$ terms,

$$O(\epsilon): \quad \left(\partial_t^{(1)} + \mathbf{v_q} \cdot \nabla^{(1)}\right) f_q^{eq} = -\frac{\hat{\omega}}{\Delta t} f_q^{(1)} \tag{5.15}$$

$$O(\epsilon^2): \quad \partial_t^{(2)} f_q^{eq} + \left(1 - \frac{\hat{\omega}}{2}\right)\left(\partial_t^{(1)} + \mathbf{v_q} \cdot \nabla^{(1)}\right) f_q^{(1)} = -\frac{\hat{\omega}}{\Delta t} f_q^{(2)} \tag{5.16}$$

Zeroth moment $\left(\sum_{q=1}^{Q}\right)$ of $O(\epsilon)$ terms in (5.15) and $O(\epsilon^2)$ terms in (5.16) respectively give

$$\partial_t^{(1)} U + \partial_{x_d}^{(1)} G^d(U) = 0, \tag{5.17}$$

$$\partial_t^{(2)} U + \left(1 - \frac{\hat{\omega}}{2}\right) \partial_{x_d}^{(1)} \left(\sum_{q=1}^{Q} v_q^{(d)} f_q^{(1)}\right) = 0. \tag{5.18}$$

From the first moment $\left(\sum_{q=1}^{Q} v_q^d\right)$ of $O(\epsilon)$ terms in (5.15), we get

$$\sum_{q=1}^{Q} v_q^d f_q^{(1)} = -\frac{\Delta t}{\hat{\omega}} \left(\partial_U G^d \left(-\partial_U G^i \partial_{x_i}^{(1)} U \right) + \partial_{x_i}^{(1)} \left(\sum_{q=1}^{Q} v_q^d v_q^i f_q^{eq} \right) \right) \tag{5.19}$$

Recombining the zeroth moment equations of $O(\epsilon)$ in (5.17) and $O(\epsilon^2)$ in (5.18), we obtain

$$\begin{aligned} &\left(\epsilon\partial_t^{(1)} + \epsilon^2\partial_t^{(2)}\right) U + \epsilon\partial_{x_d}^{(1)} G^d(U) \\ &+ \left(1 - \frac{\hat{\omega}}{2}\right) \epsilon\partial_{x_d}^{(1)} \left(-\frac{\Delta t}{\hat{\omega}} \left(\partial_U G^d \left(-\partial_U G^i \epsilon\partial_{x_i}^{(1)} U \right) + \epsilon\partial_{x_i}^{(1)} \left(\sum_{q=1}^{Q} v_q^d v_q^i f_q^{eq} \right) \right) \right) = 0 + \mathcal{O}(\epsilon^3) \end{aligned} \tag{5.20}$$

Replacing the multiple scale expansions of derivatives present in the above equation with their non-expanded form, we get

$$\partial_t U + \partial_{x_d} G^d(U) = \Delta t \left(\frac{1}{\hat{\omega}} - \frac{1}{2} \right) \partial_{x_d} \left(\partial_{x_i} \left(\sum_{q=1}^{Q} v_q^d v_q^i f_q^{eq} \right) - \partial_U G^d \partial_U G^i \partial_{x_i} U \right) \tag{5.21}$$

upto $\mathcal{O}\left(\Delta t^2\right)$.

The above equation is well known in LBM community for modelling of advection-diffusion equation, where the term on the right hand side is commonly decomposed into physical diffusion and numerical diffusion. As we are dealing with hyperbolic equations in this work, the term on the right hand side of (5.21) solely represents numerical diffusion.

5.3.2 Equilibrium function

In the previous sections, we imposed the following conditions on f_q^{eq}:

$$\sum_{q=1}^{Q} f_q^{eq} = U, \ \sum_{q=1}^{Q} v_q^d f_q^{eq} = G^d(U). \tag{5.22}$$

In this section, we present some f_q^{eq} that satisfy the above requirements.

5.3.2.1 Classical D1Q2

Consider one dimension (D=1) and 2 discrete velocities ($Q = 2$) such that $v_1^1 = \lambda$ and $v_2^1 = -\lambda$, and $\lambda = \frac{\Delta x_1}{\Delta t}$. Then,

$$f_q^{eq} = \frac{1}{2} U - (-1)^q \frac{1}{2\lambda} G^1(U) \text{ for } q \in \{1, 2\} \tag{5.23}$$

satisfies (5.22). The Chapman-Enskog expansion (5.21) in this case becomes,

$$\partial_t U + \partial_{x_1} G^1(U) = \Delta t \left(\frac{1}{\hat{\omega}} - \frac{1}{2}\right) \partial_{x_1} \left(\left(\lambda^2 I - |\partial_U G^1|^2\right) \partial_{x_1} U\right). \tag{5.24}$$

It is to be noted that the $\mathcal{O}(\Delta t)$ term on the right hand side of the above equation represents numerical diffusion. For stability, we require the numerical diffusion coefficient to be positive. Therefore, we require $\lambda^2 I > |\partial_U G^1|^2$ and $0 < \hat{\omega} < 2$.

5.3.2.2 D1Q3

Consider one dimension (D=1) and 3 discrete velocities ($Q = 3$) such that $v_1^1 = \lambda$, $v_2^1 = 0$ and $v_3^1 = -\lambda$, and $\lambda = \frac{\Delta x_1}{\Delta t}$. Then,

$$f_q^{eq} = \frac{1}{3} U + (\delta_{q1} - \delta_{q3}) \frac{1}{2\lambda} G^1(U) \text{ for } q \in \{1, 2, 3\} \tag{5.25}$$

where δ is the *Kronecker delta function*, satisfies (5.22). The Chapman-Enskog expansion (5.21) in this case becomes,

$$\partial_t U + \partial_{x_1} G^1(U) = \Delta t \left(\frac{1}{\hat{\omega}} - \frac{1}{2}\right) \partial_{x_1} \left(\left(\frac{2}{3}\lambda^2 I - |\partial_U G^1|^2\right) \partial_{x_1} U\right). \tag{5.26}$$

Enforcement of the positivity of numerical diffusion coefficient yields $\lambda^2 I > \frac{3}{2}|\partial_U G^1|^2$ and $0 < \hat{\omega} < 2$.

5.3.2.3 Upwind D$\overline{d}$Q$(2\overline{d}+1)$

Consider $D = \overline{d}$ and $Q = 2\overline{d} + 1$ with $\lambda_d = \frac{\Delta x_d}{\Delta t}$ and

$$v_q^d = \lambda_d \delta_{qd} - \lambda_d \delta_{q\left(d+\left(\overline{d}+1\right)\right)}. \tag{5.27}$$

Define

$$f_q^{eq} = \begin{cases} \frac{G^{q+}}{\lambda_q}, & \text{for } q \in \{1, 2, \ldots, \overline{d}\} \\ U - \sum_{d=1}^{\overline{d}} \left(\frac{G^{d+} + G^{d-}}{\lambda_d}\right), & \text{for } q = \overline{d} + 1 \\ \frac{G^{\left(q-\left(\overline{d}+1\right)\right)-}}{\lambda_{q-\left(\overline{d}+1\right)}}, & \text{for } q \in \{\overline{d}+2, \overline{d}+3, \ldots, 2\overline{d}+1\} \end{cases} \tag{5.28}$$

with $G^d = G^{d+} - G^{d-}$. This satisfies (5.22) and leads to the *Flux Decomposition* technique of [7]. Using an additional choice, G^{d+} and G^{d-} for a hyperbolic system can be evaluated by a suitable flux splitting method available in literature. For instance, one can use G^{d+} and

G^{d-} from commonly known flux vector splitting methods such as kinetic flux vector splitting [211], Steger-Warming flux vector splitting [286] and van Leer's flux vector splitting [313]. One can also evaluate G^{d+} and G^{d-} from some flux difference splitting methods such as Roe's approximate Riemann solver [266] and kinetic flux difference splitting [280]. If we consider scalar conservation laws (*i.e.*, $p = 1$), then we can simply use the sign of wave speed $\partial_U G^d$ to determine the split fluxes as:

$$\partial_U G^{d+} = \begin{cases} \partial_U G^d & \text{if } \partial_U G^d > 0 \\ 0 & \text{if } \partial_U G^d \leq 0 \end{cases}, \partial_U G^{d-} = \begin{cases} 0 & \text{if } \partial_U G^d > 0 \\ -\partial_U G^d & \text{if } \partial_U G^d \leq 0 \end{cases}, \tag{5.29}$$

$$G^{d\pm} = \int_0^U \partial_U G^{d\pm} dU \text{ if } G^d(U = 0) = 0. \tag{5.30}$$

The Chapman-Enskog expansion (5.21) for the case of upwind D$\overline{d}$Q$(2\overline{d}+1)$ becomes,

$$\partial_t U + \partial_{x_d} G^d(U) = \Delta t \left(\frac{1}{\hat{\omega}} - \frac{1}{2}\right) \partial_{x_d} \left(\delta_{di} \lambda_d \partial_U \left(G^{d+} + G^{d-}\right) - \partial_U G^d \partial_U G^i\right) \partial_{x_i} U. \tag{5.31}$$

For positivity of numerical diffusion coefficient, we require

$$\delta_{di} \lambda_d \partial_U \left(G^{d+} + G^{d-}\right) - \partial_U G^d \partial_U G^i > 0 \tag{5.32}$$

along with $0 < \hat{\omega} < 2$.

For all the models of equilibrium function described above, a condition relating λ_d and $\partial_U G^d$ is obtained while ensuring positivity of numerical diffusion coefficient. Such relations are known as sub-characteristic conditions as they relate the characteristic speeds of vector-kinetic equation to those of the hyperbolic conservation law.

Remark 5.1 *In all the models of equilibrium function described above, $0 < \hat{\omega} < 2$ is required for enforcing the positivity of numerical diffusion coefficient. We know that $\hat{\omega} = \omega$ and $\hat{\omega} = \tilde{\omega}$ for LBEs in (5.7) and (5.10) respectively. Thus, the stability requirement is,*

$$\textit{For LBE in (5.7)}: \quad 0 < \hat{\omega} = \omega = \frac{\Delta t}{\epsilon} < 2 \implies 0 < \Delta t < 2\epsilon, \tag{5.33}$$

$$\textit{For LBE in (5.10)}: 0 < \hat{\omega} = \tilde{\omega} = \frac{\omega}{1+\omega} = \frac{\Delta t}{\epsilon + \Delta t} < 2 \implies \Delta t > 0. \tag{5.34}$$

It is to be noted that the requirement of $0 < \tilde{\omega} < 2$ does not impose any upper-bound on Δt for the LBE in (5.10).

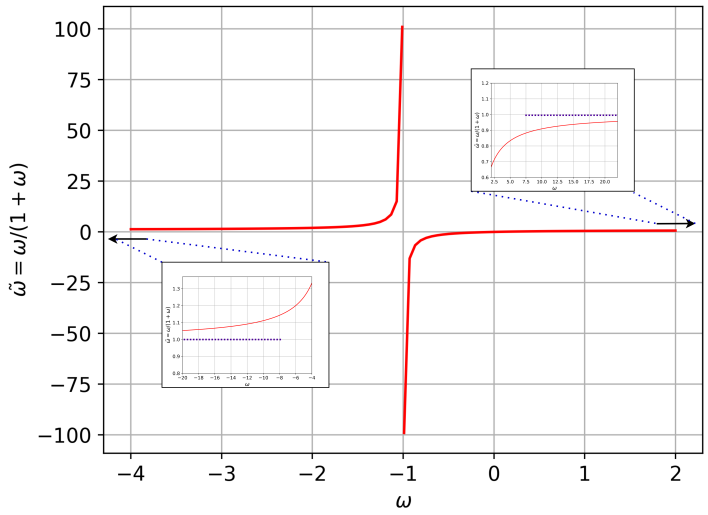


Figure 5.1: Plot of $\tilde{\omega}$ vs. ω

Remark 5.2 *For the LBE in* (5.10)*, the positivity of numerical diffusion coefficient enforces* $0 < \tilde{\omega} < 2$*. However, since* $\tilde{\omega} = \frac{\omega}{1+\omega}$ *and* $\omega = \frac{\Delta t}{\epsilon} > 0$*,* $\tilde{\omega}$ *is restricted to the interval* $(0, 1)$*. Figure 5.1 shows the plot of* $\tilde{\omega}$ *vs.* ω*, and it can be seen that* $0 < \tilde{\omega} < 1$ *for* $\omega > 0$*.*

5.4 Properties of the lattice Boltzmann equation

In this section, we discuss the properties of LBEs in (5.7) (with $\hat{\omega} = \omega$) and (5.10) (with $\hat{\omega} = \tilde{\omega}$). The properties considered are: H-inequality, macroscopic finite difference form, consistency, total variation boundedness and positivity.

5.4.1 H-inequality

We prove that an H-inequality is associated with the LBE obtained from semi-implicit discretisation of the VKE (*i.e.,* (5.10)). We also show that a constraint on ω is required to associate an H-inequality with the LBE obtained from explicit discretisation of the VKE (*i.e.,* (5.7)). For convenience, we consider scalar conservation laws (*i.e.,* $p = 1$) in the presentation of H-inequality.

Definition 5.1 *Define a function* $H_q(f_q)$ *such that:*

- $H_q(f_q)$ *is convex with respect to* f_q *(i.e.,* $\frac{\partial H_q}{\partial f_q}$ *is monotonically increasing and* $\frac{\partial^2 H_q}{\partial f_q^2}$ *is positive-definite),*
- $\sum_{q=1}^{Q} H_q(f_q^{eq}) = \eta(U)$*,*

- $\sum_{q=1}^{Q} H_q(f_q^{eq}) \leq \sum_{q=1}^{Q} H_q(f_q)$.

We consider the semi-implicit discretisation (5.8) of VKE with the notation $f_{qx_i}^{n+1} := f_q(\mathbf{x}, t+\Delta t)$, $f_{qy_i}^{n} := f_q(\mathbf{x} - \mathbf{v_q}\Delta t, t)$ and $f_{qy_i}^{eq^n} := f_q^{eq}(\mathbf{x} - \mathbf{v_q}\Delta t, t)$:

$$f_{qx_i}^{n+1} = f_{qy_i}^{n} - \omega\left(f_{qx_i}^{n+1} - f_{qy_i}^{eq^n}\right). \tag{5.35}$$

Theorem 5.1 *There exists an inequality*

$$H_q\left(f_{qx_i}^{n+1}\right) - H_q\left(f_{qy_i}^{n}\right) \leq -\omega\left(H_q\left(f_{qx_i}^{n+1}\right) - H_q\left(f_{qy_i}^{eq^n}\right)\right) \tag{5.36}$$

corresponding to the semi-implicit discretisation (5.35) of VKE with $\omega = \frac{\Delta t}{\epsilon} > 0$. *Here,* $H_q(f_q)$ *follows the definition 5.1.*

Proof: Left multiplying $\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}$ to (5.35), we get

$$\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}} \left(f_{qx_i}^{n+1} - f_{qy_i}^{n}\right) = -\omega \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}} \left(f_{qx_i}^{n+1} - f_{qy_i}^{eq^n}\right). \tag{5.37}$$

We consider the left and right hand sides of the above equation separately.
By mean value theorem, we have

$$\left.\frac{\partial H_q}{\partial f_q}\right|_{f_a} \left(f_{qx_i}^{n+1} - f_{qy_i}^{n}\right) = H_q\left(f_{qx_i}^{n+1}\right) - H_q\left(f_{qy_i}^{n}\right) \tag{5.38}$$

for some f_a lying on the line segment connecting $f_{qx_i}^{n+1}$ and $f_{qy_i}^{n}$. Further, we have the following due to the monotonicity of $\frac{\partial H_q}{\partial f_q}$:

$$f_{qx_i}^{n+1} \geq f_a \geq f_{qy_i}^{n} \implies \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}} \geq \left.\frac{\partial H_q}{\partial f_q}\right|_{f_a} \geq \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qy_i}^{n}}, \tag{5.39}$$

$$f_{qx_i}^{n+1} \leq f_a \leq f_{qy_i}^{n} \implies \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}} \leq \left.\frac{\partial H_q}{\partial f_q}\right|_{f_a} \leq \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qy_i}^{n}}. \tag{5.40}$$

Thus, we obtain the following inequality involving the term on the left hand side of (5.37):

$$H_q\left(f_{qx_i}^{n+1}\right) - H_q\left(f_{qy_i}^{n}\right) = \left.\frac{\partial H_q}{\partial f_q}\right|_{f_a} \left(f_{qx_i}^{n+1} - f_{qy_i}^{n}\right) \leq \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}} \left(f_{qx_i}^{n+1} - f_{qy_i}^{n}\right). \tag{5.41}$$

On the other hand, we also have the following by mean value theorem:

$$\left.\frac{\partial H_q}{\partial f_q}\right|_{f_b}\left(f_{qx_i}^{n+1}-f_{qy_i}^{eq^n}\right)=H_q\left(f_{qx_i}^{n+1}\right)-H_q\left(f_{qy_i}^{eq^n}\right) \tag{5.42}$$

for some f_b lying on the line segment connecting $f_{qx_i}^{n+1}$ and $f_{qy_i}^{eq^n}$. Further, due to the monotonicity of $\frac{\partial H_q}{\partial f_q}$, we have

$$f_{qx_i}^{n+1}\geq f_b\geq f_{qy_i}^{eq^n}\implies \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}\geq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_b}\geq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qy_i}^{eq^n}}, \tag{5.43}$$

$$f_{qx_i}^{n+1}\leq f_b\leq f_{qy_i}^{eq^n}\implies \left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}\leq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_b}\leq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qy_i}^{eq^n}}. \tag{5.44}$$

Thus, we obtain the following inequality involving the term on the right hand side of (5.37):

$$H_q\left(f_{qx_i}^{n+1}\right)-H_q\left(f_{qy_i}^{eq^n}\right)=\left.\frac{\partial H_q}{\partial f_q}\right|_{f_b}\left(f_{qx_i}^{n+1}-f_{qy_i}^{eq^n}\right)\leq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}\left(f_{qx_i}^{n+1}-f_{qy_i}^{eq^n}\right). \tag{5.45}$$

Therefore, from (5.41) and (5.45), we obtain

$$\begin{aligned}H_q\left(f_{qx_i}^{n+1}\right)-H_q\left(f_{qy_i}^{n}\right)&\leq\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}\left(f_{qx_i}^{n+1}-f_{qy_i}^{n}\right) &(5.46)\\ &=-\omega\left.\frac{\partial H_q}{\partial f_q}\right|_{f_{qx_i}^{n+1}}\left(f_{qx_i}^{n+1}-f_{qy_i}^{eq^n}\right) &(5.47)\\ &\leq-\omega\left(H_q\left(f_{qx_i}^{n+1}\right)-H_q\left(f_{qy_i}^{eq^n}\right)\right),\ \text{since }\omega=\frac{\Delta t}{\epsilon}>0. &(5.48)\end{aligned}$$

□

Remark 5.3 *The following can be inferred from the above theorem:*

$$H_q\left(f_{qx_i}^{n+1}\right)\leq\frac{1}{1+\omega}H_q\left(f_{qy_i}^{n}\right)+\frac{\omega}{1+\omega}H_q\left(f_{qy_i}^{eq^n}\right), \tag{5.49}$$

$$H_q\left(f_{qx_i}^{n+1}\right)\leq(1-\tilde{\omega})\,H_q\left(f_{qy_i}^{n}\right)+\tilde{\omega}H_q\left(f_{qy_i}^{eq^n}\right). \tag{5.50}$$

Since $\sum_{q=1}^{Q}H_q\left(f_{qy_i}^{eq^n}\right)\leq\sum_{q=1}^{Q}H_q\left(f_{qy_i}^{n}\right)$ *according to the definition of* H_q*, we obtain*

$$\sum_{q=1}^{Q}H_q\left(f_{qx_i}^{n+1}\right)\leq\sum_{q=1}^{Q}H_q\left(f_{qy_i}^{n}\right). \tag{5.51}$$

Thus, for the LBE obtained from semi-implicit discretisation of the VKE, the H-inequality holds without enforcing any constraint on $\omega = \frac{\Delta t}{\epsilon}$.
The following remark 5.4 presents H-inequality for general LBE, and the associated conditions. This has been presented particularly for explicit case in [69].

Remark 5.4 *Consider the general LBE,*

$$f_{q_{x_i}}^{n+1} = (1-\hat{\omega})\, f_{q_{y_i}}^{n} + \hat{\omega} f_{q_{y_i}}^{eq^n} \tag{5.52}$$

with $\hat{\omega} = \omega$ (for explicit discretisation of VKE) and $\hat{\omega} = \tilde{\omega}$ (for semi-implicit discretisation of VKE). Applying H_q on this LBE, we obtain

$$H_q\left(f_{q_{x_i}}^{n+1}\right) = H_q\left((1-\hat{\omega})\, f_{q_{y_i}}^{n} + \hat{\omega} f_{q_{y_i}}^{eq^n}\right) \tag{5.53}$$

$$\leq (1-\hat{\omega})\, H_q\left(f_{q_{y_i}}^{n}\right) + \hat{\omega} H_q\left(f_{q_{y_i}}^{eq^n}\right), \text{for } 0 < \hat{\omega} \leq 1 \tag{5.54}$$

Since $\sum_{q=1}^{Q} H_q\left(f_{q_{y_i}}^{eq^n}\right) \leq \sum_{q=1}^{Q} H_q\left(f_{q_{y_i}}^{n}\right)$, we obtain

$$\sum_{q=1}^{Q} H_q\left(f_{q_{x_i}}^{n+1}\right) \leq \sum_{q=1}^{Q} H_q\left(f_{q_{y_i}}^{n}\right). \tag{5.55}$$

Thus, the H-inequality holds for the general LBE if the constraint $0 < \hat{\omega} \leq 1$ is satisfied. It is to be noted that the H-inequality yields a stronger constraint on $\hat{\omega}$ than the positivity of numerical diffusion coefficient.
From the above remarks, the following can be inferred:

- For LBE obtained by explicit discretisation of VKE, $\hat{\omega} = \omega$. Hence, H-inequality holds corresponding to this LBE if $0 < \omega = \frac{\Delta t}{\epsilon} \leq 1$. It is to be noted that this constraint on ω is more restrictive than the constraint $0 < \omega < 2$ that enforces positivity of numerical diffusion coefficient.

- For LBE obtained by semi-implicit discretisation of VKE, $\hat{\omega} = \tilde{\omega} = \frac{\omega}{1+\omega}$. According to remark 5.4, H-inequality holds corresponding to this LBE if $0 < \tilde{\omega} = \frac{\omega}{1+\omega} \leq 1$, and this is satisfied for all $\omega = \frac{\Delta t}{\epsilon} > 0$. This also agrees with remark 5.3 which states that H-inequality holds corresponding to this LBE for all $\omega = \frac{\Delta t}{\epsilon} > 0$. Thus, the semi-implicit case of LBE is entropy-satisfying by construction without imposing any constraint.

5.4.2 Macroscopic finite difference form

In this section, we show the macroscopic finite difference form of LBEs in (5.7) (with $\hat{\omega} = \omega$) and (5.10) (with $\hat{\omega} = \tilde{\omega}$).

We briefly provide some technicalities for clarity. LBE is evolved on a fixed uniform grid with spacing Δx_d along direction d. At every time step, $\lambda_{d,n}$ is evaluated such that the sub-characteristic condition obtained by enforcing the positivity of numerical diffusion coefficient is satisfied. Thus, the discrete velocities can change with time step, and they are given by: $v_{q,n}^d = m_q|_d \, \lambda_{d,n}$, where $m_q|_d \in \mathbb{Z}$ is constant for direction d and q^{th} discrete velocity. The current time step is found by using $t_{n+1} - t_n := \Delta t_n = \frac{\Delta x_d}{\lambda_{d,n}}$. Note that in addition to satisfying the sub-characteristic condition, $\lambda_{d,n} > \Delta x_d$ is essential for upper-bounding Δt_n as $\Delta t_n < 1$. Further, ω is kept constant for all time steps, and hence ϵ is allowed to depend on n as Δt_n depends on n.

For convenience, we consider one dimension ($D = 1$) in the presentation of macroscopic finite difference form. Hence, the subscript and superscript d indicating d^{th} dimension can be ignored in all the variables. We consider the general LBE,

$$f_q\left(x_i, t_n + \Delta t_n\right) = (1 - \hat{\omega}) f_q\left(x_i - v_{q,n}\Delta t_n, t_n\right) + \hat{\omega} f_q^{eq}\left(U\left(x_i - v_{q,n}\Delta t_n, t_n\right)\right) \tag{5.56}$$

with $\hat{\omega} = \omega$ and $\hat{\omega} = \tilde{\omega}$ respectively for explicit and semi-implicit cases. For brevity, we introduce the following notations:

$f_{q_i}^{n+1} := f_q\left(x_i, t_n + \Delta t_n\right)$, $f_{q_{i-m_q}}^{n} := f_q\left(x_i - v_{q,n}\Delta t_n, t_n\right)$ and $f_{q_{i-m_q}}^{eq^n} := f_q^{eq}\left(U\left(x_i - v_{q,n}\Delta t_n, t_n\right)\right)$.

We also utilise the splitting of $f_{q_i}^n$ as equilibrium and non-equilibrium parts: $f_{q_i}^n = f_{q_i}^{eq^n} + f_{q_i}^{neq^n}, \forall i, n$. Note here that we have absorbed ϵ of $\epsilon f_{q_i}^{neq^n}$ (refer (5.5)) into $f_{q_i}^{neq^n}$ (*i.e.*, $f_{q_i}^{neq^n} = \mathcal{O}(\epsilon)$) for convenience in presentation. Further, we also assume that $f_{q_i}^0 = f_{q_i}^{eq^0}$ at the initial time. Thus, $f_{q_i}^{neq^0} = 0$.

Theorem 5.2 *The general LBE*

$$f_{q_i}^{n+1} = (1 - \hat{\omega}) f_{q_{i-m_q}}^{n} + \hat{\omega} f_{q_{i-m_q}}^{eq^n} \tag{5.57}$$

is equivalent to

$$f_{q_i}^{n+1} = \hat{\omega}\left(\sum_{k=0}^{N-1} (1 - \hat{\omega})^k f_{q_{i-(k+1)m_q}}^{eq^{n-k}}\right) + (1 - \hat{\omega})^N f_{q_{i-(N+1)m_q}}^{eq^{n-N}} \tag{5.58}$$

if $f_{q_{i-(N+1)m_q}}^{neq^{n-N}} = 0$. *Here* $N \in \mathbb{N}$.

Proof: Using $f^{n}_{q_{i-m_q}} = f^{eq^n}_{q_{i-m_q}} + f^{neq^n}_{q_{i-m_q}}$ in the general LBE (5.57), we obtain

$$f^{n+1}_{q_i} = f^{eq^n}_{q_{i-m_q}} + (1-\hat{\omega}) f^{neq^n}_{q_{i-m_q}}. \tag{5.59}$$

Using $f^{n+1}_{q_i} = f^{eq^{n+1}}_{q_i} + f^{neq^{n+1}}_{q_i}$ in the above equation yields

$$f^{neq^{n+1}}_{q_i} = -f^{eq^{n+1}}_{q_i} + f^{eq^n}_{q_{i-m_q}} + (1-\hat{\omega}) f^{neq^n}_{q_{i-m_q}}. \tag{5.60}$$

Inserting $f^{neq^{n+1}}_{q_i}$ from the above equation into $f^{neq^n}_{q_{i-m_q}}$ in (5.59) by employing the transformation $n := n'+1$, $i - m_q := i'$, we get

$$f^{n+1}_{q_i} = f^{eq^n}_{q_{i-m_q}} + (1-\hat{\omega}) \left(-f^{eq^n}_{q_{i-m_q}} + f^{eq^{n-1}}_{q_{i-2m_q}} + (1-\hat{\omega}) f^{neq^{n-1}}_{q_{i-2m_q}} \right), \tag{5.61}$$

$$= \hat{\omega} f^{eq^n}_{q_{i-m_q}} + (1-\hat{\omega}) \left(f^{eq^{n-1}}_{q_{i-2m_q}} + (1-\hat{\omega}) f^{neq^{n-1}}_{q_{i-2m_q}} \right). \tag{5.62}$$

Recursively inserting $f^{neq^{n+1}}_{q_i}$ from (5.60) into the non-equilibrium term of above equation with the transformation $n - j := n' + 1$ and $i - (j+1)m_q := i'$ where $j \in \{1, 2, \ldots, N-1\}$, we get

$$f^{n+1}_{q_i} = \hat{\omega} \left((1-\hat{\omega})^0 f^{eq^n}_{q_{i-m_q}} + (1-\hat{\omega})^1 f^{eq^{n-1}}_{q_{i-2m_q}} + \cdots + (1-\hat{\omega})^{N-1} f^{eq^{n-(N-1)}}_{q_{i-Nm_q}} \right) + (1-\hat{\omega})^N \left(f^{eq^{n-N}}_{q_{i-(N+1)m_q}} + (1-\hat{\omega}) f^{neq^{n-N}}_{q_{i-(N+1)m_q}} \right). \tag{5.63}$$

If $f^{neq^{n-N}}_{q_{i-(N+1)m_q}} = 0$, then we obtain (5.58). □

The above theorem depicts the multi-step nature of LBE by considering t_{n-N} as the initial time. That is, $f^{n+1}_{q_i}$ depends on the values of the equilibrium function in neighboring grid points at all previous time steps starting from the initial time t_{n-N}. Note that $f^{neq^{n-N}}_{q_{i-(N+1)m_q}} = 0$ as $f^{n-N}_{q_{i-(N+1)m_q}} = f^{eq^{n-N}}_{q_{i-(N+1)m_q}}$ is considered at the initial time.

Summing (5.58) over q with some form of equilibrium function discussed in section 5.3.2, we obtain the macroscopic finite difference form. In this work, we consider the upwind D$\overline{d}$Q($2\overline{d}$+1) model (*i.e.,* D1Q3 for one dimension). The equilibrium function for upwind D1Q3 model is,

$$f^{eq^n}_{1_i} = \frac{G^{+^n}_i}{\lambda_n} \tag{5.64}$$

$$f^{eq^n}_{2_i} = U^n_i - \frac{G^{+^n}_i + G^{-^n}_i}{\lambda_n} \tag{5.65}$$

$$f^{eq^n}_{3_i} = \frac{G^{-^n}_i}{\lambda_n} \tag{5.66}$$

and the corresponding velocities are $v_{1,n} = \lambda_n$, $v_{2,n} = 0$ and $v_{3,n} = -\lambda_n$. Thus, $m_1 = 1$, $m_2 = 0$ and $m_3 = -1$.

Remark 5.5 *For $k \in \{0, 1, \ldots, N\}$, we have*

$$\sum_{q=1}^{3} f^{eq^{n-k}}_{q_{i-(k+1)m_q}} = f^{eq^{n-k}}_{1_{i-(k+1)}} + f^{eq^{n-k}}_{2_i} + f^{eq^{n-k}}_{3_{i+(k+1)}} \tag{5.67}$$

$$= \frac{G^{+^{n-k}}_{i-(k+1)}}{\lambda_{n-k}} + U^{n-k}_i - \frac{G^{+^{n-k}}_i + G^{-^{n-k}}_i}{\lambda_{n-k}} + \frac{G^{-^{n-k}}_{i+(k+1)}}{\lambda_{n-k}} \tag{5.68}$$

$$= U^{n-k}_i - \frac{1}{\lambda_{n-k}} \left(\left(G^{+^{n-k}}_i - G^{+^{n-k}}_{i-(k+1)} \right) - \left(G^{-^{n-k}}_{i+(k+1)} - G^{-^{n-k}}_i \right) \right) \tag{5.69}$$

$$= U^{n-k}_i - \frac{\Delta t_{n-k}}{\Delta x} \left(\left(G^{+^{n-k}}_i - G^{+^{n-k}}_{i-(k+1)} \right) - \left(G^{-^{n-k}}_{i+(k+1)} - G^{-^{n-k}}_i \right) \right). \tag{5.70}$$

Defining the notation

$$\mathcal{U}^{n-k+1}_{i,(k+1)} := U^{n-k}_i - \frac{\Delta t_{n-k}}{\Delta x} \left(\left(G^{+^{n-k}}_i - G^{+^{n-k}}_{i-(k+1)} \right) - \left(G^{-^{n-k}}_{i+(k+1)} - G^{-^{n-k}}_i \right) \right), \tag{5.71}$$

$\sum_{q=1}^{3}$ (5.58) *becomes*

$$U^{n+1}_i = \hat{\omega} \left(\sum_{k=0}^{N-1} (1-\hat{\omega})^k \mathcal{U}^{n-k+1}_{i,(k+1)} \right) + (1-\hat{\omega})^N \mathcal{U}^{n-N+1}_{i,(N+1)}. \tag{5.72}$$

(5.72) is the macroscopic finite difference form of the LBEs in (5.7) (with $\hat{\omega} = \omega$) and (5.10) (with $\hat{\omega} = \tilde{\omega}$).

Remark 5.6 *If $\hat{\omega} = 1$, then $(1-\hat{\omega})^0 = 1$ and $(1-\hat{\omega})^k = 0$ for $k \in \{1, 2, \ldots, N\}$. In this case, the macroscopic finite difference form (5.72) becomes,*

$$U^{n+1}_i = \mathcal{U}^{n+1}_{i,1} = U^n_i - \frac{\Delta t_n}{\Delta x} \left(\left(G^{+^n}_i - G^{+^n}_{i-1} \right) - \left(G^{-^n}_{i+1} - G^{-^n}_i \right) \right) \tag{5.73}$$

which is an explicit (or forward) Euler upwind scheme for the hyperbolic system $\partial_t U + \partial_x G(U) = 0$.

Note: For $k \in \{0, 1, \ldots, N\}$, $\mathcal{U}^{n-k+1}_{i,(k+1)}$ in (5.71) is an explicit (or forward) Euler upwind discretisation of the hyperbolic system $\partial_t U + \partial_x G(U) = 0$, at time t_{n-k+1} with grid spacing $(k+1)\Delta x$. Thus, (5.72) which is the macroscopic finite difference form of LBE is simply a linear combination of upwind discretisations at varied time levels and grid spacings.

Remark 5.7 *For $0 < \hat{\omega} < 1$, $(1-\hat{\omega})^k > 0$ holds true for $k \in \{0, 1, \ldots, N\}$. Hence, in this case, numerical diffusion of the macroscopic finite difference form (5.72) has positively weighted contributions from each $\mathcal{U}_{i,(k+1)}^{n-k+1}$. Thus, when $0 < \hat{\omega} < 1$, it is expected that the numerical diffusion increases with decrease in $\hat{\omega}$ while all the parameters remain frozen.*
On the other hand, when $1 < \hat{\omega} < 2$, the sign of $(1-\hat{\omega})^k$ alternates with k. Therefore, numerical diffusion of the macroscopic finite difference form (5.72) experiences alternately signed (with respect to k) weighted contributions from $\mathcal{U}_{i,(k+1)}^{n-k+1}$.

As a consequence, the minimum (over $\hat{\omega}$) numerical diffusion in LBE obtained by semi-implicit discretisation of VKE is larger than that in the explicit case.

5.4.3 Consistency

In this section, we discuss the consistency of the macroscopic finite difference form (5.72) with the hyperbolic system $\partial_t U + \partial_x G(U) = 0$.

Theorem 5.3 *Under suitable smoothness assumptions on all involved variables, the expression (5.71) becomes*

$$\mathcal{U}_{i,(k+1)}^{n-k+1} = \begin{cases} U_i^n - \sum_{j=1}^{k} \Delta t_{n-j}\, \partial_t U|_i^n - (k+1)\Delta t_{n-k} \partial_x\, G|_i^n & \text{for } k \in \{1, 2, \ldots, N\} \\ U_i^n - (k+1)\Delta t_{n-k} \partial_x\, G|_i^n & \text{for } k = 0 \end{cases} \tag{5.74}$$

upto $\mathcal{O}\left(k(k+1)\Delta x^2\right)$, if $\Delta t_m = \mathcal{O}(\Delta x)\ \forall m$.

Proof: Taylor expanding each term in $\mathcal{U}_{i,(k+1)}^{n-k+1}$:

$$U_i^{n-k} = \begin{cases} U_i^n - \sum_{j=1}^{k} \Delta t_{n-j}\, \partial_t U|_i^n + \mathcal{O}\left((k\Delta x)^2\right) & \text{for } k \in \{1, 2, \ldots, N\} \\ U_i^n & \text{for } k = 0 \end{cases} \tag{5.75}$$

$$\text{since } \sum_{j=1}^{k} \Delta t_{n-j} = \mathcal{O}(k\Delta x) \text{ (as } \Delta t_m = \mathcal{O}(\Delta x), \forall m)$$

$$\left(G_i^{+^{n-k}} - G_{i-(k+1)}^{+^{n-k}}\right) = (k+1)\Delta x \partial_x\, G^+\big|_i^{n-k} + \mathcal{O}\left(((k+1)\Delta x)^2\right)$$
$$\left(G_{i+(k+1)}^{-^{n-k}} - G_i^{-^{n-k}}\right) = (k+1)\Delta x \partial_x\, G^-\big|_i^{n-k} + \mathcal{O}\left(((k+1)\Delta x)^2\right)$$

$$\left(G_i^{+^{n-k}} - G_{i-(k+1)}^{+^{n-k}}\right) - \left(G_{i+(k+1)}^{-^{n-k}} - G_i^{-^{n-k}}\right) = (k+1)\Delta x \partial_x \left(G^+ - G^-\right)\big|_i^{n-k}$$
$$= (k+1)\Delta x \partial_x \, G|_i^{n-k} \tag{5.76}$$
$$\text{upto } \mathcal{O}\left(((k+1)\Delta x)^2\right)$$

$$\partial_x \, G|_i^{n-k} = \begin{cases} \partial_x \, G|_i^n - \sum_{j=1}^k \Delta t_{n-j} \partial_{tx} \, G|_i^n + \mathcal{O}\left((k\Delta x)^2\right) & \text{for } k \in \{1, 2, \ldots, N\} \\ \partial_x \, G|_i^n & \text{for } k = 0 \end{cases} \tag{5.77}$$

For $k \in \{1, 2, \ldots, N\}$,

$$(k+1)\Delta x \partial_x \, G|_i^{n-k} \simeq (k+1)\Delta x \partial_x \, G|_i^n - \sum_{j=1}^k \Delta t_{n-j}(k+1)\Delta x \partial_{tx} \, G|_i^n$$
$$= (k+1)\Delta x \partial_x \, G|_i^n + \mathcal{O}\left(k(k+1)\Delta x^2\right), \tag{5.78}$$

since $\Delta t_m = \mathcal{O}(\Delta x), \forall m$.
Thus, inserting the above expressions into (5.71), we get (5.74). □

Remark 5.8 *Taylor expanding U_i^{n+1} about U_i^n, we get*

$$U_i^{n+1} = U_i^n + \Delta t_n \, \partial_t U|_i^n + \mathcal{O}\left(\Delta t_n^2\right). \tag{5.79}$$

Inserting (5.74) and the above expression into (5.72), we obtain

$$U_i^n + \Delta t_n \, \partial_t U|_i^n = \hat{\omega}\Bigg((1-\hat{\omega})^0 \left(U_i^n - \Delta t_n \partial_x \, G|_i^n\right)$$
$$+ \sum_{k=1}^{N-1} (1-\hat{\omega})^k \left(U_i^n - \sum_{j=1}^k \Delta t_{n-j} \, \partial_t U|_i^n - (k+1)\Delta t_{n-k} \partial_x \, G|_i^n \right) \Bigg)$$
$$+ (1-\hat{\omega})^N \left(U_i^n - \sum_{j=1}^N \Delta t_{n-j} \, \partial_t U|_i^n - (N+1)\Delta t_{n-N} \partial_x \, G|_i^n \right) \tag{5.80}$$

upto $\mathcal{O}\left(N(N+1)\Delta x^2\right)$. Upon simplifying the above expression, we obtain the following upto $\mathcal{O}\left(N(N+1)\Delta x^2\right)$:

$$\left(1 - \hat{\omega} \sum_{k=0}^{N-1} (1-\hat{\omega})^k - (1-\hat{\omega})^N \right) U_i^n$$
$$+ \left(\Delta t_n + \hat{\omega} \sum_{k=1}^{N-1} (1-\hat{\omega})^k \sum_{j=1}^k \Delta t_{n-j} + (1-\hat{\omega})^N \sum_{j=1}^N \Delta t_{n-j} \right) \partial_t U|_i^n$$

$$+\left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k(k+1)\Delta t_{n-k}+(1-\hat{\omega})^N(N+1)\Delta t_{n-N}\right)\partial_x\, G|_i^n=0. \quad (5.81)$$

Remark 5.9 *The coefficients of U_i^n and $\partial_t U|_i^n$ in (5.81) can be simplified as shown below:*

$$1-\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k-(1-\hat{\omega})^N=1-\hat{\omega}\left(\frac{1-(1-\hat{\omega})^N}{1-(1-\hat{\omega})}\right)-(1-\hat{\omega})^N,\ \textit{for}\ \hat{\omega}\neq 0$$
$$=0 \quad (5.82)$$

Since $1=\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k+(1-\hat{\omega})^N$, we have $\Delta t_n=\left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k+(1-\hat{\omega})^N\right)\Delta t_n$. Therefore,

$$\Delta t_n+\hat{\omega}\sum_{k=1}^{N-1}(1-\hat{\omega})^k\sum_{j=1}^{k}\Delta t_{n-j}+(1-\hat{\omega})^N\sum_{j=1}^{N}\Delta t_{n-j}=\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k\sum_{j=0}^{k}\Delta t_{n-j}+(1-\hat{\omega})^N\sum_{j=0}^{N}\Delta t_{n-j}. \quad (5.83)$$

Inserting (5.82) and (5.83) into (5.81), we obtain

$$\left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k\sum_{j=0}^{k}\Delta t_{n-j}+(1-\hat{\omega})^N\sum_{j=0}^{N}\Delta t_{n-j}\right)\partial_t U|_i^n$$
$$+\left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k(k+1)\Delta t_{n-k}+(1-\hat{\omega})^N(N+1)\Delta t_{n-N}\right)\partial_x\, G|_i^n=0. \quad (5.84)$$

upto $\mathcal{O}\left(N(N+1)\Delta x^2\right)$.

Remark 5.10 *If $\Delta t_m=\Delta t, \forall m$, then (5.84) becomes*

$$\left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k(k+1)\Delta t+(1-\hat{\omega})^N(N+1)\Delta t\right)\left(\partial_t U|_i^n+\partial_x\, G|_i^n\right)=\mathcal{O}\left(N(N+1)\Delta x^2\right),$$
$$\implies \partial_t U|_i^n+\partial_x\, G|_i^n=\mathcal{O}\left(N\Delta x\right). \quad (5.85)$$

Thus, in this case, the macroscopic finite difference form of LBE is consistent with the hyperbolic system.

Remark 5.11 *If $\hat{\omega}=1$, then $(1-\hat{\omega})^0=1$ and $(1-\hat{\omega})^k=0$ for $k\in\{1,2,\ldots,N\}$. Thus (5.84)*

becomes

$$\begin{aligned} \Delta t_n \left(\partial_t U|_i^n + \partial_x G|_i^n\right) &= \mathcal{O}\left(N(N+1)\Delta x^2\right), \\ \implies \partial_t U|_i^n + \partial_x G|_i^n &= \mathcal{O}\left(N(N+1)\Delta x\right). \end{aligned} \tag{5.86}$$

Therefore, the macroscopic finite difference form of LBE is consistent with the hyperbolic system for this case too.

Although the lattice Boltzmann algorithm is consistent for the two special cases: (i) constant time step size and (ii) $\hat{\omega} = 1$, it can be seen from (5.84) that consistency cannot be attained in the general case as $\sum_{j=0}^{k} \Delta t_{n-j} \neq (k+1)\Delta t_{n-k}$ for $k \in \{1, 2, \ldots, N\}$. However, one can choose constant Δt such that the sub-characteristic condition holds for all time steps. In this way, the algorithm will be consistent with the hyperbolic system for the choice of the time-step satisfying the sub-characteristic condition.

5.4.4 Total Variation Boundedness

The total variation boundedness (TVB) property of a numerical method for hyperbolic system ensures that the spatial variation remains bounded for all time steps. In this section, we discuss the TVB property of our lattice Boltzmann method by using its macroscopic finite difference form (5.72). This expression contains $\mathcal{U}_{i,(k+1)}^{n-k+1}$ for $k \in \{0, 1, \ldots, N\}$. For discussion of TVB property, we consider $\mathcal{U}_{i,(k+1)}^{n-k+1}$ derived by utilising upwind D1Q3 equilibrium function as in section 5.4.2.

Definition 5.2 *The total variation of any variable θ defined on a lattice structure indexed by i is given by,*

$$\mathbf{TV}(\theta) = \sum_i |\theta_{i+1} - \theta_i|$$

Theorem 5.4 *Let U_i^{n+1} given by (5.72) be the macroscopic finite difference form. Then, its total variation satisfies*

$$\mathbf{TV}\left(U^{n+1}\right) \leq \left(|\hat{\omega}| \sum_{k=0}^{N-1} |1-\hat{\omega}|^k + |1-\hat{\omega}|^N\right) \mathrm{C} \tag{5.87}$$

if $\mathbf{TV}\left(\mathcal{U}_{(k+1)}^{n-k+1}\right) \leq \mathrm{C}$, *for* $k \in \{0, 1, \ldots, N\}$.

Proof: Consider U_i^{n+1} given by (5.72). Then, $U_{i+1}^{n+1} - U_i^{n+1}$ becomes

$$U_{i+1}^{n+1} - U_i^{n+1} = \hat{\omega}\left(\sum_{k=0}^{N-1}(1-\hat{\omega})^k\left(\mathcal{U}_{i+1,(k+1)}^{n-k+1} - \mathcal{U}_{i,(k+1)}^{n-k+1}\right)\right) + (1-\hat{\omega})^N\left(\mathcal{U}_{i+1,(N+1)}^{n-N+1} - \mathcal{U}_{i,(N+1)}^{n-N+1}\right).$$

Then,

$$\left|U_{i+1}^{n+1} - U_i^{n+1}\right| \leq |\hat{\omega}|\sum_{k=0}^{N-1}|1-\hat{\omega}|^k\left|\mathcal{U}_{i+1,(k+1)}^{n-k+1} - \mathcal{U}_{i,(k+1)}^{n-k+1}\right| + |1-\hat{\omega}|^N\left|\mathcal{U}_{i+1,(N+1)}^{n-N+1} - \mathcal{U}_{i,(N+1)}^{n-N+1}\right|,$$

$$\implies \mathbf{TV}\left(U^{n+1}\right) \leq |\hat{\omega}|\sum_{k=0}^{N-1}|1-\hat{\omega}|^k\,\mathbf{TV}\left(\mathcal{U}_{(k+1)}^{n-k+1}\right) + |1-\hat{\omega}|^N\,\mathbf{TV}\left(\mathcal{U}_{(N+1)}^{n-N+1}\right).$$

Using $\mathbf{TV}\left(\mathcal{U}_{(k+1)}^{n-k+1}\right) \leq \mathrm{C}$ for $k \in \{0,1,\ldots,N\}$ in the above expression, we get (5.87). □

Remark 5.12 *If* $0 < \hat{\omega} \leq 1$, *then* (5.87) *becomes*

$$\mathbf{TV}\left(U^{n+1}\right) \leq \left(\hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k + (1-\hat{\omega})^N\right)\mathrm{C} \tag{5.88}$$

$$= C, \ \textit{since}\ \hat{\omega}\sum_{k=0}^{N-1}(1-\hat{\omega})^k + (1-\hat{\omega})^N = 1. \tag{5.89}$$

Therefore, if the underlying difference scheme $\left(\mathcal{U}_{i,(k+1)}^{n-k+1}\right)$ due to the choice of equilibrium function is TVB, then the lattice Boltzmann method induced by it is also TVB (*i.e.*, $\mathbf{TV}\left(U^{n+1}\right) \leq \mathrm{C}$) if $0 < \hat{\omega} \leq 1$.
Since upwind methods are TVB, $\mathbf{TV}\left(\mathcal{U}_{(k+1)}^{n-k+1}\right) \leq \mathrm{C}$ is true for the choice of upwind equilibrium function. Hence, the corresponding lattice Boltzmann method is also TVB if $0 < \hat{\omega} \leq 1$.

5.4.5 Positivity

Some of the variables of hyperbolic systems are positive for all time (*e.g.*, density and internal energy in Euler's system of gas dynamics, water height in shallow water system). Numerical schemes for such hyperbolic systems are expected to ensure the positivity of these variables. In this section, we show the positivity property of our lattice Boltzmann method by using its macroscopic finite difference form (5.72). $\mathcal{U}_{i,(k+1)}^{n-k+1}$ in this expression is derived by utilising upwind D1Q3 equilibrium function as in section 5.4.2.

Theorem 5.5 *Let* U_i^{n+1} *given by* (5.72) *be the macroscopic finite difference form. If* $\mathcal{U}_{i,(k+1)}^{n-k+1}$ *is positive for* $k \in \{0,1,\ldots,N\}$ *and* $0 < \hat{\omega} \leq 1$, *then* U_i^{n+1} *is positive.*

Proof: This is trivially seen from (5.72). □

Therefore, if the underlying difference scheme $\left(\mathfrak{U}_{i,(k+1)}^{n-k+1}\right)$ due to the choice of equilibrium function is positive, then the lattice Boltzmann method induced by it is also positive if $0 < \hat{\omega} \leq 1$.

Thus, we discussed some properties of our LBEs. To conclude, the stability-related properties like H-inequality, total variation boundedness, and positivity are realisable if the stronger condition $0 < \hat{\omega} \leq 1$ is satisfied (naturally satisfied in the semi-implicit case) while small numerical diffusion is realisable for $\hat{\omega} > 1$ (explicit case can be used in the interval $1 < \hat{\omega} < 2$ while ensuring positivity of numerical diffusion coefficient), depicting the trade-off between stability and accuracy.

Remark 5.13 *It is expected that the properties of LBM can be understood from its macroscopic finite difference form in (5.72) by utilising the properties of corresponding underlying difference scheme $\mathfrak{U}_{i,(k+1)}^{n-k+1}$ that occurs due to the choice of equilibrium functions. For instance, discrete conservation (with periodic boundary conditions) of LBM is evident if $\mathfrak{U}_{i,(k+1)}^{n-k+1}$ satisfies discrete conservation with periodic boundary conditions.*

Thus, in this section, novel discussions concerning LBEs derived by semi-implicit and explicit discretisations of VKE, on properties such as H-inequality, macroscopic finite difference form, consistency, total variation boundedness and positivity have been presented.

5.5 Hyperbolic conservation laws with source terms

In this section, we extend our lattice Boltzmann method to hyperbolic conservation laws with source terms. Consider

$$\partial_t U + \partial_{x_d} G^d(U) = S(U), \tag{5.90}$$

where $S(U)$ is the source term.

5.5.1 Vector-kinetic equation

To approximate (5.90), consider the vector-kinetic equation

$$\partial_t f_q + \partial_{x_d}\left(v_q^d f_q\right) = -\frac{1}{\epsilon}\left(f_q - f_q^{eq}(U)\right) + r(f_q). \tag{5.91}$$

Summing (5.91) over all q, we get

$$\partial_t \sum_{q=1}^{Q} f_q + \partial_{x_d} \sum_{q=1}^{Q} \left(v_q^d f_q\right) = -\frac{1}{\epsilon} \sum_{q=1}^{Q} \left(f_q - f_q^{eq}(U)\right) + \sum_{q=1}^{Q} r(f_q). \tag{5.92}$$

If $\sum_{q=1}^{Q} f_q = \sum_{q=1}^{Q} f_q^{eq} = U$ and $\sum_{q=1}^{Q} r(f_q) = S(U)$, the above equation becomes

$$\partial_t U + \partial_{x_d} \sum_{q=1}^{Q} \left(v_q^d f_q\right) = S(U). \tag{5.93}$$

In the limit $\epsilon \to 0$, we infer from (5.91) that $f_q \to f_q^{eq}(U)$. If $\sum_{q=1}^{Q} v_q^d f_q^{eq} = G^d(U)$, then (5.93) becomes (5.90) in the limit $\epsilon \to 0$.
Hereafter, we denote $r_q := r(f_q)$ for convenience.

5.5.2 Lattice Boltzmann equation

As in section 5.3, f_q in the collision term can be treated both explicitly and implicitly leading to LBEs with $\hat{\omega} = \omega$ and $\hat{\omega} = \tilde{\omega} = \frac{\omega}{1+\omega}$ respectively. The source term r_q is discretised in Crank-Nicolson fashion. Thus, the LBE becomes

$$\begin{aligned} f_q\left(\mathbf{x}, t+\Delta t\right) = (1-\hat{\omega}) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \hat{\omega} f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right) \\ + \frac{\Delta t}{2}\left(r_q\left(\mathbf{x}, t+\Delta t\right) + r_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right). \end{aligned} \tag{5.94}$$

The collision-streaming algorithm

$$\begin{aligned} \text{Collision: } F_q^* &= (1-\hat{\omega}) f_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) + \hat{\omega} f_q^{eq}\left(U\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right)\right) + \frac{\Delta t}{2} r_q\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) \\ \text{Streaming: } & \qquad F_q\left(\mathbf{x}, t+\Delta t\right) = F_q^*\left(\mathbf{x} - \mathbf{v_q}\Delta t, t\right) \end{aligned}$$

can be used to numerically implement the LBEs. After finding $F_q\left(\mathbf{x}, t+\Delta t\right) = f_q\left(\mathbf{x}, t+\Delta t\right) - \frac{\Delta t}{2} r_q\left(\mathbf{x}, t+\Delta t\right)$, we find $U\left(\mathbf{x}, t+\Delta t\right)$ by solving

$$\sum_q F_q\left(\mathbf{x}, t+\Delta t\right) = U\left(\mathbf{x}, t+\Delta t\right) - \frac{\Delta t}{2} S(U\left(\mathbf{x}, t+\Delta t\right)) \tag{5.95}$$

using a non-linear iterative solver (e.g., Newton's root finding method).

5.5.3 Chapman-Enskog expansion

The Chapman-Enskog expansion can be obtained by first Taylor expanding the LBE (5.94) as,

$$\left(\partial_t + \mathbf{v_q}\cdot\nabla\right) f_q = -\frac{\hat{\omega}}{\Delta t}\left(f_q - f_q^{eq}\right) + \frac{\hat{\omega}}{2}\left(\partial_t + \mathbf{v_q}\cdot\nabla\right)\left(f_q - f_q^{eq}\right) + r_q + \mathcal{O}(\Delta t^2). \tag{5.96}$$

Consider the perturbation expansions:

$$f_q = f_q^{eq} + \epsilon f_q^{(1)} + \epsilon^2 f_q^{(2)} + \ldots;\ r_q = \epsilon r_q^{(1)} + \epsilon^2 r_q^{(2)} + \ldots \tag{5.97}$$

Since $\sum_{q=1}^{Q} f_q = \sum_{q=1}^{Q} f_q^{eq} = U$, we have $\sum_{q=1}^{Q} f_q^{(i)} = 0, \forall i \in \mathbb{N}$. Multiple scale expansion of derivatives of f_q gives $\partial_t f_q = \left(\epsilon\partial_t^{(1)} + \epsilon^2\partial_t^{(2)} + ...\right) f_q$ and $\mathbf{v_q}\cdot\nabla f_q = \epsilon\mathbf{v_q}\cdot\nabla^{(1)} f_q$.
Using perturbation expansion of f_q, r_q and multiple scale expansion of derivatives of f_q in (5.96) and separating out $O(\epsilon)$ and $O(\epsilon^2)$ terms,

$$O(\epsilon): \qquad \left(\partial_t^{(1)} + \mathbf{v_q}\cdot\nabla^{(1)}\right) f_q^{eq} = -\frac{\hat{\omega}}{\Delta t} f_q^{(1)} + r_q^{(1)} \tag{5.98}$$

$$O(\epsilon^2): \ \partial_t^{(2)} f_q^{eq} + \left(1 - \frac{\hat{\omega}}{2}\right)\left(\partial_t^{(1)} + \mathbf{v_q}\cdot\nabla^{(1)}\right) f_q^{(1)} = -\frac{\hat{\omega}}{\Delta t} f_q^{(2)} + r_q^{(2)} \tag{5.99}$$

Zeroth moment $\left(\sum_{q=1}^{Q}\right)$ of $O(\epsilon)$ terms in (5.98) and $O(\epsilon^2)$ terms in (5.99) respectively give

$$\partial_t^{(1)} U + \partial_{x_d}^{(1)} G^d(U) = \sum_{q=1}^{Q} r_q^{(1)}, \tag{5.100}$$

$$\partial_t^{(2)} U + \left(1 - \frac{\hat{\omega}}{2}\right)\partial_{x_d}^{(1)}\left(\sum_{q=1}^{Q} v_q^{(d)} f_q^{(1)}\right) = \sum_{q=1}^{Q} r_q^{(2)}. \tag{5.101}$$

From the first moment $\left(\sum_{q=1}^{Q} v_q^d\right)$ of $O(\epsilon)$ terms in (5.98), we get

$$\sum_{q=1}^{Q} v_q^d f_q^{(1)} = -\frac{\Delta t}{\hat{\omega}}\left(\partial_U G^d\left(-\partial_U G^i \partial_{x_i}^{(1)} U + \sum_{q=1}^{Q} r_q^{(1)}\right) - \sum_{q=1}^{Q} v_q^d r_q^{(1)} + \partial_{x_i}^{(1)}\left(\sum_{q=1}^{Q} v_q^d v_q^i f_q^{eq}\right)\right) \tag{5.102}$$

Recombining the zeroth moment equations of $O(\epsilon)$ in (5.100) and $O(\epsilon^2)$ in (5.101) and reversing the multiple scale expansions, we get

$$\partial_t U + \partial_{x_d} G^d(U) = S(U) + \Delta t \left(\frac{1}{\hat{\omega}} - \frac{1}{2}\right) \Bigg(\underbrace{\partial_{x_d}\left(\partial_{x_i}\left(\sum_{q=1}^{Q} v_q^d v_q^i f_q^{eq}\right) - \partial_U G^d \partial_U G^i \partial_{x_i} U\right)}_{\text{Numerical Diffusion}} + \underbrace{\partial_{x_d}\left(\partial_U G^d \sum_{q=1}^{Q} r_q - \sum_{q=1}^{Q} v_q^d r_q\right)}_{\text{Spurious Numerical Convection}} \Bigg). \quad (5.103)$$

5.5.4 Spurious Numerical Convection and modelling r_q

The spurious numerical convection in (5.103) due to the discretisation of source term must by avoided in order to have a reliable numerical method. Therefore, we require r_q to satisfy

$$\sum_{q=1}^{Q} v_q^d r_q = \partial_U G^d \sum_{q=1}^{Q} r_q = \partial_U G^d S(U). \quad (5.104)$$

Thus, an r_q that satisfies (5.104) along with $\sum_{q=1}^{Q} r_q = S(U)$ is required. Note that these requirements are similar to those imposed on f_q^{eq}, and hence expressions similar to those in section 5.3.2 could be obtained for different models:

$$\textbf{Classical D1Q2}: \quad r_q = \frac{1}{2}S(U) - (-1)^q \frac{1}{2\lambda}\partial_U G^1 S(U), \text{ for } q \in \{1,2\}$$

$$\textbf{D1Q3}: \quad r_q = \frac{1}{3}S(U) + (\delta_{q1} - \delta_{q3})\frac{1}{2\lambda}\partial_U G^1 S(U), \text{ for } q \in \{1,2,3\}$$

$$\textbf{Upwind D}\overline{\textbf{d}}\textbf{Q}(\textbf{2}\overline{\textbf{d}}+\textbf{1}): r_q = \begin{cases} \frac{\partial_U G^{q+} S(U)}{\lambda_q}, & \text{for } q \in \{1,2,\ldots,\overline{d}\} \\ S(U) - \sum_{d=1}^{\overline{d}} \left(\frac{\left(\partial_U G^{d+} + \partial_U G^{d-}\right) S(U)}{\lambda_d}\right), & \text{for } q = \overline{d}+1 \\ \frac{\partial_U G^{(q-(\overline{d}+1))-} S(U)}{\lambda_{q-(\overline{d}+1)}}, & \text{for } q \in \{\overline{d}+2,\ldots,2\overline{d}+1\} \end{cases}$$

Thus, after finding $U(\mathbf{x}, t+\Delta t)$ by solving (5.95), we can find $f_q^{eq}(U(\mathbf{x}, t+\Delta t))$ (as discussed in section 5.3.2) and $r_q(U(\mathbf{x}, t+\Delta t))$ (as discussed above), before proceeding with the next time step.

Remark 5.14 *We modelled r_q such that the spurious numerical convection due to discretisation of source term is nullified. This prevents the occurrence of spurious wave speeds and incorrect locations of discontinuities, commonly encountered in literature. Thus, balancing of convection and source terms which is a crucial problem in the finite volume framework, can be easily*

handled in the lattice Boltzmann framework. Our strategy thus enforces the desired property of **well-balancing** *and, at the same time, takes care of stiffness of the source terms to a significant extent.*

Note that r_q is of $\mathcal{O}(\epsilon)$ in (5.97). Hence, our method and underlying removal of numerical convection works well for $S(U) = \sum_{q=1}^{Q} r_q = \mathcal{O}(\epsilon)$.

5.6 D2Q9 model of lattice Boltzmann method

The equilibrium function in (5.28) causes the underlying difference scheme $\left(\mathcal{U}_{i,(k+1)}^{n-k+1}\right)$ to result in pure upwinding along the coordinate directions. In this section, in addition to discrete velocities moving along coordinate directions, we also introduce discrete velocities moving along diagonal-to-coordinate directions. This enables the splitting of positive and negative fluxes even along diagonal-to-coordinate directions, thereby resulting in better multi-dimensional behavior. We consider two dimensions and a uniform lattice with equal grid spacing, with $\Delta x_1 = \Delta x_2 := \Delta x$, in our presentation.

5.6.1 Equilibrium function

We consider 9 discrete velocities: $\mathbf{v_1} = [\lambda, 0]$, $\mathbf{v_2} = [0, \lambda]$, $\mathbf{v_3} = [\lambda, \lambda]$, $\mathbf{v_4} = [-\lambda, \lambda]$, $\mathbf{v_5} = [0, 0]$, $\mathbf{v_6} = [-\lambda, 0]$, $\mathbf{v_7} = [0, -\lambda]$, $\mathbf{v_8} = [-\lambda, -\lambda]$, $\mathbf{v_9} = [\lambda, -\lambda]$, and the corresponding equilibrium functions:

$$
\begin{gathered}
f_1^{eq} = \frac{G^{\alpha+}}{\lambda},\ f_2^{eq} = \frac{G^{\beta+}}{\lambda},\ f_3^{eq} = \frac{G^{\gamma+}}{\lambda},\ f_4^{eq} = \frac{G^{\zeta+}}{\lambda}, \\
f_5^{eq} = U - \frac{1}{\lambda}\left(\left(G^{\alpha+} + G^{\beta+} + G^{\gamma+} + G^{\zeta+}\right) + \left(G^{\alpha-} + G^{\beta-} + G^{\gamma-} + G^{\zeta-}\right)\right), \\
f_6^{eq} = \frac{G^{\alpha-}}{\lambda},\ f_7^{eq} = \frac{G^{\beta-}}{\lambda},\ f_8^{eq} = \frac{G^{\gamma-}}{\lambda},\ f_9^{eq} = \frac{G^{\zeta-}}{\lambda}.
\end{gathered}
\tag{5.105}
$$

Here $G^{l+} - G^{l-} = G^l$ for $l \in \mathcal{Z} = \{\alpha, \beta, \gamma, \zeta\}$. These equilibrium functions satisfy $\sum_{q=1}^{Q=9} f_q^{eq} = U$. In order to ensure $\sum_{q=1}^{Q=9} v_q^d f_q^{eq} = G^d(U)$, we need to satisfy the following requirements:

$$
\sum_{q=1}^{Q=9} v_q^1 f_q^{eq} = G^{\alpha} + G^{\gamma} - G^{\zeta} = G^1(U), \tag{5.106}
$$

$$
\sum_{q=1}^{Q=9} v_q^2 f_q^{eq} = G^{\beta} + G^{\gamma} + G^{\zeta} = G^2(U). \tag{5.107}
$$

Thus, we have

$$G^{\gamma} = \frac{G^2 + G^1}{2} - \frac{G^{\beta} + G^{\alpha}}{2} \text{ and } G^{\zeta} = \frac{G^2 - G^1}{2} - \frac{G^{\beta} - G^{\alpha}}{2}, \ \forall G^{\alpha}, G^{\beta}. \tag{5.108}$$

In this setting, the underlying difference scheme corresponding to the equilibrium function (5.105) is:

$$\begin{aligned}
\mathcal{U}_{i,j,(k+1)}^{n-k+1} := U_{i,j}^{n-k} &- \frac{\Delta t_{n-k}}{\Delta x}\left(\left(G_{i,j}^{\alpha+^{n-k}} - G_{i-(k+1),j}^{\alpha+^{n-k}}\right) - \left(G_{i+(k+1),j}^{\alpha-^{n-k}} - G_{i,j}^{\alpha-^{n-k}}\right)\right) \\
&- \frac{\Delta t_{n-k}}{\Delta x}\left(\left(G_{i,j}^{\beta+^{n-k}} - G_{i,j-(k+1)}^{\beta+^{n-k}}\right) - \left(G_{i,j+(k+1)}^{\beta-^{n-k}} - G_{i,j}^{\beta-^{n-k}}\right)\right) \\
&- \frac{\Delta t_{n-k}}{\Delta x}\left(\left(G_{i,j}^{\gamma+^{n-k}} - G_{i-(k+1),j-(k+1)}^{\gamma+^{n-k}}\right) - \left(G_{i+(k+1),j+(k+1)}^{\gamma-^{n-k}} - G_{i,j}^{\gamma-^{n-k}}\right)\right) \\
&- \frac{\Delta t_{n-k}}{\Delta x}\left(\left(G_{i,j}^{\zeta+^{n-k}} - G_{i+(k+1),j-(k+1)}^{\zeta+^{n-k}}\right) - \left(G_{i-(k+1),j+(k+1)}^{\zeta-^{n-k}} - G_{i,j}^{\zeta-^{n-k}}\right)\right).
\end{aligned} \tag{5.109}$$

Further, the Chapman-Enskog expansion (5.21) corresponding to the equilibrium function (5.105) becomes,

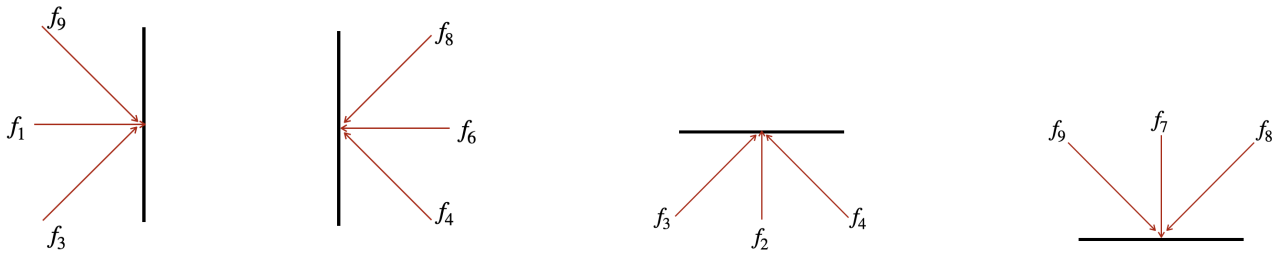
$$\begin{aligned}
\partial_t U + \partial_{x_d} G^d(U) = \Delta t &\left(\frac{1}{\hat{\omega}} - \frac{1}{2}\right) \\
&\left(\partial_{x_1}\left(\lambda\partial_U\left(G^{\alpha+} + G^{\alpha-} + G^{\gamma+} + G^{\gamma-} + G^{\zeta+} + G^{\zeta-}\right) - \left(\partial_U G^1\right)^2\right)\partial_{x_1} U\right. \\
&\quad + \partial_{x_1}\left(\lambda\partial_U\left(G^{\gamma+} + G^{\gamma-} - G^{\zeta+} - G^{\zeta-}\right) - \partial_U G^1 \partial_U G^2\right)\partial_{x_2} U \\
&\quad + \partial_{x_2}\left(\lambda\partial_U\left(G^{\gamma+} + G^{\gamma-} - G^{\zeta+} - G^{\zeta-}\right) - \partial_U G^2 \partial_U G^1\right)\partial_{x_1} U \\
&\left. + \partial_{x_2}\left(\lambda\partial_U\left(G^{\beta+} + G^{\beta-} + G^{\gamma+} + G^{\gamma-} + G^{\zeta+} + G^{\zeta-}\right) - \left(\partial_U G^2\right)^2\right)\partial_{x_2} U\right).
\end{aligned} \tag{5.110}$$

Thus, in addition to upwinding along coordinate directions, this model allows upwinding even along diagonal-to-coordinate directions.

5.6.2 Boundary conditions

In this sub-section, we present the expressions for f_q corresponding to those specific q that are unknown at the boundaries. At boundary, the macroscopic variables $U, G^{\alpha}, G^{\beta}, G^{\gamma}$ and G^{ζ} are known. From these, the split fluxes $G^{\alpha^{\pm}}, G^{\beta^{\pm}}, G^{\gamma^{\pm}}$ and $G^{\zeta^{\pm}}$ can be found. Using these split fluxes, equilibrium functions can be evaluated at the boundary. Thus, by taking $f_q^{neq} = f_q - f_q^{eq}\ \forall q \in \{1, 2, .., 9\}$, it can be inferred from the definition of conserved moment $\sum_{n=1}^{9} f_q =$

$\sum_{q=1}^{9} f_q^{eq} = U$ that, $\sum_{q=1}^{9} f_q^{neq} = 0$.



(a) Left (b) Right (c) Bottom (d) Top

Figure 5.2: Boundary conditions (Black lines indicate boundaries; red arrows indicate unknown functions at each boundary)

5.6.2.1 Left boundary

At any point on left boundary, f_2, f_4, f_5, f_6, f_7 and f_8 are known from the computational domain, as these functions from neighbouring points (in the computational domain) hop to points on left boundary. Let $\mathcal{I}$ be the set of these known functions. The unknowns at left boundary are f_1, f_3 and f_9 (as shown in figure 5.2a), as these functions must come from the outside of computational domain to left boundary, and let $\mathcal{J}$ be the set of these unknown functions. Since f_q^{eq} can be evaluated $\forall q \in \{1, 2, .., 9\}$ and f_q is known $\forall q \in \mathcal{I}$, $f_q^{neq} = f_q - f_q^{eq}$ can be found $\forall q \in \mathcal{I}$ (as $\mathcal{I} \subset \{1, 2, .., 9\}$). Then $f_q^{neq}, \forall q \in \mathcal{J}$ can be written as,

$$f_3^{neq} = -f_8^{neq} - \frac{f_2^{neq} + f_5^{neq} + f_7^{neq}}{3} \tag{5.111}$$

$$f_1^{neq} = -f_6^{neq} - \frac{f_2^{neq} + f_5^{neq} + f_7^{neq}}{3} \tag{5.112}$$

$$f_9^{neq} = -f_4^{neq} - \frac{f_2^{neq} + f_5^{neq} + f_7^{neq}}{3} \tag{5.113}$$

satisfying $\sum_{q=1}^{9} f_q^{neq} = 0$. Now, $f_q = f_q^{eq} + f_q^{neq}\ \forall q \in \mathcal{J}$ can be found to be,

$$f_3 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda} \sum_{z \in \mathbb{Z}, z \neq \beta} \left(G^{z+} + G^{z-}\right) - f_8 - \frac{f_2 + f_5 + f_7}{3} \tag{5.114}$$

$$f_1 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda} \sum_{z \in \mathbb{Z}, z \neq \beta} \left(G^{z+} + G^{z-}\right) - f_6 - \frac{f_2 + f_5 + f_7}{3} \tag{5.115}$$

$$f_9 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda} \sum_{z \in \mathbb{Z}, z \neq \beta} \left(G^{z+} + G^{z-}\right) - f_4 - \frac{f_2 + f_5 + f_7}{3} \tag{5.116}$$

5.6.2.2 Right boundary

By following the same procedure of obtaining left boundary conditions, the unknown functions at right boundary (as shown in figure 5.2b) can be found as,

$$f_4 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\beta}\left(G^{z+} + G^{z-}\right) - f_9 - \frac{f_2 + f_5 + f_7}{3} \tag{5.117}$$

$$f_6 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\beta}\left(G^{z+} + G^{z-}\right) - f_1 - \frac{f_2 + f_5 + f_7}{3} \tag{5.118}$$

$$f_8 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\beta}\left(G^{z+} + G^{z-}\right) - f_3 - \frac{f_2 + f_5 + f_7}{3} \tag{5.119}$$

5.6.2.3 Bottom boundary

The unknown functions at bottom boundary (as shown in figure 5.2c) can be found as,

$$f_3 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_8 - \frac{f_1 + f_5 + f_6}{3} \tag{5.120}$$

$$f_2 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_7 - \frac{f_1 + f_5 + f_6}{3} \tag{5.121}$$

$$f_4 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_9 - \frac{f_1 + f_5 + f_6}{3} \tag{5.122}$$

5.6.2.4 Top boundary

The unknown functions at top boundary (as shown in figure 5.2d) can be found as,

$$f_9 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_4 - \frac{f_1 + f_5 + f_6}{3} \tag{5.123}$$

$$f_7 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_2 - \frac{f_1 + f_5 + f_6}{3} \tag{5.124}$$

$$f_8 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} + \frac{U}{3} - \frac{1}{3\lambda}\sum_{z\in\mathcal{Z}, z\neq\alpha}\left(G^{z+} + G^{z-}\right) - f_3 - \frac{f_1 + f_5 + f_6}{3} \tag{5.125}$$

5.6.2.5 Bottom-left corner

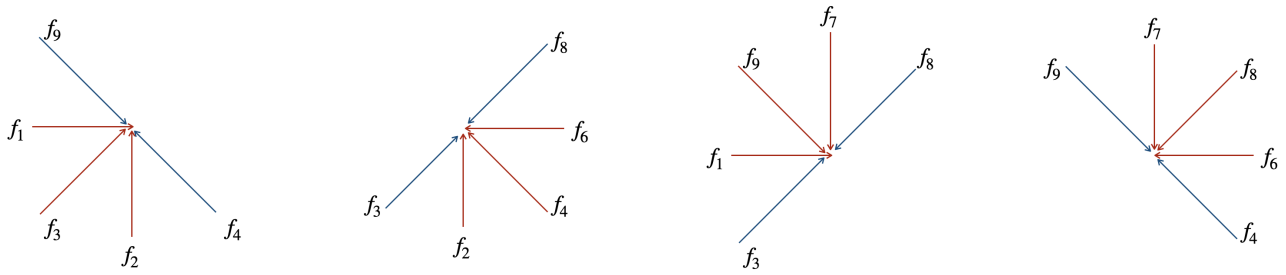
At bottom left corner, the known equilibrium functions are f_7, f_8, f_5 and f_6. The unknown equilibrium functions are f_1, f_3, f_2, f_4 and f_9. Since f_4 and f_9 do not enter or leave the compu-

tational domain, evaluation of them is not needed. Hence, it can be assumed that $f_9^{neq} + f_4^{neq} + f_5^{neq} = 0$. Then f_q^{neq} for other unknown equilibrium distribution functions can be written as,

$$f_1^{neq} = -f_6^{neq} \tag{5.126}$$
$$f_3^{neq} = -f_8^{neq} \tag{5.127}$$
$$f_2^{neq} = -f_7^{neq} \tag{5.128}$$

satisfying $\sum_{q=1}^{9} f_q^{neq} = 0$. Now, $f_q = f_q^{eq} + f_q^{neq}$ can be found to be,

$$f_1 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} - f_6 \tag{5.129}$$
$$f_3 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} - f_8 \tag{5.130}$$
$$f_2 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} - f_7 \tag{5.131}$$



(a) Bottom-left (b) Bottom-right (c) Top-left (d) Top-right

Figure 5.3: Corner conditions (Red arrows indicate unknown functions that are evaluated; Blue arrows indicate unknown functions that are not evaluated)

5.6.2.6 Bottom-right corner

By following the same procedure for obtaining bottom-left corner conditions, the bottom-right corner conditions (as shown in figure 5.3b) are found to be,

$$f_2 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} - f_7 \tag{5.132}$$
$$f_4 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} - f_9 \tag{5.133}$$
$$f_6 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} - f_1 \tag{5.134}$$

5.6.2.7 Top-left corner

The top-left corner conditions (as shown in figure 5.3c) are,

$$f_1 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} - f_6 \tag{5.135}$$

$$f_9 = \frac{G^{\zeta+} + G^{\zeta-}}{\lambda} - f_4 \tag{5.136}$$

$$f_7 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} - f_2 \tag{5.137}$$

5.6.2.8 Top-right corner

The top-right corner conditions (as shown in figure 5.3d) are,

$$f_6 = \frac{G^{\alpha+} + G^{\alpha-}}{\lambda} - f_1 \tag{5.138}$$

$$f_7 = \frac{G^{\beta+} + G^{\beta-}}{\lambda} - f_2 \tag{5.139}$$

$$f_8 = \frac{G^{\gamma+} + G^{\gamma-}}{\lambda} - f_3 \tag{5.140}$$

5.7 Numerical results

In this section, we present the numerical validation of our lattice Boltzmann methods (LBM) discussed in the previous sections. Firstly, we depict the influence of ω on numerical diffusion and order of accuracy. Then, we numerically validate our LBM for hyperbolic conservation laws with source terms, and D2Q9 model of LBM. For all the cases, the numerical results are obtained by using LBE derived by explicit discretisation of VKE. Due to the algorithmic similarity of LBEs derived by explicit and semi-implicit discretisation of VKE, the numerical results obtained by semi-implicit case for $0 < \tilde{\omega} < 1$ are same as that obtained by explicit case for $0 < \omega < 1$. Hence we only present the numerical validation of explicit case with larger interval $0 < \omega < 2$.

5.7.1 Sinusoidal initial condition

The domain of the problem is $[0, 1] \subset \mathbb{R}$. We consider inviscid Burgers' equation with flux function as $G^1(U) = \frac{1}{2}U^2$. The initial condition is $U(x_1, 0) = \sin(2\pi x_1)$, and boundary is periodic. An LBM with upwind D1Q3 equilibrium functions is utilised to obtain the numerical solution. $\lambda_1 = \frac{\Delta x_1}{\Delta t}$ is chosen such that the sub-characteristic condition in (5.32) (which simplifies in this case as $\lambda_1 \geq \sup_{i \in \Omega_g} |U_i|$, where Ω_g is the set of grid points) is satisfied. Since we expect

the numerical solution to be bounded between -1 and 1 for all times, we choose $\lambda_1 = 1$, and fix $\Delta t = \frac{\Delta x_1}{\lambda_1}$ for all time steps in order to have a consistent discretisation of the inviscid Burgers' equation (as discussed in remark 5.10). Further, we consider different values for $\omega = \frac{\Delta t}{\epsilon}$ such as, $\omega = 0.1, 0.6, 1.0, 1.4, 1.9$ and compare their numerical diffusion by freezing all the other parameters. We also consider discretisation of the domain with different number of grid points N such as, $N = 41, 81, 161, 321$ in order to study the order of convergence. The reference solution utilised in finding the L_2 error norm is obtained by evaluating the method of characteristics solution with a tolerance of 10^{-15}.

Tables 5.1 and 5.2 show the L_2 error norms and convergence orders for different values of ω at time $T = \frac{0.1}{2\pi}$ while the solution is still smooth. It is seen from the tables that for each fixed value of N, L_2 error norm of the numerical solution increases with decrease in ω, validating the remark 5.7. Further, although only first order of accuracy is expected according to Chapman-Enskog expansion (5.31), we observe more than second order accuracy for large values of ω. This increase in order of accuracy for large values of ω can be attributed to the smaller numerical diffusion for $\omega > 1$ when compared to $\omega < 1$, as mentioned in remark 5.7. We also observe that $\mathcal{O}(L_2)$ corresponding to a fixed N increases with increase in ω.

N	Δx_1	L_2, $\omega =$ 1.9	EOC, $\omega =$ 1.9	L_2, $\omega =$ 1.4	EOC, $\omega =$ 1.4	L_2, $\omega =$ 1.0	EOC, $\omega =$ 1.0
41	0.025	0.000597	-	0.000597	-	0.000597	-
81	0.0125	9.68×10^{-5}	2.626	0.000158	1.915	0.000230	1.380
161	0.00625	2.14×10^{-5}	2.175	3.88×10^{-5}	2.032	6.41×10^{-5}	1.841
321	0.003125	3.20×10^{-6}	2.744	1.20×10^{-5}	1.690	2.33×10^{-5}	1.460

Table 5.1: Sinusoidal initial condition at $T = \frac{0.1}{2\pi}$ for $\omega = 1.9, 1.4, 1.0$

5.7.2 LBM for hyperbolic conservation laws with source terms

The governing equation is of the form (5.90) with p=1 (scalar conservation law). We show that our scheme captures the discontinuities at correct locations due to the nullification of spurious numerical convection by our choice of r_q. Further, since $r_q = \mathcal{O}(\epsilon)$ is essential for such a possibility of nullification as mentioned in remark 5.14, the numerical results with correct locations of discontinuities are presented whenever $S(U) = \mathcal{O}(\epsilon)$.

N	Δx_1	$L_2,\ \omega = 0.6$	$EOC,\ \omega = 0.6$	$L_2,\ \omega = 0.1$	$EOC,\ \omega = 0.1$
41	0.025	0.000597	-	0.000597	-
81	0.0125	0.000306	0.965	0.000405	0.562
161	0.00625	0.000100	1.611	0.000161	1.325
321	0.003125	4.38×10^{-5}	1.194	0.000103	0.644

Table 5.2: Sinusoidal initial condition at $T = \frac{0.1}{2\pi}$ for $\omega = 0.6, 0.1$

5.7.2.1 One dimensional discontinuity

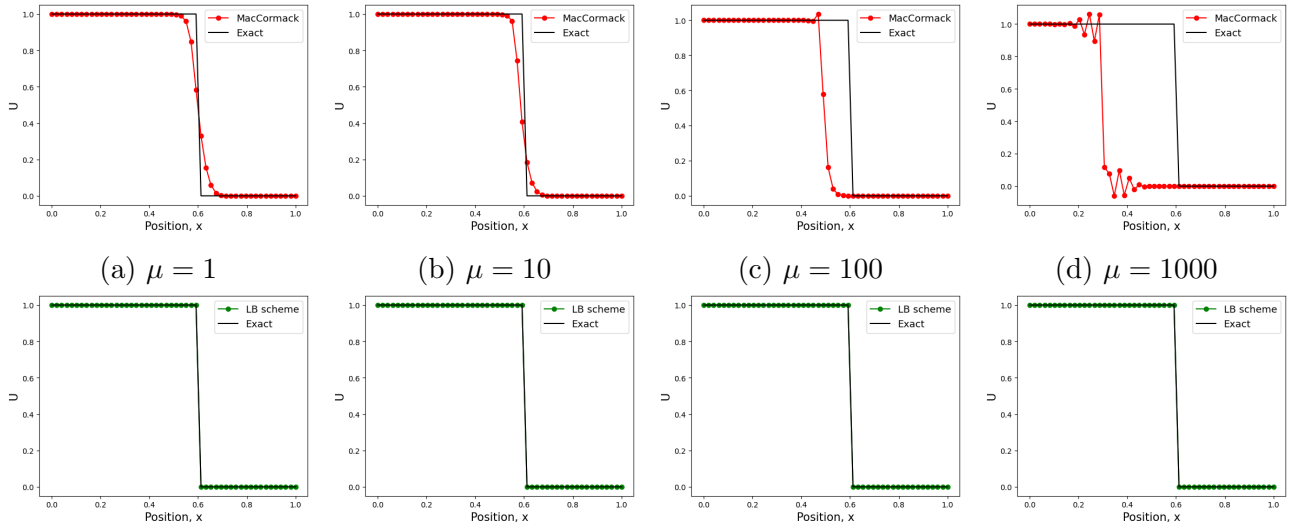
This is the test problem used by LeVeque and Yee [202] to understand the cause for incorrectness in speeds of discontinuities for stiff source terms. The domain is $[0, 1] \subset \mathbb{R}$, and is split up into 50 evenly spaced grid points. For this problem, $G^1(U) = U$ and $S(U) = -\mu U(U-1)(U-\frac{1}{2})$. Initial conditions are:

$$U(x_1, 0) = \begin{cases} 1 & \text{for } x_1 \leq 0.3 \\ 0 & \text{for } x_1 > 0.3 \end{cases}.$$

Boundary conditions are: $U(0, t) = 1$ and $U(1, t) = 0$ for $t \geq 0$. An LBM with upwind D1Q3 form for f_q^{eq} and r_q is utilised to obtain the numerical solution. $\lambda_1 = \frac{\Delta x_1}{\Delta t}$ is chosen such that the sub-characteristic condition in (5.32) (which simplifies in this case as $\lambda_1 \geq 1$) is satisfied. In particular, we use $\lambda_1 = 1$, and this incidentally results in numerical solution being the same as method of characteristics solution (even in smooth regions) since the wave-speed in the problem is also 1. Therefore, in addition to capturing discontinuities at correct locations (due to our choice of r_q), the solution is also exact in smooth regions. Further, the time step is chosen as $\Delta t = \frac{\Delta x_1}{\lambda_1}$. We also consider $\omega = 1$ for the simulation of this problem and this ensures consistency with the governing equation irrespective of the choice of Δt (as discussed in remark 5.11).

A comparison of numerical solutions reproduced from LeVeque and Yee [202] and numerical solutions obtained from our LB scheme is shown in figure 5.4 at $T = 0.3$ for different values of μ. The MacCormack's method suffers from spurious numerical convection for μ as small as 100, while our LB scheme is devoid of the effects of spurious numerical convection until $\mu = 1000$. We observe numerical convection in LB scheme for $\mu \geq 10000$ (not shown in figure), and this validates the remark 5.14 that our scheme is suitable when $S(U) = \mathcal{O}(\epsilon)$.

Hence, for this problem, we can infer that $\epsilon = \mathcal{O}(k\tilde{\mu})$ where $\tilde{\mu}$ represents the value of μ upto which the method of nullification of numerical convection works. Thus, $\epsilon = \mathcal{O}(k10^3)$ for some



(a) $\mu = 1$ (b) $\mu = 10$ (c) $\mu = 100$ (d) $\mu = 1000$

(e) $\mu = 1$ (f) $\mu = 10$ (g) $\mu = 100$ (h) $\mu = 1000$

Figure 5.4: Top: Extended MacCormack's method with limiter based on U^n(Reproduced from [202]), Bottom: Our LB scheme for hyperbolic conservation laws with source terms

constant $k < 10^{-3}$.

5.7.2.2 Two dimensional discontinuity

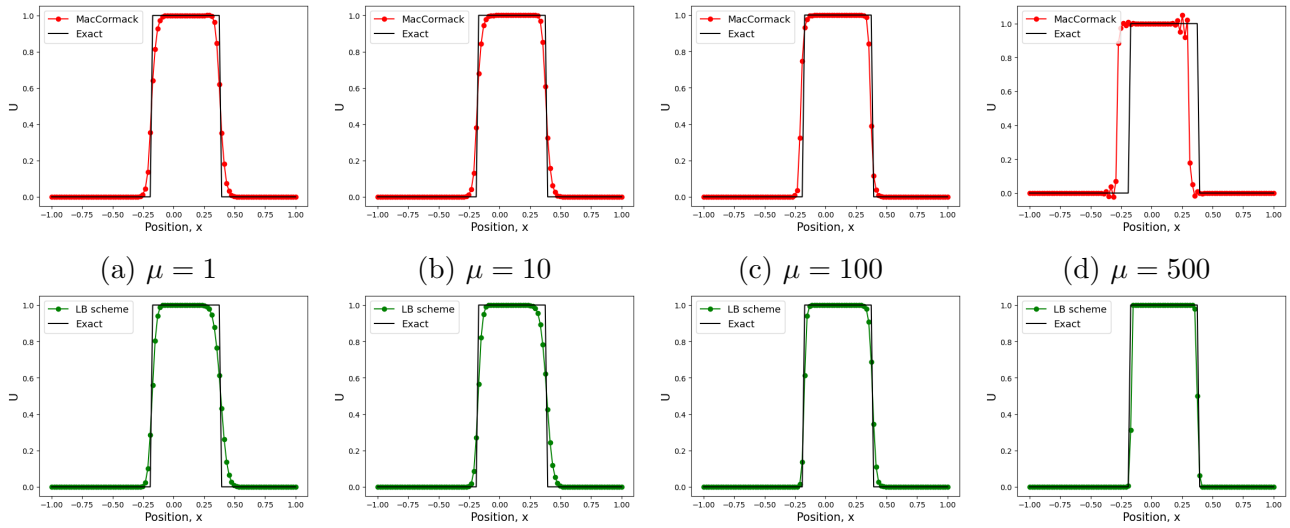
We introduce a variant of LeVeque and Yee [202]'s problem in two dimensions, to understand the effect of ϵ on numerical convection. The domain is $[-1, 1] \times [-1, 1] \subset \mathbb{R}^2$, and is split up into 100×100 grid points. Note that $\Delta x_1 = \Delta x_2 = \Delta x$ is same as the grid spacing used in the previous one dimensional problem. For this problem, $G^1(U) = G^2(U) = U$ and $S(U) = -\mu U(U-1)(U-\frac{1}{2})$. Initial conditions are:

$$U(x_1, x_2, 0) = \begin{cases} 1 & \text{for } x_1^2 + x_2^2 \leq 0.3 \\ 0 & \text{for } x_1^2 + x_2^2 > 0.3 \end{cases}.$$

Boundary conditions are: $U(\pm 1, x_2, t) = 0$ for $x_2 \in [-1, 1]$ and $t \geq 0$; $U(x_1, \pm 1, t) = 0$ for $x_1 \in [-1, 1]$ and $t \geq 0$. An LBM with upwind D2Q5 form for f_q^{eq} and r_q is utilised to obtain the numerical solution. $\lambda = \frac{\Delta x}{\Delta t}$ is chosen such that the sub-characteristic condition in (5.32) is satisfied. This simplifies in this case as

$$\det \begin{pmatrix} \lambda - 1 & -1 \\ -1 & \lambda - 1 \end{pmatrix} \geq 0 \implies \lambda \geq 0 \text{ and } \lambda \geq 2.$$

Further, the time step is chosen as $\Delta t = \frac{\Delta x}{\lambda}$. We also consider $\omega = 1$ for the simulation of this problem and this ensures consistency with the governing equation irrespective of the choice of Δt (as discussed in remark 5.11). A comparison of numerical solutions obtained from MacCormack's method and our LB scheme is shown in figure 5.5 at $T = 0.1$ for different values of μ. It can be seen that, for $\mu = 500$, the MacCormack's method suffers from spurious



(a) $\mu = 1$ (b) $\mu = 10$ (c) $\mu = 100$ (d) $\mu = 500$

(e) $\mu = 1$ (f) $\mu = 10$ (g) $\mu = 100$ (h) $\mu = 500$

Figure 5.5: Cross-sectional plot at $x_2 = 0$. **Top:** Extended MacCormack's method with limiter based on U^n, **Bottom:** Formulated LB scheme for hyperbolic conservation laws with source terms

numerical convection while our LB scheme does not.

In the following, we make an estimation of μ up to which our method will work according to remark 5.14. For this, we use subscripts $D = 1$ and $D = 2$ to compare certain variables from sections 5.7.2.1 and 5.7.2.2 respectively. Since $\lambda_{D=1} = 1$, $\lambda_{D=2} = 2$, and Δx is the same for both one and two dimensional problems, we have $\Delta t_{D=2} = \frac{\Delta t_{D=1}}{2}$. Further, since $\omega_{D=1} = \frac{\Delta t_{D=1}}{\epsilon_{D=1}}$ and $\omega_{D=2} = \frac{\Delta t_{D=2}}{\epsilon_{D=2}}$ are both equal to 1, we have $\epsilon_{D=2} = \frac{\epsilon_{D=1}}{2}$. Thus, since $\epsilon_{D=1}$ in section 5.7.2.1 is $\mathcal{O}(k10^3)$, $\epsilon_{D=2}$ for two dimensional problem is $\mathcal{O}\left(k\frac{10^3}{2}\right)$. Hence, for this problem, our LB scheme is expected to be devoid of spurious numerical convection for μ up to $\mathcal{O}(500)$, and this is validated by the numerical results shown in figure 5.5.

5.7.2.3 Three dimensional discontinuity

Here, we introduce a variant of LeVeque and Yee [202]'s problem in three dimensions. The domain is $[-1, 1] \times [-1, 1] \times [-1, 1] \subset \mathbb{R}^3$, and is split up into $100 \times 100 \times 100$ grid points. Note that $\Delta x_1 = \Delta x_2 = \Delta x_3 = \Delta x$ is same as the grid spacing used in the one dimensional

case. For this problem, $G^1(U) = G^2(U) = G^3(U) = U$ and $S(U) = -\mu U(U-1)(U-\frac{1}{2})$. Initial conditions are:

$$U(x_1, x_2, x_3, 0) = \begin{cases} 1 & \text{for } x_1^2 + x_2^2 + x_3^2 \leq 0.3 \\ 0 & \text{for } x_1^2 + x_2^2 + x_3^2 > 0.3 \end{cases}.$$

Boundary conditions are: $U(\pm 1, x_2, x_3, t) = 0$ for $(x_2, x_3) \in [-1, 1] \times [-1, 1]$ and $t \geq 0$; $U(x_1, \pm 1, x_3, t) = 0$ for $(x_1, x_3) \in [-1, 1] \times [-1, 1]$ and $t \geq 0$; $U(x_1, x_2, \pm 1, t) = 0$ for $(x_1, x_2) \in [-1, 1] \times [-1, 1]$ and $t \geq 0$. An LBM with upwind D3Q7 form for f_q^{eq} and r_q is utilised to obtain the numerical solution. $\lambda = \frac{\Delta x}{\Delta t}$ is chosen such that the sub-characteristic condition in (5.32) is satisfied. This simplifies in this case as

$$\det \begin{pmatrix} \lambda - 1 & -1 & -1 \\ -1 & \lambda - 1 & -1 \\ -1 & -1 & \lambda - 1 \end{pmatrix} \geq 0 \implies \lambda \geq 0 \text{ and } \lambda \geq 3.$$

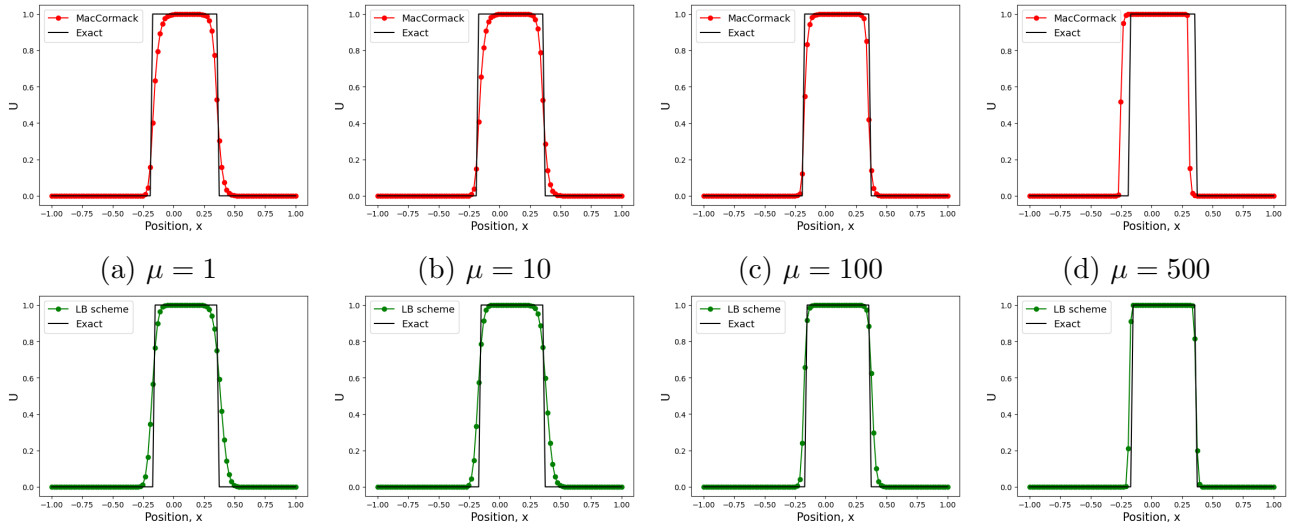
Further, the time step is chosen as $\Delta t = \frac{\Delta x}{\lambda}$. We also consider $\omega = 1$ for the simulation of this problem and this ensures consistency with the governing equation irrespective of the choice of Δt (as discussed in remark 5.11). A comparison of numerical solutions obtained from MacCormack's method and our LB scheme is shown in figure 5.6 at $T = 0.1$ for different values of μ. It can be seen that, for $\mu = 500$, the MacCormack's method suffers from spurious numerical convection while our LB scheme does not.

In the following, we use subscripts $D = 1$ and $D = 3$ to compare certain variables from sections 5.7.2.1 and 5.7.2.3 respectively. Since $\lambda_{D=1} = 1$, $\lambda_{D=3} = 3$, and Δx is the same for both one and three dimensional problems, we have $\Delta t_{D=3} = \frac{\Delta t_{D=1}}{3}$. Further, since $\omega_{D=1} = \frac{\Delta t_{D=1}}{\epsilon_{D=1}}$ and $\omega_{D=3} = \frac{\Delta t_{D=3}}{\epsilon_{D=3}}$ are both equal to 1, we have $\epsilon_{D=3} = \frac{\epsilon_{D=1}}{3}$. Thus, since $\epsilon_{D=1}$ in section 5.7.2.1 is $\mathcal{O}(k10^3)$, $\epsilon_{D=3}$ for three dimensional problem is $\mathcal{O}\left(k\frac{10^3}{3}\right)$. Hence, for this problem, our LB scheme is expected to be devoid of spurious numerical convection for μ up to $\mathcal{O}\left(\frac{10^3}{3}\right)$, and we observe nullification of numerical convection for μ up to 500 in figure 5.6.

5.7.2.4 Non-linear problem with discontinuity

This is a variant of the problem from Embid, Goodman and Majda [239]. The domain is $[0, 1] \subset \mathbb{R}$, and is split up into 100 evenly spaced grid points. The flux function $G^1(U) = \frac{1}{2}U^2$ is non-linear and $S(U) = \mu(6x - 3)U$. Boundary conditions are $U(x_1 = 0, t) = 1$ and $U(x_1 = 1, t) = -0.1$, $\forall t$. For numerical simulation of this steady problem, 500 iterations are utilised with the initialisation

$$U(x_1, 0) = \begin{cases} 1 & \text{for } x_1 \leq 0.1 \\ -1 & \text{for } x_1 > 0.1 \end{cases}.$$



(a) $\mu = 1$ (b) $\mu = 10$ (c) $\mu = 100$ (d) $\mu = 500$

(e) $\mu = 1$ (f) $\mu = 10$ (g) $\mu = 100$ (h) $\mu = 500$

Figure 5.6: Cross-sectional plot at $x_2, x_3 = 0$. **Top:** Extended MacCormack's method with limiter based on U^n, **Bottom:** Formulated LB scheme for hyperbolic conservation laws with source terms

For this problem, λ is chosen based on sub-characteristic condition and ω is fixed as 1. The numerical solutions obtained using LB scheme plotted against the exact solution, for different values of μ, are shown in fig. 5.7. It is seen that the numerical method correctly locates the discontinuities for different values of μ.

5.7.3 D2Q9 model of LBM

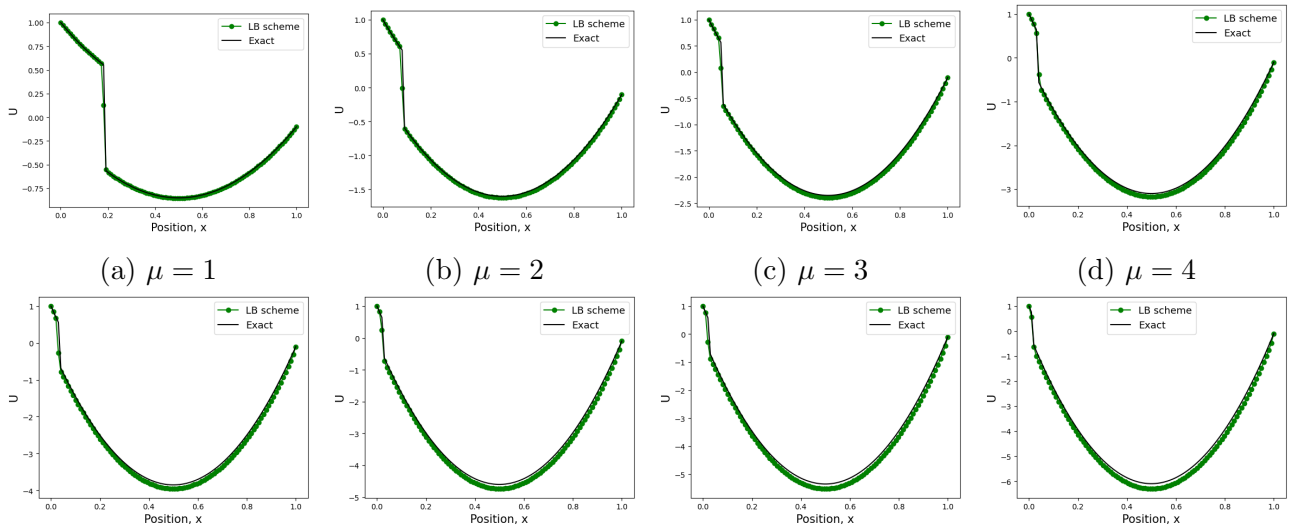
In this section, we show the diagonal upwinding nature of our D2Q9 model of LBM. For this, we consider a standard two-dimensional linear problem from [285]. The domain is $[0,1]\times[0,1] \subset \mathbb{R}^2$, and is split up into 50×50 grid points. Here $\Delta x_1 = \Delta x_2 = \Delta x$. The flux functions are $G^1(U) = aU$, $G^2(U) = bU$ where $a = \cos\theta$, $b = \sin\theta$ and $\theta \in \left(0, \frac{\pi}{2}\right)$. Boundary conditions are:

$$\begin{aligned} U(0, x_2, t) &= 1 \quad \text{for } 0 < x_2 < 1, \\ U(x_1, 0, t) &= 0 \quad \text{for } 0 < x_1 < 1, \end{aligned}, \quad \forall t.$$

Exact solution is:

$$\begin{aligned} U(x_1, x_2, t) &= 1 \quad \text{for } bx_1 - ax_2 < 0, \\ U(x_1, x_2, t) &= 0 \quad \text{for } bx_1 - ax_2 > 0, \end{aligned}, \quad \forall t.$$

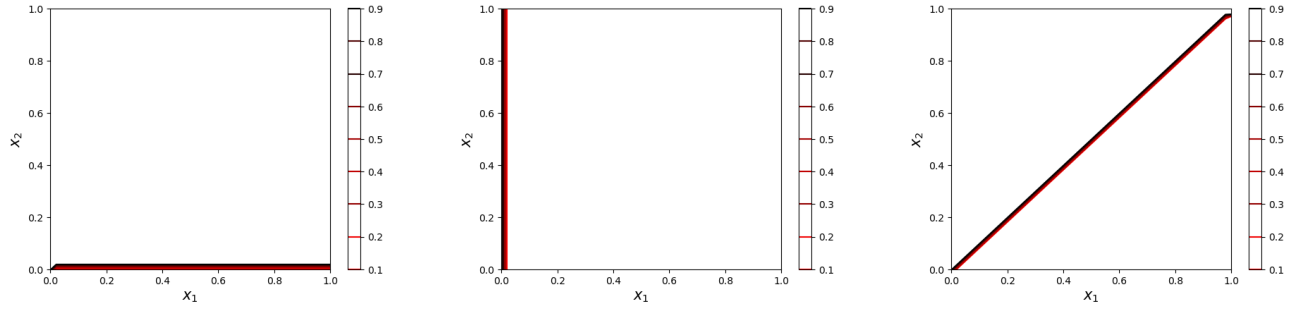
It can be noted that the problem is steady. An LBM with D2Q9 equilibrium functions (5.105) is utilised to obtain the numerical solution. For this problem, we run 1000 iterations of our



(a) $\mu = 1$ (b) $\mu = 2$ (c) $\mu = 3$ (d) $\mu = 4$

(e) $\mu = 5$ (f) $\mu = 6$ (g) $\mu = 7$ (h) $\mu = 8$

Figure 5.7: LB scheme for non-linear problem with different values of μ



(a) $\theta = 0$ (b) $\theta = \frac{\pi}{2}$ (c) $\theta = \frac{\pi}{4}$

Figure 5.8: Discontinuities along coordinate and diagonal-to-coordinate directions captured exactly due to upwinding

LBM before presenting the steady state solution. λ is chosen such that the sub-characteristic condition obtained by imposing positivity of numerical diffusion coefficient in (5.110) is satisfied. Further, we consider $\omega = 1$ for the simulation of this problem and this ensures consistency with the governing equation (as discussed in remark 5.11).

The numerical solutions for $\theta = 0$ and $\theta = \frac{\pi}{2}$ obtained by choosing the fluxes $G^{\gamma} = G^{\zeta} = 0$ (thereby replicating a standard D2Q5 upwind model), are shown in figures 5.8a and 5.8b respectively. The numerical solution for $\theta = \frac{\pi}{4}$ obtained by choosing $G^{\alpha} = G^{\beta} = 0$, is shown in figure 5.8c. It can be seen from these results that, for a specific partition of total flux between coordinate and diagonal-to-coordinate directions, the D2Q9 model captures discontinuities aligned with x_1, x_2 and diagonal directions exactly.

5.8 Summary and conclusions

The following are the major highlights of the chapter.

- An LBE is derived by semi-implicit discretisation of VKE, and its relaxation factor is compared with that of the usual LBE obtained by explicit discretisation of VKE.

- Macroscopic finite difference form of the LBEs is derived, and it is utilised in establishing consistency of LBEs with the hyperbolic system, and in showing the total variation boundedness and the positivity of LBM.

- The usual condition on $\hat{\omega}$ enforced by positivity of numerical diffusion coefficient in Chapman-Enskog expansion is $0 < \hat{\omega} < 2$. On the other hand, the properties such as H-inequality, total variation boundedness and positivity enforce the stronger constraint $0 < \hat{\omega} \leq 1$. By construction, the LBE that we derived by semi-implicit discretisation of VKE naturally satisfies this stronger condition as $\tilde{\omega}$ is in the interval $(0, 1)$ (since $\tilde{\omega} = \frac{\omega}{1+\omega}$ with $\omega = \frac{\Delta t}{\epsilon} > 0$) as explained in remark 5.2. Hence, with semi-implicit discretisation, large values of ω can be used, and $\tilde{\omega}$ will still satisfy the stability properties.

- Smaller numerical diffusion and better order of accuracy are realisable for $1 < \hat{\omega} < 2$ in the case of LBE derived by explicit discretisation of VKE.

- The LBM framework is extended to hyperbolic conservation laws with source terms and the spurious numerical convection due to imbalance between convection and source terms is removed by suitable modelling of r_q. The resulting method not only leads to *well-balancing* but also is effective for source terms of significant stiffness. Thus, our lattice Boltzmann framework can easily overcome the problem of well-balancing which is often encountered while approximating hyperbolic equations with source terms in finite volume framework, as mentioned in remark 5.14.

- A D2Q9 model of our LBM framework allows upwinding along diagonal directions, in addition to the usual upwinding along co-ordinate directions, resulting in better multidimensional behaviour.

Chapter 6

Conclusions

The objective of this thesis is to develop structure preserving numerical methods that yield relevant numerical solutions, for hyperbolic PDE systems and multiscale kinetic equations. Four novel numerical methods that satisfy some of the structures like entropy stability, asymptotic preservation and well-balancing have been presented. Each chapter has addressed specific structure preserving strategies that are of concern to a given system of governing equations.

In chapter 2, an entropy stable scheme for vector-kinetic model of hyperbolic systems is presented. It is shown that this also recovers entropy stability of the hyperbolic system in addition to ensuring entropy stability of the vector-kinetic model. As the vector-BGK model present in literature need not allow for the existence of entropy flux potentials that are crucial in the construction of entropy conserving/stable fluxes, a modification to vector-BGK model is made by enforcing positive-definiteness of Jacobians. Another important aspect behind the development of this scheme is the proof that the entropy variables for vector-kinetic model and hyperbolic system are same. This property facilitates the conservation/stability of both vector-kinetic and macroscopic (hyperbolic PDE system) entropies. The numerical results show that the entropies of both vector-kinetic model and hyperbolic system remain constant (or decay), depicting the entropy conservation (or stability) property. A potential future scope would be to ensure entropy consistency of all vector-kinetic entropies. Apart from satisfying the discrete entropy inequality, this property would ensure that the discrete form of entropy equation as a whole is satisfied. The difficulty in achieving entropy consistency lies in an appropriate choice of numerical flux.

In chapter 3, a high order asymptotic preserving (AP) scheme that is formulated through micro-macro decomposition, for diffusive-scaled linear kinetic equations is presented. The usage of a specific class of time integrator allows the applicability of the framework to problems with even non-well-prepared initial conditions. The high order AP framework based on micro-macro de-

composition is extended to advection-diffusion asymptotics and the formal proof of asymptotic preserving property required more involved definitions when compared to that of the diffusion case. The framework is also extended to inflow boundary problems by considering another type of micro-macro decomposition. The numerical results show that required high order accuracy is attained for all values of the parameter without reducing mesh and time step sizes. A potential future scope would be to develop a high order AP scheme for more involved kinetic equations applicable for plasma flows like Vlasov equations coupled with Poisson or Maxwell equations.
In chapter 4, an asymptotic preserving scheme that satisfies entropy stability for the barotropic Euler system is presented. An appropriate implicit-explicit (IMEX) time discretisation is utilised for the time semi-discrete scheme to achieve the asymptotic preserving property. Three different spatial discretisation strategies are considered and their entropy stability properties have been studied. The numerical results show that the entropy, potential energy and kinetic energy remain nearly constant or decay in different regimes of the parameter Mach number. The scheme also decomposes into a numerical scheme for incompressible barotropic Euler equations in the low Mach number limit. The potential future scope consists of extension to other hyperbolic systems like Euler's system. In this, the choice of appropriate spatial discretisation strategies for the energy fluxes would be crucial for ensuring entropy stability in different regimes of the parameter Mach number.
In chapter 5, two different lattice Boltzmann discretisations of vector-kinetic models of hyperbolic systems are compared based on some important properties likes H-inequality, total variation boundedness, and positivity. This comparison shows that the explicit discretisation can give better accuracy with relaxation factor between 1 and 2, while semi-implicit discretisation is stable irrespective of the value of relaxation factor. Further, the macroscopic finite difference form of the lattice Boltzmann discretisation shows that the lattice Boltzmann method is nothing but a multi-step finite difference scheme. If an upwind equilibrium function is used, then the lattice Boltzmann method is simply a multi-step upwind method. Further, a novel well-balancing strategy that avoids spurious numerical convection due to imbalance between convection and source terms is also presented. The numerical results validate this strategy by capturing discontinuities at correct locations. As there is a bound on the value of stiff parameter for which this strategy is applicable, a potential future research direction would be to develop a strategy to increase (or remove) the bound so that the spurious numerical convection could be avoided for many (or all) stiff equations. Another potential future research direction would be to extend this well-balancing strategy to systems of PDEs, by also imposing the inherent bounds on some special variables (like density and energy that are positive in reactive Euler's system) present in these systems.

Bibliography

[1] Albi, G., Dimarco, G., and Pareschi, L. Implicit-explicit multistep methods for hyperbolic systems with multiscale relaxation. *SIAM J. of Scientific Comput. 42* (2020), 2402–2435. 6, 49

[2] Alouges, F., Ghidaglia, J.-m., and Tajchman, M. On the Interaction of Upwinding and Forcing for Nonlinear Hyperbolic Systems of Conservation Laws. *Technical Report CMLA, 9907* (03 1999). 7

[3] Anandan, M., Boutin, B., and Crouseilles, N. High order asymptotic preserving scheme for linear kinetic equations with diffusive scaling. *arXiv 2305.13393 (math.NA)* (2023). 10

[4] Anandan, M., and Raghurama Rao, S. V. D2Q9 Model of Upwind Lattice Boltzmann Scheme for Hyperbolic Scalar Conservation Laws. In *World Congress in Computational Mechanics and ECCOMAS Congress* (2022), Scipedia SL. 114

[5] Anandan, M., and Raghurama Rao, S. V. Entropy conserving/stable schemes for a vector-kinetic model of hyperbolic systems. *Applied Mathematics and Computation 465* (2024), 128410. 5, 86, 114

[6] Anandan, M., and Raghurama Rao, S. V. On Lattice Boltzmann Methods based on vector-kinetic models for hyperbolic partial differential equations. *arXiv 2401.03952 (math.NA)* (2024). 4, 9

[7] Aregba-Driollet, D., and Natalini, R. Discrete Kinetic Schemes for Multidimensional Systems of Conservation Laws. *SIAM Journal on Numerical Analysis 37*, 6 (2000), 1973–2004. 3, 13, 114, 115, 119

[8] Arumuga Perumal, D., and Dass, A. K. A Review on the development of lattice Boltzmann computation of macro fluid flows and heat transfer. *Alexandria Engineering Journal 54*, 4 (2015), 955–971. 114

[9] Arun, K. R., and Lukáčová-Medviďová, M. A Characteristics Based Genuinely Multidimensional Discrete Kinetic Scheme for the Euler Equations. *Journal of Scientific Computing 55* (2013), 40–64. 4

[10] Arun, K. R., and Samantaray, S. Asymptotic Preserving Low Mach Number Accurate IMEX Finite Volume Schemes for the Isentropic Euler Equations. *Journal of Scientific Computing 82*, 35 (2020). 86

[11] Arun, K.R., Lukáčová–Medviďová, M., Prasad, P., and Raghurama Rao, S.V.. A second order accurate kinetic relaxation scheme for inviscid compressible flows. In *Recent developments in the numerics of nonlinear hyperbolic conservation laws*, R. Ansorge, H. Bijl, A. Meister, and T. Sonar, Eds., vol. 120 of *Notes on Numerical Fluid Mechanics and Multidisciplinary Design*. Springer-Verlag, 2013, pp. 1–24. 4

[12] Ascher, U., Ruuth, S., and Spiteri, R. Implicit-explicit Runge-Kutta methods for time dependent partial differential equations. *Appl. Numer. Math. 25* (1997), 151–167. 6, 49, 54

[13] Asinari, P., and Ohwada, T. Connection between kinetic methods for fluid-dynamic equations and macroscopic finite-difference schemes. *Comput. Math. Appl. 58* (2009), 841–861. 114

[14] Ayuso, B., Carrillo, J. A., and Shu, C.-W. Discontinuous Galerkin methods for the one-dimensional Vlasov-Poisson system. *Kinetic and Related Models 4*, 4 (2011), 955–989. 8

[15] Balsara, D. S. Higher-order accurate space-time schemes for computational astrophysics—part i: finite volume methods. *Living Reviews in Computational Astrophysics 3*, 2 (2017), 1–138. 4

[16] Banda, M., Klar, A., Pareschi, L., and Seaïd, M. Lattice-Boltzmann type relaxation systems and high order relaxation schemes for the incompressible Navier-Stokes equations. *Mathematics of Computation 77*, 262 (2008), 943–965. 114

[17] Bao, W., and Jin, S. The Random Projection Method for Hyperbolic Conservation Laws with Stiff Reaction Terms. *Journal of Computational Physics 163*, 1 (2000), 216–248. 7

[18] Bao, W., and Jin, S. The Random Projection Method for Stiff Detonation Capturing. *SIAM Journal on Scientific Computing 23*, 3 (2001), 1000–1026. 7

[19] Bardos, C., Golse, F., and Levermore, C. D. Fluid dynamic limits of kinetic equations II convergence proofs for the boltzmann equation. *Communications on Pure and Applied Mathematics 46*, 5 (1993), 667–753. 8

[20] Bardos, C., Golse, F., and Levermore, D. Fluid dynamic limits of kinetic equations. I. Formal derivations. *Journal of Statistical Physics 63* (1991), 323–344. 8

[21] Barth, T. Numerical methods for gasdynamic systems on unstructured systems. In *An introduction to recent developments in theory and numerics for conservation laws*, M. O. D. Kröner and C. Rohde, Eds. Springer, Berlin, 1999, pp. 195–285. 4, 13, 86

[22] Bell, J. B., Colella, P., and Glaz, H. M. A second-order projection method for the incompressible navier-stokes equations. *Journal of Computational Physics 85*, 2 (1989), 257–283. 6

[23] Bellotti, T. Truncation errors and modified equations for the lattice Boltzmann method via the corresponding Finite Difference schemes. *ESAIM: Mathematical Modelling and Numerical Analysis 57*, 3 (2023), 1225–1255. 114

[24] Bellotti, T., Graille, B., and Massot, M. Finite Difference formulation of any lattice Boltzmann scheme. *Numerische Mathematik 152* (2022), 1–40. 114

[25] Ben-Artzi, M. The generalized Riemann problem for reactive flows. *Journal of Computational Physics 81*, 1 (1989), 70–101. 7

[26] Bennett, S., Asinari, P., and Dellar, P. J. A lattice Boltzmann model for diffusion of binary gas mixtures that includes diffusion slip. *International Journal for Numerical Methods in Fluids 69*, 1 (2012), 171–189. 114

[27] Bennoune, M., Lemou, M., and Mieussens, L. An asymptotic preserving scheme for the Kac model of the Boltzmann equation in the diffusion limit. *Continuum Mechanics and Thermodynamics 21* (2009), 401–421. 10

[28] Benzoni-Gavage, S., and Serre, D. *Multi-dimensional hyperbolic partial differential equations. First order systems and applications*. Oxford Mathematical Monographs, Oxford University Press, 2007. 2

[29] Bernsdorf, J., Brenner, G., and Durst, F. Numerical analysis of the pressure drop in porous media flow with lattice Boltzmann (BGK) automata. *Computer physics communications 129*, 1-3 (2000), 247–255. 114

[30] Berthelin, F., and Bouchut, F. Relaxation to isentropic gas dynamics for a BGK system with single kinetic entropy. *Methods and applications of analysis 9* (2002), 313–327. 13

[31] Bird, G. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Clarendon Press, 1994. 8

[32] Birdsall, C. K., and Langdon, A. B. *Plasma Physics via Computer Simulation*. CRC press, 1991. 8

[33] Bispen, G., Arun, K. R., Lukáčová-Medvid'ová, M., and Noelle, S. IMEX Large Time Step Finite Volume Methods for Low Froude Number Shallow Water Flows. *Communications in Computational Physics 16*, 2 (2014), 307–347. 86

[34] Bispen, G., Lukáčová-Medvid'ová, M., and Yelash, L. Asymptotic preserving IMEX finite volume schemes for low Mach number Euler equations with gravitation. *Journal of Computational Physics 335* (2017), 222–248. 86

[35] Bobylev, A. V. *Kinetic Equations Volume 1: Boltzmann Equation, Maxwell Models, and Hydrodynamics beyond Navier–Stokes*. De Gruyter, Berlin, Boston, 2020. 8

[36] Boris, J. P., and Book, D. L. Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works. *Journal of Computational Physics 11*, 1 (1973), 38–69. 4, 8

[37] Boscarino, S., Pareschi, L., and Russo, G. Implicit-Explicit Runge–Kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit. *SIAM J. on Scientific Comput. 35*, 1 (2013), A22–A51. 10, 49, 50, 54

[38] Boscarino, S., Pareschi, L., and Russo, G. A unified IMEX Runge-Kutta approach for hyperbolic systems with multiscale relaxation. *SIAM J. Numer. Anal. 55*, 4 (2017), 2085–2109. 49

[39] Boscarino, S., Qiu, J.-M., Russo, G., and Xiong, T. A high order semi-implicit IMEX WENO scheme for the all-Mach isentropic Euler system. *Journal of Computational Physics 392* (2019), 594–618. 86

[40] Boscarino, S., and Russo, G. Flux-Explicit IMEX Runge–Kutta Schemes for Hyperbolic to Parabolic Relaxation Problems. *SIAM J. Numer. Anal. 51*, 1 (2013), 163–190. 49

[41] Boscarino, S., Russo, G., and Scandurra, L. All Mach Number Second Order Semi-implicit Scheme for the Euler Equations of Gas Dynamics. *Journal of Scientific Computing 77* (2018), 850–884. 86

[42] Bösch, F., Dorschner, B., and Karlin, I. Entropic multi-relaxation free-energy lattice Boltzmann model for two-phase flows. *Europhysics Letters 122*, 1 (2018), 14002. 114

[43] Bouchut, F. Construction of BGK Models with a Family of Kinetic Entropies for a Given System of Conservation Laws. *Journal of Statistical Physics 95* (1999), 113–170. 13, 17, 18, 19, 20, 114, 115

[44] Bouchut, F. Entropy satisfying flux vector splittings and kinetic BGK models. *Numerische Mathematik 94* (2003), 623–672. 13, 114

[45] Bouchut, F., Jobic, Y., Natalini, R., Occelli, R., and Pavan, V. Second-order entropy satisfying BGK-FVS schemes for incompressible Navier-Stokes equations. *The SMAI Journal of computational mathematics 4* (2018), 1–56. 4, 13

[46] Bourgat, J.-F., Le Tallec, P., and Tidriri, M. Coupling Boltzmann and Navier–Stokes Equations by Friction. *Journal of Computational Physics 127*, 2 (1996), 227–245. 9

[47] Bourgat, J. F., Tallec, P. L., Perthame, B., and Qiu, Y. Coupling Boltzmann and Euler equations without Overlapping. *Contemporary Mathematics 157* (1994). 9

[48] Broadwell, J. E. Study of rarefied shear flow by the discrete velocity method. *Journal of Fluid Mechanics 19*, 3 (1964), 401–414. 9

[49] Brown, D. L., Cortez, R., and Minion, M. L. Accurate Projection Methods for the Incompressible Navier–Stokes Equations. *Journal of Computational Physics 168*, 2 (2001), 464–499. 6

[50] Cabannes, H., Gatignol, R., and Luo, L. The Discrete Boltzmann Equation (Theory and Applications). Tech. rep., Lecture Notes, University of California at Berkeley, 2003. 9

[51] Carleman, T. Problems Mathematiques dans la Theorie Cinetique des Gaz. *Publ. Sci. Inst. Mittag-Leffler* (1957). 9

[52] Carpenter, M., and Kennedy, C. Additive Runge-Kutta schemes for convection-diffusion-reaction equations. *Appl. Numer. Math. 44* (2003), 139–181. 49, 54

[53] Carrillo, J. A., and Vecil, F. Nonoscillatory Interpolation Methods Applied to Vlasov-Based Models. *SIAM Journal on Scientific Computing 29*, 3 (2007), 1179–1206. 8

[54] Cercignani, C. *Mathematical Methods in Kinetic Theory.* Plenum Press, New York, 1969. 8

[55] Chan, J., Ranocha, H., Rueda-Ramírez, A. M., Gassner, G., and Warburton, T. On the Entropy Projection and the Robustness of High Order Entropy Stable Discontinuous Galerkin Schemes for Under-Resolved Flows. *Frontiers in Physics 10* (2022). 13

[56] Chandrashekar, P. Kinetic Energy Preserving and Entropy Stable Finite Volume Schemes for Compressible Euler and Navier-Stokes Equations. *Communications in Computational Physics 14*, 5 (2013), 1252–1286. 13, 86

[57] Chandrashekar, P., and Klingenberg, C. Entropy Stable Finite Volume Scheme for Ideal Compressible MHD on 2-D Cartesian Meshes. *SIAM Journal on Numerical Analysis 54*, 2 (2016), 1313–1340. 13, 86

[58] Chang, L. C., and D.Kwak. On the method of pseudocompressibility for numerically solving incompressible flows. *AIAA Paper 84-0252* (1984). 6

[59] Chen, S., and Doolen, G. D. Lattice Boltzmann method for fluid flows. *Annual review of fluid mechanics 30*, 1 (1998), 329–364. 114

[60] Cheng, C., and Knorr, G. The integration of the vlasov equation in configuration space. *Journal of Computational Physics 22*, 3 (1976), 330–351. 8

[61] Chizari, H., Singh, V., and Ismail, F. Cell-vertex entropy-stable finite volume methods for the system of Euler equations on unstructured grids. *Computers & Mathematics with Applications 98* (2021), 261–279. 13, 86

[62] Choi, D. H., and Merkle, C. L. Application of time-iterative schemes to incompressible flow. *AIAA Journal 23* (1984), 1518–1524. 6

[63] Chorin, A. J. A numerical method for solving incompressible viscous flow problems. *Journal of Computational Physics 2*, 1 (1967), 12–26. 6

[64] Chorin, A. J. Numerical Solution of the Navier-Stokes Equations. *Mathematics of Computation 22*, 104 (1968), 745–762. 6

[65] Chorin, A. J. On the convergence of discrete approximations to the Navier-Stokes equations. *Mathematics of Computation 23*, 106 (1969), 341–353. 6

[66] Chorin, A. J. Random choice methods with applications to reacting gas flow. *Journal of Computational Physics 25*, 3 (1977), 253–272. 7

[67] Colella, P., Majda, A., and Roytburd, V. Theoretical and Numerical Structure for Reacting Shock Waves. *SIAM Journal on Scientific and Statistical Computing 7*, 4 (1986), 1059–1080. 7

[68] Cordier, F., Degond, P., and Kumbaro, A. An Asymptotic-Preserving all-speed scheme for the Euler and Navier–Stokes equations. *Journal of Computational Physics 231*, 17 (2012), 5685–5704. 86

[69] Coulette, D., Courtès, C., Franck, E., and Navoret, L. Vectorial Kinetic Relaxation Model with Central Velocity. Application to Implicit Relaxations Schemes . *Communications in Computational Physics 27*, 4 (2020), 976–1013. 4, 114, 124

[70] Courant, R., Friedrichs, K., and Lewy, H. On the Partial Difference Equations of Mathematical Physics. *Mathematische Annalen 100* (1928), 32–74. 2

[71] Courant, R., and Hilbert, D. *Methods of Mathematical Physics Vol. 2: Partial Differential Equations*. Wiley-Interscience, New York, 1962. 1

[72] Courant, R., Isaacson, E., and Rees, M. On the solution of nonlinear hyperbolic differential equations by finite differences. *Communications on Pure and Applied Mathematics 5*, 3 (1952), 243–255. 2

[73] Crean, J., Hicken, J. E., Del Rey Fernández, D. C., Zingg, D. W., and Carpenter, M. H. Entropy-stable summation-by-parts discretization of the Euler equations on general curved elements. *Journal of Computational Physics 356* (2018), 410–438. 13, 86

[74] Crestetto, A., Crouseilles, N., Dimarco, G., and Lemou, M. Asymptotically complexity diminishing schemes (ACDS) for kinetic equations in the diffusive scaling. *J. Comput. Phys. 394* (2019), 243–262. 49

[75] Crouseilles, N., and Lemou, M. An asymptotic preserving scheme based on a micro-macro decomposition for collisional Vlasov equations: diffusion and high-field scaling limits. *Kin. Rel. Models 4* (2011), 441–477. 49, 53, 81

[76] Crouseilles, N., Mehrenberger, M., and Sonnendrücker, E. Conservative semi-Lagrangian schemes for Vlasov equations. *Journal of Computational Physics 229*, 6 (2010), 1927–1953. 8

[77] Dafermos, C. *Hyperbolic Conservation Laws in Continuum Physics*. Grundlehren der mathematischen Wissenschaften. Springer Berlin Heidelberg, 2009. 2

[78] Degond, P., and Dimarco, G. Fluid simulations with localized boltzmann upscaling by direct simulation Monte-Carlo. *Journal of Computational Physics 231*, 6 (2012), 2414–2437. 9

[79] Degond, P., Dimarco, G., and Mieussens, L. A moving interface method for dynamic kinetic–fluid coupling. *Journal of Computational Physics 227*, 2 (2007), 1176–1208. 9

[80] Degond, P., Dimarco, G., and Mieussens, L. A multiscale kinetic–fluid solver with dynamic localization of kinetic effects. *Journal of Computational Physics 229*, 13 (2010), 4907–4933. 9

[81] Degond, P., and Tang, M. All Speed Scheme for the Low Mach Number Limit of the Isentropic Euler Equations. *Communications in Computational Physics 10*, 1 (2011), 1–31. 86, 105

[82] Dellacherie, S. Analysis of Godunov type schemes applied to the compressible Euler system at low Mach number. *Journal of Computational Physics 229*, 4 (2010), 978–1016. 85

[83] Dellacherie, S. Construction and Analysis of Lattice Boltzmann Methods Applied to a 1D Convection-Diffusion Equation. *Acta Applicandae Mathematicae 131* (2014), 69–140. 114

[84] Deshmukh, R. L. *Lattice Boltzmann Relaxation Schemes for High Speed Flows*. PhD thesis, 2016. 114

[85] Deshpande, S. M. Kinetic theory based new upwind methods for inviscid compressible flows. *AIAA paper no. AIAA-1986-0275* (1986). 3

[86] Deshpande, S. M. On the Maxwellian distribution, symmetric form, and entropy conservation for the Euler equations. *NASA-TP-2583, L-16036, NAS 1.60:2583* (1986). 13

[87] Dimarco, G., and Loubere, R. Towards an ultra efficient kinetic scheme. Part I: Basics on the BGK equation. *Journal of Computational Physics 255* (2013), 680–698. 8

[88] Dimarco, G., and Loubere, R. Towards an ultra efficient kinetic scheme. Part II: The high order case. *Journal of Computational Physics 255* (2013), 699–719. 8

[89] Dimarco, G., Loubère, R., and Vignal, M.-H. Study of a New Asymptotic Preserving Scheme for the Euler System in the Low Mach Number Limit. *SIAM Journal on Scientific Computing 39*, 5 (2017), A2099–A2128. 86

[90] Dimarco, G., Loubère, R., Michel-Dansac, V., and Vignal, M.-H. Second-order implicit-explicit total variation diminishing schemes for the Euler system in the low Mach regime. *Journal of Computational Physics 372* (2018), 178–201. 86

[91] Dimarco, G., and Pareschi, L. Exponential Runge-Kutta methods for stiff kinetic equations. *SIAM J. of Numer. Anal. 49* (2011), 2057–2077. 10, 49

[92] Dimarco, G., and Pareschi, L. Asymptotic Preserving Implicit-Explicit Runge–Kutta Methods for Nonlinear Kinetic Equations. *SIAM J. on Numer. Anal. 51*, 2 (2013), 1064–1087. 10, 49, 50, 54, 57

[93] Dimarco, G., and Pareschi, L. Implicit explicit linear multistep methods for stiff kinetic equations. *SIAM J. of Numer. Anal. 55* (2017), 664–690. 49

[94] Dimarco, G., Pareschi, L., and Rispoli, V. Implicit-Explicit Runge-Kutta Schemes for the Boltzmann-Poisson System for Semiconductors. *Communications in Computational Physics 15*, 5 (2014), 1291–1319. 10

[95] Dimarco, G., Pareschi, L., and Rispoli, V. Implicit-Explicit Runge-Kutta Schemes for the Boltzmann-Poisson System for Semiconductors. *Commun. Comput. Phys. 15*, 5 (2014), 1291–1319. 49

[96] Dimarco, G., Pareschi, L., and Samaey, G. Asymptotic Preserving Monte Carlo methods for transport equations in the diffusive limit. *SIAM J. Sci. Comput. 40* (2018), 504–528. 49

[97] Ding, Z., Einkemmer, L., and Li, Q. Dynamical low-rank integrator for the linear boltzmann equation: error analysis in the diffusion limit. *SIAM J. on Numer. Anal. 59* (2021). 49

[98] Dolejší, V., and Feistauer, M. *Discontinuous Galerkin Method: Analysis and Applications to Compressible Flow*. Springer, 2015. 4, 7

[99] Donea, J., and A.Huerta. *The Finite Element Method for Flow Problems*. Wiley, 2003. 7

[100] Dubois, F. Equivalent partial differential equations of a lattice Boltzmann scheme. *Computers & Mathematics with Applications 55*, 7 (2008), 1441–1449. Mesoscopic Methods in Engineering and Science. 114

[101] Dubois, F., Février, T., and Graille, B. On the stability of a relative velocity lattice Boltzmann scheme for compressible Navier–Stokes equations. *Compt. Rend. Mécan. 343*, 10–11 (2015), 599–610. 114

[102] Dubois, F., Graille, B., and Rao, S. R. A notion of non-negativity preserving relaxation for a mono-dimensional three velocities scheme with relative velocity. *Journal of Computational Science 47* (2020), 101181. 114

[103] d'Humières, D., and Ginzburg, I. Viscosity independent numerical errors for Lattice Boltzmann models: From recurrence equations to "magic" collision numbers. *Computers & Mathematics with Applications 58*, 5 (2009), 823–840. Mesoscopic Methods in Engineering and Science. 114

[104] Einkemmer, L., Hu, J., and Wang, Y. An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation. *J. of Comput. Phys. 439* (2021), 110353. 49

[105] Ekaterinaris, J. A. High-order accurate, low numerical diffusion methods for aerodynamics. *Progress in Aerospace Sciences 41* (2005), 192–300. 4

[106] Evans, L. C. *Partial differential equations*. American Mathematical Society, Providence, R.I., 2010. 1

[107] Eymann, T. A., and Roe, P. L. Active flux schemes for systems. *AIAA Paper No. AIAA-2011-3840* (2011). 4

[108] Fainsilber, L., Kurlberg, P., and Wennberg, B. Lattice points on circles and discrete velocity models for the Boltzmann equation. *SIAM J. Math. Anal. 37* (2006), 1903–1922. 9

[109] Fattahi, E., Waluga, C., Wohlmuth, B., Rüde, U., Manhart, M., and Helmig, R. Lattice Boltzmann methods in porous media simulations: From laminar to turbulent flow. *Computers & Fluids 140* (2016), 247–259. 114

[110] Fei, L., Du, J., Luo, K. H., Succi, S., Lauricella, M., Montessori, A., and Wang, Q. Modeling realistic multiphase flows using a non-orthogonal multiple-relaxation-time lattice Boltzmann method. *Physics of Fluids 31*, 4 (2019), 042105. 114

[111] Filbet, F. On Deterministic Approximation of the Boltzmann Equation in a Bounded Domain. *Multiscale Modeling & Simulation 10*, 3 (2012), 792–817. 9

[112] Filbet, F., Hu, J., and Jin, S. A numerical scheme for the quantum Boltzmann equation with stiff collision terms. *ESAIM: Mathematical Modelling and Numerical Analysis 46*, 2 (2012), 443–463. 9

[113] Filbet, F., and Jin, S. A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources. *Journal of Computational Physics 229*, 20 (2010), 7625–7648. 10

[114] Filbet, F., and Pareschi, L. A Numerical Method for the Accurate Solution of the Fokker–Planck–Landau Equation in the Nonhomogeneous Case. *Journal of Computational Physics 179*, 1 (2002), 1–26. 9

[115] Filbet, F., Pareschi, L., and Toscani, G. Accurate numerical methods for the collisional motion of (heated) granular flows. *Journal of Computational Physics 202*, 1 (2005), 216–235. 9

[116] Filbet, F., Sonnendrücker, E., and Bertrand, P. Conservative Numerical Schemes for the Vlasov Equation. *Journal of Computational Physics 172*, 1 (2001), 166–187. 8

[117] Fjordholm, U., Mishra, S., and Tadmor, E. *Energy Preserving and Energy Stable Schemes for the Shallow Water Equations*. London Mathematical Society Lecture Note Series. Cambridge University Press, 2009, p. 93–139. 5, 37, 39, 43

[118] Fjordholm, U. S., Mishra, S., and Tadmor, E. Arbitrarily High-order Accurate Entropy Stable Essentially Nonoscillatory Schemes for Systems of Conservation Laws. *SIAM Journal on Numerical Analysis 50*, 2 (2012), 544–573. 17, 36, 37, 97

[119] Friedrichs, K. O. Symmetric positive linear differential equations. *Communications on Pure and Applied Mathematics 11*, 3 (1958), 333–418. 3

[120] Fučík, R., and Straka, R. Equivalent finite difference and partial differential equations for the lattice Boltzmann method. *Computers & Mathematics with Applications 90* (2021), 96–103. 114

[121] Gabetta, E., Pareschi, L., and Toscani, G. Relaxation schemes for nonlinear kinetic equations. *SIAM Journal on Numerical Analysis 34*, 6 (1997), 2168–2194. 10

[122] Gamba, I. M., and Tharkabhushanam, S. H. SHOCK AND BOUNDARY STRUCTURE FORMATION BY SPECTRAL-LAGRANGIAN METHODS FOR THE INHOMOGENEOUS BOLTZMANN TRANSPORT EQUATION. *Journal of Computational Mathematics 28*, 4 (2010), 430–460. 9

[123] Gassner, G. J., Svärd, M., and Hindenlang, F. J. Stability Issues of Entropy-Stable and/or Split-form High-order Schemes. *Journal of Scientific Computing 90*, 79 (2022). 13

[124] Gassner, G. J., Winters, A. R., and Kopriva, D. A. A well balanced and entropy conservative discontinuous Galerkin spectral element method for the shallow water equations. *Applied Mathematics and Computation 272* (2016), 291–308. Recent Advances in Numerical Methods for Hyperbolic Partial Differential Equations. 13, 86

[125] Gassner, G. J., Winters, A. R., and Kopriva, D. A. Split form nodal discontinuous Galerkin schemes with summation-by-parts property for the compressible Euler equations. *Journal of Computational Physics 327* (2016), 39–66. 13, 86

[126] Gatignol, R. Discretisation of the velocity - space in kinetic theory of gases. In *Proceedings of the Fourth International Conference on Numerical Methods in Fluid Dynamics* (Berlin, Heidelberg, 1975), R. D. Richtmyer, Ed., Springer Berlin Heidelberg, pp. 181–186. 9

[127] Geratz, K. J., Klein, R., Munz, C. D., and Roller, S. *Multiple Pressure Variable (MPV) Approach for Low Mach Number Flows Based on Asymptotic Analysis*. Vieweg+Teubner Verlag, Wiesbaden, 1996, pp. 340–354. 6

[128] Ginzburg, I. Truncation Errors, Exact And Heuristic Stability Analysis Of Two-Relaxation-Times Lattice Boltzmann Schemes For Anisotropic Advection-Diffusion Equation. *Communications in Computational Physics 11*, 5 (2012), 1439–1502. 114

[129] Ginzburg, I., Verhaeghe, F., and d'Humiéres, D. Two-Relaxation-Time Lattice Boltzmann Scheme: About Parametrization, Velocity, Pressure and Mixed Boundary Conditions. *Communications in Computational Physics 3*, 2 (2008), 427–478. 114

[130] Glaister, P. Flux difference splitting for the Euler equations in one spatial co-ordinate with area variation. *International Journal for Numerical Methods in Fluids 8*, 1 (1988), 97–119. 7

[131] Godlewski, E., and Raviart, P. A. Numerical Approximation of Hyperbolic Systems of Conservation Laws. In *Applied Mathematical Sciences*. Springer New York, 1996. 2

[132] Godunov, S. K. A difference scheme for numerical solution of discontinuous solution of hydrodynamic equations. *Matematicheskii Sbornik 47* (1959), 271–306. 3

[133] Goldstein, D., Sturtevant, B., and Broadwell, J. Investigations of the Motion of Discrete-Velocity Gases. In *Rarefied Gas Dynamics: Theoretical and Computational Techniques*. American Institute of Aeronautics and Astronautics, Inc., 1989, pp. 100–117. 9

[134] Golse, F., and Saint-Raymond, L. The Navier–Stokes limit of the Boltzmann equation for bounded collision kernels. *Inventiones mathematicae 155* (2004), 81–161. 8

[135] Gosse, L. A well-balanced flux-vector splitting scheme designed for hyperbolic systems of conservation laws with source terms. *Computers & Mathematics with Applications 39*, 9 (2000), 135–159. 7

[136] Graille, B. Approximation of mono-dimensional hyperbolic systems: A lattice Boltzmann scheme as a relaxation method. *Journal of Computational Physics 266* (2014), 74–88. 114

[137] Greenberg, J. M., and Leroux, A. Y. A Well-Balanced Scheme for the Numerical Processing of Source Terms in Hyperbolic Equations. *SIAM Journal on Numerical Analysis 33*, 1 (1996), 1–16. 7

[138] Gresho, P. M., and Chan, S. T. On the theory of semi-implicit projection methods for viscous incompressible flow and its implementation via a finite element method that also introduces a nearly consistent mass matrix. Part 2: Implementation. *International Journal for Numerical Methods in Fluids 11*, 5 (1990), 621–659. 107

[139] Grunau, D., Chen, S., and Eggert, K. A lattice Boltzmann model for multiphase fluid flows. *Physics of Fluids A: Fluid Dynamics 5*, 10 (1993), 2557–2562. 114

[140] Guo, Z., and Shu, C. *Lattice Boltzmann method and its application in engineering*, vol. 3. World Scientific, 2013. 114

[141] Guo, Z., and Zhao, T. Lattice Boltzmann model for incompressible flows through porous media. *Physical review E 66*, 3 (2002), 036304. 114

[142] Guo, Z., and Zhao, T. A lattice Boltzmann model for convection heat transfer in porous media. *Numerical Heat Transfer, Part B 47*, 2 (2005), 157–177. 114

[143] Haack, J., Jin, S., and Liu, J. An All-Speed Asymptotic-Preserving Method for the Isentropic Euler and Navier-Stokes Equations. *Communications in Computational Physics 12*, 4 (2012), 955–980. 86

[144] Han, M., and Ooka, R. *Turbulence Models and LBM-Based Large-Eddy Simulation (LBM-LES)*. Springer Nature Singapore, Singapore, 2023, pp. 101–113. 114

[145] Harlow, F. H., and Welch, J. E. Numerical Calculation of Time-Dependent Viscous Incompressible Flow of Fluid with Free Surface. *The Physics of Fluids 8*, 12 (12 1965), 2182–2189. 6

[146] Harten, A. On the symmetric form of systems of conservation laws with entropy. *Journal of Computational Physics 49*, 1 (1983), 151–164. 4, 12, 15

[147] Harten, A., Engquist, B., Osher, S., and Chakravarthy, S. R. Uniformly high order accurate essentially non-oscillatory schemes, III. *Journal of Computational Physics 71*, 2 (1987), 231–303. 4

[148] Harten, A., Lax, P. D., and Leer, B. V. On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws. *SIAM Review 25*, 1 (1983), 35–61. 3

[149] Hesthaven, J. S., and Warburton, T. *Nodal Discontinuous Galerkin Methods Algorithms, Analysis, and Applications*. Springer, 2008. 7

[150] Himika, T. A., Hassan, S., Hasan, M. F., and Molla, M. M. Lattice Boltzmann Simulation of MHD Rayleigh–Bénard Convection in Porous Media. *Arabian Journal for Science and Engineering 45* (2020), 9527–9547. 114

[151] Hirsch, C., and Hakimi, N. Preconditioning methods for time-marching Navier-Stokes solvers. In *In Solution Techniques for Large-Scale CFD Problems, CMAS*. Jown Wiley & Sons, 1995, pp. 333–353. 6

[152] Holdych, D. J., Noble, D. R., Georgiadis, J. G., and Buckius, R. O. Truncation error analysis of lattice Boltzmann methods. *Journal of Computational Physics 193*, 2 (2004), 595–619. 114

[153] Hosseini, S., Atif, M., Ansumali, S., and Karlin, I. Entropic lattice Boltzmann methods: A review. *Computers & Fluids 259* (2023), 105884. 114

[154] Huang, J.-J. Simplified method for simulation of incompressible viscous flows inspired by the lattice Boltzmann method. *Phys. Rev. E 103* (May 2021), 053311. 114

[155] Hughes, T. J. R. *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis*. Dover Publications Inc., 1987. 7

[156] Hughes, T. J. R., Franca, L. P., and Mallet, M. A new finite element formulation for computational fluid dynamics: I. Symmetric forms of the compressible Euler and Navier—Stokes equations and the second law of thermodynamics. *Applied Mechanics and Engineering 54* (1986), 223–234. 12

[157] Inamuro, T., Yoshino, M., and Ogino, F. Accuracy of the lattice Boltzmann method for small Knudsen number with finite Reynolds number. *Physics of Fluids 9*, 11 (1997), 3535–3542. 114

[158] Ismail, F., and Roe, P. L. Affordable, entropy-consistent Euler flux functions II: Entropy production at shocks. *Journal of Computational Physics 228*, 15 (2009), 5410–5436. 5, 13, 45, 86

[159] Jaisankar, S., and Raghurama Rao, S. V. A central Rankine–Hugoniot solver for hyperbolic conservation laws. *Journal of Computational Physics 228*, 3 (2009), 770–798. 3

[160] Jaiswal, S. An Entropy Stable Scheme for the Non-Linear Boltzmann Equation. *J. Comput. Phys. 463*, C (aug 2022). 13

[161] Jameson, A. Analysis and design of numerical schemes for gas dynamics 1: Artifical diffusion, upwind biasing, limiters and their effect on accuracy and multigrid convergence. *Internaional Journal of Computational Fluid Dynamics*, 3-4 (1995), 171–218. 4

[162] Jameson, A. Analysis and Design of Numerical Schemes for Gas Dynamics 1 Artificial Diffusion, Upwind Biasing, Limiters and Their Effect on Accuracy and Multigrid Convergence. *RIACS Technical Report 94.15, International Journal of Computational Fluid Dynamics 4* (1995), 171–218. 3

[163] Jameson, A., Schmidt, W., and Turkel, E. Numerical solution of the Euler equations by finite volume methods using Runge Kutta time stepping schemes. In *14th Fluid and Plasma Dynamics Conference*. Palo Alto,CA,U.S.A., 1981. 3, 4

[164] Jang, J., Li, F., Qiu, J.-M., and Xiong, T. Analysis of Asymptotic Preserving DG-IMEX schemes for linear kinetic transport equations in a diffusive scaling. *SIAM J. on Numer. Anal. 52*, 4 (2014), 2048–2072. 49

[165] Jang, J., Li, F., Qiu, J.-M., and Xiong, T. High order asymptotic preserving DG-IMEX schemes for discrete-velocity kinetic equations in a diffusive scaling. *J. of Comput. Phys. 281* (2015), 199–224. 49, 50, 54, 62

[166] Jiang, G.-S., and Shu, C.-W. Efficient Implementation of Weighted ENO Schemes. *Journal of Computational Physics 126*, 1 (1996), 202–228. 4

[167] Jin, S. Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. *SIAM J. on Scientific Comput. 21* (1999), 441–454. 49, 86

[168] Jin, S. A steady-state capturing method for hyperbolic systems with geometrical source terms. *ESAIM: Mathematical Modelling and Numerical Analysis 35*, 4 (2001), 631–645. 7

[169] Jin, S. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. *Riv. Mat. Univ. Parma* (2012), 177–216. 10, 49, 86

[170] Jin, S., and Levermore, D. The discrete-ordinate method in diffusive regimes. *Transport Theory Stat. Phys. 22* (1993), 739–791. 59

[171] Jin, S., Pareschi, L., and Toscani, G. Diffusive relaxation schemes for multiscale discrete-velocity kinetic equations. *SIAM Journal on Numerical Analysis 35*, 6 (1998), 2405–2439. 62

[172] Jin, S., Pareschi, L., and Toscani, G. Uniformly accurate diffusive relaxation schemes for multiscale transport equations. *SIAM J. Num. Anal. 38* (2000), 913–936. 10, 49

[173] Jin, S., and Xin, Z. The relaxation schemes for systems of conservation laws in arbitrary space dimensions. *Communications on Pure and Applied Mathematics 48*, 3 (1995), 235–276. 4, 114

[174] Jobic, Y., Kumar, P., Topin, F., and Occelli, R. Determining permeability tensors of porous media: A novel 'vector kinetic' numerical approach. *International Journal of Multiphase Flow 110* (2019), 198–217. 114

[175] Junk, M., Klar, A., and Luo, L.-S. Asymptotic analysis of the lattice Boltzmann equation. *Journal of Computational Physics 210*, 2 (2005), 676–704. 114

[176] Junk, M., and Yang, Z. Asymptotic Analysis of Lattice Boltzmann Boundary Conditions. *Journal of Statistical Physics 121* (2005), 3–35. 114

[177] Kaiser, K., Schütz, J., Schöbel, R., and Noelle, S. A New Stable Splitting for the Isentropic Euler Equations. *Journal of Scientific Computing 70* (2017), 1390–1407. 39, 40

[178] Klainerman, S., and Majda, A. Singular limits of quasilinear hyperbolic systems with large parameters and the incompressible limit of compressible fluids. *Communications on Pure and Applied Mathematics 34*, 4 (1981), 481–524. 85

[179] Klainerman, S., and Majda, A. Compressible and incompressible fluids. *Communications on Pure and Applied Mathematics 35*, 5 (1982), 629–651. 85

[180] Klar, A. Asymptotic-induced domain decomposition methods for kinetic and drift diffusion semiconductor equations. *SIAM J. on Scientific Comput. 19*, 6 (1998), 2032–2050. 49, 81

[181] Klar, A. Asymptotic-induced scheme for nonstationary transport equations in the diffusive limit. *SIAM J. Numer. Anal. 35* (1998), 1073–1094. 49

[182] Klar, A., and Schmeiser, C. Numerical passage from radiative heat transfer to nonlinear diffusion models. *Math. Models Methods Appl. Sci. 11* (2001), 749–767. 49

[183] Klein, R. Semi-implicit extension of a godunov-type scheme based on low mach number asymptotics I: One-dimensional flow. *Journal of Computational Physics 121*, 2 (1995), 213–237. 6

[184] Koda, Y., and Lien, F.-S. The Lattice Boltzmann Method Implemented on the GPU to Simulate the Turbulent Flow Over a Square Cylinder Confined in a Channel. *Flow, Turbulence and Combustion 94* (2015), 495–512. 114

[185] Kolobov, V., Arslanbekov, R., Aristov, V., Frolova, A., and Zabelok, S. Unified solver for rarefied and continuum flows with adaptive mesh and algorithm refinement. *Journal of Computational Physics 223*, 2 (2007), 589–608. 9

[186] Kurganov, A. An accurate deterministic projection method for for hyperbolic systems with stiff source term. In *Hyperbolic problems: Theory, Numerics, Applications. Proceedings of the* 9^{th} *international conference held in CalTech, Mar 2022* (2003), Springer, pp. 665–674. 7

[187] Kurganov, A., and Tadmor, E. New High-Resolution Central Schemes for Nonlinear Conservation Laws and Convection–Diffusion Equations. *Journal of Computational Physics 160*, 1 (2000), 241–282. 3

[188] Kuzmin, D., and Jämäläinen, J. *Finite Element Methods for Computational Fluid Dynamics: A Practical Guide*. SIAM, 2014. 7

[189] Lafitte, P., and Samaey, G. Asymptotic-preserving projective integration schemes for kinetic equations in the diffusion limit. *SIAM J. Sci. Comp. 34* (2012), 579–602. 49

[190] Lallemand, P., and Luo, L.-S. Theory of the lattice Boltzmann method: Dispersion, dissipation, isotropy, Galilean invariance, and stability. *Phys. Rev. E 61* (Jun 2000), 6546–6562. 114

[191] Laney, C. B. *Computational Gasdynamics*. Cambridge University Press, 1998. 4

[192] Larsen, E., and Keller, J. Asymptotic solution of neutron transport problems for small mean free paths. *J. Math. Phys. 15* (1974), 75–81. 49

[193] Lax, P., and Wendroff, B. Systems of conservation laws. *Communications on Pure and Applied Mathematics 13*, 2 (1960), 217–237. 2, 3

[194] Lax, P. D. *Selected Papers Volume I*. Springer, 2005. 3

[195] Lax, P. D. *Hyperbolic Partial Differential Equations*. Courant Lecture Notes Vol. 14. American Mathematical Society, Courant Institute of Mathematical Sciences, 2006. 2

[196] LeFloch, P. G., Mercier, J. M., and Rohde, C. Fully Discrete, Entropy Conservative Schemes of Arbitrary Order. *SIAM Journal on Numerical Analysis 40*, 5 (2002), 1968–1992. 17, 26

[197] LeFloch, P. G., and Ranocha, H. Kinetic Functions for Nonclassical Shocks, Entropy Stability, and Discrete Summation by Parts. *Journal of Scientific Computing 87* (2020). 13

[198] Lemou, M. Relaxed micro-macro schemes for kinetic equations. *Comptes Rendus Mathematique 348* (2010), 455–460. 49, 52, 53

[199] Lemou, M., and Méhats, F. Micro-macro schemes for kinetic equations including boundary layers. *SIAM J. on Scientific Comput. 34*, 6 (2012), B734–B760. 50, 69, 70, 71, 73, 79, 81

[200] Lemou, M., and Mieussens, L. A New Asymptotic Preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. *SIAM J. on Scientific Comput. 31*, 1 (2008), 334–368. 10, 49, 52, 59, 60, 69, 81

[201] Leray, J. *Hyperbolic differential equations*. Institute for Advanced Study, 1953. 2

[202] LeVeque, R. J., and Yee, H. C. A Study of Numerical Methods for Hyperbolic Conservation laws with Stiff Source terms. *J. Comput. Phys. 86* (1990), 187–210. xiii, 7, 144, 145, 146

[203] Lewis, E. E., and Miller, W. F. *Computational methods of neutron transport*. United States, 1 1984. 8

[204] Li, Q., Du, D., Fei, L., and Luo, K. H. Three-dimensional non-orthogonal MRT pseudopotential lattice Boltzmann model for multiphase flows. *Computers & Fluids 186* (2019), 128–140. 114

[205] Liboff, R. L. *Introduction to the Theory of Kinetic Equations*. New York: John Wiley and Sons, Inc., 1969. 8

[206] Lions, P. L., Perthame, B., and Tadmor, E. A Kinetic Formulation of Multidimensional Scalar Conservation Laws and Related Equations. *Journal of the American Mathematical Society 7*, 1 (1994), 169–191. 13

[207] Lions, P. L., Perthame, B., and Tadmor, E. Kinetic formulation of the isentropic gas dynamics and p-systems. *Communications in Mathematical Physics 163* (1994), 415–431. 13

[208] Liou, M.-S., and Steffen, C. J. A New Flux Splitting Scheme. *Journal of Computational Physics 107*, 1 (1993), 23–39. 3

[209] Liu, X.-D., and Tadmor, E. Third order nonoscillatory central scheme for hyperbolic conservation laws. *Numerische Mathematik 79*, 3 (1998), 397–425. 3

[210] MacCormack, R. W. The Effect of Viscosity in Hypervelocity Impact Cratering. *Journal of Spacecraft and Rockets 40* (1969), 757–763. 3

[211] Mandal, J., and Deshpande, S. Kinetic flux vector splitting for euler equations. *Computers & Fluids 23*, 2 (1994), 447–478. 3, 120

[212] Manzanero, J., Rubio, G., Kopriva, D. A., Ferrer, E., and Valero, E. An entropy–stable discontinuous Galerkin approximation for the incompressible Navier–Stokes equations with variable density and artificial compressibility. *Journal of Computational Physics 408* (2020), 109241. 13

[213] Markowich, P. A., Ringhofer, C. A., and Schmeiser, C. *Semiconductor Equations*. Springer Vienna, 1990. 8

[214] Martínez, D. O., Chen, S., and Matthaeus, W. H. Lattice Boltzmann magnetohydrodynamics. *Physics of Plasmas 1*, 6 (1994), 1850–1867. 114

[215] McHugh, P. R., and Ramshaw, J. D. Damped artificial compressibility iteration scheme for implicit calculations of unsteady incompressible flow. *International Journal for Numerical Methods in Fluids 21*, 2 (1995), 141–153. 6

[216] Mieussens, L. Discrete-Velocity Models and Numerical Schemes for the Boltzmann-BGK Equation in Plane and Axisymmetric Geometries. *Journal of Computational Physics 162*, 2 (2000), 429–466. 9

[217] Miniati, F., and Colella, P. A modified higher order Godunov's scheme for stiff source conservative hydrodynamics. *Journal of Computational Physics 224*, 2 (2007), 519–538. 7

[218] Mischler, S. Convergence of Discrete-Velocity Schemes for the Boltzmann Equation. *Archive for Rational Mechanics and Analysis 140* (1997), 53–77. 9

[219] Mishra, S. C., and Roy, H. K. Solving transient conduction and radiation heat transfer problems using the lattice Boltzmann method and the finite volume method. *Journal of Computational Physics 223*, 1 (Apr. 2007), 89–107. 114

[220] Mohamad, A. *Lattice boltzmann method*, vol. 70. Springer, 2011. 114

[221] Munz, C.-D., Roller, S., Klein, R., and Geratz, K. The extension of incompressible flow solvers to the weakly compressible regime. *Computers & Fluids 32*, 2 (2003), 173–196. 6

[222] Naldi, G., and Pareschi, L. Numerical schemes for kinetic equations in diffusive regimes. *Appl. Math. Lett. 11* (1998), 29–35. 49

[223] Nanbu, K. Direct Simulation Scheme Derived from the Boltzmann Equation. I. Monocomponent Gases. *Journal of the Physical Society of Japan 49*, 5 (1980), 2042–2049. 8

[224] Nandagiri, V. R., and Raghurama Rao, S. V. A Boltzmann scheme with physically relevant discrete velocities for Euler equations. *International Journal of Advances in Engineering Sciences and Applied Mathematics 13*, 2-3 (2021), 305–328. 3

[225] Naseri Nia, S., Rabiei, F., Rashidi, M., and Kwang, T. Lattice Boltzmann simulation of natural convection heat transfer of a nanofluid in a L-shape enclosure with a baffle. *Results in Physics 19* (2020), 103413. 114

[226] Natalini, R. A Discrete Kinetic Approximation of Entropy Solutions to Multidimensional Scalar Conservation Laws. *Journal of Differential Equations 148*, 2 (1998), 292–317. 4, 13

[227] Nessyahu, H., and Tadmor, E. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics 87*, 2 (1990), 408–463. 3

[228] Noelle, S., Bispen, G., Arun, K. R., Lukáčová-Medviďová, M., and Munz, C.-D. A Weakly Asymptotic Preserving Low Mach Number Scheme for the Euler Equations of Gas Dynamics. *SIAM Journal on Scientific Computing 36*, 6 (2014), B989–B1024. 86

[229] Osher, S., and Solomon, F. Upwind difference schemes for hyperbolic systems of conservation laws. *Mathematics of Computation 38* (1982), 339–374. 3

[230] Panferov, V. A., and Heintz, A. G. A new consistent discrete-velocity model for the Boltzmann equation. *Mathematical Methods in the Applied Sciences 25*, 7 (2002), 571–593. 9

[231] Pappou, T., and Tsangaris, S. Development of an artificial compressibility methodology using flux vector splitting. *International Journal for Numerical Methods in Fluids 25*, 5 (1997), 523–545. 6

[232] Pareschi, L., and Perthame, B. A Fourier spectral method for homogeneous boltzmann equations. *Transport Theory and Statistical Physics 25*, 3-5 (1996), 369–382. 9

[233] Pareschi, L., and Russo, G. Numerical solution of the Boltzmann equation I: Spectrally accurate approximation of the collision operator. *SIAM journal on numerical analysis 37*, 4 (2000), 1217–1245. 9

[234] Pareschi, L., and Russo, G. Implicit-Explicit Runge-Kutta methods and applications to hyperbolic systems with relaxation. *J. of Scientific Comput. 25* (2005), 129–155. 6, 49

[235] Pareschi, L., and Russo, G. Efficient asymptotic preserving deterministic methods for the Boltzmann equation. Tech. rep., FERRARA UNIV (ITALY), 2011. 10

[236] Parisot, M. Entropy-satisfying scheme for a hierarchy of dispersive reduced models of free surface flow. *International Journal for Numerical Methods in Fluids 91*, 10 (2019), 509–531. 13, 86

[237] Park, J. H., and Munz, C.-D. Multiple pressure variables methods for fluid flow at all Mach numbers. *International Journal for Numerical Methods in Fluids 49*, 8 (2005), 905–931. 6

[238] Pattison, M., Premnath, K., Morley, N., and Abdou, M. Progress in lattice Boltzmann methods for magnetohydrodynamic flows relevant to fusion applications. *Fusion Engineering and Design 83*, 4 (2008), 557–572. 114

[239] P.Embid, J.Goodman, and A.Majda. Multiple Steady States for 1-D Transonic Flow. *SIAM Journal on Scientific and Statistical Computation 5*, 1 (1984), 21–41. 147

[240] Peng, Z., Cheng, Y., Qiu, J.-M., and Li, F. Stability-enhanced AP IMEX-LDG schemes for linear kinetic transport equations under a diffusive scaling. *J. of Comput. Phys. 415* (2020), 109485. 49, 50

[241] Peng, Z., Cheng, Y., Qiu, J.-M., and Li, F. Stability-enhanced AP IMEX1-LDG method: Energy-based stability and rigorous AP property. *SIAM J. of Numer. Anal. 59*, 2 (2021), 925–954. 49, 50, 54, 74

[242] Peng, Z., and Li, F. Asymptotic Preserving IMEX-DG-S schemes for linear kinetic transport equations based on Schur complement. *SIAM J. of Scientific Comput. 43*, 2 (2021), A1194–A1220. 49, 50, 56, 69, 74

[243] Perthame, B. Boltzmann Type Schemes for Gas Dynamics and the Entropy Property. *SIAM Journal on Numerical Analysis 27*, 6 (1990), 1405–1421. 3

[244] Perthame, B., and Simeoni, C. A kinetic scheme for the Saint-Venant system¶with a source term. *CALCOLO 38* (2001), 201–231. 7

[245] Perthame, B., and Tadmor, E. A kinetic equation with kinetic entropy functions for scalar conservation laws. *Communications in Mathematical Physics 136* (1991), 501–517. 13

[246] Pradipto, and Purqon, A. Accuracy and Numerical Stabilty Analysis of Lattice Boltzmann Method with Multiple Relaxation Time for Incompressible Flows. *Journal of Physics: Conference Series 877*, 1 (jul 2017), 012035. 114

[247] Premnath, K. N., and Abraham, J. Three-dimensional multi-relaxation time (MRT) lattice-Boltzmann models for multiphase flow. *Journal of Computational Physics 224*, 2 (2007), 539–559. 114

[248] Prendergast, K. H., and Xu, K. Numerical Hydrodynamics from Gas-Kinetic Theory. *Journal of Computational Physics 109*, 1 (1993), 53–66. 3

[249] Pullin, D. Direct simulation methods for compressible inviscid ideal-gas flow. *Journal of Computational Physics 34*, 2 (1980), 231–244. 3

[250] Puppo, G., and Semplice, M. Entropy and the numerical integration of conservation laws. *Physics Procedia 00* (2011), 1–28. 13

[251] Puppo, G., and Semplice, M. Numerical Entropy and Adaptivity for Finite Volume Schemes. *Communications in Computational Physics 10*, 5 (2011), 1132–1160. 13, 86

[252] Qiu, J.-M., and Shu, C.-W. Positivity preserving semi-Lagrangian discontinuous Galerkin formulation: Theoretical analysis and application to the Vlasov–Poisson system. *Journal of Computational Physics 230*, 23 (2011), 8386–8409. 8

[253] Raghurama Rao, S. V., and Balakrishna, K. An Accurate Shock Capturing Algorithm with a Relaxation System for Hyperbolic Conservation Laws. *AIAA paper no. AIAA-2003-4115* (2003). 4

[254] Raghurama Rao, S. V., and Deshmukh, Rohan Kotnala, S. A Lattice Boltzmann Relaxation Scheme for Inviscid Compressible Flows. arXiv preprint, 2015. arXiv:1504.04089 [math.NA]. 114

[255] Raghurama Rao, S. V., and Deshpande, S. M. Peculiar Velocity based Upwind Method for Inviscid Compressible Flows. *Computational Fluid Dynamics Journal 3*, 4 (1995), 415–432. 3

[256] Raghurama Rao, S. V., and Subba Rao, M. V. A simple multi-dimensional relaxation scheme for hyperbolic conservation laws. *AIAA paper no. AIAA-2003-3535* (2003). 4

[257] Ranocha, H., Dalcin, L., and Parsani, M. Fully discrete explicit locally entropy-stable schemes for the compressible Euler and Navier–Stokes equations. *Computers & Mathematics with Applications 80*, 5 (2020), 1343–1359. 13

[258] Ranocha, H., Sayyari, M., Dalcin, L., Parsani, M., and Ketcheson, D. I. Relaxation Runge–Kutta Methods: Fully Discrete Explicit Entropy-Stable Schemes for the Compressible Euler and Navier–Stokes Equations. *SIAM Journal on Scientific Computing 42*, 2 (2020), A612–A638. 31

[259] Ray, D., and Chandrashekar, P. Entropy stable schemes for compressible Euler equations. *International Journal of Numerical Analysis and Modeling 4*, 4 (2013), 335 – 352. 13, 86

[260] Ray, D., Chandrashekar, P., Fjordholm, U. S., and Mishra, S. Entropy Stable Scheme on Two-Dimensional Unstructured Grids for Euler Equations. *Communications in Computational Physics 19*, 5 (2016), 1111–1140. 13, 86

[261] Reitz, R. D. One-dimensional compressible gas dynamics calculations using the Boltzmann equation. *Journal of Computational Physics 42*, 1 (1981), 108–123. 3

[262] Rheinländer, M. K. *Analysis of lattice-Boltzmann methods: asymptotic and numeric investigation of a singularly perturbed system.* PhD thesis, 2007. 114

[263] Ricchiuto, M., and Torlo, D. Analytical travelling vortex solutions of hyperbolic equations for validating very high order schemes. *ArXiv abs/2109.10183* (2021). 42

[264] Richtmyer, R. D., and Morton, K. W. *Difference Methods for Initial Value Problems, second Edition.* Wiley-Interscience, New York, 1967. 3

[265] Rieper, F. A low-Mach number fix for Roe's approximate Riemann solver. *Journal of Computational Physics 230*, 13 (2011), 5263–5287. 107

[266] Roe, P. Approximate Riemann solvers, parameter vectors, and difference schemes. *Journal of Computational Physics 43*, 2 (1981), 357–372. 3, 120

[267] Roe, P. L. Upwind differencing schemes for hyperbolic conservation laws with source terms. In *Nonlinear Hyperbolic Problems* (Berlin, Heidelberg, 1987), C. Carasso, D. Serre, and P.-A. Raviart, Eds., Springer Berlin Heidelberg, pp. 41–51. 7

[268] Roe, P. L. *A Brief Introduction to High-Resolution Schemes.* Springer, 1997, pp. 9–28. 4

[269] Rogier, F., and Schneider, J. A direct method for solving the Boltzmann equation. *Transport Theory and Statistical Physics 23*, 1-3 (1994), 313–338. 9

[270] Ruhi, A., Raghurama Rao, S. V., and Muddu, S. A lattice Boltzmann relaxation scheme for incompressible fluid flows. *International Journal of Advances in Engineering Sciences and Applied Mathematics 14*, 1 (2022), 34–47. 114

[271] Rusanov, V. V. *Calculation of interaction of non-steady shock waves with obstacles.* National Research Council of Canada, Ottawa, 1962. 3

[272] Russo, G., and Filbet, F. Semilagrangian schemes applied to moving boundary problems for the BGK model of rarefied gas dynamics. *Kinetic and Related Models 2* (2009), 231–250. 8

[273] RUSSO, G., SANTAGATI, P., and YUN, S.-B. CONVERGENCE OF A SEMI-LAGRANGIAN SCHEME FOR THE BGK MODEL OF THE BOLTZMANN EQUATION. *SIAM Journal on Numerical Analysis 50*, 3 (2012), 1111–1135. 8

[274] Saint-Raymond, L. Hydrodynamic Limits of the Boltzmann Equation. In *Lecture Notes in Mathematics*. Springer Berlin, Heidelberg, 2009. 8

[275] Sanders, R. H., and Prendergast, K. H. The Possible Relation of the 3-KILOPARSEC Arm to Explosions in the Galactic Nucleus. *Astrophysical journal 188* (1974), 489–500. 3

[276] Schneider, J. Direct coupling of fluid and kinetic equations. *Transport Theory and Statistical Physics 25*, 6 (1996), 681–698. 9

[277] Schochet, S. Fast Singular Limits of Hyperbolic PDEs. *Journal of Differential Equations 114*, 2 (1994), 476–512. 85

[278] Shashi Shekar Roy, and Raghurama Rao, S. V. A kinetic scheme with variable velocities and relative entropy. *Computers & Fluids 265* (2023), 106016. 3

[279] Sheikholeslami, M., Gorji-Bandpy, M., and Vajravelu, K. Lattice Boltzmann simulation of magnetohydrodynamic natural convection heat transfer of Al2O3–water nanofluid in a horizontal cylindrical enclosure with an inner triangular cylinder. *International Journal of Heat and Mass Transfer 80* (2015), 16–25. 114

[280] Shrinath, K. S., Maruthi, N. H., Raghurama Rao, S. V., and Vasudeva Rao, V. A Kinetic Flux Difference Splitting method for compressible flows. *Computers & Fluids 250* (2023), 105702. 3, 13, 120

[281] Shu, C.-W., and Osher, S. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics 77*, 2 (1988), 439–471. 30

[282] Simonis, S., Frank, M., and Krause, M. J. On relaxation systems and their relation to discrete velocity Boltzmann models for scalar advection–diffusion equations. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 378*, 2175 (2020), 20190400. 114

[283] Simonis, S., Frank, M., and Krause, M. J. Constructing relaxation systems for lattice Boltzmann methods. *Applied Mathematics Letters 137* (2023), 108484. 114

[284] Sonnendrücker, E., Roche, J., Bertrand, P., and Ghizzo, A. The Semi-Lagrangian Method for the Numerical Resolution of the Vlasov Equation. *Journal of Computational Physics 149*, 2 (1999), 201–220. 8

[285] Spekreijse, S. Multigrid solution of monotone second-order discretizations of hyperbolic conservation laws. *Mathematics of Computation 49*, 179 (1987), 135–155. 148

[286] Steger, J. L., and Warming, R. Flux vector splitting of the inviscid gasdynamic equations with application to finite-difference methods. *Journal of Computational Physics 40*, 2 (1981), 263–293. 3, 120

[287] Strang, G., and Fix, J. *An Analysis of the Finite Element Method.* Prentice-Hall, Englewood Cliffs, 1973. 7

[288] Strauss, W. A. *Partial Differential Equations: An Introduction, second edition.* Wiley, 2008. 1

[289] Strikwerda, J. C. Finite Difference Methods for the Stokes and Navier–Stokes Equations. *SIAM Journal on Scientific and Statistical Computing 5*, 1 (1984), 56–68. 6

[290] Succi, S. *The lattice Boltzmann equation: for fluid dynamics and beyond.* Oxford university press, 2001. 114

[291] Suga, S. An Accurate Multi-level Finite Difference Scheme for 1D Diffusion Equations Derived from the Lattice Boltzmann Method. *Journal of Statistical Physics 140* (2010), 494–503. 114

[292] Svärd, M., and Mishra, S. Implicit–explicit schemes for flow equations with stiff source terms. *Journal of Computational and Applied Mathematics 235*, 6 (2011), 1564–1577. 7

[293] Tadmor, E. The Numerical Viscosity of Entropy Stable Schemes for Systems of Conservation Laws. I. *Mathematics of Computation 49*, 179 (1987), 91–103. 12, 13, 15, 16, 17, 86, 96

[294] Tadmor, E. Entropy stability theory for difference approximations of nonlinear conservation laws and related time-dependent problems. *Acta Numerica 12* (2003), 451–512. 5, 12, 13, 15, 16, 17, 25, 86, 96

[295] Tadmor, E. Entropy stable schemes. In *Handbook of Numerical Methods for Hyperbolic Problems*, R. Abgrall and C.-W. Shu, Eds., vol. 17 of *Chapter 18 in Handbook of Numerical Analysis*. Elsevier, 2016, pp. 467–493. 15, 86, 96

[296] Tamamidis, P., Zhang, G., and Assanis, D. N. Comparison of Pressure-Based and Artificial Compressibility Methods for Solving 3D Steady Incompressible Viscous Flows. *Journal of Computational Physics 124*, 1 (1996), 1–13. 6

[297] Tang, M. Second order all speed method for the isentropic Euler equations. *Kinetic and Related Models 5*, 1 (2012), 155–184. 86

[298] Temam, R. *Navier-Stokes Equations*. North-Holland, Amsterdam, 1984. 6

[299] ter Haar, D. The Transport Equation in the Case of Coulomb Interactions. In *Collected Papers of L. D. Landau* (1969), pp. 163–170. 8

[300] Thorsten Pöschel, N. B. Granular Gas Dynamics. In *Lecture Notes in physics*. Springer Berlin, Heidelberg, 2003. 8

[301] Tiwari, S. Coupling of the Boltzmann and Euler Equations with Automatic Domain Decomposition. *Journal of Computational Physics 144*, 2 (1998), 710–726. 9

[302] Tiwari, S., and Klar, A. An adaptive domain decomposition procedure for Boltzmann and Euler equations. *Journal of Computational and Applied Mathematics 90*, 2 (1998), 223–237. 9

[303] Tong, Z., Li, M., Xie, T., and Gu, Z. Lattice Boltzmann Method for Conduction and Radiation Heat Transfer in Composite Materials. *Journal of Thermal Science 31* (2022), 777–789. 114

[304] Toro, E. *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction, Third Edition*. Springer, 2009. 4

[305] Toro, E., and Vázquez-Cendón, M. Flux splitting schemes for the Euler equations. *Computers & Fluids 70* (2012), 1–12. 3

[306] Toro, E. F., Spruce, M., and Speares, W. Restoration of the contact surface in the HLL-Riemann solver. *Shock Waves 4* (1994), 25–34. 3

[307] Turkel, E. Review of preconditioning methods for fluid dynamics. *Applied Numerical Mathematics 12*, 1 (1993), 257–284. SPECIAL ISSUE. 6

[308] Turkel, E. Preconditioning Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics 31* (Jan. 1999), 385–416. 6

[309] Uehling, E. A., and Uhlenbeck, G. E. Transport Phenomena in Einstein-Bose and Fermi-Dirac Gases. I. *Phys. Rev. 43* (Apr 1933), 552–561. 8

[310] Van den Akker, H. E. Lattice Boltzmann simulations for multi-scale chemical engineering. *Current Opinion in Chemical Engineering 21* (2018), 67–75. 114

[311] van Leer, B. Towards the ultimate conservative difference scheme. II. Monotonicity and conservation combined in a second-order scheme. *Journal of Computational Physics 14*, 4 (1974), 361–370. 4

[312] van Leer, B. Towards the ultimate conservative difference scheme. IV. A New Approach to Numerical Convection. *Journal of Computational Physics 23*, 3 (1977), 276–299. 4

[313] van Leer, B. Flux-vector splitting for the euler equations. In *Eighth International Conference on Numerical Methods in Fluid Dynamics* (Berlin, Heidelberg, 1982), E. Krause, Ed., Springer Berlin Heidelberg, pp. 507–512. 3, 120

[314] van Leer, B. On the Relation Between the Upwind-Differencing Schemes of Godunov, Engquist–Osher and Roe. *SIAM Journal on Scientific and Statistical Computing 5*, 1 (1984), 1–20. 7

[315] van Leer, B. *ICASE and the History of High-Resolution Schemes*. Springer, 1997, pp. 1–7. 4

[316] van Leer, B., Lee, W. T., and Roe., P. L. Characteristic time-stepping or local preconditioning of the Euler equations. *AIAA Paper 91-1552* (1991). 6

[317] Villani, C. A Review of Mathematical Topics in Collisional Kinetic Theory. In *Handbook of Mathematical Fluid Dynamics, Vol. I*. North-Holland, 2002, pp. 71–305. 8

[318] VonNeumann, J., and Richtmyer, R. D. A Method for the Numerical Calculation of Hydrodynamic Shocks. *Journal of Applied Physics 21*, 3 (1950), 232–237. 2

[319] Wang, Z. Spectral (finite) volume method for conservation laws on unstructured grids - basic formulation. *Journal of Computational Physics 178* (2002), 210–251. 4

[320] Wang, Z. A perspective on high-order methods in computational fluid dynamics. *SCIENCE CHINA: Physics, Mechanics & Astronomy 59*, 1 (2017), 614701:1–6. 4

[321] Wang, Z. J., Fidkowski, K., Abgrall, R., Bassi, F., Caraeni, D., Cary, A., Deconinck, H., Hartmann, R., Hillewaert, K., Huynh, H. T., Kroll, N., May, G., Persson, P.-O., van Leer, B., and Visbal, M. High-order CFD methods: current status and perspective. *International Journal for Numerical Methods in Fluids 72* (2013), 811–845. 4

[322] Weinan, E., and Liu, J. Projection method iii: Spatial discretization on the staggered grid. *Mathematics of Computation 71*, 237 (2002), 27–47. 6

[323] Wintermeyer, N., Winters, A. R., Gassner, G. J., and Kopriva, D. A. An entropy stable nodal discontinuous Galerkin method for the two dimensional shallow water equations on unstructured curvilinear meshes with discontinuous bathymetry. *Journal of Computational Physics 340* (2017), 200–242. 13, 86

[324] Wu, L., White, C., Scanlon, T. J., Reese, J. M., and Zhang, Y. Deterministic numerical solutions of the Boltzmann equation using the fast spectral method. *Journal of Computational Physics 250* (2013), 27–52. 9

[325] Y. Liu, M. V., and Wang, Z. Spectral difference method for unstructured grids i: basic formulation. *Journal of Computational Physics 216* (2006), 780–801. 4

[326] Yamaleev, N. K., Del Rey Fernández, D. C., Lou, J., and Carpenter, M. H. Entropy stable spectral collocation schemes for the 3-D Navier-Stokes equations on dynamic unstructured grids. *Journal of Computational Physics 399* (2019), 108897. 13

[327] Yan, B., and Yan, G. A steady-state lattice Boltzmann model for incompressible flows. *Computers & Mathematics with Applications 61*, 5 (2011), 1348–1354. 114

[328] Yan, G., Kaur, S., Banks, J. W., and Hicken, J. E. Entropy-stable discontinuous Galerkin difference methods for hyperbolic conservation laws. *Journal of Computational and Applied Mathematics 422* (2023), 114885. 13, 86

[329] Yong, W.-A., Zhao, W., and Luo, L.-S. Theory of the Lattice Boltzmann method: Derivation of macroscopic equations via the Maxwell iteration. *Phys. Rev. E 93* (Mar 2016), 033310. 114

[330] Yu, B., Li, L., Zhang, B., and Wang, J. An Approach to Obtain the Correct Shock Speed for Euler Equations with Stiff Detonation. *Communications in Computational Physics 22*, 1 (2017), 259–284. 7

[331] Zakerzadeh, H., and Noelle, S. A note on the stability of implicit-explict flux-splittings for stiff systems of hyperbolic conservation laws. *Commun. Math. Sci. 16*, 1 (2018), 1–15. 86

[332] Zeifang, J., Schütz, J., Kaiser, K., Beck, A., Lukáčová-Medvid'ová, M., and Noelle, S. A Novel Full-Euler Low Mach Number IMEX Splitting. *Communications in Computational Physics 27*, 1 (2019), 292–320. 86

[333] Zha, G.-C., and Bilgen, E. Numerical solutions of Euler equations by using a new flux vector splitting scheme. *International Journal for Numerical Methods in Fluids 17*, 2 (1993), 115–144. 3

[334] Zinkiewicz, O., Taylor, R., and Nithiarasu, P. *The Finite Element Method for Fluid Flows*, 7^{th} ed. Elsevier, 2014. 7

[335] Zu, Y. Q., and He, S. Phase-field-based lattice Boltzmann model for incompressible binary fluid systems with density and viscosity contrasts. *Phys. Rev. E 87* (2013), 043301. 114