HIGH ORDER INVARIANT DOMAIN PRESERVING FINITE VOLUME SCHEMES FOR NONLINEAR HYPERBOLIC CONSERVATION LAWS

A Dissertation

by

YUCHEN HUA

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Bojan Popov
Committee Members,	Jean-Luc Guermond
	Raytcho Lazarov
	Jean Ragusa
Head of Department,	Emil Straube

August 2019

Major Subject: Mathematics

ABSTRACT

In this dissertation we develop high order invariant domain preserving schemes for general hyperbolic systems.

The schemes are based on the general central schemes of formally second, third and fourth order accuracy. The invariant domain property is modified as the quasiconcave constraint and is enforced via a so-called convex limiting technique. There are two classes of schemes developed. One is based on the invariant domain satisfying nonlinear reconstruction and the other method is made to be invariant domain preserving via the convex flux limiting. The main theoretical results are Theorem 4.3.1 and Theorem 4.3.2. The convex limiting process could sufficiently reduce the oscillations of the numerical solutions at discontinuities like shocks, while it does not deteriorate the order of the underlying central scheme. The numerical performance of the methods is tested on a variety of benchmark problems.

To my family.

ACKNOWLEDGMENTS

First and foremost, I want to express the depth of my gratitude to my advisor Dr. Bojan Popov, for his mentoring and guidance of my research work. His sagacious, patience and encouragement always support me to face the difficulties in my work and make it possible for me to finish this dissertation. I will always show my deepest respect to him for what he has done for me.

I want to acknowledge Dr. Jean-Luc Guermond from the bottom of my heart, for all the discussions with him, his brilliant inspirations and constructive comments for my project. His serious working attitude and high demands in academia will always affect me in my future work.

I sincerely appreciate Dr. Ignacio Tomas for his help. It was a pleasurable experience to work with him. He always spared no effort to help me and I learned a lot of useful things from him which to a great extent increased my work productivity.

I'm also grateful to Dr. Murtazo Nazarov. He is always happy to offer help. The discussion with him has enlightened me a lot in my research work.

Most importantly, none of this would have been possible without the support of my family. I am deeply grateful to my mother Xinhua and father Hongjin for the constant love, care and support from overseas for all these years.

I'm also thankful to former and current student Dr. Wenyu Lei, Dr. Justin Owen, Dr. Mohmood Ettehad and Dr Srinivas Subramanian for being great office mates.

Last but not least, I would like to thank the faculty and staff at the Mathematics Department of Texas A&M University for providing an effective and enjoyable learning environment.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor Bojan Popov [advisor], Professor Jean-Luc Guermond and Professor Raytcho Lazarov of the Department of Mathematics and Professor Jean Ragusa of the Department of Nuclear Engineering.

All the work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University and a dissertation research fellowship from the National Science Foundation grants DMS–1619892.

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION TO NONLINEAR HYPERBOLIC SYSTEMS

1.1 Convention for Vector and Tensor Operations

First, we introduce the usual convention for vector and tensor operations. That is, for any column vectors $\boldsymbol{a} = (a_1, \cdots, a_n)$, $\boldsymbol{b} = (b_1, \cdots, b_n)$, order 2 tensors $\boldsymbol{f}$ with entries f_{ij}, where $i = 1, \cdots, n, j = 1, \cdots, m$ and $\boldsymbol{g}$ with entries g_{ij}, where $i = 1, \cdots, m, j = 1, \cdots, n$, we have the following standard identities, which are going to be used throughout this dissertation:

$$\begin{aligned} &\boldsymbol{a} \otimes \boldsymbol{b} = \boldsymbol{a}\boldsymbol{b}^\top, \qquad &&(\boldsymbol{a} \cdot \boldsymbol{f})_i = \sum_{j=1}^{n} a_i f_{ji}, \quad &&(\boldsymbol{g} \cdot \boldsymbol{a})_i = \sum_{j=1}^{n} f_{ij} a_j, \\ &(\nabla \cdot \boldsymbol{a})_{ij} = \frac{\partial a_j}{\partial x_i}, \qquad &&(\nabla \cdot \boldsymbol{f})_i = \sum_{j=1}^{n} \frac{\partial f_{ji}}{\partial x_j}, \quad &&\nabla \cdot (\boldsymbol{a} \otimes \boldsymbol{b}) = \boldsymbol{a} \cdot \nabla \boldsymbol{b} + \boldsymbol{b} \nabla \cdot \boldsymbol{a}. \end{aligned}$$

1.2 Hyperbolic Conservation Laws

1.2.1 Physical Interpretation of Conservation Laws

In this subsection we introduce the systems of nonlinear, divergence structure first-order hyperbolic partial differential equations (PDEs), which arise as models of physical phenomena, the reader is referred to [14, §4,§11] and [5] for more details.

Let d and m be positive integers and consider the following vector function

$$\boldsymbol{u} = \boldsymbol{u}(\boldsymbol{x}, t) = (u_1(\boldsymbol{x}, t), \dots, u_m(\boldsymbol{x}, t))^\top \in \mathbb{R}^m,$$

the components of which are the densities of various conserved quantities in a physical system. Given any compact set $\Omega \in \mathbb{R}^d$, which has a piecewise smooth boundary $\partial\Omega$, note that the integral

$$\int_\Omega \boldsymbol{u} dx$$

represents the total amount of these quantities within Ω and time t. Physically conservation laws assert that the rate of change of $\boldsymbol{u}$ within Ω is governed by a nonlinear convection flux function $\boldsymbol{f} : \mathbb{R}^m \mapsto (\mathbb{R}^m)^d$, which controls the rate of loss or increase of $\boldsymbol{u}$ through the boundary of the domain $\partial\Omega$, where $\boldsymbol{f}$ is a matrix with entries $f_{ij}(\boldsymbol{u})$, $1 \leq i \leq m$, $1 \leq j \leq d$. Otherwise stated, we have that

$$\frac{d}{dt}\int_{\Omega} \boldsymbol{u} \mathrm{d}\Omega + \int_{\partial\Omega} \boldsymbol{f}(\boldsymbol{u}) \cdot \boldsymbol{n} \mathrm{d}\Gamma = 0, \tag{1.1}$$

for any $n = (n_1, \dots, n_d)^\top \in \mathbb{R}^d$ denoting the outward unit normal along Ω.

Now we assume that $\boldsymbol{u}$ and $\boldsymbol{f}$ are sufficiently smooth, i.e., $\boldsymbol{u}$ and $\boldsymbol{f}$ are all continuously differentiable on Ω, then using the divergence theorem we could rewrite (1.1) as follows

$$\int_{\Omega} \partial_t \boldsymbol{u} \mathrm{d}\Omega + \int_{\Omega} \nabla\cdot \boldsymbol{f}(\boldsymbol{u}) \mathrm{d}\Omega = 0. \tag{1.2}$$

As the region Ω is arbitrary, we could derive from (1.2) the initial value problem for a general system of conservation laws:

$$\partial_t \boldsymbol{u} + \nabla\cdot \boldsymbol{f}(\boldsymbol{u}) = \mathbf{0} \qquad \text{for } (\boldsymbol{x}, t) \in \mathbb{R}^d \times \mathbb{R}_+ \tag{1.3a}$$

$$\boldsymbol{u} = \boldsymbol{u}_0(x) \qquad \text{for } x \in \mathbb{R}^d \tag{1.3b}$$

where (1.3b) describes the initial distribution of $\boldsymbol{u}$.

1.2.2 Connection Between Conservation Laws and Hyperbolic Systems

We consider the conservation law (1.3a) and denote $A^j(\boldsymbol{u}) \doteq D\boldsymbol{f}_j$ to be the Jacobian matrix of $\boldsymbol{f}_j$ at $\boldsymbol{u}$, where $\boldsymbol{f}_j = (f_{1,j}, \cdots, f_{m,j})^\top$, $j = 1, \cdots, d$. If for all $\alpha_1, \cdots, \alpha_d \in \mathbb{R}$, we have that the matrix $A := \alpha_1 A^1 + \cdots + \alpha_d A^d$ has only real eigenvalues and is diagonalizable, then system (1.3a) is hyperbolic and could be written in the following quasilinear form

$$\boldsymbol{u}_t + \sum_{j=1}^{d} A^j(\boldsymbol{u}) \cdot \boldsymbol{u}_{x_j} = 0 \tag{1.4}$$

If the matrix A has m distinct real eigenvalues, it follows that it is diagonalizable, then we say that system (1.3a) is strictly hyperbolic. Note that in this dissertation, all problems we will consider are hyperbolic.

1.3 Examples of Conservation Laws

In this section, we introduce several examples of conservation laws: the linear transport equation and the Burgers equation; the one dimensional isentropic gas dynamics (the so called P-system); and the compressible Euler equation. All of these examples are classical and of great importance in the study of hyperbolic conservation laws. In the following sections(§1.4-§1.6), we will illustrate more important properties of hyperbolic conservation laws and its solutions with these examples. One should notice that in this section we only explain the derivation of the equations and the physical interpretation of each equation, no discussion on initial or boundary condition is included here.

1.3.1 Scalar Conservation Law

First, we consider scalar conservation laws with one space dimension, which in general describe the movement of a traveling wave, see [5]. In this subsection we will introduce two typical examples of scalar conservation law: the linear transport equation and the Burgers equation.

Let $d = 1$ and $m = 1$ in system (1.3a), we obtain a conservation law in one space dimension of the form:

$$u_t + f(u)_x = 0, \tag{1.5}$$

where u denotes the conserved quantity while f is the flux. Now assume that f is differentiable of u with $f'(u) = a(u)$, using the chain rule, equation (1.5) could be written in the quasilinear form:

$$u_t + a(u)u_x = 0, \tag{1.6}$$

where $a(u)$ measures the speed of the wave which depends on u.

1.3.1.1 Conservation Principle

Here we explain the conservation of equation (1.5) from the physical point of view. Let's integral equation 1.5 on a given interval $[a, b]$, we could obtain

$$\begin{aligned}\frac{d}{dt}\int_a^b u(x,t)dx &= -\int_a^b f(u(x,t))_x dx \\ &= f(u(a,t)) - f(u(b,t)) = [\text{inflow at a}] - [\text{outflow at b}].\end{aligned} \tag{1.7}$$

which indicates that there is no creation or destroy implemented on the quantity of u: the change of the total amount of u restricted in the interval $[a, b]$ is the result of the flow across the boundary.

As stated above, $a(u)$ in equation (1.6) denotes the velocity of the transported wave. Now we give two examples of scalar conservation law by taking different $a(u)$ separately.

1.3.1.2 Linear Transport Equation

In the linear case, a single wave in space is just translated while maintaining its shape. Here the dependent variable u denotes the height of the wave and it is the conserved quantity. Thus we could apply equation (1.6) to describe this movement. By assumption the shape of the wave is maintained during the transport. Hence, the wave speed $a(u)$ is a constant function of u, which is denoted by c. Thus we obtain the following equation

$$\frac{\partial u}{\partial t} + c\frac{\partial u}{\partial x} = 0, \tag{1.8}$$

which is known as the linear transport equation.

1.3.1.3 Inviscid Burgers Equation

Burgers' equation is a fundamental PDE occurring in various areas of applied mathematics, such as fluid mechanics and gas dynamics. The equation was first introduced by Harry Bateman in 1915, see [3, 62] and later studied by Johannes Martinus Burgers in 1948, see [6]. Here we introduce the inviscid form of Burgers' equation, which is a special case of nonlinear wave equations.

To obtain the inviscid Burgers equation, we set the wave speed $a(u) = u$ in equation (1.6) and obtain the following equation:

$$u_t + uu_x = 0. \tag{1.9}$$

Equivalently, by taking $f(u) = \frac{u^2}{2}$ in equation (1.5), we could rewrite equation (1.9) to be

$$\frac{\partial}{\partial t}u + \frac{\partial}{\partial x}(\frac{u^2}{2}) = 0. \tag{1.10}$$

1.3.2 The P-system

Here we discuss the one-dimensional motion of an isentropic gas, which is modeled by the so-called P-system, see [61, 57, 64]. To start with, we consider the equations of gas dynamics for an inviscid, non-heat conducting gas. In Lagrangian coordinates the system is written as follows:

$$\partial_t v - \partial_x u = 0, \tag{1.11a}$$

$$\partial_t u + \partial_x p(v) = 0, \tag{1.11b}$$

where $(x, t) \in \mathbb{R} \times \mathbb{R}_+$, v is the specific volume, with $v = \frac{1}{\rho}$ and ρ is the density, u denotes the velocity. The mapping $v \mapsto p(v)$ is the pressure which depends on he particular gas under consideration, and is assumed to be of class $C^2(\mathbb{R}^+; \mathbb{R})$, which satisfies the following property.

$$p' < 0, \qquad p'' > 0, \tag{1.12}$$

which implies that the system is strictly hyperbolic and genuinely nonlinear, see Example 1.6.1 for more details. A typical example of the pressure used in P-system is the so-called gamma-law, where we set $p(v) = rv^{-\gamma}$ with $r > 0$ and $\gamma \geq 1$. Typically we take $r = \frac{(\gamma-1)^2}{4\gamma}$, see [50].

Remark. In gas dynamics, an isentropic process is such that the entropy of the system remains unchanged in time. More details about thermodynamical entropy function will be introduced in the following sections.

1.3.2.1 Conservation Principle

To explain the physical meaning of equation (1.11), we are going to rewrite it in Eulerian coordinates. This is the so-called system of isentropic gas dynamics:

$$\partial_t \rho + \partial_x(\rho u) = 0, \tag{1.13a}$$

$$\partial_t(\rho u) + \partial_x(\rho u^2 + p) = 0. \tag{1.13b}$$

Notice that equation (1.13) is equivalent to a subsystem of the Euler system (1.22). Therefore, we will only give a brief explanation here. More details will be discussed in §1.3.3 for the compressible Euler system.

(i) The conservation of mass. Integrating equation (1.13a) on an interval $[a, b]$ and applying the fundamental theorem of calculus, we have:

$$\frac{\partial}{\partial t}\int_a^b \rho dx = -(\rho u)|_a^b. \tag{1.14}$$

Here the left term $\frac{\partial}{\partial t}\int_a^b \rho dx$ denotes the rate of the change of the mass in $[a, b]$, while the right term $-(\rho u)|_a^b$ measures the incoming mass minus the outgoing mass per unit time, where the term ρu denotes the mass flux.

(ii) The conservation of momentum. Similarly, integrating equation (1.13b) on $[a, b]$ we have:

$$\frac{\partial}{\partial t}\int_a^b \rho u dx = -(\rho u^2)|_a^b - \int_a^b \frac{\partial}{\partial x} p dx. \tag{1.15}$$

On the left-hand side, the term $\frac{\partial}{\partial t}\int_a^b \rho u dx$ denotes the rate of the change of the momentum in $[a, b]$. On the right-hand side, the first term $-(\rho u^2)|_a^b$ measures the net rates at which the momentum enters $[a, b]$, where the term ρu^2 is the momentum flux. The second term $-\int_a^b \frac{\partial}{\partial x} p dx$ on the right-hand side measures the total force imposed to the material on $[a,b]$ and has the unit (momentum/time-volume), thus could be viewed as total rate of change of momentum within $[a,b]$ due to pressure gradients. Therefore, the right-hand side of the equation describes the rate of change of momentum due to the boundary flux and pressure gradients.

Remark. As mentioned above, the P-system could be considered as a subsystem of the Euler system, thus the description of the conservation principle will be similar to the Euler system, see §1.3.3 for more details. However, the temperature of the P-system is a constant, so we don't have the conservation of specific energy for the P-system.

1.3.3 Compressible Euler System

In this subsection we state the Euler equation of compressible gas dynamics, see [59]. The governing equation in Eulerian coordinates has the following conservation form:

$$\partial_t \rho + \nabla\cdot \boldsymbol{m} = 0, \tag{1.16a}$$

$$\partial_t \boldsymbol{m} + \nabla\cdot(\boldsymbol{m} \otimes \frac{\boldsymbol{m}}{\rho} + p\mathbb{I}) = 0, \tag{1.16b}$$

$$\partial_t E + \nabla\cdot(\frac{\boldsymbol{m}}{\rho}(E + p)) = 0, \tag{1.16c}$$

Here ρ is the material density with unit(mass/volume). $\boldsymbol{m} = \rho \boldsymbol{u}$ is the momentum, where $\boldsymbol{u}$ is the particle velocity with unit(length/time). p is the pressure with unit (force/area). $E = \rho(\frac{1}{2}\rho||\boldsymbol{u}||_{L_2}^2 + e)$ is the total energy of the system per unit volume, where e is the specific internal energy with unit(energy/mass), which is determined by a caloric Equation of State (EOS). For ideal gas, we have the following expression:

$$e = e(\rho, p) = \frac{p}{(\gamma - 1)\rho}, \qquad T = (\gamma - 1)e, \tag{1.17}$$

where T is the temperature, $\gamma = \frac{c_p}{c_v}$ denoting the ratio of constant pressure and constant volume heat capacities. $\mathbb{I}$ is an identity matrix in $\mathbb{R}^d$ and d is the dimension number.

Here we briefly introduce some other thermodynamic variables and their relationships for the Euler system and refer to [59] for more details.The basic variables are the pressure p and the specific volume v, which are characterized by the temperature T and satisfy the following relationship for thermally ideal gas:

$$pv = RT, \tag{1.18}$$

where R is a constant depends the particular gas under consideration. The First Law of Thermodynamics states that the change of internal energy is given by

$$dQ = de + pdv, \tag{1.19}$$

where Q is the the heat transmitted to the system and $-pdv$ is the work done on the system by the pressure. Furthermore, the Second Law of Thermodynamics introduces a new variable s, called entropy, which satisfies the relation:

$$Tds = de + pdv. \tag{1.20}$$

The last variable of fundamental interest is the speed of sound, which is defined as

$$a = \sqrt{\frac{\gamma p}{\rho}}. \tag{1.21}$$

Remark. For the Euler system with a γ-law equation of state (1.17), from the Second Law of Thermodynamics (1.20) we derive that the physical specific entropy s is given by $s = \log(e^{\frac{1}{\gamma-1}}\rho^{-1})$.

1.3.3.1 Conservation Principle

Now we explain the physical meaning of the Euler system. We rewrite equations (1.16) in the following form:

$$\partial_t \rho + \nabla\cdot(\rho \boldsymbol{u}) = 0, \tag{1.22a}$$

$$\partial_t(\rho \boldsymbol{u}) + \nabla\cdot(\rho \boldsymbol{u} \otimes \boldsymbol{u} + p\mathbb{I}) = 0, \tag{1.22b}$$

$$\partial_t(\rho e + \frac{1}{2}\rho \boldsymbol{u}^2) + \nabla\cdot(\boldsymbol{u}(\rho e + \frac{1}{2}\rho \boldsymbol{u}^2 + p)) = 0. \tag{1.22c}$$

The equations (1.22) represent conservation of mass, momentum and energy respectively. Notice that the first two equations represent the same principles as the once in the isentropic gas dynamics in (1.13).

(i) The conservation of mass. Integrate equation (1.22a) on a control volume V and apply the integral theorem, we have:

$$\frac{\partial}{\partial t}\int_V \rho dV = -\oint \rho \boldsymbol{u} \cdot \boldsymbol{n} dA, \tag{1.23}$$

where $\boldsymbol{n}$ is the unit normal outward the surface. Here the left term $\frac{\partial}{\partial t}\int_V \rho dV$ denotes the rate of the change of the mass within V, while the right term $-\oint \rho \boldsymbol{u} \cdot \boldsymbol{n} dA$ measures the incoming mass over the outgoing mass per unit time, where the term $\rho \boldsymbol{u}$ is the mass flux.

(ii) The conservation of momentum. Integrate equation (1.22b) on a control volume V and apply the integral theorem, then for each direction i of V we have:

$$\frac{\partial}{\partial t}\int_V \rho u_i dV = -\oint \rho u_i \boldsymbol{u} \cdot \boldsymbol{n} dA - \int_V \frac{\partial}{\partial x_i} p dV, \tag{1.24}$$

where $\boldsymbol{u} = (u_1, \cdots, u_d)$.

On the left-hand side, the term $\frac{\partial}{\partial t}\int_V \rho u_i dV$ denotes the rate of the change of the momentum in direction i within V, where $m_i = \rho u_i$ is the ith component of the momentum. On the right-hand side, the first term $-\oint \rho u_i \boldsymbol{u} \cdot \boldsymbol{n} dA$ measures the ith component net rates at which the momentum enters V, where the term $\rho u_i \boldsymbol{u}$ is the momentum flux in direction i. For the second term $-\int_V \frac{\partial}{\partial x_i} p dV$ on the right-hand, spatial derivative of the pressure $\frac{\partial}{\partial x_i} p$ has a unit(force/volume), where the force has the unit (momentum/time). Therefore the term $-\int_V \frac{\partial}{\partial x_i} p dV$ measures the total rate of change of momentum within V due to the pressure gradients. As a conclusion, the whole equation describes the rate of change of momentum due to the boundary flux and pressure gradients.

(iii) The conservation of total energy. Similarly, integrating equation (1.22c) within a control volume V gives us:

$$\frac{\partial}{\partial t} \int_V (\rho e + \frac{1}{2} \rho \boldsymbol{u}^2) dV = -\oint (\rho e + \frac{1}{2} \rho \boldsymbol{u}^2) \boldsymbol{u} \cdot \boldsymbol{n} dA - \int_V \nabla \cdot (p \boldsymbol{u}) dV. \qquad (1.25)$$

On the left-hand side, the term $\frac{\partial}{\partial t} \int_V (\rho e + \frac{1}{2} \rho \boldsymbol{u}^2) dV$ has a unit(energy/time) and represents the rate of the change of the total energy within V. On the right-hand side, the first term $-\oint (\rho e + \frac{1}{2} \rho \boldsymbol{u}^2) \boldsymbol{u} \cdot \boldsymbol{n} dA$ measures the net rates at which the total energy enters $[a, b]$, where the term $\rho e \boldsymbol{u}$ and $(\frac{1}{2} \rho \boldsymbol{u}^2) \boldsymbol{u}$ represents the internal energy flux and kinetic energy respectively. For the second term $-\int_V \nabla \cdot (p \boldsymbol{u}) dV$ on the right-hand side, by integration on part we obtain:

$$-\int_V \nabla \cdot (p \boldsymbol{u}) dV = -\int_V \nabla p \cdot \boldsymbol{u} dV - \int_V p \nabla \cdot \boldsymbol{u} dV. \qquad (1.26)$$

For the first term of equation (1.26), $-\nabla p$ is the force per unit implement on the material within V, it follows that the term $-\int_V \nabla p \cdot \boldsymbol{u} dV$ measures the total change of the kinetic energy due to the pressure within V. For the second term of equation (1.26), $p \nabla \cdot \boldsymbol{u} dV$ has a unit (volume/time) and thus represents the total volume change due to the material movement. Therefore, the term $-\int_V p \nabla \cdot \boldsymbol{u} dV$ represents the change of internal energy due to the compression or expansion of the material. As a conclusion, the right-hind side of equation (1.25) represents the rate of change of the total energy due to the boundary fluxes of internal and kinetic energy plus the changes of kinetic energy due to pressure gradients and changes of the internal energy from the effect of the body force from the material.

1.4 Characteristics

We start to discuss the solutions to initial problems of conservation laws (1.3a) in this section. To simplify the discussion, we set $d = 1$ in system (1.3a) and consider the following one dimensional Cauchy problem:

$$\partial_t \boldsymbol{u} + \partial_x \boldsymbol{f}(\boldsymbol{u}) = 0, \qquad \text{for } (x,t) \in \mathbb{R} \times \mathbb{R}_+, \tag{1.27a}$$

$$\boldsymbol{u} = \boldsymbol{u}_0(x), \qquad \text{for } x \in \mathbb{R}. \tag{1.27b}$$

If $\boldsymbol{f}$ is smooth enough, we have the quasilinear form of (1.27a):

$$\boldsymbol{u}_t + A(\boldsymbol{u}) \cdot \boldsymbol{u}_x = 0, \tag{1.28}$$

where $A(\boldsymbol{u}) \doteq \boldsymbol{f}(\boldsymbol{u})$ is the Jacobian matrix of $\boldsymbol{f}$ at $\boldsymbol{u}$. We will explain in the following subsections that the properties of the solution to system (1.27a) will be strongly correlated to the characteristic curves determined by the eigenvalues of $A(\boldsymbol{u})$. For simplicity, we only state the conclusion here without proof, and we refer the reader to [14, 58, 59] for more details.

1.4.1 Characteristics Curves

Assuming that $A(u)$ has real eigenvalues $(\lambda_1(x,t), \cdots, \lambda_m(x,t))$, we could define the characteristic curves for system (1.27a) to be the solutions of the following differential equations:

$$\frac{dx}{dt} = \lambda_i(x,t), \qquad i = 1, \cdots, m \tag{1.29}$$

One could prove that the solution $\boldsymbol{u}$ is a constant along each curve.

Remark. From the discussion above we could see that, physically the eigenvalues $\{\lambda_i\}$

represent speeds of propagation of profiles of m waves decoupled from system (1.28) separately. Typically, these speeds will be measured positive in the direction of increasing x and negative otherwise.

1.4.2 Loss of Regularity

Here we consider the solutions defined on the regions between two characteristic curves, which are highly correlated with the features of these curves as t increases: regions where different characteristic curves approach each other correspond to compression waves, and regions where the different characteristic curves move away from each other correspond to expansion waves. When two characteristic curves intersect, it lead to a multivalued solution which is not physical. Thus one has to define a discontinuous solution according to some physical considerations. These discontinuities are called shocks. A typical example is shown as follows.

Example 1.4.1. We consider the inviscid Burgers equation (1.9) with the following initial condition:

$$u(x,0) = \frac{1}{1+x^2}. \tag{1.30}$$

By equation (1.29) the characteristic curve is determined by the solution of equation $\frac{dx}{dt} = u$, along which the directional derivative of $u(t,x)$ vanishes. It then follows that u is a constant along the characteristic lines in the $t-x$ plane:

$$t \mapsto x(t) = x + tu(x,0) = x + \frac{t}{1+x^2}. \tag{1.31}$$

Thus the solution to the Cauchy problem is given implicitly by

$$u(t, x + \frac{t}{1+x^2}) = \frac{1}{1+x^2}. \tag{1.32}$$

when these lines do not intersect. When t increase, the map in equation (1.31) is no longer one-to-one, hence these characteristic lines will intersect and produce multivalued solutions.

1.4.3 Riemann Problem

A typical example to illustrate the characteristics of Cauchy problem (1.27a),(1.27b) is the so-called Riemann Problem with a piecewise constant initial condition

$$\boldsymbol{u}_0(x) = \boldsymbol{u}(x,0) = \begin{cases} \boldsymbol{u}_L & \text{if } x < 0, \\ \boldsymbol{u}_R & \text{if } x > 0. \end{cases} \tag{1.33}$$

The solution to the Riemann problem is an important building block in the theory and approximation of conservation laws. A simple but instructive case is when the flux linear. Then, the Jacobian of the flux is a constant matrix $Ain\mathbb{R}^m \times \mathbb{R}^m$. Let r_i be the eigenvectors of A, $i \in \{1, \cdots, m\}$. The solution with initial condition (1.33) can be obtained as follows, one could see [5] for more details.

We write the vector $\boldsymbol{u}_R - \boldsymbol{u}_L$ as a linear combination of r_i:

$$\boldsymbol{u}_R - \boldsymbol{u}_L = \sum_{i=1}^{m} c_i r_i, \tag{1.34}$$

and define

$$\boldsymbol{w}_i \doteq \boldsymbol{u}_L + \sum_{j<=i} c_j r_j, \qquad i = 0, \cdots, m. \tag{1.35}$$

Then the solution takes the form

$$\boldsymbol{u}(t,x) = \begin{cases} \boldsymbol{w}_0 = \boldsymbol{u}_L & \text{for } x/t < \lambda_1, \\ & \dots \\ \boldsymbol{w}_i & \text{for } \lambda_i < x/t < \lambda_{i+1}, \\ & \dots \\ \boldsymbol{w}_n = \boldsymbol{u}_R & \text{for } x/t > \lambda_n, \end{cases} \tag{1.36}$$

The corresponding $t-x$ plane which describes the distribution of the solution is known as the so-called Riemann fan, see figure 1.1.

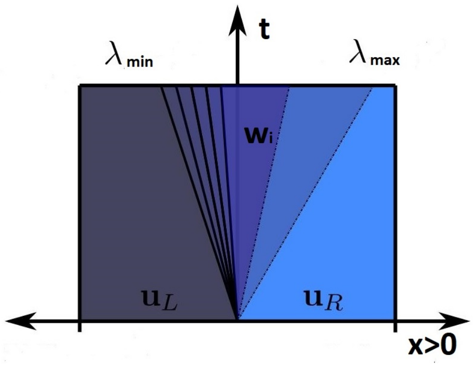


Figure 1.1: The Riemann Fan

Remark. The solutions of nonlinear Riemann problems have a much more complex form constructed of elementary wave solutions, which will be explained in §1.6.2

1.5 Weak Solutions

As shown in example 1.4.1, in the case of genuinely nonlinear cases, even smooth initial condition will possibly lead to a discontinuous solution at finite time. In order to construct solutions globally in time, we thus interpret the equations in a distributional sense. Still, our discussion is given for the one dimensional case while the results for multi dimensional cases are quite similar.

1.5.1 Basic Definition

Definition 1.5.1. A measurable function $\boldsymbol{u} : \mathbb{R} \times [0, +\infty) \mapsto \mathbb{R}^m$ is said to be a distributional solution of hyperbolic system (1.3a), if for every C^1-function $\phi(\boldsymbol{x}, t)$ with compact support within $\Omega = \mathbb{R} \times [0, +\infty)$, we have

$$\iint_\Omega \left(\frac{\partial \phi}{\partial t} \cdot \boldsymbol{u} + \frac{\partial \phi}{\partial x} \cdot \boldsymbol{f}_i(\boldsymbol{u})\right) dxdt = 0 \tag{1.37}$$

Remark. One should notice that no continuity assumption is made on $\boldsymbol{u}$, instead we only require that $\boldsymbol{u}$ and $\boldsymbol{f}(\boldsymbol{u})$ to be in $L^1_{loc}(\mathbb{R} \times [0, +\infty))$, i.e., locally integrable on $\Omega = \mathbb{R} \times [0, +\infty)$.

Definition 1.5.2. A function $\boldsymbol{u} : \mathbb{R} \times [0, +\infty) \mapsto \mathbb{R}^m$ is a weak solution of Cauchy problem (1.27a),(1.27b), if $\boldsymbol{u}$ is a continuous function from $[0, T]$ into L^1_{loc}, satisfies the initial condition $\boldsymbol{u}(x, 0) = \boldsymbol{u}_0(x)$ and the restriction of $\boldsymbol{u}$ to the open strip $]0, T[\times\mathbb{R}$ is a distributional solution of (1.27a).

There is a useful result for weak solution, known as the famous Rankine-Hugoniot jump conditions, which could be used to check whether a piecewise constant function of the following form

$$\boldsymbol{u}(t,x) = \begin{cases} \boldsymbol{u}_L & \text{if } x < \lambda t, \\ \boldsymbol{u}_R & \text{if } x \geq \lambda t, \end{cases} \tag{1.38}$$

is a weak solutions of system (1.27a). One could read [48, 59] for more details.

Lemma 1.5.1. If $\boldsymbol{u}(t, x)$ defined by (1.38) is a weak solution of conservation laws (1.27a), we have

$$\lambda(\boldsymbol{u}_R - \boldsymbol{u}_L) = \boldsymbol{f}(\boldsymbol{u}_R) - \boldsymbol{f}(\boldsymbol{u}_l). \tag{1.39}$$

As a result of Lemma 1.5.1 for scalar equations($m = 1$), we could define the shock speed λ as

$$\lambda = \frac{\boldsymbol{f}(\boldsymbol{u}_R) - \boldsymbol{f}(\boldsymbol{u}_L)}{\boldsymbol{u}_R - \boldsymbol{u}_L}, \tag{1.40}$$

which describes the movement of the shocks.

Example 1.5.1. We consider the Riemann problem for Burgers equation (1.10) with initial condition

$$u(x,0) = \begin{cases} 0 & \text{if } x < 0, \\ 1 & \text{if } x \geq 0. \end{cases} \tag{1.41}$$

For $0 < \alpha < 1$, we consider the following equations

$$u_\alpha(x,t) = \begin{cases} 0 & \text{if } x < \dfrac{\alpha t}{2}, \\ \alpha & \text{if } \dfrac{\alpha t}{2} \leq x < \dfrac{(1+\alpha)t}{2}, \\ 1 & \text{if } x \geq \dfrac{(1+\alpha)t}{2}. \end{cases} \tag{1.42}$$

Clearly all equations defined by (1.42) satisfy Burgers equation (1.10) in smooth regions and has two shocks of speed $\lambda = \frac{\alpha}{2}$ and $\lambda = \frac{1+\alpha}{2}$. By Lemma 1.5.1, equation (1.42) defines infinitely many weak solutions for Cauchy problem (1.10) and (1.41).

1.5.2 Admissibility Conditions

In example 1.5.1 we notice that there is no guarantee of uniqueness and continuity of weak solutions, therefore further admissible conditions must be added to the system. Motivated by the considerations that whether the solution is physical, some of the conditions are presented as follows.

1.5.2.1 Varnishing Viscosity

A weak solution $\boldsymbol{u}$ of system (1.3a) is admissible in the sense of vanishing viscosity if there exists a sequence of smooth solutions $\boldsymbol{u}^\epsilon$ to the parabolic system:

$$\partial_t \boldsymbol{u}^\epsilon + \partial_x \boldsymbol{f}(\boldsymbol{u}^\epsilon) = \epsilon \partial_{xx} \boldsymbol{u}^\epsilon \tag{1.43}$$

which converges to $\boldsymbol{u}$ in L^1_{loc} as $\epsilon \to 0^+$. Meanwhile, if equation (1.43) has a sequence of continuous differentiable solutions which converge in L^1_{loc}, it is proved that the limit is a solution of equation (1.37). Related results are established for $d = 1$, $m = 1$, see [46, 66] and for $d = 1$, $m = 2$, see [7, 12, 13].

1.5.2.2 Entropy Inequalities

In general it is difficult to give uniform estimates of the solutions to equation (1.43) or to discrete versions of viscosity approximations, thus we have to deduce other conditions characterizing the vanishing viscosity limit mentioned above. To start with, we introduce the concept of entropy, which is a generalization of thermodynamic entropy.

Definition 1.5.3. A continuous differential function $\eta : \mathbb{R}^m \mapsto \mathbb{R}$ is called an entropy of conservation law (1.27a) with entropy flux $\boldsymbol{F} : \mathbb{R}^m \mapsto \mathbb{R}$ if it holds that:

(i) η is a convex function of $\boldsymbol{u}$ and

(ii) For all $\boldsymbol{u} \in \mathbb{R}^m$ we have the following equation which measures the entropy production:

$$D\boldsymbol{F}(\boldsymbol{u}) = \eta'(\boldsymbol{u})^\top D\boldsymbol{f}(\boldsymbol{u}). \tag{1.44}$$

An immediate consequence of (1.44) is that, assume that $\boldsymbol{u}$ is a C^1-solution of system (1.27a), then we have:

$$\eta(\boldsymbol{u})_t + \boldsymbol{F}(\boldsymbol{u})_x = 0, \tag{1.45}$$

which is called the entropy conservation function since it represents the conservation of entropy of the system.

Remark. It is known that the entropy satisfies a conservation equation only in the regions where the solution is smooth, thus we could check the smoothness by computing the entropy production. An application of this is to construct the slope limiter for high order central scheme, see §3.5.

Now we use Definition 1.5.3 to derive the so-called entropy inequality, which is originally introduced by [66, 45].

Assume $\eta, \boldsymbol{F} \in C^2$, multiplying equation(1.43) with $D\eta(\boldsymbol{u}^\epsilon)$ leads to

$$[\eta(\boldsymbol{u}^\epsilon)]_t + [\boldsymbol{F}(\boldsymbol{u}^\epsilon)]_x = \epsilon D\eta(\boldsymbol{u}^\epsilon)\boldsymbol{u}^\epsilon_{xx} = \epsilon\{[\eta(\boldsymbol{u})]_{xx} - D^2\eta(\boldsymbol{u}^\epsilon)\cdot(\boldsymbol{u}^\epsilon_x \otimes \boldsymbol{u}^\epsilon_x)\}. \tag{1.46}$$

Since η is convex, its second derivative at any $\boldsymbol{u}^\epsilon$ is a positive definite quadratic form, it follows that the term $D^2\eta(\boldsymbol{u}^\epsilon)\cdot(\boldsymbol{u}^\epsilon_x \otimes \boldsymbol{u}^\epsilon_x)$ satisfies

$$D^2\eta(\boldsymbol{u}^\epsilon)\cdot(\boldsymbol{u}^\epsilon_x \otimes \boldsymbol{u}^\epsilon_x) = \sum_{i,j=1}^{m} \frac{\partial^2\eta(\boldsymbol{u}^\epsilon)}{\partial\boldsymbol{u}_i\partial\boldsymbol{u}_j}\frac{\partial\boldsymbol{u}^\epsilon_i}{\partial x}\frac{\partial\boldsymbol{u}^\epsilon_j}{\partial x} \geq 0. \tag{1.47}$$

Therefore, multiplying equation (1.47) by a nonnegative smooth function ϕ with a compact support and integrating by parts, we have

$$\iint \{\eta(\boldsymbol{u}^\epsilon)\phi_t + F(\boldsymbol{u}^\epsilon)\phi_x\}dxdt \geq -\epsilon \iint \eta(\boldsymbol{u}^\epsilon)\phi_{xx}dxdt. \tag{1.48}$$

Let $\boldsymbol{u}^\epsilon \to \boldsymbol{u}$ in L^1 as $\epsilon \to 0$, then equation (1.48) yields

$$\iint \{\eta(\boldsymbol{u})\phi_t + F(\boldsymbol{u})\phi_x\} dxdt \geq 0, \tag{1.49}$$

for all $\phi \in C_c^1, \phi > 0$. In general this is restated as $\eta(\boldsymbol{u})_t + \boldsymbol{F}(\boldsymbol{u})_x \leq 0$ in the sense of distribution. Therefore we introduce the following admissible condition.

Definition 1.5.4 (Entropy Inequality)**.** A weak solution $\boldsymbol{u}$ of equation (1.27a) is entropy admissible if

$$\eta(\boldsymbol{u})_t + \boldsymbol{F}(\boldsymbol{u})_x \leq 0 \tag{1.50}$$

in the sense of distribution.

Remark. It is well known that for the scalar case(i.e., $m = 1$), the Cauchy problem (1.27a),(1.27b) has a unique entropy solution which satisfies the entropy inequality for any entropy pairs. See [2, 66] for details.

Generally, scalar equations has many entropy pairs, while most physical system has at least one entropy pair which satisfy the entropy inequality. Here we give some examples of entropy pairs for hyperbolic systems explained in §1.3.

(i) Linear Transport Equation. Assuming u is a smooth solution to equation (1.8), it is clear that taking

$$\eta(u) = u, \qquad F(u) = cu \tag{1.51}$$

satisfies (1.45) and thus is an entropy pair of the equation.

(ii) Inviscid Burgers Equation. Similarly, assuming u is a smooth solution to equation (1.10), then taking

$$\eta(u) = \frac{u^2}{2}, \qquad F(u) = \frac{u^3}{3} \tag{1.52}$$

satisfies (1.45) and thus is an option of the entropy pair of the equation.

(iii) The P-system. We consider the P-system (1.11) with gamma-law $p(v) = rv^{-\gamma}$. Assume that $\boldsymbol{u} = (v, u)^\top$ is the smooth solution. Then we multiply (1.11a) with $p(v)$ and (1.11b) with u and add the two to obtain

$$\partial_t(\frac{u^2}{2}) + \partial_x(vu + \frac{r\gamma}{1-\gamma}v^{-\gamma+1}) = 0. \tag{1.53}$$

Thus we could set the entropy pair of the P-system to be

$$\eta(\boldsymbol{u}) = \frac{u^2}{2}, \qquad F(\boldsymbol{u}) = vu + \frac{r\gamma}{1-\gamma}v^{-\gamma+1}. \tag{1.54}$$

Apparently η is a convex function of $\boldsymbol{u}$.

(iv) Compressible Euler System. We consider the Euler system (1.22) and follow the process in [32]. As mentioned in §1.3.3, we take the specific entropy to be the physical specific entropy

$$s = \log(e^{\frac{1}{\gamma-1}}\rho^{-1}) = \frac{1}{\gamma-1}\log e - \log\rho. \tag{1.55}$$

Then its derivatives with respect to density and energy are

$$\frac{\partial s}{\partial \rho} = -\frac{1}{\rho}, \qquad \frac{\partial s}{\partial e} = \frac{1}{(\gamma-1)e}. \tag{1.56}$$

Assuming a smooth solution, take a dot product of (1.22b) with $\boldsymbol{u}$ and subtracting (1.22c) we obtain

$$\frac{\partial}{\partial t}(\rho e) + \nabla\cdot(\boldsymbol{u}\rho e) + p\nabla\cdot\boldsymbol{u} = 0. \tag{1.57}$$

Now multiply (1.22a) with $\frac{\partial s}{\partial \rho}$, (1.57) with $\frac{\partial s}{\partial e}$ and use the EOS (1.17), some mathematical computation give us

$$\frac{\partial s}{\partial t} + \boldsymbol{u}\cdot\nabla s = 0. \tag{1.58}$$

Then for any scalar differentiable function f, we multiple (1.58) with $f'(s)$ to obtain

$$\frac{\partial f(s)}{\partial t} + \boldsymbol{u} \cdot \nabla f(s) = 0. \tag{1.59}$$

For the last step, we multiply (1.59) by ρ, (1.22a) by $f(s)$ and add them up, which leads to

$$\frac{\partial}{\partial t}(\rho f(s)) + \nabla(\boldsymbol{u}\rho f(s)) = 0. \tag{1.60}$$

Therefore taking

$$\eta(\boldsymbol{u}) = -\rho f(s), \qquad F(\boldsymbol{u}) = -\boldsymbol{u} \cdot \nabla f(s), \tag{1.61}$$

satisfy entropy production (1.45). Notice that η is convex function of the conserved variables $\boldsymbol{u}$, we require that, see [31]

$$f'(s) > 0, \qquad \frac{\gamma - 1}{\gamma} f'(s) + f''(s) > 0. \tag{1.62}$$

Remark. From the discussion above, we know that there are families of functional which satisfy condition (1.62). In our work, we typically take either $\eta(\boldsymbol{u}) = -\rho s$ or the following *limit* entropy $f(s) = (\gamma - 1)^{\frac{1}{\gamma}} \exp(s)^{\frac{\gamma-1}{\gamma}}$ (note that for this choice we have $\frac{\gamma-1}{\gamma} f'(s) + f''(s) = 0$) and the limit entropy pair is

$$\eta(\boldsymbol{u}) = p^{\frac{1}{\gamma}}, \qquad F(\boldsymbol{u}) = vp^{\frac{1}{\gamma}}. \tag{1.63}$$

1.6 Elementary Wave Solutions to Riemann Problems

Following [5], we construct the basic solution of Riemann problem mentioned in §1.4.3. For convenience of our discussion, we restate the conservation equation here

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0, \tag{1.64}$$

with a piecewise initial condition

$$\boldsymbol{u}_0(x) = \boldsymbol{u}(x,0) = \begin{cases} \boldsymbol{u}_L & \text{if } x < 0, \\ \boldsymbol{u}_R & \text{if } x > 0. \end{cases} \tag{1.65}$$

Still we take $A(\boldsymbol{u}) \doteq Df(\boldsymbol{u})$ to be the Jacobian matrix.

Throughout our analysis, we adopt the following standard assumption, see [44]. Let $\lambda_i(\boldsymbol{u})$ be the eigenvalues of $A(\boldsymbol{u})$ and $r_i(\boldsymbol{u})$ be corresponding eigenvectors, with $i = 1, \cdots, m$. Then the i-th field is either genuinely nonlinear with $D\lambda_i(\boldsymbol{u}) \cdot r_i(u) > 0$ for all $\boldsymbol{u}$, or linearly degenerate, with $D\lambda_i(\boldsymbol{u}) \cdot r_i(u) = 0$ for all u. In other word, we don't allow partly increase and decrease on the integral curves so there doesn't exist any local extremes. We will explain that the solutions of the Riemann problem can only be rarefaction waves, shocks and contact discontinuities under this assumption.

1.6.1 Elementary Curves

For the conservation laws (1.64) and a fixed state $\boldsymbol{u}_0$ in $\mathbb{R}^m-$space, we introduce two types of elementary curves connected with $\boldsymbol{u}_0$: the rarefaction curve and the shock curve. All settings we are going to use here are the same to the descriptions in §1.4, $A(\boldsymbol{u})$ denotes the Jacobian matrix of $\boldsymbol{f}$, $\lambda_i(\boldsymbol{u})$ are eigenvalues of A, l_i and $r_i(\boldsymbol{u})$ are corresponding left and right eigenvectors with $l_i \cdot r_j = \delta_{ij}$ be the Kronecker value.

1.6.1.1 The Rarefaction curve

For $\boldsymbol{u}_0 \in \mathbb{R}^m$, the integral curve of vector field r_i through u_0 is determined by solving the following Cauchy problem

$$\frac{d\boldsymbol{u}}{d\sigma} = r_i(\boldsymbol{u}), \qquad \boldsymbol{u}(0) = \boldsymbol{u}_0. \tag{1.66}$$

We denote this curve as

$$\sigma \mapsto R_i(\sigma)(\boldsymbol{u}_0). \tag{1.67}$$

1.6.1.2 The Shock curve

For $\boldsymbol{u}_0 \in \mathbb{R}^m$ and $i \in \{1, \cdots, m\}$, we consider the curve connected to the right of $\boldsymbol{u}_0$ by an $i-$shock. which satisfy the Rankine-Hugoniot condition (1.39)

$$\lambda(\boldsymbol{u} - \boldsymbol{u}_0) = \boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{u}_0). \tag{1.68}$$

which could be rewritten as

$$A(\boldsymbol{u}, \boldsymbol{u}_0)(\boldsymbol{u} - \boldsymbol{u}_0) = \boldsymbol{f}(\boldsymbol{u}) - \boldsymbol{f}(\boldsymbol{u}_0), \tag{1.69}$$

where

$$A(\boldsymbol{u}, \boldsymbol{u}_0) \doteq \int_0^1 A(s\boldsymbol{u} + (1-s)\boldsymbol{u}_0)ds \tag{1.70}$$

is the average matrix with $\boldsymbol{u} - \boldsymbol{u}_0$ be the $i-$th eigenvector of $A(\boldsymbol{u}, \boldsymbol{u}_0)$. It has been proved (1.69) hold if and only if $\boldsymbol{u} - \boldsymbol{u}_0$ is orthogonal to every left $j-$eigenvector of $A(\boldsymbol{u}, \boldsymbol{u}_0)$ with $j \neq i$. Thus (1.69) is equivalent to

$$l_j(\boldsymbol{u}, \boldsymbol{u}_0) \cdot (\boldsymbol{u} - \boldsymbol{u}_0) = 0 \qquad \text{for all } j \neq i, \tag{1.71}$$

where $l_j(\boldsymbol{u}, \boldsymbol{u}_0)$ are the left eigenvectors of $A(\boldsymbol{u}, \boldsymbol{u}_0)$. Linearizing (1.71) give us

$$l_j(\boldsymbol{u}_0) \cdot (\boldsymbol{w} - \boldsymbol{u}_0) = 0 \qquad \text{for all } j \neq i, \tag{1.72}$$

which has solutions $\boldsymbol{w} = \boldsymbol{u}_0 + cr_i(\boldsymbol{u}_0), c \in \mathbb{R}$. Consequently this leads to the so-called $i-$shock curve, which is denoted by

$$\sigma \mapsto S_i(\sigma)(\boldsymbol{u}_0). \tag{1.73}$$

Remark. One can show that the two curves R_i and S_i have a second order contact at $\boldsymbol{u}_0$, i.e., we have that

$$|R_i(\sigma)(\boldsymbol{u}_0) - S_i(\sigma)(\boldsymbol{u}_0)| = \mathcal{O}(1) \cdot \sigma^3. \tag{1.74}$$

1.6.2 Solutions to Riemann Problems

With the curves introduced above, we are going to construct the general solution to the Riemann problem.

1.6.2.1 Three Special Cases

Here we introduce three special cases which compose the general solution of the Riemann problem.

(i) Centered Rarefaction Waves. Let the $i-$th field be genuinely nonlinear, and assume that $\boldsymbol{u}_R$ lies on the positive $i-$rarefaction curve through $\boldsymbol{u}_L$, i.e., $\boldsymbol{u}_R = R_i(\sigma)(u_L)$ for some $\sigma > 0$. For each $s \in [0, \sigma]$, the characteristic speed is defined as$\lambda_i(s) = \lambda_i(R_i(s)(\boldsymbol{u}_L))$. By genuinely nonlinearity, the map $s \mapsto \lambda_i(s)$ is strictly increasing, so for every $\lambda \in [\lambda_i(\boldsymbol{u}_L), \lambda_i(\boldsymbol{u}_R)]$, there exists unique $s \in [0, \sigma]$ such that $\lambda = \lambda_i(s)$. Therefore, for $t \geq 0$, the function

$$\boldsymbol{u}(t,x) = \begin{cases} \boldsymbol{u}_L & \text{if } x/t < \lambda_i(\boldsymbol{u}_L), \\ R_i(s)(\boldsymbol{u}_L) & \text{if } x/t = \lambda_i(s) \in [\lambda_i(\boldsymbol{u}_L), \lambda_i(\boldsymbol{u}_R)], \\ \boldsymbol{u}_R & \text{if } x/t > \lambda_i(\boldsymbol{u}_R), \end{cases} \tag{1.75}$$

is a piecewise smooth solution of the Riemann problem.

(ii) Shocks. Let again the $i-$th field be genuinely nonlinear, and $\boldsymbol{u}_R$ is connected to $\boldsymbol{u}_L$ by an $i-$shock, i.e., $\boldsymbol{u}_R = S_i(\sigma)(u_L)$. Then the function

$$\boldsymbol{u}(t,x) = \begin{cases} \boldsymbol{u}_L & \text{if } x < \lambda t, \\ \boldsymbol{u}_R & \text{if } x > \lambda t, \end{cases} \tag{1.76}$$

defines a piecewise constant solution to the Riemann problem, where $\lambda \doteq \lambda_i(\boldsymbol{u}_L, \boldsymbol{u}_R)$

is the Rankine-Hugoniot speed of the shock, which satisfies

$$\lambda_i(\boldsymbol{u}_L) > \lambda_i(\boldsymbol{u}_L, \boldsymbol{u}_R) > \lambda_i(\boldsymbol{u}_R). \tag{1.77}$$

(iii) Contact Discontinuities. Assume that the $i-$th field is linearly degenerate and the state $\boldsymbol{u}_R$ lies on the $i-$th rarefaction curve through $\boldsymbol{u}_L$, i.e., $\boldsymbol{u}_R = R_i(\sigma)(\boldsymbol{u}_L)$ for some σ. By assumption, the $i-$th characteristic speed λ_i is constant along this curve. Choosing $\lambda = \lambda(\boldsymbol{u}_L)$, the piecewise constant function (1.76) provides a solution to the Riemann problem since the Rankine-Hugoniot conditions hold at the point of the jump

$$\begin{aligned} f(\boldsymbol{u}_R) - f(\boldsymbol{u}_L) &= \int_0^\sigma Df(R_i(s)(\boldsymbol{u}_L))r_i(R_i(s)(\boldsymbol{u}_L))ds \\ &= \int_0^\sigma \lambda(\boldsymbol{u}_L)r_i(R_i(s)(\boldsymbol{u}_L))ds = \lambda_i(\boldsymbol{u}_L)\cdot[R_i(\sigma)(\boldsymbol{u}_L) - \boldsymbol{u}_L]. \end{aligned} \tag{1.78}$$

It then follows by (1.78) that for linearly degenerate fields the shock and rarefaction curves actually coincide, i.e., $S_i(\sigma(\boldsymbol{u}_0)) = R_i(\sigma(\boldsymbol{u}_0))$ for all σ.

1.6.2.2 The General Solution

The results of §1.6.2.1 could be summarized as follows. For a fixed state $\boldsymbol{u}_L$ and $i \in \{1, \cdots, m\}$, we define the mixed curve

$$\Psi_i(\sigma)(\boldsymbol{u}_L) = \begin{cases} R_i(\sigma)(\boldsymbol{u}_L) & \text{if } \sigma \geq 0, \\ S_i(\sigma)(\boldsymbol{u}_L) & \text{if } \sigma < 0. \end{cases} \tag{1.79}$$

Then for $\boldsymbol{u}_R = \Psi_i(\sigma)(\boldsymbol{u}_L)$, the solution to the Riemann problem (1.64),(1.65) contains the elementary waves: a rarefaction wave, a shock or a contact discontinuities. Similar to the discussion in §1.4.3, we construct the solution by finding the intermediate states $\boldsymbol{w}_0 = \boldsymbol{u}_L, \boldsymbol{w}_1, \cdots, \boldsymbol{w}_n = \boldsymbol{u}_R$ such that each pair of adjacent states $\boldsymbol{w}_{i-1}, \boldsymbol{w}_i$ are connected by an elementary wave, i.e.,

$$\boldsymbol{w}_i = \Psi(\sigma_i)(\boldsymbol{w}_{i-1}). \tag{1.80}$$

Thus the original problem decompose into n Riemann problems

$$\boldsymbol{u}_t + \boldsymbol{f}(\boldsymbol{u})_x = 0, \qquad \boldsymbol{u}(0,x) = \begin{cases} \boldsymbol{w}_{i-1}, & \text{if } x < 0, \\ \boldsymbol{w}_i, & \text{if } x > 0. \end{cases} \tag{1.81}$$

By construction, each of these problems has an entropy admissible solution consisting of a simple wave of the $i-$th characteristic family.

(i) The $i-$th characteristic field is genuinely nonlinear and $\sigma_i > 0$. The solution of (1.81) consists of a centered rarefaction wave. The $i-$th characteristic speeds range over the interval $[\lambda_i^-, \lambda_i^+]$ with $\lambda_i^- \doteq \lambda_i(\boldsymbol{w}_{i-1})$ and $\lambda_i^+ \doteq \lambda_i(\boldsymbol{w}_i)$.

(ii) Either the $i-$th characteristic field is genuinely nonlinear with $\sigma_i \leq 0$, or the $i-$th characteristic field is linearly degenerate with arbitrary σ_i. Then the solution of (1.81) consists of an admissible shock or a contact discontinuity, traveling with Rankine-Hugoniot speed $\lambda_i^- \doteq \lambda_i(\boldsymbol{w}_{i-1}, \boldsymbol{w}_i)$.

As a conclusion, the piecewise smooth solution of system (1.64),(1.65) is defined as

$$\boldsymbol{u}(t,x) = \begin{cases} \boldsymbol{w}_0 = \boldsymbol{u}_L & \text{if } x/t \in]-\infty, \lambda_1^-[, \\ R_i(s)(\boldsymbol{w}_{i-1}) & \text{if } x/t = \lambda_i(R_i(s)(\boldsymbol{w}_{i-1})) \in [\lambda_i^-, \lambda_i^+[, \\ \boldsymbol{w}_i & \text{if } x/t \in [\lambda_i^+, \lambda_{i-1}^-[, \\ \boldsymbol{w}_m = \boldsymbol{u}_R & \text{if } x/t \in [\lambda_m^+, \infty[. \end{cases} \tag{1.82}$$

if all characteristics defined above are finite.

Example 1.6.1. To illustrate the discussion above, we consider the Riemann problem of the P-system (1.11). Let $\boldsymbol{u} = (v, u)^\top$, it is easy to compute

$$A(\boldsymbol{u}) = \begin{pmatrix} 0 & -1 \\ p'(v) & 0 \end{pmatrix}. \tag{1.83}$$

The eigenvalues and eigenvectors are

$$\lambda_1 = -\sqrt{-p'(v)}, \qquad \lambda_2 = \sqrt{-p'(v)}. \tag{1.84}$$

$$r_1 = \begin{pmatrix} 1 \\ \sqrt{-p'(v)} \end{pmatrix}, \qquad r_2 = \begin{pmatrix} -1 \\ \sqrt{-p'(v)} \end{pmatrix}. \tag{1.85}$$

Thus the system is hyperbolic if $p'(v) < 0$ for all $v > 0$. Also the system is genuinely nonlinear if $p''(v) > 0$ for all $v > 0$ since $D\lambda_1 \cdot r1_{=} D\lambda_2 \cdot r_2 = \frac{p''(v)}{2\sqrt{-p'(v)}}$.

The rarefaction waves through $\boldsymbol{u}_L$ are given by

$$R_1 = \{(v, u); \quad u - u_L = \int_{v_L}^{v} \sqrt{-p'(s)} ds\}, \tag{1.86}$$

and

$$R_2 = \{(v, u); \quad u - u_L = -\int_{v_L}^{v} \sqrt{-p'(s)} ds\}. \tag{1.87}$$

The shock curves are computed as

$$S_1 = \{(v, u); \quad -(u - u_L)^2 = (v - v_L)(p(v) - p(v_L)), \quad \lambda \doteq -\frac{u - u_L}{v - v_L} < 0\}, \tag{1.88}$$

and

$$S_1 = \{(v, u); \quad -(u - u_L)^2 = (v - v_L)(p(v) - p(v_L)), \quad \lambda \doteq -\frac{u - u_L}{v - v_L} > 0\}. \tag{1.89}$$

These four curves depart the $\mathbb{R}^2$ into four areas:

$$\begin{aligned} &\Omega_1, \quad \text{bordering on } R_1, S_2, \qquad &&\Omega_1, \quad \text{bordering on } R_1, R_2, \\ &\Omega_1, \quad \text{bordering on } S_1, S_2, \qquad &&\Omega_1, \quad \text{bordering on } S_1, R_2. \end{aligned} \tag{1.90}$$

It has been proved that the general solution is determined by the location of $\boldsymbol{u}_R \in \mathbb{R}^2$:

(i) If $\boldsymbol{u} \in \Omega_1$, the solution consists of two rarefaction waves.

(ii) If $\boldsymbol{u} \in \Omega_2$, the solution consists of a 1-rarefaction wave and a 2-shock.

(iii) If $\boldsymbol{u} \in \Omega_3$, the solution consists of two shocks.

(iv) If $\boldsymbol{u} \in \Omega_4$, the solution consists of a 1-shock wave and a 2-rarefaction.

Remark (Riemann Invariants)**.** For the system of isentropic gas dynamics we could introduce the concepts of a Riemann invariant, a function that is constant along a characteristic curve of a fixed family (one or two) defined by (1.66), see for examples in [51]. Specifically in our study of the P-system, we take the families of Riemann invariants to be

$$w_1(\boldsymbol{u}) = u + \int_v^\infty d\mu, \quad \text{and} \quad w_2(\boldsymbol{u}) = u - \int_v^\infty d\mu, \tag{1.91}$$

with the notation $d\mu := \sqrt{-p'(s)}ds$. For general hyperbolic systems one could read [15] for more details.

2. INTRODUCTION TO INVARIANT DOMAIN

In considering the well-posedness of solutions to hyperbolic systems, some certain types of restrictions must be enforced to guarantee that the solutions are physical. An typical example to understand such restrictions is the Euler system, for which we enforce the positivity of density, internal energy and minimum principle of the specific entropy. In this chapter we will introduce the concepts of invariant domain for general hyperbolic systems, while examples for the systems discussed in § 1.3 will be explained specifically.

2.1 Integral Average of Riemann Solution

To start with, some important results of the solution to a Riemann problem will be recalled here, one could read [21] for more details.

We consider Riemann problem (1.64),(1.65) and assume that there is a clear notion for the solution of it. Namely, we assume that there exists an admissible set $\mathcal{A} \subset \mathbb{R}^m$ such that the following one-dimensional Riemann problem is (uniquely) solvable

$$\partial_t \boldsymbol{u} + \partial_x(\boldsymbol{f}(\boldsymbol{u})) = 0, \qquad (x,t) \in \mathbb{R} \times \mathbb{R}_+, \qquad \boldsymbol{u}(x,0) = \begin{cases} \boldsymbol{u}_L & \text{if } x < 0 \\ \boldsymbol{u}_R & \text{if } x > 0 \end{cases} \tag{2.1}$$

for any Riemann pair $(\boldsymbol{u}_L, \boldsymbol{u}_R)$ in $\mathcal{A}^2$. We define the characteristic in the same way of § 1.6.2.2, which gives us

$$\lambda_1^- \leq \lambda_1^+ \leq \cdots \leq \lambda_m^- \leq \lambda_m^+. \tag{2.2}$$

Now we assume that the maximum speed of propagation $\lambda_{\text{max}}(\boldsymbol{u}_L, \boldsymbol{u}_R) := \max(|\lambda_1^-|, |\lambda_m^+|)$ is a finite number, then by the result of (1.82) it follows that for $t \geq 0$ we have

$$\boldsymbol{u}(x,t) = \begin{cases} \boldsymbol{u}_L, & \text{if } x \leq -t\lambda_{\text{max}}(\boldsymbol{u}_L, \boldsymbol{u}_R), \\ \boldsymbol{u}_R, & \text{if } x \geq t\lambda_{\text{max}}(\boldsymbol{u}_L, \boldsymbol{u}_R), \end{cases} \tag{2.3}$$

We assume also that there exists a convex subset A of $\mathcal{A}$, which we call invariant set, such that for any Riemann pair in $A \times A$, the average of the Riemann solution over the Riemann fan is also in A for all $(x,t) \in \mathbb{R}\times\mathbb{R}_+$. The existence of such set has been established by [9] on a very large class of hyperbolic systems.

Instead of considering the Riemann solution directly, we give the following result which describes the integral average of solution (2.3).

Lemma 2.1.1. Let $\boldsymbol{u}_L, \boldsymbol{u}_R \in \mathcal{A}$, let $\boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)$ be the unique Riemann solution to (2.1), let $\bar{\boldsymbol{u}}(t, \boldsymbol{u}_L, \boldsymbol{u}_R) := \int_{-\frac{1}{2}}^{\frac{1}{2}} \boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)(x,t)dx$ and assume that $t\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R) \leq \frac{1}{2}$, then

$$\bar{\boldsymbol{u}}(t, \boldsymbol{u}_L, \boldsymbol{u}_R) = \frac{1}{2}(\boldsymbol{u}_L + \boldsymbol{u}_R) - t(\boldsymbol{f}(\boldsymbol{u}_R) - \boldsymbol{f}(\boldsymbol{u}_L)). \tag{2.4}$$

Proof. Since $t\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R) \leq \frac{1}{2}$, we have that

$$\begin{aligned} \bar{\boldsymbol{u}}(t, \boldsymbol{u}_L, \boldsymbol{u}_R) &= \int_{-\frac{1}{2}}^{\frac{1}{2}} \boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)(x,t)dx \\ &= (\frac{1}{2} - t\lambda_{\max})(\boldsymbol{u}_L + \boldsymbol{u}_R) + \int_{-t\lambda_{\max}}^{t\lambda_{\max}} \boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)(x,t)dx. \end{aligned} \tag{2.5}$$

On the other hand, we integrate the equation (2.1) on $[-t\lambda_{\max}, t\lambda_{\max}] \times [0,t]$, which leads to

$$\begin{aligned} 0 &= \int_0^t \int_{-t\lambda_{\max}}^{t\lambda_{\max}} \boldsymbol{u}_t dx d\tau + \int_0^t \int_{-t\lambda_{\max}}^{t\lambda_{\max}} \boldsymbol{f}(\boldsymbol{u})_x dx d\tau \\ &= \int_{-t\lambda_{\max}}^{t\lambda_{\max}} (\boldsymbol{u}(x,t) - \boldsymbol{u}(x,0))dx + \int_0^t (\boldsymbol{f}(\boldsymbol{u}(t\lambda_{\max}, t)) - \boldsymbol{f}(\boldsymbol{u}(-t\lambda_{\max}, t)))d\tau \\ &= \int_{-t\lambda_{\max}}^{t\lambda_{\max}} \boldsymbol{u}(x,t)dx - \lambda(\boldsymbol{u}_L + \boldsymbol{u}_R) + t(\boldsymbol{f}(\boldsymbol{u}_R) - \boldsymbol{f}(\boldsymbol{u}_L)). \end{aligned} \tag{2.6}$$

which implies that

$$\int_{-t\lambda_{\max}}^{t\lambda_{\max}} \boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)(x,t)dx = t\lambda_{\max}(\boldsymbol{u}_L + \boldsymbol{u}_R) - t(\boldsymbol{f}(\boldsymbol{u}_R) - \boldsymbol{f}(\boldsymbol{u}_L)). \tag{2.7}$$

Plug (2.7) into (2.5) proves out result. □

2.2 Invariant Sets and Invariant Domain

Following the discussion in [21], we introduce the notions of invariant sets and invariant domains. The definitions here are slightly different from those in [9], [15], [34], and [33]. We will associate invariant sets only with solutions of Riemann problems and define invariant domains only for an approximation process.

Definition 2.2.1 (Invariant Set)**.** We say that a set $A \subset \mathcal{A} \subset \mathbb{R}^m$ is invariant for (1.3) if for any pair $(\boldsymbol{u}_L, \boldsymbol{u}_R) \in A \times A$ and any $t > 0$, the integral average of the entropy solution of the Riemann problem (2.1) over the Riemann fan, say, $\frac{1}{t(\lambda_m^+ - \lambda_1^-)} \int_{\lambda_1^- t}^{\lambda_m^+ t} \boldsymbol{u}(\boldsymbol{u}_L, \boldsymbol{u}_R)(x,t)dx$, remains in A.

We now introduce the notion of invariant domain for an approximation process. Let $\boldsymbol{X}_h \subset L^1(\mathbb{R}^d, \mathbb{R}^m)$ be a finite-dimensional approximation space and let $S_h : \boldsymbol{X}_h \ni uh \mapsto S_h(\boldsymbol{u}_h) \in \boldsymbol{X}_h$ be a discrete process over $\boldsymbol{X}_h$. Henceforth we abuse the language by saying that a member of $\boldsymbol{X}_h$, say $\boldsymbol{u}_h$, is in the set $A \subset \mathbb{R}^m$ when actually we mean that $\{\boldsymbol{u}_h(\boldsymbol{x}) | x \in \mathbb{R}\} \subset A$.

Definition 2.2.2 (Invariant Domain)**.** A convex invariant set $A \subset \mathcal{A} \subset \mathbb{R}^m$ is said to be an invariant domain for the process S_h if and only if for any state $\boldsymbol{u}_h$ in A, the state $S_h(\boldsymbol{u}_h)$ is also in A.

For scalar conservation equations the notions of invariant sets and invariant domains are closely related to the maximum principle, see § 2.3.1. In the case of nonlinear systems, the notion of maximum principle does not apply and must be replaced by the notion of invariant domain. For example, the invariant domain theory when $m = 2$ and $d = 1$ relies on the existence of global Riemann invariants, the best known examples are the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian form, i.e., the so-called P-system, see [51]. For general hyperbolic systems, results of invariant domain property for various finite volume schemes has been established, we refer the reader to [9, 15, 34, 55].

2.3 Examples of Invariant Domains of Hyperbolic Systems

Here, we will illustrate the abstract notions of invariant sets and invariant domains with the examples discussed in § 1.3. At the same time, for each example we are going to give a brief description of the local speed of propagation, which is the only needed information of the to be introduced central scheme, see § 3.1.

2.3.1 Invariant Domain of Scalar Equations

Assume $m = 1$ and d is arbitrary, i.e., (1.3) is a scalar equation. Provided $\boldsymbol{f} \in \mathsf{Lip}(\mathbb{R};\mathbb{R}^d)$, any bounded interval is an admissible set for (1.3). For any Riemann data u_L, u_R, the maximum speed of propagation in (2.3) is bounded by $\lambda_{\max}(u_L,u_R) := ||\boldsymbol{f}||_{\mathsf{Lip}(u_{\min},u_{\max})}$ where $u_{\min} = \min(u_L,u_R)$ and $u_{\max} = \max(u_L,u_R)$. If $\boldsymbol{f}$ is convex and is of class C^1, we have $\lambda_{\max}(u_L,u_R) = \max(|\boldsymbol{f}'(u_L)|,|\boldsymbol{f}'(u_R)|)$ if $f(u_L) \leq f(u_R)$ and $\lambda_{\max}(u_L,u_R) = (\boldsymbol{f}(u_L) - \boldsymbol{f}(u_R))/(u_L - u_R)$ otherwise. Any interval $[a,b] \subset \mathbb{R}$ is admissible and is an invariant set for (1.3), i.e., if $u_R, u_L \in [a,b]$, then $a \leq u(u_L,u_R) \leq b$ for all $t \geq 0$, i.e., any interval $[a,b]$ is an invariant domain for any numerical scheme which satisfies a local maximum principle property, see [10] for example.

2.3.2 Invariant Domain of P-system

Here we consider the P-system (1.11) with the gamma-law. Recall that using the notation $d\mu := \sqrt{-p'(s)}ds$, and assuming $\int_1^\infty d\mu < \infty$, the system has two families of global Riemann invariants:

$$w_1(\boldsymbol{u}) = u + \int_v^\infty d\mu, \quad \text{and} \quad w_2(\boldsymbol{u}) = u - \int_v^\infty d\mu. \tag{2.8}$$

Note that $\int_1^\infty d\mu < \infty$ if $\gamma > 1$. If $\gamma = 1$, we take $w_1(\boldsymbol{u}) = u - \sqrt{r}\log(v)$ and $w_2(\boldsymbol{u}) = u + \sqrt{r}\log(v)$ instead. Let $a, b \in \mathbb{R}$, then it can be shown that any set $A_{ab} \in \mathbb{R}_+ \times \mathbb{R}$ of the form

$$A_{ab} := \{\boldsymbol{u} \in \mathbb{R}_+ \times \mathbb{R} | a \leq w_2(\boldsymbol{u}), w_1(\boldsymbol{u}) \leq b\} \tag{2.9}$$

is an invariant set for the system (1.11), see [34, 64]. Moreover, A_{ab} is an invariant domain for the Lax-Frederich scheme, see [34, 33], and the Guaranteed Maximum Speed (GMS) scheme of [21].

Following [21], we give an estimation of the maximum speed in the following result.

Lemma 2.3.1. Let $(v_L, v_R), (u_L, u_R) \in \mathbb{R} \times \mathbb{R}$ with $v_L, v_R < \infty$. Then

$$\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R) = \begin{cases} \sqrt{-p'(\min(u_L, u_R))}, & \text{if } u_L - u_R > \sqrt{(v_L - v_R)(p(v_R) - p(v_L)}, \\ \sqrt{-p'(v^*)}, & \text{otherwise,} \end{cases} \tag{2.10}$$

where v^* is the unique solution of $\phi(v) = f_L(v) + f_R(v) + u_L - u_R$ and

$$f_z(v) = \begin{cases} -\sqrt{(p(v) - p(v_z)(v_z - v)} & \text{if } v \le v_z, \\ \displaystyle\int_{v_z}^{v} d\mu & \text{if } v > v_z. \end{cases} \tag{2.11}$$

By setting $w_1^{\max} = \max(w_1(u_L), w_1(u_R))$ and $w_2^{\min} = \min(w_2(u_L), w_2(u_R))$ we have $v^0 \le \min(v_L, v_R, v^*)$, i.e., $\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R) \le \sqrt{-p'(v^0)}$ where

$$v^0 = (\gamma r)(\frac{4}{(\gamma - 1)(w_1^{\max} - w_2^{\min})})^{\frac{2}{\gamma - 1}}. \tag{2.12}$$

Remark. The proof of the lemma is established using the characteristic curves described in Example 1.6.1, one could read [21] for details. It should be noticed that the exact value of v^* is computed with v^0 while the invariant domain property guarantees that $v^0 \le v^*$. In practice we will use v^0 as the estimation of the maximum speed in all numerical tests of the P-system.

2.3.3 Invariant Domain of Euler System

Here we consider the compressible Euler system (1.16). Taking s to be the physical specific entropy, it is known that

$$A_r := \{(\rho, \boldsymbol{m}, E) |\, \rho > 0, e > 0, s \ge r\} \tag{2.13}$$

is an invariant set for the Euler system for any $r \in \mathbb{R}$. For example, it is shown that A_r is convex and is an invariant domain for the staggered Lax-Friedrichs scheme, see [15] and the non-staggered Lax-Friedrichs scheme, see [11]. The Guaranteed Maximum Speed (GMS) first order scheme from [21] are also invariant domain preserving.

Following [21, 26], the estimation of the local speed of propagation is given as follows.

Let $\boldsymbol{c}_L := (\rho_L, \boldsymbol{m}_L, \mathcal{E}_L)$ and $\boldsymbol{c}_R := (\rho_R, \boldsymbol{m}_R, \mathcal{E}_R)$, where $\mathcal{E}_z = E_z - \frac{1}{2}\frac{||\boldsymbol{m}_z^\perp||_{\updownarrow 2}^2}{\rho_z}$, $z \in \{L, R\}$. Also, we set $a_z = \sqrt{\frac{\gamma p_z}{\rho_z}}$ to be the sound speed, $A_z := \frac{2}{(\gamma+1)\rho_z}$, $B_z := \frac{\gamma-1}{\gamma+1}p_z$, where again z is either L or R. Then the maximum wave speed is given by

$$\lambda_{\max}(\boldsymbol{c}_L, \boldsymbol{c}_R) = \max(|\lambda_1^-(\boldsymbol{c}_L, \boldsymbol{c}_R)|, |\lambda_3^+(\boldsymbol{c}_L, \boldsymbol{c}_R)|), \tag{2.14}$$

where

$$|\lambda_1^-(\boldsymbol{c}_L, \boldsymbol{c}_R)| = u_L - a_L(1 + \frac{\gamma+1}{2\gamma}(\frac{p^* - p_L}{p_L})_+)^{\frac{1}{2}}, \tag{2.15a}$$

$$|\lambda_3^+(\boldsymbol{c}_L, \boldsymbol{c}_R)| = u_R + a_R(1 + \frac{\gamma+1}{2\gamma}(\frac{p^* - p_R}{p_R})_+)^{\frac{1}{2}}, \tag{2.15b}$$

with $z_+ := \max(0, z)$ and p^* is the solution to the following function

$$\phi(p) := f(p, L) + f(p, R) + u_R - u_L, \tag{2.16a}$$

$$f(p, z) := \begin{cases} (p - p_z)(\frac{A_z}{p+B_z})^{\frac{1}{2}} & \text{if } p \geq p_z, \\ \frac{2a_z}{\gamma-1}((\frac{p}{p_z})^{\frac{\gamma-1}{2\gamma}} - 1) & \text{if } p < p_z. \end{cases} \tag{2.16b}$$

Remark. In practice, it may take a large number of iterations to compute p^* with traditional techniques and one should note that we only need an upper bound on $\lambda_{\max}$. A fast algorithm described in [26] gives the guaranteed upper bound to any prescribed accuracy ϵ of the type $\lambda_{\max} \leq \tilde{\lambda}_{\max} \leq (1 + \epsilon)\lambda_{\max}$ and it is shown that

$$\tilde{p^*} = \left(\frac{a_L + a_R - \frac{\gamma-1}{2}(u_R - u_L)}{a_L p_L^{-\frac{\gamma-1}{2\gamma}} + a_R p_R^{-\frac{\gamma-1}{2\gamma}}} \right)^{\frac{2\gamma}{\gamma-1}} \tag{2.17}$$

is an upper bound for p^* for $\gamma \in (1, \frac{5}{3}]$.

3. CENTRAL SCHEMES FOR HYPERBOLIC SYSTEMS

In this chapter we will introduce central schemes of approximation orders one, two, three and four in space. We describe the fully discrete form of these schemes and some general types of the local polynomial reconstruction used. We have made the following contributions:

(i) We adopt to the central scheme framework the first order invariant domain preserving GMS-scheme (see [21]) and thus create a first order central scheme which satisfies all the invariant domain properties of the underlying PDE.

(ii) We derive an entropy based indicator, which could be used to detect the local smoothness of the solution.

(iii) We apply the smoothness indicator to design a novel local slope reconstruction for the second order KT-scheme in [41].

(iv) We apply the smoothness indicator to design an adaptive local polynomial reconstruction for a general high order central scheme up to order four in space.

Notice that in this dissertation all theoretical results and numerical computations will be made on uniform space discretizations with non-staggered rectangular meshes, i.e., in one space dimension we take $x_j := j\Delta x,\ j \in \mathbb{Z}$ to be the center of cell $I_{x_j} := [x_{j-1/2}, x_{j+1/2}]$, where $x_{j\pm1/2} := x_j \pm \frac{\Delta x}{2}$ denotes the cell interface and Δx is the mesh size. All variable are defined on the cell centers x_j. Also, $t^n := n\Delta t$ denotes the $n-$step of the time discretization. Our presentation is done in the case of one and two space dimensions, as typically done in the central schemes literature. However, it is not difficult to derive analogous results in arbitrary space dimension on rectangular meshes. For further details on central schemes we refer the reader to [41, 42].

3.1 A Brief Review of Classical Central Schemes

Central schemes are widely used in approximations of nonlinear hyperbolic conservation laws (1.3) because they don't rely on the specific eigenstructure of the system and do not require exact or approximate Riemann solvers. In particular, we do not have to involve the characteristic decomposition of the flux $\boldsymbol{f}$, and the computation of the Jacobian of $\boldsymbol{f}$ is not required as well. Actually, the only needed information to build a central scheme is the local speed of propagation $\lambda_{\max}$. One should notice that in practice we are not going to use the exact value of $\lambda_{\max}$. We will only need an upper bound on the speed, see the discussion in §2.3.

The study of central scheme originated from the first-order Lax-Friedrichs scheme (the LxF-scheme) which, see [46, 16], owns the following form in one dimensional case:

$$\boldsymbol{u}_j^{n+1} = \frac{\boldsymbol{u}_{j+1}^n + \boldsymbol{u}_{j-1}^n}{2} - \frac{\lambda}{2}[\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)], \tag{3.1}$$

where $\boldsymbol{u}_j^n$ is an cell average of $\boldsymbol{u}$ at the grid point x_j and t^n, $\lambda := \Delta t/\Delta x$ is the mesh ratio and. It is known that the central LxF scheme (3.1) enjoys an advantage over the canonical upwind Godunov type scheme, see [17], that it does not require exact Riemann solvers in its construction. However, on the other hand the LxF scheme will create a large numerical dissipation, which prevents sharp resolution at shocks or rarefaction tips.

One natural extension of the LxF-scheme is the Nessyahu-Tadmor scheme (the NT-scheme), see [53], which has a higher resolution than the LxF-scheme while retains the simplicity of a Riemann solver free approach. The idea to construct the NT-scheme is to replace the piecewise constant solution in the LxF-scheme with the MUSCL piecewise linear second order approximation, see [60], which could efficiently reduce the the excessive numerical dissipations and therefore provide a higher approximation accuracy. Still the LxF solver is applied, which avoids the complex construction of the Riemann solver. The NT-scheme in one dimensional case is derived as follows. See [53].

We assume that an approximation to the solution of system(1.3) at time t^n is piecewise linear approximation of the form

$$\widetilde{\boldsymbol{u}}(x,t^n) := \sum_j [\bar{\boldsymbol{u}}_j^n + (\boldsymbol{u}_x)_j^n(x-x_j)]\mathbf{1}_{[x_{j-1/2},x_{j+1/2}]}, \tag{3.2}$$

where $\bar{\boldsymbol{u}}_j^n \approx \int_{I_{x_j}} \boldsymbol{u}(\xi,t^n)d\xi/\Delta x$ denotes the compute cell average, $(\boldsymbol{u}_x)_j^n$ are approximations to the exact derivatives, which are reconstructed from the cell averages computed above. The flux integrals of the piecewise-linear interpolant $\bar{\boldsymbol{u}}_j^n$ on $[x_j, x_{j+1}]$ leads to the NT-scheme, see [41]:

$$\bar{\boldsymbol{u}}_{j+1/2}^{n+1} = \frac{\bar{\boldsymbol{u}}_j^n + \bar{\boldsymbol{u}}_{j+1}^n}{2} + \frac{\Delta x}{8}((\boldsymbol{u}_x)_j^n - (\boldsymbol{u}_x)_{j+1}^n) - \lambda[\boldsymbol{f}(\boldsymbol{u}_{j+1}^{n+1/2}) - \boldsymbol{f}(\boldsymbol{u}_j^{n+1/2})], \tag{3.3}$$

where the interface values are

$$\boldsymbol{u}_j^{n+1/2} = \boldsymbol{u}_j^n - \frac{\Delta t}{2}(\boldsymbol{f}_x)_j^n. \tag{3.4}$$

Therefore, the approximate solution $\boldsymbol{u}_j^{n+1}$ at the next time level $t = t^{n+1}$ is obtained by evaluating the cell averages of $\bar{\boldsymbol{u}}_{j+1/2}^{n+1}$ on the staggered meshes, which leads to

$$\frac{\boldsymbol{u}_j^{n+1} - \boldsymbol{u}_j^n}{\Delta t} = -\frac{\boldsymbol{f}(\boldsymbol{u}_{j+1}^-) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^-)}{2\Delta x} + \frac{1}{2\Delta t}(\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n) - \frac{1}{2\Delta t}(\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n), \tag{3.5}$$

where $\boldsymbol{u}_j^- = \boldsymbol{u}_j^n - \frac{1}{2}\lambda(\boldsymbol{u}_x)_j$.

Remark. If we set $(\boldsymbol{u}_x)_j^n \equiv 0$ in particular, the NT-scheme will reduce to the first-order LxF-scheme in the staggered form .

3.2 First Order GMS-Scheme

Here we introduce the derive of a first order invariant domain preserving scheme. The idea is to adopt the invariant domain preserving scheme in [21] to the central scheme framework and one will see that this new scheme coincides with the fully discrete first order central scheme in [41]. Fully discrete form for both one dimensional and two dimensional cases will be presented in this section, and a proof of the invariant domain preserving property will be provided.

3.2.1 One Dimensional Case

We first consider the one space dimensional case of system(1.3).

$$\begin{cases} \partial_t \boldsymbol{u} + \partial_x \boldsymbol{f}(\boldsymbol{u}) = 0, & \text{for } (x,t) \in \mathbb{R} \times \mathbb{R}_+ \\ \boldsymbol{u}(x,0) = \boldsymbol{u}_0(x) \end{cases} \tag{3.6}$$

Using the assumption that the space discretization is uniform, we set the cell centers to be $x_j := j\Delta x$, $j \in \mathbb{Z}$, and assume that the approximate solution $\widetilde{\boldsymbol{u}}^n(x) \approx \boldsymbol{u}(x, t^n)$ at time t^n is a piecewise constant function

$$\widetilde{\boldsymbol{u}}^n(x) := \sum_j \boldsymbol{u}_j^n \mathbf{1}_{[x_{j-1/2}, x_{j+1/2}]}, \qquad x_{j\pm 1/2} := x_j \pm \frac{\Delta x}{2}. \tag{3.7}$$

The values $\boldsymbol{u}_j^n$, $j \in \mathbb{Z}$, are the cell averages of the approximate solution at time t^n. The time step $\Delta t_n := t^{n+1} - t^n$ is generic, determined by a CFL condition for each $n \geq 0$, and we will denote $t^{n+1} = t^n + \Delta t$ where we abuse the notation and drop the dependence on n in the time step. Then by adopting the GMS-scheme in [21] to the finite volume framework, the invariant domain preserving first order scheme could be written as the following form:

$$\frac{\boldsymbol{u}_j^{n+1} - \boldsymbol{u}_j^n}{\Delta t} = -\frac{\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)}{2\Delta x} + \frac{\lambda_{j+1/2}^n}{2\Delta x}(\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n) - \frac{\lambda_{j-1/2}^n}{2\Delta x}(\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n), \tag{3.8}$$

where the quantity $\lambda_{j+1/2}^n := \lambda_{\max}(\boldsymbol{u}_j^n, \boldsymbol{u}_{j+1}^n, \boldsymbol{f})$ denotes the maximum speed of propagation of the Riemann problem with initial left state $\boldsymbol{u}_j^n$, right state $\boldsymbol{u}_{j+1}^n$ and flux $\boldsymbol{f}$. One should notice that only the upper bound of the exact speed is needed, see the examples in § 2.3. Similar to the discussion in [21], we provide the following result which states that the invariant domain preserving property of the scheme given by (3.8) holds if the following CFL condition holds for all $n \geq 0$

$$\max_j \frac{\Delta t \lambda^n_{j+1/2}}{\Delta x} \leq \frac{1}{2}. \tag{3.9}$$

Theorem 3.2.1. Let $A \subset \mathcal{A}$ be an invariant set for (3.6) in the sense of Definition (2.2.2). Assume that A is convex and that for any admissible states $\boldsymbol{u}_L$, $\boldsymbol{u}_R$, the maximum speed of propagation $\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R, \boldsymbol{f})$ is finite. Assume that $\boldsymbol{u}^0_h \in A$ and the CFL condition (3.9) holds. Then we have:

(i) A is an invariant domain for the process $\boldsymbol{u}^n_h \mapsto \boldsymbol{u}^{n+1}_h$ where $\boldsymbol{u}^{n+1}_h$ is computed with scheme (3.8) for all $n \geq 0$.

(ii) Given $n \geq 0$ and $j \in \{1 : I\}$, let $B \subset A$ be a convex invariant set such that $\boldsymbol{u}^n_j \in B$ and $\boldsymbol{u}^n_{j\pm 1} \in B$, then $\boldsymbol{u}^{n+1}_j \in B$.

Proof. The proof is similar to the proof in [21] that we try to express the update given by (3.8) in the form of a convex combination of values which belongs to the invariant domain. For this sake, we introduce the following auxiliary value

$$\bar{\boldsymbol{u}}^{n+1}_{j+1/2} := \frac{1}{2}(\boldsymbol{u}^n_j + \boldsymbol{u}^n_{j+1}) - \frac{1}{2\lambda^n_{j+1/2}}(\boldsymbol{f}(\boldsymbol{u}^n_{j+1}) - \boldsymbol{f}(\boldsymbol{u}^n_j)), \tag{3.10}$$

which under the CFL-condition (3.9) are integral averages of the exact solution of the Riemann problem with a left state $\boldsymbol{u}^n_j$, a right state $\boldsymbol{u}^n_{j+1}$ and a flux $\boldsymbol{f}$. Then by Lemma 2.1.1 it follows that these states are naturally in the local invariant set of the system.

Next by (3.8) we have

$$
\begin{aligned}
\boldsymbol{u}_j^{n+1} =&\boldsymbol{u}_j^n - (\frac{\Delta t}{2\Delta x})(\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)) \\
&+ (\frac{\Delta t}{2\Delta x})\lambda_{j+1/2}^n(\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n) \\
&- (\frac{\Delta t}{2\Delta x})\lambda_{j-1/2}^n(\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n) \\
=&\boldsymbol{u}_j^n - (\frac{\Delta t}{2\Delta x})(\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_j^n) + \boldsymbol{f}(\boldsymbol{u}_j^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)) \\
&+ (\frac{\Delta t}{2\Delta x})\lambda_{j+1/2}^n(\boldsymbol{u}_{j+1}^n + \boldsymbol{u}_j^n - 2\boldsymbol{u}_j^n) \\
&- (\frac{\Delta t}{2\Delta x})\lambda_{j-1/2}^n(-\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n + 2\boldsymbol{u}_j^n) \\
=&(1 - \frac{\Delta t}{\Delta x}(\lambda_{j+1/2}^n + \lambda_{j-1/2}^n))\boldsymbol{u}_j^n \\
&+ (\frac{\Delta t}{\Delta x}\lambda_{j+1/2}^n)\bar{\boldsymbol{u}}_{j+1/2}^{n+1} + (\frac{\Delta t}{\Delta x}\lambda_{j-1/2}^n)\bar{\boldsymbol{u}}_{j-1/2}^{n+1}.
\end{aligned}
\tag{3.11}
$$

Then by the CFL condition(3.9) it follows that $\boldsymbol{u}_j^{n+1}$ is a convex combination of $\boldsymbol{u}_j^n$ and bar states $\bar{\boldsymbol{u}}_{j+1/2}^{n+1}, \bar{\boldsymbol{u}}_{j-1/2}^{n+1}$. Note that the quantity $\bar{\boldsymbol{u}}_{j+1/2}^{n+1}$ is exactly of the form $\bar{\boldsymbol{u}}(t, \boldsymbol{u}_j^n, \boldsymbol{u}_{j+1}^n)$ with the flux $\boldsymbol{f}$ and a fake time $t = \frac{1}{2\lambda_{j+1/2}^n}$, this proves that $\bar{\boldsymbol{u}}_{j+1/2}^{n+1} := \bar{\boldsymbol{u}}(t, \boldsymbol{u}_j^n, \boldsymbol{u}_{j+1}^n) \in B$ for all $j \in \{1 : I\}$. Then by convexity of B we have that $\boldsymbol{u}_j^{n+1} \in B$ if we have $\boldsymbol{u}_j^n \in B$ and $\boldsymbol{u}_{j\pm1}^n \in B$. As a consequence, A is a invariant domain for scheme(3.8), which completes the proof. □

Remark. It is easy to verify that using forward Euler time stepping to the semi-discrete form of the first order central scheme in [41, Eqn. (4.8)] will result in the same discrete method of (3.8). Therefore, in the one-dimensional case, the fully discrete first order central scheme derived from [41] coincides with the invariant domain preserving scheme when the CFL condition is defined as above and the Euler time stepping is used.

Now we rewrite the fully discrete scheme(3.8) in flux form using the notation $\boldsymbol{u}_j^{L,n+1}$, which indicates the first order method.

$$
\frac{\boldsymbol{u}_j^{L,n+1} - \boldsymbol{u}_j^n}{\Delta t} = -\frac{L_{j+1/2}^n - L_{j-1/2}^n}{\Delta x}
\tag{3.12}
$$

where $L^n_{j+1/2}$ is the first order interface numerical flux of the form

$$L^n_{j+1/2} = \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^n_{j+1}) + \boldsymbol{f}(\boldsymbol{u}^n_j)) - \frac{1}{2}\lambda^n_{j+1/2}(\boldsymbol{u}^n_{j+1} - \boldsymbol{u}^n_j). \tag{3.13}$$

Remark. The Euler time stepping in (3.12) can be upgraded to any Strong Stability Preserving (SSP) Runge-Kutta (RK) schemes. In practice we take the following version of the SSP-RK schemes, see [18, 19, 47]. That is, for a system of ODEs

$$\boldsymbol{u}_t = \boldsymbol{L}(\boldsymbol{u}), \tag{3.14}$$

an optimal third order total variation diminishing(TVD) Runge-Kutta scheme is given by

$$\begin{aligned} \boldsymbol{u}^{(1)} &= \boldsymbol{u}^n + \Delta \boldsymbol{L}(\boldsymbol{u}^n), \\ \boldsymbol{u}^{(2)} &= \frac{3}{4}\boldsymbol{u}^n + \frac{1}{4}\boldsymbol{u}^{(1)} + \frac{1}{4}\Delta \boldsymbol{L}(\boldsymbol{u}^{(1)}), \\ \boldsymbol{u}^{n+1} &= \frac{1}{3}\boldsymbol{u}^n + \frac{2}{3}\boldsymbol{u}^{(2)} + \frac{2}{3}\Delta \boldsymbol{L}(\boldsymbol{u}^{(2)}). \end{aligned} \tag{3.15}$$

Notice that the SSP-RK schemes has a stronger property than TVD, that is, it could preserve any convex functional bound that is satisfied under the forward Euler integration. Therefore it could be used for any invariant domain preserving schemes. One could see the discussion in [25, § 4.5] for more details and references on SSP-RK schemes.

3.2.2 Two Dimensional Case

Now we consider the case of two space dimension in (1.3) with $\boldsymbol{x} := (x, y)$

$$\begin{cases} \partial_t \boldsymbol{u} + \partial_x \boldsymbol{f}(\boldsymbol{u}) + \partial_y \boldsymbol{g}(\boldsymbol{u}) = 0, \qquad \text{for } (\boldsymbol{x}, t) \in \mathbb{R}^2 \times \mathbb{R}_+, \\ \boldsymbol{u}(\boldsymbol{x}, 0) = \boldsymbol{u}_0(\boldsymbol{x}). \end{cases} \tag{3.16}$$

Still, we use uniform rectangular mesh with cell centers $(x_j, y_k) := (j\Delta x, k\Delta y), j, k \in \mathbb{Z}$, and take the approximate solution $\widetilde{\boldsymbol{u}}^n(x, y) \approx \boldsymbol{u}(x, y, t^n)$ at time t^n to be a piecewise constant function

$$\widetilde{\boldsymbol{u}}^n(x,y) := \sum_{j,k} \boldsymbol{u}^n_{j,k} \mathbf{1}_{[x_{j-1/2},x_{j+1/2}]\times[y_{k-1/2},y_{k+1/2}]}, \tag{3.17}$$

where the values $\boldsymbol{u}^n_{j,k}$, $j,k \in \mathbb{Z}$, are the cell averages of the approximate solution at time t^n. The time step is determined in the same way as in the one dimensional case. Namely, we denote $t^{n+1} = t^n + \Delta t$ where we abuse the notation and drop the dependence on n in the time step. The fully discrete scheme is given as follows

$$\begin{aligned}\frac{\boldsymbol{u}^{n+1}_{j,k} - \boldsymbol{u}^n_{j,k}}{\Delta t} =& -\frac{\boldsymbol{f}(\boldsymbol{u}^n_{j+1,k}) - \boldsymbol{f}(\boldsymbol{u}^n_{j-1,k})}{2\Delta x} \\ &+ \frac{\lambda^n_{j+1/2,k}}{2\Delta x}(\boldsymbol{u}^n_{j+1,k} - \boldsymbol{u}^n_{j,k}) - \frac{\lambda^n_{j-1/2,k}}{2\Delta x}(\boldsymbol{u}^n_{j,k} - \boldsymbol{u}^n_{j-1,k}) \\ &- \frac{\boldsymbol{f}(\boldsymbol{u}^n_{j,k+1}) - \boldsymbol{f}(\boldsymbol{u}^n_{j,k-1})}{2\Delta y} \\ &+ \frac{\lambda^n_{j,k+1/2}}{2\Delta y}(\boldsymbol{u}^n_{j,k+1} - \boldsymbol{u}^n_{j,k}) - \frac{\lambda^n_{j,k-1/2}}{2\Delta x}(\boldsymbol{u}^n_{j,k} - \boldsymbol{u}^n_{j,k-1}).\end{aligned} \tag{3.18}$$

Similar to the one dimensional case, we have the following result of the invariant domain preserving property if the local speeds are given by $\lambda^{n,x}_{j+1/2,k} = \lambda_{\max}(\boldsymbol{u}^n_{j,k}, \boldsymbol{u}^n_{j+1,k}, \boldsymbol{f})$, $\lambda^{n,y}_{j,k+1/2} = \lambda_{\max}(\boldsymbol{u}^n_{j,k}, \boldsymbol{u}^n_{j,k+1}, \boldsymbol{g})$ and the following CFL condition holds for all $n \geq 0$

$$\max_{j,k}\Big(\frac{\Delta t \lambda^{n,x}_{j+1/2,k}}{\Delta x}, \frac{\Delta t \lambda^{n,y}_{j,k+1/2}}{\Delta y}\Big) \leq \frac{1}{4} \tag{3.19}$$

Theorem 3.2.2. Let $A \subset \mathcal{A}$ be an invariant set for (3.16) in the sense of Definition (2.2.2). Assume that A is convex and that for any admissible states $\boldsymbol{u}_L$, $\boldsymbol{u}_R$, the maximum speed of propagation $\lambda_{\max}(\boldsymbol{u}_L, \boldsymbol{u}_R, \boldsymbol{F})$ is finite, where $\boldsymbol{F} = (\boldsymbol{f}, \boldsymbol{g})^\top$. Assume that $\boldsymbol{u}^0_h \in A$ and the CFL condition (3.19) holds. Then we have:

(i) A is an invariant domain for the process $\boldsymbol{u}^n_h \mapsto \boldsymbol{u}^{n+1}_h$ where $\boldsymbol{u}^{n+1}_h$ is computed with scheme (3.18) for all $n \geq 0$.

(ii) Given $n \geq 0$ and $j,k \in \{1 : I\}$, let $B \subset A$ be a convex invariant set such that

$\boldsymbol{u}^n_{j,k} \in B$ and $\boldsymbol{u}^n_{j\pm1,k}, \boldsymbol{u}^n_{j,k\pm1} \in B$, then $\boldsymbol{u}^{n+1}_{j,k} \in B$.

Remark. Similar to the one-dimensional case, the proof of Theorem 3.2.2 relies on the introduction of the intermediate bar states

$$\bar{\boldsymbol{u}}^{n+1}_{j+1/2,k} = \frac{1}{2}(\boldsymbol{u}^n_{j,k} + \boldsymbol{u}^n_{j+1,k}) - \frac{1}{2\lambda_{j+1/2,k}}(\boldsymbol{f}(\boldsymbol{u}^n_{j+1,k}) - \boldsymbol{f}(\boldsymbol{u}^n_{j,k})), \tag{3.20}$$

and

$$\bar{\boldsymbol{u}}^{n+1}_{j,k+1/2} = \frac{1}{2}(\boldsymbol{u}^n_{j,k} + \boldsymbol{u}^n_{j,k+1}) - \frac{1}{2\lambda_{j,k+1/2}}(\boldsymbol{f}(\boldsymbol{u}^n_{j,k+1}) - \boldsymbol{f}(\boldsymbol{u}^n_{j,k})), \tag{3.21}$$

which are the integral averages of exact solutions of the Riemann problems under the CFL-condition (3.19), therefore, these states are naturally in the local invariant set of the problem due to the definition. Then it follows by convexity that the statement of Theorem(3.2.2) hold.

In order to implement a convex limiting process, the fully discrete first order central scheme of the flux form is given as follows

$$\frac{\boldsymbol{u}^{L,n+1}_{j,k} - \boldsymbol{u}^{L,n}_{j,k}}{\Delta t} = -\frac{L^{n,x}_{j+1/2,k} - L^{n,x}_{j-1/2,k}}{\Delta x} - \frac{L^{n,y}_{j,k+1/2} - L^{n,y}_{j,k-1/2}}{\Delta y}, \tag{3.22}$$

where the first-order interface fluxes are defined as follows:

$$L^{n,x}_{j+1/2,k} = \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^n_{j+1,k}) + \boldsymbol{f}(\boldsymbol{u}^n_{j,k})) - \frac{1}{2}\lambda^{n,x}_{j+1/2,k}(\boldsymbol{u}^n_{j+1,k} - \boldsymbol{u}^n_{j,k}), \tag{3.23a}$$

$$L^{n,y}_{j,k+1/2} = \frac{1}{2}(\boldsymbol{g}(\boldsymbol{u}^n_{j,k+1}) + \boldsymbol{g}(\boldsymbol{u}^n_{j,k})) - \frac{1}{2}\lambda^{n,y}_{j,k+1/2}(\boldsymbol{u}^n_{j,k+1} - \boldsymbol{u}^n_{j,k}). \tag{3.23b}$$

Remark. In the general multidimensional case ($d \geq 2$), we have similar results as Theorem 3.2.1 and Theorem 3.2.2 which will hold when the constant in the CFL condition is $\frac{1}{2d}$, i.e.,

$$\max_{j_1,\cdots,j_d}\left(\frac{\Delta t \lambda^n_{j_1+1/2,\cdots,j_d}}{\Delta x_i}, \cdots, \frac{\Delta t \lambda^n_{j_1,\cdots,j_i+1/2,\cdots,j_d}}{\Delta x_i}, \cdots, \frac{\Delta t \lambda^n_{j_1,\cdots,j_d+1/2}}{\Delta x_i}\right) \leq \frac{1}{2d}, \tag{3.24}$$

where $\lambda^n_{j_1,\cdots,j_i+1/2,\cdots,j_d} = \lambda_{\max}(\boldsymbol{u}^n_{j_1,\cdots,j_i,\cdots,j_d}, \boldsymbol{u}^n_{j_1,\cdots,j_i+1,\cdots,j_d}, \boldsymbol{f}_i)$ Thus one could see that this scheme is based on a guaranteed bound on the local maximum wave speed and is henceforth referred to as the GMS-scheme (guaranteed maximum speed), see [22] for more details on GMS-schemes.

3.3 Second Order KT-Scheme

In this section we recall the second order central scheme from [41], which we call the KT-scheme in the rest of the dissertation. It is known that the NT-scheme scheme we introduced in § 3.1 enjoys a high resolution since it could efficiently reduce the numerical dissipation. However, the influence caused by the numerical dissipations will also accumulate over the time steps and then be observed in certain numerical tests, see [41]. One solution is to apply a semi-discrete formulation which is known to enjoy a smaller numerical viscosity that is proportional to the vanishing time step Δt. The KT-scheme, which is an improvement of the NT-scheme, is designed to have a smaller numerical dissipation, and are known as the first fully discrete Godunov-type central scheme which admit a semi-discrete form, see [41]. Same to the discussion in § 3.2, one and two dimensional cases will be explained separately here.

3.3.1 One Dimensional Case

We first recall the classical NT-scheme in one space dimensional, which is constructed by computing the integral average over the Riemann fans based over the staggered control volumes $[x_j, x_{j+1}] \times [t^n, t^{n+1}]$. The main improvement of the KT-scheme is that a more precise information of the local propagation of discontinuities is considered, and to compute the integral average over new control volumes $[x^{n,-}_{j+1/2}, x^{n,+}_{j+1/2}] \times [t^n, t^{n+1}]$ centered at $x_{j+1/2} \times [t^n, t^{n+1}]$. Here $x^{n,-}_{j+1/2} := x_{j+1/2} - \lambda^n_{j+1/2}\Delta t, x^{n,+}_{j+1/2} := x_{j+1/2} + \lambda^n_{j+1/2}\Delta t$ are selected to separate the smooth and non-smooth regions of the Riemann fan, where $\lambda^n_{j+1/2}$ denotes the maximum local propagation speed of discontinuities. Thus the spatial width of the narrower control volume is $2\lambda^n_{j+1/2}\Delta$ t. Consequently, evaluation of cell averages at t^{n+1} are computed on non-smooth regions $[x^{n,-}_{j+1/2}, x^{n,+}_{j+1/2}]$ and smooth regions

$[x^{n,+}_{j-1/2}, x^{n,-}_{j+1/2}]$ separately. Finally, to evaluate the cell averages over the original uniformly non-staggered meshes $[x_{j-1/2}, x_{j+1/2}]$, a piecewise-linear reconstruction will be applied over the nonuniform cells $[x^{n,-}_{j+1/2}, x^{n,+}_{j+1/2}]$, which will lead to the KT-scheme as a result. Now we state the semi-discrete form of the scheme as follows, one could read [41] for more details.

We assume the same setup of space and time discretization as for first order GMS-scheme, and assume the approximate solution $\widetilde{\boldsymbol{u}}^n = \boldsymbol{u}(x, t^n)$ at time $t = t^n$ to be piecewise linear

$$\widetilde{\boldsymbol{u}}^n := \sum_j [\boldsymbol{u}^n_j + (\boldsymbol{u}_x)^n_j (x - x_j)] \mathbf{1}_{[x_{j-1/2}, x_{j+1/2}]}, \qquad x_{j\pm 1/2} := x_j \pm \frac{\Delta x}{2}, \tag{3.25}$$

where the values $\boldsymbol{u}^n_j$ are cell averages of approximate solutions and $(\boldsymbol{u}_x)^n_j$ are approximations of exact derivatives $\boldsymbol{u}_x$ at (x_j, t^n). The semi-discrete form of the KT-scheme is given as follows, see [41].

$$\begin{aligned}\frac{d}{dt}\boldsymbol{u}_j(t) = &- \frac{(\boldsymbol{f}(\boldsymbol{u}^+_{j+1/2}(t)) + \boldsymbol{f}(\boldsymbol{u}^-_{j+1/2}(t))) - (\boldsymbol{f}(\boldsymbol{u}^+_{j-1/2}(t)) + \boldsymbol{f}(\boldsymbol{u}^-_{j-1/2}(t)))}{2\Delta x} \\ &+ \frac{\lambda_{j+1/2}(t)(\boldsymbol{u}^+_{j+1/2}(t) - \boldsymbol{u}^-_{j+1/2}(t)) - \lambda_{j-1/2}(t)(\boldsymbol{u}^+_{j-1/2}(t) - \boldsymbol{u}^-_{j-1/2}(t))}{2\Delta x}\end{aligned} \tag{3.26}$$

where $\boldsymbol{u}^+_{j+1/2} := \boldsymbol{u}_{j+1}(t) - \frac{\Delta x}{2}(\boldsymbol{u}_x)_{j+1}(t)$, $\boldsymbol{u}^-_{j+1/2} := \boldsymbol{u}_j(t) + \frac{\Delta x}{2}(\boldsymbol{u}_x)_j(t)$ are the interface values and $\lambda_{j+1/2}(t) = \lambda_{\max}(\boldsymbol{u}^-_{j+1/2}(t), \boldsymbol{u}^+_{j+1/2}(t), \boldsymbol{f})$ denotes the local maximum speed.

By setting the second order numerical flux to be

$$H^n_{j+1/2} := \frac{\boldsymbol{f}(\boldsymbol{u}^{n,+}_{j+1/2}) + \boldsymbol{f}(\boldsymbol{u}^{n,-}_{j+1/2})}{2} - \frac{\lambda^n_{j+1/2}}{2}(\boldsymbol{u}^{n,+}_{j+1/2} - \boldsymbol{u}^{n,-}_{j+1/2}). \tag{3.27}$$

and using a forward Euler in time we obtain the fully discrete KT-scheme:

$$\frac{\boldsymbol{u}^{H,n+1}_j - \boldsymbol{u}^n_j}{\Delta t} = -\frac{H^n_{j+1/2} - H^n_{j-1/2}}{\Delta x}. \tag{3.28}$$

Remark. There are a lot of numerical tests which indicates that the KT-scheme given by (3.26) is not invariant domain preserving. However, by setting the numerical derivative $(\boldsymbol{u}_x)_j$ to be zero in (3.26), the semi-discrete form of KT-scheme will reduce to its first order form (3.8), which satisfies the invariant domain preserving property.

Remark. For the fully discrete form of the KT-scheme (3.27),(3.28), in order to retain the high order accuracy, we will apply the third order SSP-RK3 scheme for the forward Euler time stepping as in the first order case, see [41] for more discussions.

Remark. Similar to the first order case, we are not going to compute the exact estimation for local speed $\lambda_{j+1/2}(t)$, which relies on the characteristic structure of the system. Instead the upper bound computed from § 2.3 will be applied in practice. At the same time, in order to apply a convex flux limiting process to the KT-scheme, we require the CFL condition to be determined by the first order GMS-scheme.

Remark. As we mentioned before, the term $(\boldsymbol{u}_x)_j^n$ denotes the approximation of the derivative locally at (x_j, t^n), which is a main affection of the performance of the numerical solutions in the sense of reducing the oscillations. There is a library of recipes for non-oscillatory reconstructions In § 3.3.3 we will explain how to decide such reconstructions in order to obtain a solution which enjoys a high resolution and be entropy consistent.

3.3.2 Two Dimensional Case

In the case of two space dimensions, we use the same rectangular cell and the same time discretization as for the first order scheme in §3.2.2. The approximate solution $\widetilde{\boldsymbol{u}}^n = \boldsymbol{u}(x, y, t^n)$ is a piecewise linear function given by

$$\widetilde{\boldsymbol{u}}^n(x,y) := \sum_{j,k} [\boldsymbol{u}_{j,k}^n + (\boldsymbol{u}_x)_{j,k}^n (x - x_j) + (\boldsymbol{u}_y)_{j,k}^n (y - y_k)] \mathbf{1}_{[x_{j-1/2}, x_{j+1/2}] \times [y_{k-1/2}, y_{k+1/2}]}, \tag{3.29}$$

where $x_{j\pm1/2} := x_j \pm \frac{\Delta x}{2}$, $y_{k\pm1/2} := y_k \pm \frac{\Delta y}{2}$. The values $\boldsymbol{u}_{j,k}^n$ are the cell averages of the approximate solutions and $((\boldsymbol{u}_x)_{j,k}^n, (\boldsymbol{u}_y)_{j,k}^n)$ is the approximate gradient on cell

$[x_{j-1/2}, x_{j+1/2}] \times [y_{k-1/2}, y_{k+1/2}]$ at time $t = t^n$. Following [41], we set the numerical fluxes to be

$$H^{n,x}_{j+1/2,k} := \frac{\boldsymbol{f}(\boldsymbol{u}^{n,+}_{j+1/2,k}) + \boldsymbol{f}(\boldsymbol{u}^{n,-}_{j+1/2,k})}{2} - \frac{\lambda^{n,x}_{j+1/2,k}}{2}(\boldsymbol{u}^{n,+}_{j+1/2,k} - \boldsymbol{u}^{n,-}_{j+1/2,k}), \tag{3.30a}$$

$$H^{n,y}_{j,k+1/2} := \frac{\boldsymbol{g}(\boldsymbol{u}^{n,+}_{j,k+1/2}) + \boldsymbol{g}(\boldsymbol{u}^{n,-}_{j,k+1/2})}{2} - \frac{\lambda^{n,y}_{j,k+1/2}}{2}(\boldsymbol{u}^{n,+}_{j,k+1/2} - \boldsymbol{u}^{n,-}_{j,k+1/2}), \tag{3.30b}$$

where $\boldsymbol{u}^{n,\pm}_{j+1/2,k} := \boldsymbol{u}^n_{j+1,k} \mp \frac{\Delta x}{2}(\boldsymbol{u}^n_x)_{j+1/2\pm1/2,k}$ and $\boldsymbol{u}^{n,\pm}_{j,k+1/2} = \boldsymbol{u}^n_{j,k+1} \mp \frac{\Delta x}{2}(\boldsymbol{u}^n_x)_{j,k+1/2\pm1/2}$, with $\lambda^{n,x}_{j+1/2,k} = \lambda_{\max}(\boldsymbol{u}^{n,-}_{j+1/2,k}(t), \boldsymbol{u}^{n,+}_{j+1/2,k}(t), \boldsymbol{f})$, $\lambda^{n,y}_{j,k+1/2} = \lambda_{\max}(\boldsymbol{u}^{n,-}_{j,k+1/2}(t), \boldsymbol{u}^{n,+}_{j,k+1/2}(t), \boldsymbol{g})$ as the local speeds. Then, a forward Euler time step of the semi-discrete KT-scheme can be written as follows

$$\frac{\boldsymbol{u}^{H,n+1}_{j,k} - \boldsymbol{u}^n_{j,k}}{\Delta t} = -\frac{H^{n,x}_{j+1/2,k} - H^{n,x}_{j-1/2,k}}{\Delta x} - \frac{H^{n,y}_{j,k+1/2} - H^{n,y}_{j,k-1/2}}{\Delta y}. \tag{3.31}$$

3.3.3 Piecewise Linear Reconstructions For KT-Schemes

In order to completely describe the KT-scheme (3.26), the slope reconstructions in (3.25) and (3.29) are to be determined. It is well known that the unlimited central slope has many disadvantages in regions which contains discontinuities, therefore a nonlinear slope reconstruction is needed. A common approach is to use a nonlinear slope limiter and we recall in this section some of the generally used options. One of our main results here is to develop a new type of adaptive slope limiter which relies on a entropy based smoothness indicator and provides a desirable improvement of the performance of the KT-scheme in non-smooth regions. This technique will also be applied to the polynomial reconstruction of the for the general high order central scheme. Furthermore instead of enforcing the local maximum principle, we will take a different aspect of view on the invariant domain property and briefly introduce the idea of constructing a limiter which is invariant domain satisfied. For simplicity, we only present the reconstruction process in one space dimension. The multidimensional case is completed by splitting the local gradient reconstruction into one dimensional steps.

To start with we are going to review several slope limiters generally used in central schemes. In practice, one judge the performance of these limiters by the properties including the approximation accuracy, the computation complexity, the convergence to an entropy satisfied solution and so on. We will try to briefly discuss these properties for the limiters and therefore introduce two new types of slope limiters: the adaptive MAPR limiter and the invariant domain preserving limiter. In our discussion, we will denote the slope limiter with notation $\sigma_j^a(\boldsymbol{u})$, where a indicates the type of limiter we use.

(i) We start with the unlimited central slope:

$$\sigma_j^{\mathrm{c}}(\boldsymbol{u}) = \frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_{j-1}}{2\Delta x}. \tag{3.32}$$

The unlimited central slope enjoys a linear complexity to construct and is able to recover a fully second order accuracy in the regions where $\boldsymbol{u}$ is smooth. However, applying the central slope will also create undesirable oscillations at discontinuities as shocks and violation of the local maximum principle, which may lead to a failure in producing the numerical solution.

(ii) We consider a classical slope reconstructions based on the so-called minmod limiter which is given by

$$\sigma_j^{\mathrm{m}}(\boldsymbol{u}) = (\boldsymbol{u}_x)_j := \mathrm{m}(\frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_j}{\Delta x}, \frac{\boldsymbol{u}_j - \boldsymbol{u}_{j-1}}{\Delta x}), \tag{3.33}$$

where the minmod operator is defined as follows

$$\mathrm{m}(x_1, x_2, \dots, x_n) = \begin{cases} \min_{1\le j\le n} x_j, & \text{if } x_j > 0\ \forall j, \\ \max_{1\le j\le n} x_j, & \text{if } x_j < 0\ \forall j, \\ 0, & \text{otherwise.} \end{cases} \tag{3.34}$$

The minmod limiter is a total variation diminishing(TVD) reconstruction, see [28, 54, 60], which could efficiently reduce the oscillation at shocks. Also, the minmod limiter can provide enough artificial viscosity that guarantees the numerical solution converges stably to the unique entropy satisfied solution for composite waves, see the KPP-test in [43]. On the other hand, the minmod limiter tends to be more diffusive and less accurate than other limiters. For example, the slope computed via minmod limiter may diminish at local extrema, thus leads to the so-called clipping phenomenon, see § 5.11. Similar phenomenon are also observed at contact discontinuities in the Mach 3 test, see § 5.12.

(iii) We introduce a θ-dependent family of minmod-type limiters with $1 \leq \theta \leq 2$, which is stated as

$$\sigma_j^{\mathrm{m},\theta}(\boldsymbol{u}) = (\boldsymbol{u}_x)_j := \mathrm{m}\big(\theta\frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_j}{\Delta x}, \frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_{j-1}}{2\Delta x}, \theta\frac{\boldsymbol{u}_j - \boldsymbol{u}_{j-1}}{\Delta x}\big), \tag{3.35}$$

see for example [36, p.1900]. The range of θ ($1 \leq \theta \leq 2$) guarantees a local maximum principle of KT-scheme for scalar equations, which is proved in [41, Cor. 5.1]. However, when computing composite waves, the minmod-θ limiter may produce a stable wrong shock caused by the overshot, which leads to an wrong solution, see for example the KPP-test in [43] where the minmod-2 limiter(i.e., $\theta = 2$ in (3.35)) is applied. The clipping phenomenon is also observed for minmod-θ limiter in the test of isentropic vortex, see § 5.11.

(iv) There are many other types of second order reconstructions, most notably the so called Uniformly Non-Oscillatory reconstruction(UNO), introduced by [29], are defined as follows

$$\sigma_j^{\mathrm{u}}(\boldsymbol{u}) = \mathrm{m}(\Delta\boldsymbol{u}_{j-1} + \frac{1}{2}\mathrm{m}(\Delta^2\boldsymbol{u}_{j-1}, \Delta^2\boldsymbol{u}_j), \Delta\boldsymbol{u}_j - \frac{1}{2}\mathrm{m}(\Delta^2\boldsymbol{u}_j, \Delta^2\boldsymbol{u}_{j+1})), \tag{3.36}$$

where $\Delta\boldsymbol{u}_j = \boldsymbol{u}_{j+1} - \boldsymbol{u}_j$, $\Delta^2\boldsymbol{u}_j = \boldsymbol{u}_{j+1} - 2\boldsymbol{u}_j + \boldsymbol{u}_j$. In general, the UNO-limiter could provide more accuracy than the classical minmod limiter as well as efficiently reducing the oscillation at shocks. However, the UNO limiter is definitely of a high computation complexity.

At last We refer the reader to [43] for more nonlinear slope reconstructions, and when is appropriate to apply them.

In order to avoid the clipping phenomenon, another gradient reconstruction, appropriate for unstructured meshes, was introduced in [8]. It is constructed using the so-called minimum-angle plane reconstruction(MAPR) as follows

$$\sigma_j^{\text{mapr},\theta}(\boldsymbol{u}) = (\boldsymbol{u}_x)_j := \text{mapr}(\theta_j \frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_j}{\Delta x}, \frac{\boldsymbol{u}_{j+1} - \boldsymbol{u}_{j-1}}{2\Delta x}, \theta_j \frac{\boldsymbol{u}_j - \boldsymbol{u}_{j-1}}{\Delta x}), \tag{3.37}$$

where

$$\text{mapr}(x_1, x_2, \ldots, x_n) = \{x_i \,|\, \text{where } |x_i| = \min_{1 \le j \le n} |x_j|\} \tag{3.38}$$

is the MAPR operator and $1 \le \theta_j$ is a number specified by the user. By taking $\theta_j = 1$ one could recover the MAPR reconstruction from [8] and the generalized MAPR ($1 \le \theta_j \le 4$) is a natural analog of the minmod-θ limiter. Apparently our purpose is to apply larger θ_j in smooth regions and smaller θ_j in non-smooth regions. Therefore, we introduce a variable R_j, which describes the smoothness of the solution and set $\theta_j^n = 2 - R_j^n$ in (3.37). Here we take R_j^n to be the entropy commutator defined in § 3.5. Simultaneously, in order to enjoy the high accuracy of central slope in smooth regions, we will use θ_j as a threshold to construct a adaptive limiter. That is, we apply the MAPR limiter (3.37) in the regions of non-smooth flow, where non-smooth is defined by $\theta_j^n \le 1.5$. If $\theta_j^n > 1.5$ we choose the central slope, which is defined in (3.32). Note that in the regions of smooth flow the entropy commutator is almost zero, see § 3.5, therefore $\theta_j^n \le 1.5$ could be a good cutoff for the limiter and one could apply other numbers in practice.

Other than constructing a slope limiter via $\boldsymbol{u}_j^n$s directly, we can also define an abstract limiter which guarantees that the second order reconstruction does not violate the local invariant domain property. In the scalar case this property coincides with the well known local maximum principle and the minmod/minmod-θ limiter both satisfy this property. Note that, however, for nonlinear systems the invariant domain depends on the characteristic structure of the systems, as the Riemann invariant for P-system and the minimum principle of specific entropy for the Euler system, see § 2.3, therefore, the minmod/minmod-θ are not an appropriate choice anymore. We call such a limiter an invariant domain preserving limiter, denoted with σ_j^{inv}. Namely σ_j^{inv} is defined such that the interface values $\boldsymbol{u}_{j+1/2}^{n,-} := \boldsymbol{u}_j^n + \frac{\sigma_j^{\text{inv}}\Delta x}{2}$ and $\boldsymbol{u}_{j-1/2}^{n,+} := \boldsymbol{u}_j^n - \frac{\sigma_j^{\text{inv}}\Delta x}{2}$ in (3.26) remain in the local invariant sets defined by the user. More precisely, we have the following definition for the local abstract limitation.

Definition 3.3.1. Let $A_{j-1/2}$ be an invariant set of (1.27a) which contains the states $\boldsymbol{u}_{j-1}^n$ and $\boldsymbol{u}_j^n$ and $A_{j+1/2}$ be an invariant set of (1.27a) containing $\boldsymbol{u}_j^n$ and $\boldsymbol{u}_{j+1}^n$. Then the invariant slope σ_j^{inv} corresponding to the invariant sets $A_{j-1/2}$ and $A_{j+1/2}$ is defined as $\sigma_j^{\text{inv}} = \ell\frac{\boldsymbol{u}_{j+1}-\boldsymbol{u}_{j-1}}{2\Delta x}$ where ℓ is the largest number in $[0,1]$ such that $\boldsymbol{u}_{j-1/2}^{n,+} \in A_{j-1/2}$ and $\boldsymbol{u}_{j+1/2}^{n,-} \in A_{j+1/2}$.

The exact computation of such limiter will be given later in § 4.3, see Algorithm 2 while for the time being we still stay in this abstract setting. The key difference between minmod-type slope limiting (or any other classical limiting) and the invariant domain slope limiting is that the former one intends to impose a local maximum principle (or reduce oscillations in the physical space) while the invariant domain slope limiting only limits the slopes so that the invariant domain property is restricted to be in the phase space at the interfaces.

As a conclusion of the discussion above, we will present two new types of local slope reconstructions that we will apply for all our numerical tests.

(i) The first type of reconstruction uses the adaptive slope limiter combined with the unlimited central slope (3.32) and the MAPR slope limiter (3.37). That is, we apply the MAPR limiter for $\theta_j^n \leq 1.5$ and the central slope for $\theta_j^n > 1.5$

$$\sigma_j^{\mathrm{a}}(\boldsymbol{u}) = \begin{cases} \sigma_j^{\mathrm{mapr},\theta}(\boldsymbol{u}) & \text{if } \theta_j^n \leq 1.5, \\ \sigma_j^{\mathrm{c}}(\boldsymbol{u}) & \text{if } \theta_j^n > 1.5, \end{cases} \tag{3.39}$$

where "a" refers to adaptive. The value of θ is computed via $\theta_j^n = 2 - R_j^n$ where R_j^n is the entropy based smoothness indicator defined in § 3.5. The invariant domain preserving property of this slope reconstruction is guaranteed by applying the convex flux limiting, see Algorithm 1. In the rest of this dissertation We refer to this method as the MAPR-EV-CL method, where "EV" refers the entropy viscosity commutator used to determine the number θ_j^n and "CL" refers to the convex flux limiting process used to impose the invariant domain preserving property.

(ii) The second type of reconstruction uses the invariant domain preserving limiter defined by Definition 3.3.1, which is guaranteed to be invariant domain preserving under a standard CFL condition, see discussions in Theorem 4.3.1 and Theorem 4.3.4; In the rest of this dissertation we refer to this method as the SO-INV-CL method, where SO refers to second order, INV refers to the invariant domain preserving property and CL refers to the convex slope limiting process used to generate this invariant domain preserving slope limiter.

Consequently, any other method we used for simulations will be shown as a standard comparison to these two schemes listed above in all our numerical tests.

3.4 General High Order Central Scheme

The second order KT-scheme (3.26) could be naturally generalized to higher order cases by replacing the local linear reconstruction with polynomials of a higher degree. These polynomials are determined by the values of $\boldsymbol{u}$ on a nodal group that relate to the point of interest, which we call the stencils. Therefore the degree of the polynomial relies on the size of the stencils. For example, a polynomial defined on a stencil of three grid points is of degree two and the corresponding central scheme has a third order accuracy. However, a higher order reconstructions may be dangerous at discontinuities since it may cause considerable oscillations. In this section we first introduce the fully discrete form for the general high order central scheme. Then we will give a brief review of one classical kind of non-oscillatory reconstruction technique, i.e., the Compact WENO reconstruction. Also, we will introduce two novel kinds of polynomial reconstructions at the end of this section.

3.4.1 One Dimensional Case

Here we give the general form of the fully discrete form of the $n-$th order central scheme in one dimensional case. Let us consider the hyperbolic system (3.6) of one space dimension. Without loss of generality, we still take same settings for first order GMS-scheme and second order KT-scheme, i.e., we take uniform cell $I_j = [x_{j-1/2}, x_{j+1/2}]$ with cell centers to be $x_j = j\Delta x$ and set $t^n = n\Delta t$ for time discretization. Then we assume the approximate solution $\widetilde{\boldsymbol{u}}^n = \boldsymbol{u}(x, t^n)$ at time $t = t^n$ to be a piecewise polynomial

$$\widetilde{\boldsymbol{u}}^n := \sum_j p_j(x)\mathbf{1}_{[x_{j-1/2},x_{j+1/2}]}, \qquad x_{j\pm1/2} := x_j \pm \frac{\Delta x}{2}. \tag{3.40}$$

where $p_j(x)$ are a collection of degree $(n-1)$ polynomials which are built such that

$$p_j(x) = \boldsymbol{u}(x) + \mathcal{O}(\Delta x^n), \qquad \forall x \in I_j, \tag{3.41}$$

therefore the order of the approximation accuracy of $\widetilde{\boldsymbol{u}}$ is n. Also we require that the approximation is conservative, i.e., $p_j(x)$ should satisfy

$$\int_{I_j} p_j(x)dx = \boldsymbol{u}_j^n, \qquad \forall j \in \mathbb{Z} \tag{3.42}$$

Now we assume that $p_j(x)$ is of the following form

$$p_j(x) := \widehat{\boldsymbol{u}}_j^n + \sum_{i=1}^{n-1} \frac{1}{i!} \boldsymbol{u}_j^{(i)} (x - x_j)^i. \tag{3.43}$$

By (3.42) we have

$$\widehat{\boldsymbol{u}}_j^n = \boldsymbol{u}_j^n - \sum_{i=1}^{n-1} \frac{(1 - (-1)^{i+1})\Delta x^i}{(i+1)! 2^{i+1}} \boldsymbol{u}_j^{(i)}. \tag{3.44}$$

Remark. In equation (3.43), $\boldsymbol{u}_j^{(i)}$ denotes the $i-$th derivative of $\boldsymbol{u}(x)$ at x_j, which is computed with the values of $\boldsymbol{u}^n$ on local stencils. We could obtain different local reconstructions by taking different $\boldsymbol{u}_j^{(i)}$, such as the to be explained CTO-WENO reconstruction.

Remark. In this dissertation we only consider the following two case, i.e.,

$$p_j(x) = \widehat{\boldsymbol{u}}_j^n + \boldsymbol{u}_j'(x - x_j) + \frac{1}{2}\boldsymbol{u}_j''(x - x_j)^2 \qquad \text{for } n = 3, \tag{3.45}$$

and

$$p_j(x) = \widehat{\boldsymbol{u}}_j^n + \boldsymbol{u}_j'(x - x_j) + \frac{1}{2}\boldsymbol{u}_j''(x - x_j)^2 + \frac{1}{6}\boldsymbol{u}_j'''(x - x_j)^3 \qquad \text{for } n = 4. \tag{3.46}$$

The corresponding $\widehat{\boldsymbol{u}}_j^n$ defined by (3.44) are

$$\widehat{\boldsymbol{u}}_j^n = \boldsymbol{u}_j^n - \frac{\Delta x^2}{24}\boldsymbol{u}_j'' \tag{3.47}$$

for both $n = 3$ and 4.

Now we give the fully discrete form of the $n-$th order central scheme as follows, which is similar to the second order case (3.27)-(3.28). That is, we take the $n-$th order numerical flux to be

$$H_{j+1/2}^n := \frac{\boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,+}) + \boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,-})}{2} - \frac{\lambda_{j+1/2}^n}{2}(\boldsymbol{u}_{j+1/2}^{n,+} - \boldsymbol{u}_{j+1/2}^{n,-}). \tag{3.48}$$

and the fully discrete scheme is given by

$$\frac{\boldsymbol{u}_j^{H,n+1} - \boldsymbol{u}_j^n}{\Delta t} = -\frac{H_{j+1/2}^n - H_{j-1/2}^n}{\Delta x}. \tag{3.49}$$

The only difference to (3.27)-(3.28) is that the interface values we used in (3.48) are computed with

$$\boldsymbol{u}_{j+1/2}^{n,+} = p_{j+1}(x_{j+1/2}), \qquad \boldsymbol{u}_{j+1/2}^{n,-} = p_j(x_{j+1/2}). \tag{3.50}$$

Remark. The local speed of propagation to construct the numerical flux is still computed via $\lambda_{j+1/2}^n = \lambda_{\max}(\boldsymbol{u}_{j+1/2}^-, \boldsymbol{u}_{j+1/2}^+, \boldsymbol{f})$, and the upper bound in § 2.3 will be used instead of the exact evaluation in practice. To advance in time, a SSP method of at least order three will be applied and we use the SSP-RK3 (3.15) for $n = 3$ in all our numerical experiments. For $n = 4$, we will apply the SSP-RK4 scheme from [18], which is shown as follows. We consider a system of ODEs

$$\boldsymbol{u}_t = \boldsymbol{L}(\boldsymbol{u}), \tag{3.51}$$

an optimal fourth order total variation diminishing(TVD) Runge-Kutta scheme is given by

$$\begin{aligned}
\boldsymbol{u}^{(1)} =& \boldsymbol{u}^n + 0.391752226571890\Delta \boldsymbol{L}(\boldsymbol{u}^n),\\
\boldsymbol{u}^{(2)} =& 0.444370493651235\boldsymbol{u}^n + 0.555629506348765\boldsymbol{u}^{(1)}\\
&+ 0.368410593050371\Delta \boldsymbol{L}(\boldsymbol{u}^{(1)}),\\
\boldsymbol{u}^{(3)} =& 0.620101851488403\boldsymbol{u}^n + 0.379898148511597\boldsymbol{u}^{(2)}\\
&+ 0.251891774271694\Delta \boldsymbol{L}(\boldsymbol{u}^{(2)}),\\
\boldsymbol{u}^{(4)} =& 0.178079954393132\boldsymbol{u}^n + 0.0.821920045606868\boldsymbol{u}^{(3)}\\
&+ 0.544974750228521\Delta \boldsymbol{L}(\boldsymbol{u}^{(3)}),\\
\boldsymbol{u}^{(4)} =& 0.517231671970585\boldsymbol{u}^{(2)}\\
&+ 0.096059710526147\boldsymbol{u}^{(3)} + 0.063692468666290\Delta \boldsymbol{L}(\boldsymbol{u}^{(3)})\\
&+ 0.386708617503269\boldsymbol{u}^{(4)} + 0.226007483236906\Delta \boldsymbol{L}(\boldsymbol{u}^{(4)}).
\end{aligned} \tag{3.52}$$

Like the SSP-RK3 scheme, the SSP-RK4 scheme is also an invariant domain preserving update in time. Furthermore, The CFL condition is determined by the corresponding GMS-scheme in order to apply the convex limiting process.

3.4.2 Two Dimensional Case

Now we consider the case of two space dimensions, which is described in (3.16). Still, we use uniform rectangular mesh with cell centers $(x_j, y_k) := (j\Delta x, k\Delta y)$, $j, k \in \mathbb{Z}$, and take the approximate solution $\widetilde{\boldsymbol{u}}^n(x, y) \approx \boldsymbol{u}(x, y, t^n)$ at time t^n to be a piecewise polynomial function.

$$\widetilde{\boldsymbol{u}}^n(x, y) := \sum_{j,k} p_{j,k}(x, y)\mathbf{1}_{[x_{j-1/2}, x_{j+1/2}]\times[y_{k-1/2}, y_{k+1/2}]}, \tag{3.53}$$

where $p_{j,k}(x, y)$ are degree $(k-1)-$ polynomial, which satisfies

$$\int_{[x_{j-1/2}, x_{j+1/2}]\times[y_{k-1/2}, y_{k+1/2}]} p_{j,k}(x, y)dxdy = \boldsymbol{u}^n_{j,k}, \qquad \forall j, k \in \mathbb{Z}. \tag{3.54}$$

Now we assume that $p_{j,k}(x, y)$ is of the form

$$p_{j,k}(x, y) := \widehat{\boldsymbol{u}}^n_{j,k} + \sum_{l=1}^{k-1}\sum_{i=0}^{l} \frac{1}{i!(l-i)!}\boldsymbol{u}^{(i,l-i)}(x - x_j)^i(y - y_k)^{l-i}, \tag{3.55}$$

where $\boldsymbol{u}^{(i,l-i)}_{j,k} := \frac{\partial^l \boldsymbol{u}}{\partial x^i \partial y^{l-i}}(x_j, y_k)$ is the partial derivative of $\boldsymbol{u}$ at (x_j, y_k). By (3.54) we have that

$$\widehat{\boldsymbol{u}}^n_{j,k} = \boldsymbol{u}^n_{j,k} - \sum_{l=1}^{n-1}\sum_{i=0}^{l} \frac{(1-(-1)^{i+1})(1-(-1)^{l-i+1})\Delta x^i \Delta y^{l-i}}{(i+1)!(l-i+1)!2^{l+2}}\boldsymbol{u}^{(i,l-i)}_{j,k}. \tag{3.56}$$

Remark. In this dissertation we only consider the following two cases, i.e.,

$$\begin{aligned} p_{j,k}(x,y) =& \widehat{\boldsymbol{u}}^n_{j,k} + (\boldsymbol{u}^n_x)_{j,k}(x-x_j) + (\boldsymbol{u}^n_y)_{j,k}(y-y_k) \\ &+ \frac{1}{2}(\boldsymbol{u}^n_{xx})_{j,k}(x-x_j)^2 + \frac{1}{2}(\boldsymbol{u}^n_{yy})_{j,k}(y-y_k)^2, \\ &+ (\boldsymbol{u}^n_{xy})_{j,k}(x-x_j)(y-y_k), \end{aligned} \tag{3.57}$$

where $n = 3$ and

$$\begin{aligned} p_{j,k}(x,y) =& \widehat{\boldsymbol{u}}^n_{j,k} + (\boldsymbol{u}^n_x)_{j,k}(x-x_j) + (\boldsymbol{u}^n_y)_{j,k}(y-y_k) \\ &+ \frac{1}{2}(\boldsymbol{u}^n_{xx})_{j,k}(x-x_j)^2 + \frac{1}{2}(\boldsymbol{u}^n_{yy})_{j,k}(y-y_k)^2, \\ &+ (\boldsymbol{u}^n_{xy})_{j,k}(x-x_j)(y-y_k) \\ &+ \frac{1}{6}(\boldsymbol{u}^n_{xxx})_{j,k}(x-x_j)^3 + \frac{1}{6}(\boldsymbol{u}^n_{yyy})_{j,k}(y-y_k)^3 \\ &+ \frac{1}{2}(\boldsymbol{u}^n_{xxy})_{j,k}(x-x_j)^2(y-y_k) + \frac{1}{2}(\boldsymbol{u}^n_{xyy})_{j,k}(x-x_j)(y-y_k)^2, \end{aligned} \tag{3.58}$$

where $n = 4$. The corresponding $\widehat{\boldsymbol{u}}^n_j$ defined by (3.56) are

$$\widehat{\boldsymbol{u}}^n_{j,k} = \boldsymbol{u}^n_{j,k} - \frac{1}{24}(\Delta x^2(\boldsymbol{u}^n_{xx})_{j,k} + \Delta y^2(\boldsymbol{u}^n_{yy})_{j,k}) \tag{3.59}$$

for both $n = 3$ and 4.

Finally we give fully discrete form of the semi-discrete form of the $n-$th order central scheme in two space dimensions, which is similar to the case of KT-scheme.

$$\frac{\boldsymbol{u}^{H,n+1}_{j,k} - \boldsymbol{u}^n_{j,k}}{\Delta t} = -\frac{H^{n,x}_{j+1/2,k} - H^{n,x}_{j-1/2,k}}{\Delta x} - \frac{H^{n,y}_{j,k+1/2} - H^{n,y}_{j,k-1/2}}{\Delta y}. \tag{3.60}$$

The numerical fluxes are defined as, see [40]:

$$\begin{aligned} H^{n,x}_{j+1/2,k} :=& \frac{1}{12}\big(\boldsymbol{f}(\boldsymbol{u}^{NW}_{j+1,k}) + \boldsymbol{f}(\boldsymbol{u}^{NE}_{j,k}) \\ &+ 4(\boldsymbol{f}(\boldsymbol{u}^{CW}_{j+1,k}) + \boldsymbol{f}(\boldsymbol{u}^{CE}_{j,k})) + \boldsymbol{f}(\boldsymbol{u}^{SW}_{j+1,k}) + \boldsymbol{f}(\boldsymbol{u}^{SE}_{j,k})\big) \end{aligned} \tag{3.61}$$

$$- \quad \frac{\lambda^{n,x}_{j+1/2,k}}{12}\big(\boldsymbol{u}^{NW}_{j+1,k} - \boldsymbol{u}^{NE}_{j,k} + 4(\boldsymbol{u}^{CW}_{j+1,k} - \boldsymbol{u}^{CE}_{j,k}) + \boldsymbol{u}^{SW}_{j+1,k} - \boldsymbol{u}^{SE}_{j,k}\big),$$

$$\begin{aligned} H^{n,y}_{j,k+1/2} :=& \frac{1}{12}\big(\boldsymbol{g}(\boldsymbol{u}^{SW}_{j,k+1}) + \boldsymbol{g}(\boldsymbol{u}^{NW}_{j,k}) \\ &+ 4(\boldsymbol{g}(\boldsymbol{u}^{CS}_{j,k+1}) + \boldsymbol{g}(\boldsymbol{u}^{CN}_{j,k})) + \boldsymbol{g}(\boldsymbol{u}^{SE}_{j,k+1}) + \boldsymbol{g}(\boldsymbol{u}^{NE}_{j,k})\big) \\ -& \quad \frac{\lambda^{n,y}_{j,k+1/2}}{12}\big(\boldsymbol{u}^{SW}_{j,k+1} - \boldsymbol{u}^{NW}_{j,k} + 4(\boldsymbol{u}^{CS}_{j,k+1} - \boldsymbol{u}^{CN}_{j,k}) + \boldsymbol{u}^{SE}_{j,k+1} - \boldsymbol{u}^{NE}_{j,k}\big), \end{aligned} \tag{3.62}$$

where $\boldsymbol{u}_{j,k} := p_{j,k}(x_j, y_k)$ and

$$\begin{aligned} &\boldsymbol{u}^{NC}_{j,k} := p_{j,k}(x_j, y_{k+1/2}), \quad \boldsymbol{u}^{SC}_{j,k} := p_{j,k}(x_j, y_{k-1/2}), \\ &\boldsymbol{u}^{CE}_{j,k} := p_{j,k}(x_{j+1/2}, y_k), \quad \boldsymbol{u}^{CW}_{j,k} := p_{j,k}(x_{j-1/2}, y_k), \\ &\boldsymbol{u}^{NE}_{j,k} := p_{j,k}(x_{j+1/2}, y_{k+1/2}), \quad \boldsymbol{u}^{NW}_{j,k} := p_{j,k}(x_{j-1/2}, y_{k+1/2}), \\ &\boldsymbol{u}^{SE}_{j,k} := p_{j,k}(x_{j+1/2}, y_{k-1/2}), \quad \boldsymbol{u}^{SW}_{j,k} := p_{j,k}(x_{j-1/2}, y_{k-1/2}), \end{aligned} \tag{3.63}$$

are the interface values with

$$\begin{aligned} \lambda^{n,x}_{j+1/2,k} &= \max(\lambda^{N}_{j+1/2,k}, \lambda^{C}_{j+1/2,k}, \lambda^{S}_{j+1/2,k}), \\ \lambda^{n,y}_{j,k+1/2} &= \max(\lambda^{E}_{j,k+1/2}, \lambda^{C}_{j,k+1/2}, \lambda^{W}_{j,k+1/2}), \end{aligned} \tag{3.64}$$

where $\lambda^{X}_{j+1/2,k} := \lambda_{\max}(\boldsymbol{u}^{XE}_{j,k}, \boldsymbol{u}^{XW}_{j+1,k}, \boldsymbol{f})$ and $\lambda^{Y}_{j,k+1/2} := \lambda_{\max}(\boldsymbol{u}^{NY}_{j,k}, \boldsymbol{u}^{SY}_{j,k+1}, \boldsymbol{g})$ for $X \in \{S, C, N\}$ and $Y \in \{W, C, E\}$ separately. SSP-RK3 and SSP-RK4 schemes will be applied for the forward Euler step and the CFL-condition is determined by the first order GMS-scheme.

3.4.3 Piecewise Polynomial Reconstruction For High Order Central Scheme

It is well known that by using the polynomial reconstruction in central scheme we can obtain a numerical solution of arbitrary high order accuracy. However, this approach may also produce oscillations at discontinuities as shocks. In this section we first introduce a classical type of polynomial reconstruction which is naturally non-oscillatory, i.e., the WENO reconstruction. Then similar to what we did for the KT-scheme, we explain two novel types of polynomial reconstruction which will also reduce the oscillation efficiently. For simplicity, our discussion will be restricted to one dimensional case.

We start by introducing the essentially non-oscillatory schemes(ENO-scheme), which, first introduced in [30], generalizes the Godunov type scheme and the second-order accurate MUSCL reconstruction by designing an essentially non-oscillatory piecewise polynomial reconstruction of the solution from its cell averages. The ENO-scheme successfully increases the order of accuracy of numerical solutions for hyperbolic conservation laws which are possibly piecewise smooth with large jump discontinuities by approximating the solution to a high degree of accuracy in smooth parts, while avoiding Gibbs oscillations near the discontinuities. Following [1, Chapter 5,Chapter 6], the main idea of the ENO reconstruction is shown as follows. We consider the hyperbolic system (1.3). Taking the same space discretization as in § 3.2 and § 3.3, i.e., we take $x_j = j\Delta x$ and take the non-staggered control mesh to be $[x_{j-1/2}, x_{j+1/2}]$. In order to obtain a $n-$th order accuracy reconstruction of $\boldsymbol{u}$, we take $p_j(x)$ to be a degree $n-1$ polynomial defined by (3.43)-(3.44). Simultaneously, the reconstruction is also required to be as non-oscillatory as possible. Technically the properties of accuracy (3.41) and the conservation property (3.42) will be automatically satisfied if p_j interpolates $\boldsymbol{u}_i$ over the adaptive stencil of grid points:

$$S_l := \{j-n+l, \cdots, j-1+l\}, \tag{3.65}$$

where $1 \leq l \leq n$ is a certain number such that the original function $\boldsymbol{u}$ is smooth in S_l. In other words, p_j is an order $(n-1)$ polynomial which is defined such that the following conservation property holds

$$\int_{I_i} p_j(x)dx = \boldsymbol{u}_i^n, \qquad \text{for } i = j-n+l, \cdots, j-1+l, \tag{3.66}$$

for a certain stencil index $l = l_i \in \{1, \cdots, n\}$, where $\boldsymbol{u}_i^n$ is the cell average of $\boldsymbol{u}$ on I_i at time t^n.

Remark. It has been proved in [30] that if the assumption that $\boldsymbol{u}$ is smooth in a stencil S_l with a fixed l holds, then there exits unique order $(n-1)$ polynomial $p_j^{(l)}$ which satisfies (3.66).

In order to find such $p_j^{(l)}$, two approaches are introduced in [30]. The first option is called the reconstruction via primitive function approach(RP), for which we define the following primitive function of $\boldsymbol{u}(x)$:

$$\boldsymbol{U}(x) = \int_{-\infty}^{x} \boldsymbol{u}(x) dx. \tag{3.67}$$

Thus the point values of the primitive function $\boldsymbol{U}(x)$ at cell interface $x_{j+1/2}$ could be evaluated as

$$\boldsymbol{U}_{j+1/2} := \boldsymbol{U}(x_{j+1/2}) = \sum_{i \leq j} \Delta x \boldsymbol{u}_i^n, \tag{3.68}$$

where $\boldsymbol{u}_i^n$ denotes the cell average. If we take P_j to be the unique $n-$th order polynomial which interpolates the interface values of $\boldsymbol{U}$, which is defined by (3.68), over the stencil consisted of points $\{x_{l-1/2}, \cdots, x_{l+n-1/2}\}$, then the order $(n-1)$ polynomial $p_j(x) = \frac{d}{dx} P_j(x)$ will uniquely satisfy (3.66). Another option is called the reconstruction via deconvolution approach(RD), which consider (3.42) as a convolution of $\boldsymbol{u}(x)$ with the indicator function over I_j. By computing the Taylor expansion of $\boldsymbol{u}(x)$ and substituting the result in (3.42), one could obtain an upper triangular linear system for computing p_j. By solving the linear system we could obtain the unique interpolation of p_j. One could read [30] for more discussions on ENO-schemes.

From the discussion above, we notice that the purpose of ENO-reconstruction is to avoid including the discontinuous cells in the stencil. The WENO-scheme, which is first introduced in [35, 52], is an extension of the ENO-scheme and enjoys the advantages of high order accuracy by obtaining more smoothness in the numerical fluxes. The main idea of WENO reconstruction is as follows, see [35, 52, 1]: instead of using only one of the candidate stencils to form the reconstruction, one uses a convex combination of all n stencils $S_1, \cdots, S_n$, where $S_l := \{I_{j-n+l}, \cdots, I_{j-1+l}\}$ with $l \in \{1, \cdots, n\}$. If we denote S to be a big stencil combined with all these stencils, then a reconstruction polynomial $p(x)$ of degree at most $n+1$ could be obtained, which satisfy

$$p(x_{j+1/2}) = \boldsymbol{u}(x_{j+1/2}) + \mathcal{O}(\Delta x^{n+1}) \tag{3.69}$$

and

$$\int_{I_j} p(x)dx = \boldsymbol{u}_j^n \tag{3.70}$$

in S. An explicit form of p is given by

$$p(x) := \sum_{l=1}^{n} w_l p_l(x). \tag{3.71}$$

The coefficient are the weights of the reconstruction, which lead to the term "WENO". If the function $\boldsymbol{u}(x)$ is smooth all over S, we have that w_i are all constant, and thus be called as linear weights. If $\boldsymbol{u}(x)$ is not smooth on S, we use the so-called non-linear weights to adaptively avoid including the discontinuous cell in the stencil. In general, the non-linear weight w_i is determined by a smoothness indicator IS_l, which measures the relative smoothness of the $\boldsymbol{u}(x)$ in stencil S_l. In each cell I_j, IS_l is defined as follows

$$IS_l = \sum_{i=1}^{k} \int_{x_{j-1/2}}^{x_{j+1/2}} \Delta x^{2i-1} (\frac{d^i}{dx^i} p_j(x))^2 dx, \qquad l \in \{1, \cdots, n\}. \tag{3.72}$$

A large IS_l indicates that the function $\boldsymbol{u}(x)$ is less smooth in the stencil S_l. For the computation of IS_l one could read [35] for reference.

In our work, we will adopt the compact third order WENO-scheme(CTO-WENO) introduced in [49]. The main idea is to introduce one quadratic polynomial and two linear reconstructions for the interpolants and the weights of the convex combination are set specifically as to obtain a third-order accuracy in smooth regions. In regions with discontinuities or large gradients, the smoothness indicator will allow the weights to change automatically such that the WENO reconstruction will switch to a second-order linear reconstruction. In the reconstruction, only the compact stencils will be used.

Remark. In order to avoid division by zero, a parameter ϵ will be introduced to the smoothness indicator, see (3.82). It is proved in [38] that ϵ should be chosen as $k\Delta x^q$, which is proportional to the $q-$th order of the mesh size Δx in order to achieve the optimal order of accuracy, where $q \leq 3$ and $pq \geq 2$ with $p \geq 1$ be the exponent used in computing the smoothness indicator, k is defined to be $||f||^2$ such that the reconstruction is invariant under the scaling of f. In all numerical report of this dissertation, we take $p = 2$ and $q = 2$.

For simplicity, here we only give the CTO-WENO reconstruction in one dimensional case without any derivation. One could read [49, 38] for more technical details. We take the approximation order to be $n = 3$ in (3.71) and use $l \in H := \{L, C, R\}$ instead of $\{1, 2, 3\}$ to be the stencil index, thus we have

$$p_j(x) = \sum_{l \in H} w_l p_l(x), \tag{3.73}$$

where $p_C(x)$ is a quadratic polynomial and $p_L(x), p_R(x)$ are both linear on a single side. One should notice that (3.73) is a convex combination, therefore we have $w_l \geq 0$ for $l \in H$ and $\sum_{l \in H} w_l = 1$. Applying (3.66) on $l = L, R$, it follows that

$$p_L(x) = \boldsymbol{u}_j^n + \frac{\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n}{\Delta x}(x - x_j), \tag{3.74a}$$

$$p_R(x) = \boldsymbol{u}_j^n + \frac{\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n}{\Delta x}(x - x_j). \tag{3.74b}$$

To compute $p_C(x)$, we have to introduce an optimal polynomial

$$p_{\text{opt}}(x) = \sum_{l \in H} c_l p_l(x), \tag{3.75}$$

where c stands for "central". Similar to $p_j(x)$, (3.75) is also a convex combination with $c_l \geq 0$ for $l \in H$ and $\sum_{l\in H} c_l = 1$. Here we say $p_{\text{opt}}(x)$ is optimal in the sense of approximation accuracy that

$$p_{\text{opt}}(x) = \boldsymbol{u}(x) + \mathcal{O}(\Delta x^3), \qquad \forall x \in I_j, \tag{3.76}$$

holds for $\boldsymbol{u}(x)$ sufficiently smooth on I_j. By applying (3.66) to $p_{\text{opt}}(x)$ on $I_{j-1}, I_j, I_{j+1} \in S_C$ we obtain the following parabola:

$$p_{\text{opt}}(x) = \widehat{\boldsymbol{u}}_j^n + (\boldsymbol{u}_x^n)_j(x - x_j) + \frac{1}{2}(\boldsymbol{u}_{xx}^n)_j(x - x_j)^2, \tag{3.77}$$

where

$$\widehat{\boldsymbol{u}}_j^n = \boldsymbol{u}_j^n - \frac{1}{24}(\boldsymbol{u}_{j+1}^n - 2\boldsymbol{u}_j^n + \boldsymbol{u}_{j-1}^n), \tag{3.78a}$$

$$(\boldsymbol{u}_x^n)_j = \frac{\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_{j-1}^n}{2\Delta x}, \tag{3.78b}$$

$$(\boldsymbol{u}_{xx}^n)_j = \frac{\boldsymbol{u}_{j+1}^n - 2\boldsymbol{u}_j^n + \boldsymbol{u}_{j-1}^n}{\Delta x^2}. \tag{3.78c}$$

Remark. We could prove that a polynomial which satisfies (3.66) is uniquely determined by (3.77)-(3.78). Notice that $p_{\text{opt}}(x)$ is also obtained by taking (3.45)-(3.47) with additional restrictions on $(\boldsymbol{u}_x^n)_j$ and $(\boldsymbol{u}_{xx}^n)_j$. Finally, one could prove that (3.76) holds for $p_{\text{opt}}(x)$.

Now by setting $c_L = c_R = \frac{1}{4}$ and $c_C = \frac{1}{2}$, it follows from (3.75),(3.77) and (3.78) that

$$\begin{aligned} p_C(x) =& \boldsymbol{u}_j^n - \frac{1}{12}(\boldsymbol{u}_{j+1}^n - 2\boldsymbol{u}_j^n + \boldsymbol{u}_{j-1}^n) \\ &+ \frac{\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_{j-1}^n}{2\Delta x}(x - x_j) + \frac{\boldsymbol{u}_{j+1}^n - 2\boldsymbol{u}_j^n + \boldsymbol{u}_{j-1}^n}{\Delta x^2}(x - x_j)^2. \end{aligned} \tag{3.79}$$

In order to complete $p(x)$ for arbitrary $\boldsymbol{u}$, i.e., there is no assumption of smoothness, we need to compute coefficients w_l for $l \in H = \{L, C, R\}$. Following [49, 38], we first compute the smoothness indicator IS_l defined in (3.72) to be

$$IS_L = (\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n)^2, \tag{3.80a}$$

$$IS_C = \frac{13}{12c_C^2}(\boldsymbol{u}_{j+1}^n - 2\boldsymbol{u}_j^n + \boldsymbol{u}_{j-1}^n)^2 + \frac{1}{4}(\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_{j-1}^n)^2, \tag{3.80b}$$

$$IS_R = (\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n)^2. \tag{3.80c}$$

and then define

$$w_l = \frac{a_l}{\sum_{k\in H} a_k}, \tag{3.81}$$

where

$$a_l := \frac{c_l}{(\epsilon(\Delta x) + IS_l)^p}, \qquad \text{for } l \in H = \{L, C, R\}. \tag{3.82}$$

As discussed before, for the optimal accuracy in approximation, we take $\epsilon(\Delta x) = k\Delta x^q$ with $q = 2, p = 2$ and $k = ||f||^2$.

Remark. Let $p(x)$ be defined by (3.73),(3.74),(3.79),(3.80) and (3.81). It is proved that

(i) If $\boldsymbol{u}(x)$ is smooth in S_C, then we have

$$\boldsymbol{u}(x) - p_j(x) = \mathcal{O}(\Delta x^3) \tag{3.83}$$

(ii) If $\boldsymbol{u}(x)$ is smooth in S_L or S_R and has a discontinuity in the other stencil, then we have

$$\boldsymbol{u}(x) - p_j(x) = \mathcal{O}(\Delta x^2) \tag{3.84}$$

We just gave a brief review of the CTO-WENO reconstruction, which could be used to build the third order central scheme. The compact WENO reconstruction enjoys a high accuracy with small stencils. Also it can capture the shocks efficiently with the help of the smoothness indicator. However, one could see that even for compact WENO, the computation cost is still very large. Also, it has been observed in [43] that the WENO reconstruction can lead to a wrong stable shock in the KPP-test. Thus we are going to introduce the following two approaches which could be applied to arbitrary polynomials reconstructions. Note that each of these two approaches allow us to get a numerical solution which converges to the unique entropy satisfied solution for composite waves. Still, our discussion will be in one space dimension, the result is easy to generalize to multidimensional cases.

Now let us assume that the underlying polynomial p_j which we use in (3.40) is defined by (3.43), the two approaches is then given as follows.

(i) The first approach is to apply a same kind of threshold θ_j^n used in § 3.3.3, which is used to determine the smoothness of the region. Our new reconstruction is built by using the optimal polynomial if $\theta_j^n \leq 1.5$ and a MAPR based linear reconstruction if $\theta_j^n > 1.5$, which is given as follows.

$$p_j^{\mathrm{a}}(x) = \begin{cases} p_{\mathrm{opt}}(x) & \text{if } \theta_j^n \leq 1.5, \\ p_{\mathrm{mapr}}(x) & \text{if } \theta_j^n > 1.5, \end{cases} \tag{3.85}$$

where "a" refers to adaptive and

$$p_{\mathrm{mapr}}(x) = \boldsymbol{u}_j^n + \sigma_j^{\mathrm{mapr},\theta}(x - x_j) \tag{3.86}$$

with $\sigma_j^{\mathrm{mapr},\theta}$ defined by (3.37). Still, the value of θ is computed via $\theta_j^n = 2 - R_j^n$ where R_j^n is the entropy commutator defined in § 3.5. The invariant domain preserving property is also imposed via a flux limiting process, see Algorithm 1. In the rest of this dissertation We refer to this method as the POL-EV-CL method, where "POL" refers to the polynomial we use, "EV" refers the entropy viscosity commutator and "CL" refers to the convex flux limiting process. In practice, the MAPR slope limiter could be replaced by any second order slope limiters which is able to capture the correct shocks in composite waves.

(ii) Similar to the abstract slope limiter defined in § 3.3.3, we introduce an abstract polynomial which is of third order accuracy and satisfies the invariant domain property. Such polynomial is called the invariant domain preserving polynomial, which is denoted with p_j^{inv}. Similar to the invariant domain preserving slope limiter, p_j^{inv} is defined such that the interface values $\boldsymbol{u}_{j+1/2}^{n,-} := p_j^{\text{inv}}(x_{j+1/2})$ and $\boldsymbol{u}_{j-1/2}^{n,+} := p_j^{\text{inv}}(x_{j-1/2})$ used in (3.26) and the state $\widehat{\boldsymbol{u}}_j^n$ defined by (3.44) all remain in the local invariant sets of the system. For this reason we introduce an auxiliary function p_j^r such that

$$p_j(x) = \boldsymbol{u}_j^n + p_j^r(x) \qquad \text{for all } x \in I_j, \tag{3.87}$$

where "r" refers to reconstruction, then we could give the definition of the abstract polynomial as follows.

Definition 3.4.1. Let $A_{j-1/2}$ be an invariant set of (1.27a) which contains the states $\boldsymbol{u}_{j-1}^n$ and $\boldsymbol{u}_j^n$ and $A_{j+1/2}$ be an invariant set of (1.27a) containing $\boldsymbol{u}_j^n$ and $\boldsymbol{u}_{j+1}^n$. Then the invariant polynomial p_j^{inv} corresponding to the invariant sets $A_{j-1/2}$ and $A_{j+1/2}$ is defined as $p_j^{\text{inv}} = \boldsymbol{u}_j^n(x) + \ell p_j^r(x)$ where ℓ is the largest number in $[0, 1]$ such that $p_j^{\text{inv}}(x_{j-1/2}) \in A_{j-1/2}$ and $p_j^{\text{inv}}(x_{j+1/2}) \in A_{j+1/2}$.

Remark. One should notice that the limited polynomial p_j^{inv} satisfies (3.42), i.e., p_j^{inv} is conservative on I_j and therefore could be used for the local construction for the high order central scheme.

Thus we could give the second approach of reconstruction uses the invariant domain preserving limiter defined with Definition 3.4.1, for which the invariant domain preserving property is guaranteed under a standard CFL condition, see discussions in Theorem 4.3.5 and Theorem 4.3.6; Similar to the second order reconstruction, in the rest of this dissertation we refer to this method as the POL-INV-CL method, where "POL" refers to the polynomial we use, "INV" refers to the invariant domain preserving property and "CL" refers to the convex slope limiting.

At the end of this section let us recall the optimal function defined by (3.77)-(3.78). From the discussion in § 3.4 we know that if we take the local reconstruction to be $p_j(x) :=$

$p_{\text{opt}}(x)$, the we have

$$p_j(x) - \boldsymbol{u}(x) = \mathcal{O}(\Delta x^3), \qquad \forall x \in I_j, \tag{3.88}$$

holds if $\boldsymbol{u}(x)$ are sufficiently smooth on I_j. Also the computation cost of p_{opt} is much cheaper than the compact WENO reconstructions, therefore we will always use $p_{\text{opt}}(x)$ as the underlying polynomial for the two approaches explained above. Therefore we will call the name of these approaches as OPT-EV-CL method and OPT-INV-CL method instead, where the term "OPT" refers to the optimal function we use.

Remark. One should notice that we apply $p_{\text{opt}}(x)$ in our approach due to its low complexity of computation. In fact, our approach is robust to be applied to any polynomial reconstructions such as WENO scheme. The performance of discontinuities will all be improved via the to be explained convex limiting algorithm. For reference, the second order optimal polynomial defined in two dimensional case is shown as follows, see [49].

$$\begin{aligned} p_{\text{opt}}(x,y) =& \widehat{\boldsymbol{u}}^n_{j,k} + (\boldsymbol{u}^n_x)_{j,k}(x-x_j) + (\boldsymbol{u}^n_y)_{j,k}(y-y_k) \\ &+ \frac{1}{2}(\boldsymbol{u}^n_{xx})_{j,k}(x-x_j)^2 + \frac{1}{2}(\boldsymbol{u}^n_{yy})_{j,k}(y-y_k)^2, \\ &+ (\boldsymbol{u}^n_{xy})_{j,k}(x-x_j)(y-y_k), \end{aligned} \tag{3.89}$$

where

$$\widehat{\boldsymbol{u}}^n_{j,k} = \boldsymbol{u}^n_{j,k} - \frac{1}{24}(\Delta x^2(\boldsymbol{u}^n_{xx})_{j,k} + \Delta y^2(\boldsymbol{u}^n_{yy})_{j,k}), \tag{3.90a}$$

$$(\boldsymbol{u}^n_x)_{j,k} = \frac{\boldsymbol{u}^n_{j+1,k} - \boldsymbol{u}^n_{j-1,k}}{2\Delta x}, \quad (\boldsymbol{u}^n_y)_{j,k} = \frac{\boldsymbol{u}^n_{j,k+1} - \boldsymbol{u}^n_{j,k-1}}{2\Delta y}, \tag{3.90b}$$

$$(\boldsymbol{u}^n_{xx})_{j,k} = \frac{\boldsymbol{u}^n_{j+1,k} - 2\boldsymbol{u}^n_{j,k} + \boldsymbol{u}^n_{j-1,k}}{\Delta x^2}, \quad (\boldsymbol{u}^n_{yy})_{j,k} = \frac{\boldsymbol{u}^n_{j,k+1} - 2\boldsymbol{u}^n_{j,k} + \boldsymbol{u}^n_{j,k-1}}{\Delta y^2}, \tag{3.90c}$$

$$(\boldsymbol{u}^n_{xy})_{j,k} = \frac{\boldsymbol{u}^n_{j+1,k+1} - \boldsymbol{u}^n_{j+1,k-1} - \boldsymbol{u}^n_{j-1,k+1} + \boldsymbol{u}^n_{j-1,k-1}}{4\Delta x \Delta y}. \tag{3.90d}$$

3.5 Entropy Based Smoothness Indicator

In § 3.3.3 and § 3.4.3 we introduce a new technique for local reconstruction, that is, we use a more accurate central unlimited slope/optimal polynomial in smooth regions and a nonlinear minmod-type/MAPR limited slope(or equivalently a first order polynomial) in the non-smooth regions. What we want is that the change between the two types of reconstructions should happen when a physical discontinuity forms. Similar to [23], the approach we take to detect a discontinuity is to measure an entropy production. Our objective is to construct a second/third order method that is entropy consistent and at the same time close to satisfying the invariant domain preserving property. Counter-examples of schemes that are invariant domain preserving but entropy violating could be seen in [22, Lemma 3.2, Lemma 4.6, § 6.1] and [21, § 5.1]. However, we do not want to rely totally on the yet to be explained limiting process to enforce the entropy consistency since the limiting process should be considered as a light post-processing applied to a method which is already entropy consistent and almost invariant domain preserving. In [23, 24], a high-order graph viscosity which guarantees the entropy consistency was introduced. However, we do not want the time discretization to interfere with the estimation of the residual, so we follow the entropy viscosity commutator approach proposed in [27]. For simplicity, our discussion of the entropy commutator will be in the one space dimension.

Let $(\eta(\boldsymbol{u}), \boldsymbol{F}(\boldsymbol{u}))$ be the entropy pair of system (1.27a), which is defined by Definition 1.5.3, i.e., η is a convex function of the conserved variables $\boldsymbol{u}$, $\boldsymbol{F}$ is the entropy flux, such that $D\boldsymbol{F}(\boldsymbol{u}) = \eta'(\boldsymbol{u})^\top D\boldsymbol{f}(\boldsymbol{u})$. Following [27, §3.4] we measure the discrepancy in the chain rule as follows

$$\Delta_j^n = \boldsymbol{F}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{F}(\boldsymbol{u}_{j-1}^n) - \eta'(\boldsymbol{u}_j^n)^\top(\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)). \tag{3.91}$$

Now we set

$$C_j^n = |\boldsymbol{F}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{F}(\boldsymbol{u}_{j-1}^n)| + |\eta'(\boldsymbol{u}_j^n)^\top| \cdot |\boldsymbol{f}(\boldsymbol{u}_{j+1}^n) - \boldsymbol{f}(\boldsymbol{u}_{j-1}^n)|, \tag{3.92}$$

to be a normalizing coefficient, where we denote $|\boldsymbol{f}| := \|\boldsymbol{f}\|_{\ell_2}$ for any vector function $\boldsymbol{f}$. Note that C_j^n could be very close to or even equals to zero in smooth regions. Therefore, in order to avoid division by zero, we introduce

$$\alpha_j^n = \max(|\boldsymbol{F}(\boldsymbol{u}_{j+1}^n)|, |\boldsymbol{F}(\boldsymbol{u}_j^n)|, |\boldsymbol{F}(\boldsymbol{u}_{j-1}^n)|), \tag{3.93a}$$

$$\beta_j^n = |\eta'(\boldsymbol{u}_j^n)^\top| \cdot \lambda_j^{\max,n} \cdot (|\boldsymbol{u}_{j+1}^n - \boldsymbol{u}_j^n| + |\boldsymbol{u}_j^n - \boldsymbol{u}_{j-1}^n|), \tag{3.93b}$$

where $\lambda_j^{\max,n} := \max(\lambda_{j+1/2}^n, \lambda_{j-1/2}^n)$ is the global maximum speed of propagation at time t^n and define the normalized entropy viscosity commutator to be

$$R_j^n = \frac{|\Delta_j^n|}{\max(C_j^n, \epsilon\alpha_j^n, \epsilon\beta_j^n)}, \tag{3.94}$$

where ϵ is a small number, which is typically taken as the square root of the machine error in practice. By definition we have that $R_j^n \in (0, 1]$ since $|\Delta_j^n| \leq C_j^n$. Moreover, one could prove that $R_j^n \sim \mathcal{O}(\Delta x)$ in smooth regions and $R_j^n \sim 1$ at shocks by computing the Taylor expansion of $\boldsymbol{F}' = \eta' \cdot \boldsymbol{f}'$. One could read [27, §3.4] for more details on entropy commutators and its applications. Consequently, we could set

$$\theta_j^n = 2 - R_j^n. \tag{3.95}$$

to be the local weights used in the adaptive limiter/polynomial discussed in § 3.3.3 and § 3.4.3.

4. QUASICONCAVITY BASED LIMITING ALGORITHM

In this chapter we introduce two novel techniques which are used to modify an existing high order scheme and make it to be locally invariant domain preserving. These techniques are all based on the so called *convex limiting*, which is first introduced in [27]. Both of these limitations will upgrade the high order scheme with a slightly polishing process and yet create a limited scheme which is invariant domain preserving and numerically preserves the accuracy of the original scheme. The first approach is called convex flux limiting process, see Algorithm 1, which will impose a modification on the numerical flux computed with the adaptive reconstruction σ_j^{a} and p_j^{a}. The second approach is called convex slope/polynomial limiting, which is used to modify the local reconstruction directly and thus build the limited reconstructions σ_j^{inv} and p_j^{inv}, such that the interface values stay in the local invariant set, see Algorithm 2. Also, we've proved in Theorem 3.2.1 and Theorem 3.2.2 that the first order scheme are always invariant domain preserving. Therefore, all these limiting process will rely on the first order central scheme (3.12) and (3.28). For simplicity, the discussion will be focus on one dimensional case while the result for two dimensional case will also be presented without any derivation.

4.1 Invariant Domains via Quasiconcave Constraints

In order to unify into a single framework all the bounds in phase space that we want to enforce onto the high order solutions, the notion of quasiconcavity are to be recalled here.

Definition 4.1.1 (Quasiconcavity)**.** Given a convex set $\mathcal{A} \subset \mathbb{R}^m$, we say that a function $\Psi : \mathcal{A} \to \mathbb{R}$ is quasiconcave if every upper level set of Ψ is convex; that is, the set $L_\lambda(\Psi) := \{\boldsymbol{u} \in \mathcal{A} | \Psi(\boldsymbol{u}) \geq \lambda\}$ is convex in the range of Ψ for any $\lambda \in \mathbb{R}$.

One should notice that concavity implies quasiconcavity. In all problems we consider, we assume that the invariant domain could be described as an intersection of quasiconcave constraints of the form

$$\Psi^z(\boldsymbol{u}) \geq 0, \tag{4.1}$$

where z is the notation of the variable to be constrained. We will enforce such quasiconcave constraints via both the flux limiting and the slope limiting process. Moreover, in practice all quasiconcave constraints will be modified to concave constraints in order to simplify the computation of the limiting process which is much simpler in concave case. Now we describe the quasiconcave constraints for all examples discussed in § 1.3.

4.1.1 Scalar Equations

For scalar equation described in §1.3.1, the local invariant domain is an interval based on the local maximum principle and we enforce it by imposing the following constraints:

$$u_j^{n,\min} := \min(u_j^n, u_{j\pm1}^n, \bar{u}_{j\pm1/2}^{n+1}), \quad u_j^{n,\max} := \max(u_j^n, u_{j\pm1}^n, \bar{u}_{j\pm1/2}^{n+1}). \tag{4.2}$$

Theorem 3.2.1 guarantees that the local maximum principle is satisfied by the first order method: $u_j^{n,\min} \leq u_j^{n+1,L} \leq u_j^{n,\max}$. So we have to enforce $u_j^{n,\min} \leq u_j^{n+1} \leq u_j^{n,\max}$ via a convex limiting process to guarantee the invariant domain preserving property. By setting $\Psi_j^1(u) = u - u_j^{n,\min}$ and $\Psi_j^2(u) = u_j^{n,\max} - u$, to imposing the local maximum principle is transfered to imposing the following two linear (therefore quasiconcave) constraints:

$$\Psi_j^1(u) \geq 0, \qquad \Psi_j^2(u) \geq 0. \tag{4.3}$$

4.1.2 The P-system

Now we consider the p-system (1.11), see §1.3.2. The invariant domain is constructed using the Riemann invariants (2.8), we define

$$w_{1,j}^{n,\max} := \max(w_{1,j}^n, w_{1,j\pm1}^n, \bar{w}_{1,j\pm1/2}^{n+1}), \quad w_{2,j}^{n,\min} := \min(w_{2,j}^n, w_{2,j\pm1}^n, \bar{w}_{2,j\pm1/2}^{n+1}), \tag{4.4}$$

where $w_{1,j}^n := w_1(\boldsymbol{u}_j^n)$, $w_{2,j}^n := w_2(\boldsymbol{u}_j^n)$, $\bar{w}_{1,j\pm1/2}^{n+1} := w_1(\bar{\boldsymbol{u}}_{j\pm1/2}^{n+1})$, $\bar{w}_{2,j\pm1/2}^{n+1} := w_2(\bar{\boldsymbol{u}}_{j\pm1/2}^{n+1})$. Theorem 3.2.1 guarantees that $w_{2,i}^{n,\min} \leq w_{2,i}^{L,n+1} \leq w_{1,i}^{L,n+1} \leq w_{1,i}^{n,\max}$. Therefore, the local invariant domain to be enforced is an intersection of two concave constraints

$$w_{2,j}^{n,\min} \le w_{2,j}^{n+1}, \qquad w_{1,j}^{n+1} \le w_{1,j}^{n,\max}. \tag{4.5}$$

By setting $\Psi_j^1(\boldsymbol{u}) = w_{1,j}^{n,\max} - w_1(\boldsymbol{u})$ and $\Psi_j^2(\boldsymbol{u}) = w_2(\boldsymbol{u}) - w_{2,j}^{n,\min}$, we are going to enforce the following two concave constraints:

$$\Psi_j^1(\boldsymbol{u}) \ge 0, \qquad \Psi_j^2(\boldsymbol{u}) \ge 0. \tag{4.6}$$

4.1.3 Euler Equations

Now we consider the Euler system (1.16), see § 1.3.3. By definition the specific entropy is a quasiconcave function of the conserved variables, which implies that $\Phi(\boldsymbol{u}) := s(\rho, e)$ is quasiconcave. Theorem 3.2.1 guarantees that the first order solution $\boldsymbol{u}^{L,n+1}$ satisfies the invariant domain property

$$\rho_j^{\max} \ge \rho_j^{L,n+1}, \quad \rho_j^{L,n+1} \ge \rho_j^{\min}, \quad e_j^{L,n+1} \ge 0, \quad s_j^{L,n+1} \ge s_j^{\min}, \tag{4.7}$$

where

$$\begin{aligned} \rho_j^{n,\min} &:= \min(\rho_j^n, \rho_{j\pm1}^n, \bar{\rho}_{j\pm1/2}^{n+1}), \quad \rho_j^{n,\max} := \max(\rho_j^n, \rho_{j\pm1}^n, \bar{\rho}_{j\pm1/2}^{n+1}), \\ e_j^{n,\min} &:= \min(e_j^n, e_{j\pm1}^n, \bar{e}_{j\pm1/2}^{n+1}), \\ s_j^{n,\min} &:= \min(\Phi(\boldsymbol{u}_j^n), \Phi(\boldsymbol{u}_{j\pm1}^n), \Phi(\bar{\boldsymbol{u}}_{j\pm1/2}^{n+1})). \end{aligned} \tag{4.8}$$

Therefore, the invariant domain to be enforced to the high order solution $\boldsymbol{u}^{H,n+1}$ lies in the intersection of the following four quasiconcave constraints:

$$\rho_j^{n,\max} \ge \rho, \quad \rho \ge \rho_j^{n,\min}, \quad e \ge e_j^{n,\min}, \quad s \ge s_j^{n,\min}. \tag{4.9}$$

It is clear that the two constraints on density are linear and thus concave. However, the other two constraints $e \ge e_j^{n,\min}$ and $s - s_j^{n,\min} \ge 0$ are not concave. One can modify the constraints by assuming that the density is already positive, and thus make them to be concave. For example, the mathematical entropy ρs and the total internal energy ρe are concave functions of the conserved variables. It follows that the modified constraints $\rho e - \rho e_j^{n,\min} \geq 0$ and $\rho s - \rho s_j^{n,\min} \geq 0$ are all concave.

In the case of taking the EOS to be the γ-law, we take the physical specific entropy to be $s = \log(e^{\frac{1}{\gamma-1}}\rho^{-1})$, see § 1.5.2.2. Therefore one can impose the the invariant domain by enforcing the following constraints

$$\rho_j^{n,\max} - \rho \geq 0, \quad \rho - \rho_j^{n,\min} \geq 0, \quad \rho e - c_j^{n,\min}\rho^\gamma \geq 0. \tag{4.10}$$

where $c_j^{n,\min} = \exp((\gamma-1)s_j^{n,\min})$. Note that constraints $\rho > 0$ and $\rho e - c_j^{n,\min}(\rho_j^{n+1})^\gamma \geq 0$ guarantees that the internal energy e is greater than $e^{\min}$. By setting $\Psi_j^1(\boldsymbol{u}) = \rho_j^{n,\max} - \rho$, $\Psi_j^2(\boldsymbol{u}) = \rho - \rho_j^{n,\min}$ and $\Psi_j^3(\boldsymbol{u}) = \rho e - c_j^{n,\min}\rho^\gamma$, we enforce the following constraints for the invariant domain property

$$\Psi_j^1(\boldsymbol{u}) \geq 0, \qquad \Psi_j^2(\boldsymbol{u}) \geq 0, \qquad \Psi_j^3(\boldsymbol{u}) \geq 0. \tag{4.11}$$

4.2 Invariant Domain via Flux Limiting

In this section we develop a novel limiting technique to enforce the quasiconcave constraints by adopting the methodology of [27] to the central scheme framework. To our knowledge, simple linear constraints as the local maximum principle of density can be easily enforced using the Flux Corrected Transport technique(FCT), see for example [4] and [65]. However, the FCT approach is designed for box-like limitation and is hard to be modified to enforce more general convex constraints without losing the approximation accuracy. In our flux limiting technique, the so-called bar states defined by (3.10), (3.20) and (3.21) is essential to construct the local invariant constraints, see examples of the invariant bounds (4.2), (4.4) and (4.8). Our discussion will be mainly focused on one dimensional case while the result of two dimensional case will be presented directly. Also, one should notice that both the KT-scheme and the CTO-WENO scheme have a similar fully discrete form while the only difference is the computation of the interface values. Without abuse of notation, in all our discussions we will use the notations and settings in § 3.3 to represent fully discrete form of the high order scheme, which could be either the KT-scheme or the CTO-WENO scheme.

4.2.1 One Dimensional Case

We subtract the first-order update (3.12) from the high-order update (3.28) and get:

$$\boldsymbol{u}_j^{H,n+1} = \boldsymbol{u}_j^{L,n+1} - \frac{\Delta t}{\Delta x}(H_{j+1/2}^n - H_{j-1/2}^n - L_{j+1/2}^n + L_{j-1/2}^n). \tag{4.12}$$

By setting $G_{j+1/2}^n := \frac{2\Delta t}{\Delta x}(H_{j+1/2}^n - L_{j+1/2}^n)$ to be the high/low order numerical flux difference, we could rewrite (4.12) as the following convex splitting form

$$\boldsymbol{u}_j^{H,n+1} = \frac{1}{2}(\boldsymbol{u}_j^{L,n+1} - G_{j+1/2}^n) + \frac{1}{2}(\boldsymbol{u}_j^{L,n+1} + G_{j-1/2}^n). \tag{4.13}$$

Following [27, §4.2], we introduce a pair of scalar limiting parameters (l_j^+, l_j^-) and create a limited high-order update

$$\boldsymbol{u}_j^{n+1}(l_j^+, l_j^-) := \frac{1}{2}(\boldsymbol{u}_j^{L,n+1} - l_j^+ G_{j+1/2}^n) + \frac{1}{2}(\boldsymbol{u}_j^{L,n+1} + l_j^- G_{j-1/2}^n). \tag{4.14}$$

which is supposed to satisfy the invariant domain property. Similar to the FCT scheme, we will recover the first-order solution if $l_j^+ = l_j^- = 0$ and the high-order solution if $l_j^+ = l_j^- = 1$.

We set Ψ_j^z to be a quasiconcave function where z is one of the constraints describing the corresponding local invariant set at cell j and denote with A_j^z the zero level set of Ψ_j^z. It then follows by definition that A_j^z is a convex set. For example, we take $\Psi_j^{\rho_{\max}} = \rho_j^{\max} - \rho$ and $A_j^{\rho_{\max}} = \{\boldsymbol{u} \,|\, \rho_j^{\max} - \rho \geq 0\}$ for the Euler system. Our goal is to find the largest positive numbers $\ell_j^{\pm} \leq 1$ such that $\boldsymbol{u}_j^{n+1}(l_j^+, l_j^-)$ remains in A_j^z, that is, we have $\Psi_j^z(\boldsymbol{u}_j^{n+1}(l_j^+, l_j^-)) \geq 0$ for any $0 \leq l_j^+ \leq \ell_j^+$ and $0 \leq l_j^- \leq \ell_j^-$. In order to simplify the notations, we denote $\boldsymbol{u}_j^+(l) := \boldsymbol{u}_j^{L,n+1} - lG_{j+1/2}^n$ and $\boldsymbol{u}_j^-(l) := \boldsymbol{u}_j^{L,n+1} + lG_{j-1/2}^n$ for any $l \in \mathbb{R}$. Consequently, the following two lemmas describe the idea of the flux limiting process for a certain constraint z.

Lemma 4.2.1. Let $\Psi_j^z : \mathcal{A} \to \mathbb{R}$ be the quasiconcave function mentioned above. Assume that $\ell_j^{z,\pm} \in [0,1]$ are such that $\Psi_j^z(\boldsymbol{u}_j^+(\ell_j^{z,+})) \geq 0$ and $\Psi_j^z(\boldsymbol{u}_j^-(\ell_j^{z,-})) \geq 0$, we have that $\Psi_j^z(\boldsymbol{u}_j^{n+1}(\ell_j^{z,+}, \ell_j^{z,-})) \geq 0$.

Proof. Let $A_0(\Psi) = \{\boldsymbol{u} \in \mathcal{A} | \Psi(\boldsymbol{u}) \geq 0\}$ be the zero level set of Ψ. By definition we have that inequalities $\Psi_j^z(\boldsymbol{u}_j^+(\ell_j^{z,+})) \geq 0$ and $\Psi_j^z(\boldsymbol{u}_j^-(\ell_j^{z,-})) \geq 0$ imply that $\boldsymbol{u}_j^+(\ell_j^{z,+}), \boldsymbol{u}_j^-(\ell_j^{z,-}) \in A_0(\Psi)$. Since $\boldsymbol{u}_j^{n+1}(\ell_j^{z,+}, \ell_j^{z,-})$ is a convex combination of $\boldsymbol{u}_j^+(\ell_j^{z,+})$ and $\boldsymbol{u}_j^-(\ell_j^{z,-})$, it follows that $\boldsymbol{u}_j^{n+1}(\ell_j^{z,+}, \ell_j^{z,-}) \in A_0(\Psi)$ as well, which leads to $\Psi_j^z(\boldsymbol{u}_j^{n+1}(\ell_j^{z,+}, \ell_j^{z,-})) \geq 0$ and thus complete the proof. □

Lemma 4.2.2. Let's define $\ell_j^{z,\pm}$ to be

$$\ell_j^{z,+} = \begin{cases} 1 & \text{if } \Psi_j^z(\boldsymbol{u}_j^+(1)) \geq 0, \\ \max\{\ell \in [0,1] | \Psi_j^z(\boldsymbol{u}_j^+(\ell)) \geq 0\} & \text{otherwise.} \end{cases} \tag{4.15}$$

$$\ell_j^{z,-} = \begin{cases} 1 & \text{if } \Psi_j^z(\boldsymbol{u}_j^-(1)) \geq 0, \\ \max\{\ell \in [0,1] | \Psi_j^z(\boldsymbol{u}_j^-(\ell)) \geq 0\} & \text{otherwise.} \end{cases} \tag{4.16}$$

Now we set $\ell_{j+1/2}^z = \min(\ell_j^{z,+}, \ell_{j+1}^{z,-})$, we have that $\Psi_j^z(\boldsymbol{u}_j^{n+1}(l_{j+1/2}^z, l_{j-1/2}^z)) \geq 0$ for all $l_{j+1/2}^z \in [0, \ell_{j+1/2}^z]$.

Proof. We use the same setting of $A_0(\Psi)$ in Lemma 4.2.1. By definition (4.15) we have that $\Psi_j^z(\ell_j^{z,+}) \geq 0$, which implies that $\ell_j^{z,+} \in A_0(\Psi_j^z)$. Also we have $0 \leq l_{j+1/2}^z \leq \ell_{j+1/2}^z \leq \ell_j^{z,+}$, it then follows by the convexity of $A_0(\Psi_j^z)$ that $l_{j+1/2}^z \in A_0(\Psi_j^z)$, which implies

$$\Psi_j^z(l_{j+1/2}^z) \geq 0. \tag{4.17}$$

At the same time, a similar computation using (4.16) give us

$$\Psi_j^z(l_{j-1/2}^z) \geq 0. \tag{4.18}$$

Now apply Lemma 4.2.2 on (4.17) and (4.18) gives that $\Psi_j^z(\boldsymbol{u}_j^{n+1}(l_{j+1/2}^z, l_{j-1/2}^z)) \geq 0$, which completes our proof. □

The second-order update $\boldsymbol{u}_j^{n+1}$ defined by (4.14) is a convex combination of $\boldsymbol{u}_j^+$ and $\boldsymbol{u}_j^-$, both of which satisfy the constraint z due to our settings. Therefore, by Lemma 4.2.1 and Lemma 4.2.2, the limited update

$$\boldsymbol{u}_j^{z,n+1} = \frac{1}{2}\boldsymbol{u}_j^+(\ell_{j+1/2}^z) + \frac{1}{2}\boldsymbol{u}_j^-(\ell_{j-1/2}^z). \tag{4.19}$$

will satisfy the constraint z. Now we are able to describe the full flux limiting process for all local constraints in the following algorithm.

Algorithm 1 Convex flux limiting

```
Input: $\boldsymbol{u}_j^{L,n+1}$, $G_{j+1/2}^n$, $k^{\max}$, $z_1, \dots, z_q$.
Output: $\boldsymbol{u}_j^{n+1}$
for $i = 1$ to $k^{\max}$ do
    for $z = 1$ to $z_q$ do
        Compute limiting parameters $\ell_{j+1/2}^z$ via Lemma 4.2.1 and Lemma 4.2.2.
    end for
    Set $\ell_{j+1/2} := \min_{z \in \{z_1,\dots,z_q\}} \ell_{j+1/2}^z$.
    Update $\boldsymbol{u}_j^{n+1} = \boldsymbol{u}_j^{n+1}(\ell_{j-1/2}, \ell_{j+1/2})$ via (4.14).
    Update $G_{j+1/2}^{n+1} = \frac{2\Delta t}{\Delta x}(H_{j+1/2}^{n+1} - L_{j+1/2}^{n+1})$
end for
Return $\boldsymbol{u}_j^{n+1}$.
```

Remark. The number $k^{\max}$ in Algorithm 1 refers to the loops of the limiting process. We are supposed to recover a higher approximation accuracy for a larger $k^{\max}$ as we get closer to the bound of $A_0(\Psi)$. In all our numerical experiments reported in § 5, we take $k^{\max} = 2$.

Remark. The computational cost of searching for $\ell_j^{z,\pm}$ for a given j can be reduced by setting $\ell_j^+ = \ell_j^- := \ell_j$ in (4.14) and denote $\boldsymbol{u}_j^{n+1}(\ell) := \boldsymbol{u}_j^{n+1}(\ell, \ell)$. Then ℓ_j is computed with one single line search

$$\ell_j = \begin{cases} 1 & \text{if } \Psi_j^z(\boldsymbol{u}_j^{n+1}(1)) \geq 0, \\ \max\{\ell \in [0,1] | \Psi_j^z(\boldsymbol{u}_j^{n+1}(\ell)) \geq 0\} & \text{otherwise.} \end{cases} \tag{4.20}$$

If $\Psi_j^z(\boldsymbol{u}_j^\pm(\ell_j)) \geq 0$, we could set $\ell_{j+1/2} = \min(\ell_j, \ell_{j+1})$ and replace **step** 3 with the computation via (4.20) in Algorithm 1. In practice, the single search is successful most of the time and therefore we make one single search instead of $2d$ searches, where d denotes the space dimension.

4.2.2 Two Dimensional Case

In the case of two space dimensions we will take a similar approach. By subtracting (3.22) from (3.31) and setting $G_{j+1/2,k}^{n,x} := \frac{4\Delta t}{\Delta x}(H_{j+1/2,k}^{n,x} - L_{j+1/2,k}^{n,x})$, $G_{j,k+1/2}^{n,y} := \frac{4\Delta t}{\Delta y}(H_{j,k+1/2}^{n,y} - L_{j,k+1/2}^{n,y})$, we obtain the convex splitting form for the high-order solution

$$\begin{aligned} \boldsymbol{u}_{j,k}^{H,n+1} &= \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} - G_{j+1/2,k}^{n,x}) + \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} + G_{j-1/2,k}^{n,x}) \\ &+ \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} - G_{j,k+1/2}^{n,y}) + \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} + G_{j,k-1/2}^{n,y}). \end{aligned} \tag{4.21}$$

By introducing four scalar limiting parameters $l_{j,k}^{x,\pm}$ and $l_{j,k}^{y,\pm}$, the limited second-order update is given by

$$\begin{aligned} \boldsymbol{u}_{j,k}^{n+1}(l_{j,k}^{x,\pm}, l_{j,k}^{y,\pm}) :=& \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} - l_{j,k}^{x,+}G_{j+1/2,k}^{n,x}) + \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} + l_{j,k}^{x,-}G_{j-1/2,k}^{n,x}) \\ &+ \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} - l_{j,k}^{y,+}G_{j,k+1/2}^{n,y}) + \frac{1}{4}(\boldsymbol{u}_{j,k}^{L,n+1} + l_{j,k}^{y,-}G_{j,k-1/2}^{n,y}) \\ :=& \frac{1}{4}\boldsymbol{u}_{j,k}^{x,+}(l_{j,k}^{x,+}) + \frac{1}{4}\boldsymbol{u}_{j,k}^{x,-}(l_{j,k}^{x,-}) + \frac{1}{4}\boldsymbol{u}_{j,k}^{y,+}(l_{j,k}^{y,+}) + \frac{1}{4}\boldsymbol{u}_{j,k}^{y,-}(l_{j,k}^{y,-}). \end{aligned} \tag{4.22}$$

where $l_{j,k}^{x,\pm}$ and $l_{j,k}^{y,\pm}$ are computed such that each of the four states $\boldsymbol{u}_{j,k}^{x,+}(l_{j,k}^{x,+})$, $\boldsymbol{u}_{j,k}^{x,-}(l_{j,k}^{x,-})$, $\boldsymbol{u}_{j,k}^{y,+}(l_{j,k}^{y,+})$, $\boldsymbol{u}_{j,k}^{y,-}(l_{j,k}^{y,-})$ satisfies the invariant domain property. Similar to the one dimensional case, we describe a given local constraint z at a cell (j,k) as the zero level set $A^z_{j,k}$ of a quasiconcave function $\Psi^z_{j,k}$ and look for the largest positive numbers $\ell^{x,\pm}_{j,k}, \ell^{y,\pm}_{j,k} \leq 1$ such that the four states mentioned above are in $A^z_{j,k}$ for any $0 \leq l^{x,+}_{j,k} \leq \ell^{x,+}_{j,k}$, $0 \leq l^{x,-}_{j,k} \leq \ell^{x,-}_{j,k}$ and $0 \leq l^{y,+}_{j,k} \leq \ell^{y,+}_{j,k}$, $0 \leq l^{y,-}_{j,k} \leq \ell^{y,-}_{j,k}$. Analogous to the one dimensional case, these limiters are computed via line searches and one can use a single line search instead of four by setting $l = l^{x,+}_{j,k} = l^{x,-}_{j,k} = l^{y,+}_{j,k} = l^{y,-}_{j,k}$, and this will work most of the time.

4.3 Invariant Domain via Slope/Polynomial Limiting

It is well known in literatures that other than the flux limiting, one can also reduce oscillations via the slope limiting. Both limiting process are different but give similar numerical results. In this section we describe a convex limiting procedure using slope limiting for the KT-scheme, which could be extended to the more general polynomial limiting for general high order schemes. The key difference is that the local invariant sets to be enforced are now located at cell interfaces and are different from the local invariant sets at cell centers used in the flux limiting, see § 4.1 and § 4.2.

4.3.1 Slope limiting For Second Order KT-Scheme

Here we will consider the second order KT-scheme. Recall that in the fully discrete form of the KT-scheme (3.28) in one dimensional case and (3.31) in two dimensional case, the numerical flux are all computed using the interface values. The core idea of the slope limiting is to enforce these interface values to stay in a local invariant set via making modification of the slope limiters which are used to determine these values. Still the discussion will focus on the one dimensional case and the result of the two dimensional case.

4.3.1.1 One Dimensional Case

We start with the one dimensional case. Let's consider the fully discrete form of the KT-scheme (3.27) and (3.28). Instead of enforcing $\boldsymbol{u}^{n+1}_j$ to be in the invariant set, we will limit the interface values $\boldsymbol{u}^{n,\pm}_{j-1/2}$ and $\boldsymbol{u}^{n,\pm}_{j+1/2}$, which are given by the local linear reconstructions. In general we know that the second order KT-scheme is not invariant domain preserving. However, we will prove a theorem that if the KT-scheme are modified such that those interface values mentioned above are in the given local invariant set, then the upgraded KT-scheme is going to be invariant domain preserving under a new CFL-condition.

Theorem 4.3.1. Let A be a convex invariant set of (1.27), $n \geq 0$ and $j \in \mathbb{Z}$ be such that $\boldsymbol{u}_j^n$ and all interface values $\boldsymbol{u}_{j\pm 1/2}^{n,\pm}$ are in A. Assume that the second-order solution $\boldsymbol{u}_j^{n+1}$ is computed with the KT-scheme (3.27)-(3.28), where $\lambda_{j-1/2}^n := \lambda_{\max}(\boldsymbol{u}_{j-1/2}^{n,-}, \boldsymbol{u}_{j-1/2}^{n,+}, \boldsymbol{f})$ and $\lambda_{j+1/2}^n := \lambda_{\max}(\boldsymbol{u}_{j+1/2}^{n,-}, \boldsymbol{u}_{j+1/2}^{n,+}, \boldsymbol{f})$. Let $\lambda_{j,\pm}^n := \lambda_{\max}(\boldsymbol{u}_{j-1/2}^{n,+}, \boldsymbol{u}_{j+1/2}^{n,-}, \boldsymbol{f})$ be the in cell local speed and we define the maximum local speed by $\lambda_j^{\max} := \max(\lambda_{j-1/2}^n, \lambda_{j+1/2}^n, \lambda_{j,\pm}^n)$. Then under the CFL condition $\frac{\Delta t \lambda_j^{\max}}{\Delta x} \leq \frac{1}{4}$, we have that $\boldsymbol{u}_j^{n+1} \in A$.

Proof. Using the definition of $\lambda_{j+1/2}^n$ we define the following bar state

$$\bar{\boldsymbol{u}}_{j+1/2,\pm}^{n+1} := \frac{\boldsymbol{u}_{j+1/2}^{n,+} + \boldsymbol{u}_{j+1/2}^{n,-}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,+}) - \boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,-})}{2\lambda_{j+1/2}^n}. \tag{4.23}$$

By assumption we have that $\boldsymbol{u}_{j+1/2}^{n,+}$ and $\boldsymbol{u}_{j+1/2}^{n,-}$ are all in A, then it follows by Lemma 2.1.1 that $\bar{\boldsymbol{u}}_{j+1/2,\pm}^{n+1}$ is in the invariant set A. Similarly, we can prove that the bar state defined by

$$\bar{\boldsymbol{u}}_{j-1/2,\pm}^{n+1} := \frac{\boldsymbol{u}_{j-1/2}^{n,+} + \boldsymbol{u}_{j-1/2}^{n,-}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}_{j-1/2}^{n,+}) - \boldsymbol{f}(\boldsymbol{u}_{j-1/2}^{n,-})}{2\lambda_{j-1/2}^n}, \tag{4.24}$$

is also in A. With this notations, we could rewrite the KT-scheme update as follows

$$\begin{aligned} \boldsymbol{u}_j^{n+1} &= \boldsymbol{u}_j^n + \frac{\Delta t}{\Delta x}\lambda_{j+1/2}^n \bar{\boldsymbol{u}}_{j+1/2,\pm}^{n+1} + \frac{\Delta t}{\Delta x}\lambda_{j-1/2}^n \bar{\boldsymbol{u}}_{j-1/2,\pm}^{n+1} \\ &+ \frac{2\Delta t}{\Delta x}\Big(- \frac{\lambda_{j+1/2}^n \boldsymbol{u}_{j+1/2}^{n,-}}{2} - \frac{\lambda_{j-1/2}^n \boldsymbol{u}_{j-1/2}^{n,+}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,-}) - \boldsymbol{f}(\boldsymbol{u}_{j-1/2}^{n,+})}{2}\Big). \end{aligned} \tag{4.25}$$

We now define another bar state

$$\bar{\boldsymbol{u}}_{j,\pm}^{n+1} := \frac{\boldsymbol{u}_{j-1/2}^{n,+} + \boldsymbol{u}_{j+1/2}^{n,-}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}_{j+1/2}^{n,-}) - \boldsymbol{f}(\boldsymbol{u}_{j-1/2}^{n,+})}{2\lambda_j^{\max}} \tag{4.26}$$

which is also in the invariant set A since $\lambda_j^{\max} \geq \lambda_{\max}(\boldsymbol{u}_{j-1/2}^{n,+}, \boldsymbol{u}_{j+1/2}^{n,-}, \boldsymbol{f})$, see Lemma 2.1.1.

Plugging (4.26) into (4.25) and using the fact that $\frac{\boldsymbol{u}^{n,+}_{j-1/2}+\boldsymbol{u}^{n,-}_{j+1/2}}{2} = \boldsymbol{u}^n_j$, we obtain

$$
\begin{aligned}
\boldsymbol{u}^{n+1}_j &= \big(1 - 4\frac{\Delta t}{\Delta x}\lambda^{\max}_j\big)\boldsymbol{u}^n_j \\
&+ \frac{\Delta t}{\Delta x}\lambda^n_{j+1/2}\bar{\boldsymbol{u}}^{n+1}_{j+1/2,\pm} + \frac{\Delta t}{\Delta x}\lambda^n_{j-1/2}\bar{\boldsymbol{u}}^{n+1}_{j-1/2,\pm} + \frac{2\Delta t}{\Delta x}\lambda^{\max}_j\bar{\boldsymbol{u}}^{n+1}_{j,\pm} \\
&+ \frac{\Delta t}{\Delta x}(\lambda^{\max}_j - \lambda^n_{j-1/2})\boldsymbol{u}^{n,+}_{j-1/2} + \frac{\Delta t}{\Delta x}(\lambda^{\max}_j - \lambda^n_{j+1/2})\boldsymbol{u}^{n,+}_{j-1/2}.
\end{aligned}
\tag{4.27}
$$

Under the CFL-condition $\frac{\Delta t \lambda^{\max}_j}{\Delta x} \leq \frac{1}{4}$, we have that $\boldsymbol{u}^{n+1}_j$ is a convex combination of the cell center states $\boldsymbol{u}^n_j$, the interface states $\boldsymbol{u}^{n,+}_{j-1/2}$, $\boldsymbol{u}^{n,+}_{j-1/2}$ and the bar states $\bar{\boldsymbol{u}}^{n+1}_{j-1/2,\pm}$, $\bar{\boldsymbol{u}}^{n+1}_{j+1/2,\pm}$, $\bar{\boldsymbol{u}}^{n+1}_{j,\pm}$. By assumption we have that the interface states $\boldsymbol{u}^{n,+}_{j-1/2}$, $\boldsymbol{u}^{n,+}_{j-1/2}$ are in the invariant set A. Also we have shown that all bar states defined above are in A because of the definition of $\lambda^{\max}_j$, then it follows by convexity that $\boldsymbol{u}^{n+1}_j \in A$. □

Remark. In order to obtain the invariant domain preserving solution we need to design a limited piecewise linear reconstruction so that the interface values $\boldsymbol{u}^{n,\pm}_{j\pm 1/2}$ are all in the local invariant set A. If the local slope is set to be zero we will recover the first-order solution, which must be in A by Theorem 3.2.1.

Now we are able to describe a convex slope limiting process which will guarantee that the interface states $\boldsymbol{u}^{n,-}_{j-1/2}$ and $\boldsymbol{u}^{n,+}_{j-1/2}$ defined above are restricted to be in a given local invariant set A. As before, we are going to impose a finite number of quasiconcave constraints $\Psi^z_{j-1/2}$, where $z \in \{z_1, \dots, z_q\}$ describes a certain type of constraint and we assume that enforcing these constraints will also guarantee that the interface states $\boldsymbol{u}^{n,-}_{j-1/2}$ and $\boldsymbol{u}^{n,+}_{j-1/2}$ are in A as well. Similar to the flux limiting we denote by $A^z_{j+1/2}$ the zero level set of $\Psi^z_{j+1/2}$, therefore enforcing $\boldsymbol{u} \in A^z_{j+1/2}$ is equivalent to enforcing $\Psi^z_{j+1/2}(\boldsymbol{u}) \geq 0$. By definition we have $\boldsymbol{u}^{n,+}_{j-1/2} = \boldsymbol{u}^n_j - \frac{\Delta x}{2}(\boldsymbol{u}^n_x)_j$ and $\boldsymbol{u}^{n,-}_{j-1/2} = \boldsymbol{u}^n_{j-1} + \frac{\Delta x}{2}(\boldsymbol{u}^n_x)_{j-1}$. In general we take $(\boldsymbol{u}^n_x)_j$ to be the unlimited central slope or any classical slope limiters, see § 3.3.3. Here we take a novel approach by setting $(\boldsymbol{u}^n_x)_j = l_j\sigma^{\mathrm{a}}_j$ for any $j \in \mathbb{Z}$, where $\sigma^{\mathrm{a}}_j := \frac{\boldsymbol{u}^n_{j+1} - \boldsymbol{u}^n_{j-1}}{2\Delta x}$ is the unlimited central slope and $l_j \in [0, 1]$ is a slope limiter. Thus we define the limited interface values as

$$\boldsymbol{u}_{j-1/2}^{n,+}(l_j) = \boldsymbol{u}_j^n - \frac{\Delta x}{2} l_j \sigma_j^{\mathrm{a}}, \qquad \boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}) = \boldsymbol{u}_{j-1}^n + \frac{\Delta x}{2} l_{j-1} \sigma_{j-1}^{\mathrm{a}}. \tag{4.28}$$

Similar to the flux limiting limitation, we need to find the largest positive values $l_j^{z,-}$ and $l_{j-1}^{z,+}$ such that for a given j both interface values $\boldsymbol{u}_{j-1/2}^{n,+}(l_j^{z,-})$ and $\boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}^{z,+})$ are in $A_{j-1/2}^z$. This is described in the following lemma.

Lemma 4.3.2. Let's define $\ell_j^{z,-}$ and $\ell_{j-1}^{z,+}$ to be

$$\ell_j^{z,-} = \begin{cases} 1 & \text{if } \Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,+}(1)) \geq 0, \\ \max\{\ell_j \in [0,1] | \Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,+}(\ell_j)) \geq 0\} & \text{otherwise.} \end{cases} \tag{4.29}$$

$$\ell_{j-1}^{z,+} = \begin{cases} 1 & \text{if } \Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,-}(1)) \geq 0, \\ \max\{\ell_{j-1} \in [0,1] | \Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,+}(\ell_{j-1})) \geq 0\} & \text{otherwise.} \end{cases} \tag{4.30}$$

Then for all $l_j^{z,-} \in [0, \ell_j^{z,-}]$ and $l_{j-1}^{z,+} \in [0, \ell_{j-1}^{z,+}]$, it holds that $\Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,+}(l_j^{z,-})) \geq 0$ and $\Psi_{j-1/2}^z(\boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}^{z,+})) \geq 0$.

Remark. The proof of Lemma 4.3.2 is similar to the proof of Lemma 4.2.2, thus we omit it here.

Let's denote $\boldsymbol{u}_j^{n+1}(l_{j-1}, l_j, l_{j+1})$ to be the limited second-order update computed with interface values $\boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}), \boldsymbol{u}_{j-1/2}^{n,+}(l_j), \boldsymbol{u}_{j+1/2}^{n,-}(l_j), \boldsymbol{u}_{j+1/2}^{n,+}(l_{j+1})$. Note that we will recover the first-order solution if $l_{j-1} = l_j = l_{j+1} = 0$ and the second-order solution if $l_{j-1} = l_j = l_{j+1} = 1$. Our goal is to find a set of local limiters such that the corresponding solutions will satisfy the invariant domain preserving property. A straightforward application of Theorem 4.3.1 and Lemma 4.3.2 gives the following result.

Lemma 4.3.3. Let $\ell_{j-1}^{z,+}, l_j^{z,-}$ be the slope limiters computed via Lemma 4.3.2 for any $j \in \mathbb{Z}$ and $n \geq 0$. If we set $\ell_j^z = \min(\ell_j^{z,-}, \ell_j^{z,+})$ for $j \in \mathbb{Z}$, then for all $l_j^z \in [0, \ell_j^z]$ we have that $\boldsymbol{u}_j^{n+1}(l_{j-1}^z, l_j^z, l_{j+1}^z) \in A_{j-1/2}^z \cup A_{j+1/2}^z$.

Remark. The proof of Lemma 4.3.3 relies on the convexity of $A^z_{j-1/2}$ and $A^z_{j+1/2}$ and is similar to the proof of Lemma 4.2.1, thus we omit it here.

Remark. Note that, the underlying assumption is that both sets $A^z_{j-1/2}$ and $A^z_{j+1/2}$ are in a local invariant set A, which guarantees that the solution computed via Lemma 4.3.3 is in A. Typically we pick up $A^z_{j-1/2}$ and $A^z_{j+1/2}$ specifically such that $A^z_{j-1/2} \cup A^z_{j+1/2} \subset A_j$, which is the invariant set for the flux limiting.

We now describe the slope limiting algorithm for all local constraints as follows.

Algorithm 2 Convex slope limiting

```
Input: $\boldsymbol{u}^n_j$, $\boldsymbol{u}^{n,+}_{j-1/2}$, $\boldsymbol{u}^{n,-}_{j+1/2}$, $z_1, \dots, z_q$.
Output: $\boldsymbol{u}^{n+1}_j$
for $z = 1$ to $z_q$ do
    Compute limiting parameters $\ell^z_j$ via Lemma 4.3.2 and Lemma 4.3.3.
end for
Set $\ell_j := \min_{z \in \{z_1,\dots,z_q\}} \ell^z_j$ for all $j \in \mathbb{Z}$.
Update $\boldsymbol{u}^{n+1,+}_{j-1/2}$ and $\boldsymbol{u}^{n+1,-}_{j+1/2}$.
Update $\boldsymbol{u}^{n+1}_j = \boldsymbol{u}^{n+1}_j(\ell_{j-1}, \ell_j, \ell_{j+1})$.
Return $\boldsymbol{u}^{n+1}_j$.
```

4.3.1.2 Two Dimensional Case

In the case of two space dimensions, we limit the second-order solution via a similar approach. We consider the KT-scheme (3.30)-(3.31), where $\boldsymbol{u}^{n,\pm}_{j+1/2,k}$ and $\boldsymbol{u}^{n,\pm}_{j,k+1/2}$ are the interface values computed via local linear reconstructions. The following result establishes the theory for slope limiting in two dimensional case.

Theorem 4.3.4. Let A be a convex invariant set of system (1.3) for $d = 2$, $n \geq 0$ and $j, k \in \mathbb{Z}$ be such that $\boldsymbol{u}^n_{j,k}$ and all interface values $\boldsymbol{u}^{n,\pm}_{j\pm1/2,k}$ and $\boldsymbol{u}^{n,\pm}_{j,k\pm1/2}$ are all in A. Assume that the second-order solution $\boldsymbol{u}^{n+1}_{j,k}$ is computed with the KT-scheme (3.30)-(3.31), where $\lambda^{n,x}_{j-1/2,k} := \lambda_{\max}(\boldsymbol{u}^{n,-}_{j-1/2,k}, \boldsymbol{u}^{n,+}_{j-1/2,k}, \boldsymbol{f})$, $\lambda^{n,x}_{j+1/2,k} := \lambda_{\max}(\boldsymbol{u}^{n,-}_{j+1/2,k}, \boldsymbol{u}^{n,+}_{j+1/2,k}, \boldsymbol{f})$, $\lambda^{n,y}_{j,k-1/2} := \lambda_{\max}(\boldsymbol{u}^{n,-}_{j,k-1/2}, \boldsymbol{u}^{n,+}_{j,k-1/2}, \boldsymbol{g})$ and $\lambda^{n,y}_{j,k+1/2} := \lambda_{\max}(\boldsymbol{u}^{n,-}_{j,k+1/2}, \boldsymbol{u}^{n,+}_{j,k+1/2}, \boldsymbol{g})$ are the interface local speeds. In addition we set the in cell local speeds to be $\lambda^{n,x}_{j,k,\pm} := \lambda_{\max}(\boldsymbol{u}^{n,+}_{j-1/2,k}, \boldsymbol{u}^{n,-}_{j+1/2,k}, \boldsymbol{f})$ and $\lambda^{n,y}_{j,k,\pm} := \lambda_{\max}(\boldsymbol{u}^{n,+}_{j,k-1/2}, \boldsymbol{u}^{n,-}_{j,k+1/2}, \boldsymbol{g})$ and define the maximum local speeds in the $x-$ and the $y-$direction respectively by

$$\lambda^{\max,x}_{j,k} := \max(\lambda^{n,x}_{j-1/2,k}, \lambda^{n,x}_{j+1/2,k}, \lambda^{n,x}_{j,k,\pm}), \tag{4.31}$$

$$\lambda^{\max,y}_{j,k} := \max(\lambda^{n,y}_{j,k-1/2}, \lambda^{n,y}_{j,k+1/2}, \lambda^{n,y}_{j,k,\pm}). \tag{4.32}$$

Then under the CFL condition $\frac{\Delta t \lambda^{\max,x}_{j,k}}{\Delta x} + \frac{\Delta t \lambda^{\max,y}_{j,k}}{\Delta y} \leq \frac{1}{4}$, we have that $\boldsymbol{u}^{n+1}_{j,k} \in A$.

Remark. The proof for Theorem 4.3.4 could be obtained by repeat the proof for Theorem 4.3.1 in both $x-$ and $y-$ directions, thus we omit it here.

Now we describe the slope limiting algorithm in two dimensional case. First, we define the limited interface values by

$$\begin{aligned}
\boldsymbol{u}^{n,+}_{j-1/2,k}(l^x_{j,k}) &= \boldsymbol{u}^n_{j,k} - \frac{\Delta x}{2} l^x_{j,k}\sigma^{\mathrm{a},x}_{j,k}, \quad \boldsymbol{u}^{n,-}_{j-1/2,k}(l^x_{j-1,k}) = \boldsymbol{u}^n_{j-1,k} + \frac{\Delta x}{2} l^x_{j-1,k}\sigma^{\mathrm{a},x}_{j-1,k}, \\
\boldsymbol{u}^{n,+}_{j,k-1/2}(l^y_{j,k}) &= \boldsymbol{u}^n_{j,k} - \frac{\Delta y}{2} l^y_{j,k}\sigma^{\mathrm{a},y}_{j,k}, \quad \boldsymbol{u}^{n,-}_{j,k-1/2}(l^y_{j,k-1}) = \boldsymbol{u}^n_{j,k-1} + \frac{\Delta y}{2} l^y_{j,k-1}\sigma^{\mathrm{a},y}_{j,k-1}.
\end{aligned}$$

where $\sigma^{\mathrm{a},x}_{j,k}$ and $\sigma^{\mathrm{a},y}_{j,k}$ are the two-dimensional central slopes defined by $\sigma^{\mathrm{a},x}_{j,k} := \frac{\boldsymbol{u}^n_{j+1,k} - \boldsymbol{u}^n_{j-1,k}}{2\Delta x}$ and $\sigma^{\mathrm{a},y}_{j,k} := \frac{\boldsymbol{u}^n_{j,k+1} - \boldsymbol{u}^n_{j,k-1}}{2\Delta y}$, $l^x_{j,k}, l^y_{j,k} \in [0,1]$ are the to be determined slope limiters. For a given constraint z, let $A^{z,x}_{j\pm1/2,k}$ and $A^{z,y}_{j,k\pm1/2}$ be the local invariant sets at cell interfaces $[x_{j-1/2}, x_{j+1/2}] \times [y_{k-1/2}, y_{k+1/2}]$ such that $A^{z,x}_{j\pm1/2,k}$ and $A^{z,y}_{j,k\pm1/2}$ are all in A. Using Lemma 4.3.2 and Lemma 4.3.3, we could find largest positive $l^x_{j,k}, l^y_{j,k} \in [0,1]$ such that $\boldsymbol{u}^{n,+}_{j-1/2,k}(l^x_{j,k}), \boldsymbol{u}^{n,-}_{j-1/2,k}(l^x_{j-1,k}) \in A^{z,x}_{j-1/2,k}$ and $\boldsymbol{u}^{n,+}_{j,k-1/2}(l^y_{j,k}), \boldsymbol{u}^{n,-}_{j,k-1/2}(l^y_{j,k-1}) \in A^{z,y}_{j,k-1/2}$ for all $j, k \in \mathbb{Z}$. Therefore, we obtain the limited solution $\boldsymbol{u}^{n+1}_{j,k}$ which is in A.

4.3.2 Polynomial Limiting For Third/Fourth Order Central Schemes

As explained in § 4.3.1, the core idea of slope limiting is to limit the interface values computed via a local slope reconstruction. This idea could be naturally extended to a central scheme using a polynomial reconstruction of any order by applying a limiter to the polynomial. We've seen in § 3.3 and § 3.4 that a high order central scheme has a similar fully discrete form to the KT-scheme while the only difference is the numerical flux due to the local reconstructions of different orders. Therefore, a new CFL-condition is required for the limited solution to be invariant domain preserving for a general high order central scheme. Our discussion will be restricted to the third order reconstruction (3.45)-(3.47) and the fourth order reconstruction (3.57)-(3.59) in one dimensional case. Notice that, in order to complete the method, we also require that the auxiliary state $\widehat{\boldsymbol{u}}$ defined by (3.47) and (3.59) are in the corresponding invariant set. For two dimensional case we only present the result without any derivation.

4.3.2.1 One Dimensional Case

We start by considering the one dimensional case and give the following result, which is analogous to Theorem 4.3.1

Theorem 4.3.5. Let A be a convex invariant set of system (1.27). Assume that the general high order solution $\boldsymbol{u}_j^{n+1}$ is computed via (3.48)-(3.49), with the local reconstruction defined by (3.45) (third order) or (3.46) (fourth order) and the auxiliary state $\widehat{\boldsymbol{u}}_j^n$ defined by (3.47). Let $n \geq 0$ and $j \in \mathbb{Z}$ be such that $\boldsymbol{u}_j^n$, all interface values $\boldsymbol{u}_{j\pm1/2}^{n,\pm}$ and $\widehat{\boldsymbol{u}}_j^n$ are in A. Also we take $\lambda_{j-1/2}^n := \lambda_{\max}(\boldsymbol{u}_{j-1/2}^{n,-}, \boldsymbol{u}_{j-1/2}^{n,+}, \boldsymbol{f})$ and $\lambda_{j+1/2}^n := \lambda_{\max}(\boldsymbol{u}_{j+1/2}^{n,-}, \boldsymbol{u}_{j+1/2}^{n,+}, \boldsymbol{f})$ to be the interface local speeds and let $\lambda_{j,\pm}^n := \lambda_{\max}(\boldsymbol{u}_{j-1/2}^{n,+}, \boldsymbol{u}_{j+1/2}^{n,-}, \boldsymbol{f})$ be the in cell local speed. Then we define the maximum local speed by $\lambda_j^{\max} := \max(\lambda_{j-1/2}^n, \lambda_{j+1/2}^n, \lambda_{j,\pm}^n)$. Therefore under the CFL condition $\frac{\Delta t \lambda_j^{\max}}{\Delta x} \leq \frac{1}{12}$, we have that $\boldsymbol{u}_j^{n+1} \in A$.

Proof. The proof is similar to Theorem 4.3.1. Using the definition of $\lambda_{j+1/2}^n$ and $\lambda_{j+1/2}^n$

we define the following bar states

$$\bar{\boldsymbol{u}}^{n+1}_{j+1/2,\pm} := \frac{\boldsymbol{u}^{n,+}_{j+1/2} + \boldsymbol{u}^{n,-}_{j+1/2}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{n,+}_{j+1/2}) - \boldsymbol{f}(\boldsymbol{u}^{n,-}_{j+1/2})}{2\lambda^n_{j+1/2}}, \tag{4.33a}$$

$$\bar{\boldsymbol{u}}^{n+1}_{j-1/2,\pm} := \frac{\boldsymbol{u}^{n,+}_{j-1/2} + \boldsymbol{u}^{n,-}_{j-1/2}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{n,+}_{j-1/2}) - \boldsymbol{f}(\boldsymbol{u}^{n,-}_{j-1/2})}{2\lambda^n_{j-1/2}}, . \tag{4.33b}$$

By assumption we have that $\boldsymbol{u}^{n,\pm}_{j+1/2}$ and $\boldsymbol{u}^{n,\pm}_{j-1/2}$ are all in A, then it follows by Lemma 2.1.1 that $\bar{\boldsymbol{u}}^{n+1}_{j+1/2,\pm}$ and $\bar{\boldsymbol{u}}^{n+1}_{j-1/2,\pm}$ are both in the invariant set A. Similarly, we could define another bar state

$$\bar{\boldsymbol{u}}^{n+1}_{j,\pm} := \frac{\boldsymbol{u}^{n,-}_{j+1/2} + \boldsymbol{u}^{n,+}_{j-1/2}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{n,-}_{j+1/2}) - \boldsymbol{f}(\boldsymbol{u}^{n,+}_{j-1/2})}{2\lambda^{\max}_j}. \tag{4.34}$$

Using the fact that $\lambda^{\max}_j \geq \lambda^n_{j,\pm}$, by Lemma 2.1.1 we have that $\bar{\boldsymbol{u}}^{n+1}_{j,\pm}$ is also in A. Now by plugging (3.48) into (3.49) and use the bar states defined above, we could rewrite the fully discrete update as follows

$$\begin{aligned}
\boldsymbol{u}^{n+1}_j = \boldsymbol{u}^n_j &- \frac{4\Delta t}{\Delta x}\lambda^{\max}_j\Big(\frac{\boldsymbol{u}^-_{j+1/2} + \boldsymbol{u}^+_{j-1/2}}{2}\Big) \\
&+ \frac{\Delta t}{\Delta x}\lambda^n_{j+1/2}\bar{\boldsymbol{u}}^{n+1}_{j+1/2,\pm} + \frac{\Delta t}{\Delta x}\lambda^n_{j-1/2}\bar{\boldsymbol{u}}^{n+1}_{j-1/2,\pm} \\
&+ \frac{2\Delta t}{\Delta x}\lambda^{\max}_j\bar{\boldsymbol{u}}^{n+1}_{j,\pm} \\
&+ \frac{\Delta t}{\Delta x}(\lambda^{\max}_j - \lambda^n_{j-1/2})\boldsymbol{u}^{n,+}_{j-1/2} + \frac{\Delta t}{\Delta x}(\lambda^{\max}_j - \lambda^n_{j+1/2})\boldsymbol{u}^{n,+}_{j-1/2}.
\end{aligned} \tag{4.35}$$

Using the fact that $\boldsymbol{u}^-_{j+1/2} = p_j(x_{j+1/2})$ and $\boldsymbol{u}^+_{j-1/2} = p_j(x_{j-1/2})$ with $p_j(x)$ defined by (3.45), we have that

$$\frac{\boldsymbol{u}^-_{j+1/2} + \boldsymbol{u}^+_{j-1/2}}{2} = \widehat{\boldsymbol{u}}^n_j + \frac{\Delta x^2}{8}\boldsymbol{u}''_j. \tag{4.36}$$

Also by (3.47) we have that

$$\boldsymbol{u}''_j = \frac{24}{\Delta x^2}(\boldsymbol{u}^n_j - \widehat{\boldsymbol{u}}^n_j). \tag{4.37}$$

Now plug (4.36) and (4.37) into (4.35) we obtain

$$\begin{aligned}\boldsymbol{u}_j^{n+1} = &(1-\frac{12\Delta t}{\Delta x}\lambda_j^{\max})\boldsymbol{u}_j^n + \frac{8\Delta t}{\Delta x}\lambda_j^{\max}\widehat{\boldsymbol{u}}_j^n\\ &+\frac{\Delta t}{\Delta x}\lambda_{j+1/2}^n\bar{\boldsymbol{u}}_{j+1/2,\pm}^{n+1}+\frac{\Delta t}{\Delta x}\lambda_{j-1/2}^n\bar{\boldsymbol{u}}_{j-1/2,\pm}^{n+1}\\ &+\frac{2\Delta t}{\Delta x}\lambda_j^{\max}\bar{\boldsymbol{u}}_{j,\pm}^{n+1}\\ &+\frac{\Delta t}{\Delta x}(\lambda_j^{\max}-\lambda_{j-1/2}^n)\boldsymbol{u}_{j-1/2}^{n,+}+\frac{\Delta t}{\Delta x}(\lambda_j^{\max}-\lambda_{j+1/2}^n)\boldsymbol{u}_{j-1/2}^{n,+}.\end{aligned} \tag{4.38}$$

Therefore, under the CFL-condition $\frac{\Delta t\lambda_j^{\max}}{\Delta x} \leq \frac{1}{12}$, we have that $\boldsymbol{u}_j^{n+1}$ is a convex combination of the cell center states $\boldsymbol{u}_j^n$, the auxiliary state $\widehat{\boldsymbol{u}}_j^n$, the interface states $\boldsymbol{u}_{j-1/2}^{n,+}$, $\boldsymbol{u}_{j-1/2}^{n,+}$ and the bar states $\bar{\boldsymbol{u}}_{j-1/2,\pm}^{n+1}$, $\bar{\boldsymbol{u}}_{j+1/2,\pm}^{n+1}$, $\bar{\boldsymbol{u}}_{j,\pm}^{n+1}$. By assumption we have that the auxiliary state $\widehat{\boldsymbol{u}}_j^n$ and the interface states $\boldsymbol{u}_{j-1/2}^{n,+}$, $\boldsymbol{u}_{j-1/2}^{n,+}$ are all in the invariant set A. Also we have shown that all bar states defined above are in A because of the definition of $\lambda_j^{\max}$, then it follows by convexity that $\boldsymbol{u}_j^{n+1} \in A$. □

Now we describe the convex polynomial limiting algorithm which enforces the interface states $\boldsymbol{u}_{j-1/2}^{n,-}$, $\boldsymbol{u}_{j-1/2}^{n,+}$ and the auxiliary value $\widehat{\boldsymbol{u}}_j^n$ to be in a given local invariant set A.

For interface states $\boldsymbol{u}_{j-1/2}^{n,-}$ and $\boldsymbol{u}_{j-1/2}^{n,+}$, let's take the similar settings to the second order case, that is, we impose a finite number of quasiconcave constraints $\Psi_{j-1/2}^z$, where each $z \in \{z_1, \ldots, z_q\}$ denotes a certain constraint and we assume that enforcing these constraints will guarantee that the interface states $\boldsymbol{u}_{j-1/2}^{n,-}$ and $\boldsymbol{u}_{j-1/2}^{n,+}$ are in A. Still we denote by $A_{j+1/2}^z$ the zero level set of $\Psi_{j+1/2}^z$, so enforcing $\boldsymbol{u} \in A_{j+1/2}^z$ is equivalent to enforcing $\Psi_{j+1/2}^z(\boldsymbol{u}) \geq 0$.

For the auxiliary state $\widehat{\boldsymbol{u}}_j^n$ defined by (3.47), similarly we impose a set of quasiconcave constraints Ψ_j^z for the same $z \in \{z_1, \cdots, z_q\}$ mentioned above and we assume that enforcing these constraints implies that $\widehat{\boldsymbol{u}}_j^n$ is in A. Still we denote by A_j^z the zero level set of Ψ_j^z, therefore enforcing $\boldsymbol{u} \in A_j^z$ is equivalent to enforcing $\Psi_j^z(\boldsymbol{u}) \geq 0$.

Now assume that $p_j(x)$ is defined by (3.45) (third order case) or (3.46) (fourth order case). Similar to the discussion in § 3.4.3, we introduce an auxiliary function $p_j^r(x)$ such that

$$p_j(x) = \boldsymbol{u}_j^n + p_j^r(x) \qquad \text{for all } x \in I_j, \tag{4.39}$$

and we can prove that

$$\int_{I_j} p_j^r(x) = 0. \tag{4.40}$$

Then by introducing $l_j \in [0,1]$ to be the polynomial limiter, we have that the limited polynomial reconstruction is given by

$$p_j^l(x) := \boldsymbol{u}_j^n + l_j p_j^r(x). \tag{4.41}$$

Notice that by (4.40) we have

$$\int_{I_j} p_j^l(x) = \boldsymbol{u}_j^n, \tag{4.42}$$

which implies that $p_j^l(x)$ is conservative on I_j. Now by setting

$$p_j^{r,-} := p_j^r(x_{j-1/2}), \qquad p_{j-1}^{r,+} := p_{j-1}^r(x_{j-1/2}) \tag{4.43}$$

we could define the limited interface values as

$$\boldsymbol{u}_{j-1/2}^{n,+}(l_j) = p_j^l(x_{j-1/2}) = \boldsymbol{u}_j^n + l_j p_j^{r,-}, \tag{4.44a}$$

$$\boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}) = p_{j-1}^l(x_{j-1/2}) = \boldsymbol{u}_{j-1}^n + l_{j-1} p_{j-1}^{r,+}. \tag{4.44b}$$

Also by (3.47) we have that

$$\widehat{\boldsymbol{u}}_j^n = \boldsymbol{u}_j^n - \frac{\Delta x^2}{24}\boldsymbol{u}_j'', \tag{4.45}$$

therefore we could set the limited auxiliary value to be

$$\widehat{\boldsymbol{u}}_j^n(l_j) = \boldsymbol{u}_j^n + l_j \boldsymbol{u}_j^r, \tag{4.46}$$

where $\boldsymbol{u}_j^r = -\frac{\Delta x^2}{24}\boldsymbol{u}_j''$.

Similar to the KT-scheme, for a given constraint z and index j, let $A_{j+1/2}^z$ and $A_{j-1/2}^z$ be the local invariant sets at cell interfaces $x_{j-1/2}$ and $x_{j+1/2}$ separately, such that $A_{j+1/2}^z$ and $A_{j-1/2}^z$ are all in A. Also let A_j^z be the local invariant sets at cell centers x_j, which is also in A. Then using Lemma 4.3.2 and Lemma 4.3.3, we could find largest positive $l_j \in [0,1]$ such that $\boldsymbol{u}_{j-1/2}^{n,+}(l_j), \boldsymbol{u}_{j-1/2}^{n,-}(l_{j-1}) \in A_{j-1/2}^z$ and $\widehat{\boldsymbol{u}}_j^n(l_j) \in A_j^z$ for all $j \in \mathbb{Z}$. By Theorem 4.3.5 we claim that the limited solution $\boldsymbol{u}_j^{n+1}$ is also in A. This procedure can also be described by Algorithm 2, the only difference is the computation of ℓ_j^z in **Step** 2.

4.3.2.2 Two Dimensional Case

In the case of two space dimensions, we limit the high-order solution via a similar approach. We consider the high order central scheme (3.60)-(3.62), where the interface values are defined by (3.63). The following result establishes the theory for polynomial limiting in two dimensional case.

Theorem 4.3.6. Let A be a convex invariant set of system (1.3) with $d = 2$. Assume that the high order solution $\boldsymbol{u}_{j,k}^{n+1}$ is computed with the high order central scheme (3.60)-(3.62) with local reconstruction (3.57) (third order) or (3.58) (fourth order) and auxiliary state $\widehat{\boldsymbol{u}}_{j,k}^n$ defined by (3.59). Let $n \geq 0$ and $j,k \in \mathbb{Z}$ be such that $\boldsymbol{u}_{j,k}^n$, all interface values defined by (3.63) and $\widehat{\boldsymbol{u}}_{j,k}^n$ are in A. The local speed of propagation $\lambda_{j+1/2,k}^{n,x}$ and $\lambda_{j,k+1/2}^{n,y}$ are defined by (3.64). Also, we define the in cell local speeds to be

$$\lambda_{j,k}^{n,x} = \max(\lambda_{j,k}^N, \lambda_{j,k}^C, \lambda_{j,k}^S), \quad \lambda_{j,k,\pm}^{n,y} = \max(\lambda_{j,k}^E, \lambda_{j,k}^C, \lambda_{j,k}^W) \tag{4.47}$$

where $\lambda_{j,k}^X := \lambda_{\max}(\boldsymbol{u}_{j,k}^{XW}, \boldsymbol{u}_{j,k}^{XE}, \boldsymbol{f})$ and $\lambda_{j,k}^Y := \lambda_{\max}(\boldsymbol{u}_{j,k}^{SY}, \boldsymbol{u}_{j,k}^{NY}, \boldsymbol{g})$ for $X \in \{S,C,N\}$ and $Y \in \{W,C,E\}$ separately. By setting the local speeds to be $\widetilde{\lambda}_{j,k}^{\max,x}$ and $\widetilde{\lambda}_{j,k}^{\max,y}$ such that $\widetilde{\lambda}_{j,k}^{\max,x} \geq \max(\lambda_{j-1/2,k}^{n,x}, \lambda_{j+1/2,k}^{n,x}, \lambda_{j,k}^{n,x})$ and $\widetilde{\lambda}_{j,k}^{\max,y} \geq \max(\lambda_{j,k-1/2}^{n,y}, \lambda_{j,k+1/2}^{n,y}, \lambda_{j,k}^{n,y})$, we define the maximum local speeds in the $x-$ and the $y-$direction respectively by

$$\lambda_{j,k}^{\max,x} = \theta_{j,k}\Delta x, \qquad \lambda_{j,k}^{\max,y} = \theta_{j,k}\Delta y, \tag{4.48}$$

where $\theta_{j,k} := \max(\frac{\widetilde{\lambda}^{\max,x}_{j,k}}{\Delta x}, \frac{\widetilde{\lambda}^{\max,y}_{j,k}}{\Delta y})$. Then under the CFL condition $\Delta t\theta_{j,k} \leq \frac{1}{16}$, we have that $\boldsymbol{u}^{n+1}_{j,k} \in A$.

Proof. The proof for Theorem 4.3.6 is analogs to the proof of Theorem 4.3.5. First we consider the $x-$direction. By definition we have $\lambda^{\max,x}_{j,k} \geq \widetilde{\lambda}^{\max,x}_{j,k}$. Therefore, for $X \in \{S, C, N\}$, we have

$$\begin{aligned}
&\frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^{XW}_{j,k}) + \boldsymbol{f}(\boldsymbol{u}^{XE}_{j-1,k}) - \lambda^{n,x}_{j-1/2,k}(\boldsymbol{u}^{XW}_{j,k} - \boldsymbol{u}^{XE}_{j-1,k}))\\
&\quad - \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^{XW}_{j+1,k}) + \boldsymbol{f}(\boldsymbol{u}^{XE}_{j,k}) - \lambda^{n,x}_{j+1/2,k}(\boldsymbol{u}^{XW}_{j+1,k} - \boldsymbol{u}^{XE}_{j,k}))\\
=&\lambda^{n,x}_{j-1/2,k}\bar{\boldsymbol{u}}^{X}_{j-1/2,k} + \lambda^{n,x}_{j+1/2,k}\bar{\boldsymbol{u}}^{X}_{j+1/2,k} + 2\lambda^{\max,x}_{j,k}\bar{\boldsymbol{u}}^{X}_{j,k}\\
&\quad + (\lambda^{\max,x}_{j,k} - \lambda^{n,x}_{j-1/2,k})\boldsymbol{u}^{XW}_{j,k} + (\lambda^{\max,x}_{j,k} - \lambda^{n,x}_{j+1/2,k})\boldsymbol{u}^{XE}_{j,k}\\
&\quad - 2\lambda^{\max,x}_{j,k}(\boldsymbol{u}^{XW}_{j,k} + \boldsymbol{u}^{XE}_{j,k}),
\end{aligned} \tag{4.49}$$

where the bar states are defined by

$$\bar{\boldsymbol{u}}^{X}_{j-1/2,k} := \frac{\boldsymbol{u}^{XE}_{j-1,k} + \boldsymbol{u}^{XW}_{j,k}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{XW}_{j,k}) - \boldsymbol{f}(\boldsymbol{u}^{XE}_{j-1,k})}{2\lambda^{n,x}_{j-1/2,k}}, \tag{4.50a}$$

$$\bar{\boldsymbol{u}}^{X}_{j+1/2,k} := \frac{\boldsymbol{u}^{XE}_{j,k} + \boldsymbol{u}^{XW}_{j+1,k}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{XW}_{j+1,k}) - \boldsymbol{f}(\boldsymbol{u}^{XE}_{j,k})}{2\lambda^{n,x}_{j+1/2,k}}, \tag{4.50b}$$

$$\bar{\boldsymbol{u}}^{X}_{j,k} := \frac{\boldsymbol{u}^{XW}_{j,k} + \boldsymbol{u}^{XE}_{j,k}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{XE}_{j,k}) - \boldsymbol{f}(\boldsymbol{u}^{XW}_{j,k})}{2\lambda^{\max,x}_{j,k}}. \tag{4.50c}$$

Next we consider the $y-$direction. By definition we have $\lambda^{\max,y}_{j,k} \geq \widetilde{\lambda}^{\max,y}_{j,k}$. Therefore, for $Y \in \{W, C, E\}$ similarly we have that

$$\begin{aligned}
&\frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^{SY}_{j,k}) + \boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k-1}) - \lambda^{n,y}_{j,k-1/2}(\boldsymbol{u}^{SY}_{j,k} - \boldsymbol{u}^{NY}_{j,k-1}))\\
&\quad - \frac{1}{2}(\boldsymbol{f}(\boldsymbol{u}^{SY}_{j,k+1}) + \boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k}) - \lambda^{n,y}_{j,k+1/2}(\boldsymbol{u}^{SY}_{j,k+1} - \boldsymbol{u}^{NY}_{j,k}))\\
=&\lambda^{n,y}_{j,k-1/2}\bar{\boldsymbol{u}}^{Y}_{j,k-1/2} + \lambda^{n,y}_{j,k+1/2}\bar{\boldsymbol{u}}^{Y}_{j,k+1/2} + 2\lambda^{\max,y}_{j,k}\bar{\boldsymbol{u}}^{Y}_{j,k}\\
&\quad + (\lambda^{\max,y}_{j,k} - \lambda^{n,y}_{j,k-1/2})\boldsymbol{u}^{SY}_{j,k} + (\lambda^{\max,y}_{j,k} - \lambda^{n,y}_{j,k+1/2})\boldsymbol{u}^{NY}_{j,k}\\
&\quad - 2\lambda^{\max,y}_{j,k}(\boldsymbol{u}^{SY}_{j,k} + \boldsymbol{u}^{NY}_{j,k}),
\end{aligned} \tag{4.51}$$

where

$$\bar{\boldsymbol{u}}^{Y}_{j,k-1/2} := \frac{\boldsymbol{u}^{NY}_{j,k-1} + \boldsymbol{u}^{SY}_{j,k}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{SY}_{j,k}) - \boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k-1})}{2\lambda^{n,y}_{j,k-1/2}}, \tag{4.52a}$$

$$\bar{\boldsymbol{u}}^{Y}_{j,k+1/2} := \frac{\boldsymbol{u}^{NY}_{j,k} + \boldsymbol{u}^{SY}_{j,k+1}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{SY}_{j,k+1}) - \boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k})}{2\lambda^{n,y}_{j,k+1/2}}, \tag{4.52b}$$

$$\bar{\boldsymbol{u}}^{Y}_{j,k} := \frac{\boldsymbol{u}^{SY}_{j,k} + \boldsymbol{u}^{NY}_{j,k}}{2} - \frac{\boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k}) - \boldsymbol{f}(\boldsymbol{u}^{NY}_{j,k})}{2\lambda^{\max,y}_{j,k}}. \tag{4.52c}$$

Now plug (4.49)-(4.52) into the high order central scheme (3.60)-(3.62), we obtain

$$\begin{aligned}
\boldsymbol{u}^{n+1}_{j,k} &= \boldsymbol{u}^{n}_{j,k} \\
&\quad - \frac{1}{3}\frac{\Delta t}{\Delta x}\left[\lambda^{\max,x}_{j,k}((\boldsymbol{u}^{SW}_{j,k} + \boldsymbol{u}^{SE}_{j,k}) + 4(\boldsymbol{u}^{CW}_{j,k} + \boldsymbol{u}^{CE}_{j,k}) + (\boldsymbol{u}^{NW}_{j,k} + \boldsymbol{u}^{NE}_{j,k}))\right] \\
&\quad - \frac{1}{3}\frac{\Delta t}{\Delta y}\left[\lambda^{\max,y}_{j,k}((\boldsymbol{u}^{SW}_{j,k} + \boldsymbol{u}^{NW}_{j,k}) + 4(\boldsymbol{u}^{SC}_{j,k} + \boldsymbol{u}^{NC}_{j,k}) + (\boldsymbol{u}^{SE}_{j,k} + \boldsymbol{u}^{NE}_{j,k}))\right] \\
&\quad + \frac{1}{6}\frac{\Delta t}{\Delta x}\left[\lambda^{n,x}_{j-1/2,k}(\bar{\boldsymbol{u}}^{S}_{j-1/2,k} + 4\bar{\boldsymbol{u}}^{C}_{j-1/2,k} + \bar{\boldsymbol{u}}^{N}_{j-1/2,k})\right. \\
&\qquad\qquad + \lambda^{n,x}_{j+1/2,k}(\bar{\boldsymbol{u}}^{S}_{j+1/2,k} + 4\bar{\boldsymbol{u}}^{C}_{j+1/2,k} + \bar{\boldsymbol{u}}^{N}_{j+1/2,k}) \\
&\qquad\qquad + 2\lambda^{\max,x}_{j,k}(\bar{\boldsymbol{u}}^{S}_{j,k} + 4\bar{\boldsymbol{u}}^{C}_{j,k} + \bar{\boldsymbol{u}}^{N}_{j,k}) \\
&\qquad\qquad + (\lambda^{\max,x}_{j,k} - \lambda^{n,x}_{j-1/2,k})(\boldsymbol{u}^{SW}_{j,k} + 4\boldsymbol{u}^{CW}_{j,k} + \boldsymbol{u}^{NW}_{j,k}) \\
&\qquad\qquad \left. + (\lambda^{\max,x}_{j,k} - \lambda^{n,x}_{j+1/2,k})(\boldsymbol{u}^{SE}_{j,k} + 4\boldsymbol{u}^{CE}_{j,k} + \boldsymbol{u}^{NE}_{j,k})\right] \\
&\quad + \frac{1}{6}\frac{\Delta t}{\Delta y}\left[\lambda^{n,y}_{j,k-1/2}(\bar{\boldsymbol{u}}^{W}_{j,k-1/2} + 4\bar{\boldsymbol{u}}^{C}_{j,k-1/2} + \bar{\boldsymbol{u}}^{E}_{j,k-1/2})\right. \\
&\qquad\qquad + \lambda^{n,y}_{j,k+1/2}(\bar{\boldsymbol{u}}^{W}_{j,k+1/2} + 4\bar{\boldsymbol{u}}^{C}_{j,k+1/2} + \bar{\boldsymbol{u}}^{E}_{j,k+1/2}) \\
&\qquad\qquad + 2\lambda^{\max,y}_{j,k}(\bar{\boldsymbol{u}}^{W}_{j,k} + 4\bar{\boldsymbol{u}}^{C}_{j,k} + \bar{\boldsymbol{u}}^{E}_{j,k}) \\
&\qquad\qquad + (\lambda^{\max,y}_{j,k} - \lambda^{n,y}_{j,k-1/2})(\boldsymbol{u}^{SW}_{j,k} + 4\boldsymbol{u}^{SC}_{j,k} + \boldsymbol{u}^{SE}_{j,k}) \\
&\qquad\qquad \left. + (\lambda^{\max,y}_{j,k} - \lambda^{n,y}_{j,k+1/2})(\boldsymbol{u}^{NW}_{j,k} + 4\boldsymbol{u}^{NC}_{j,k} + \boldsymbol{u}^{NE}_{j,k})\right]
\end{aligned} \tag{4.53}$$

Using interface value defined by (3.63) and the polynomial defined by (3.57)-(3.59), we have that

$$\begin{aligned} &(\boldsymbol{u}_{j,k}^{SW}+\boldsymbol{u}_{j,k}^{SE})+4(\boldsymbol{u}_{j,k}^{CW}+\boldsymbol{u}_{j,k}^{CE})+(\boldsymbol{u}_{j,k}^{NW}+\boldsymbol{u}_{j,k}^{NE}) \\ =\ & 12\widehat{\boldsymbol{u}}_{j,k}^{n}+\frac{3\Delta x^2}{2}(\boldsymbol{u}_{xx}^{n})_{j,k}+\frac{\Delta y^2}{2}(\boldsymbol{u}_{yy}^{n})_{j,k}. \end{aligned} \tag{4.54}$$

and

$$\begin{aligned} &(\boldsymbol{u}_{j,k}^{SW}+\boldsymbol{u}_{j,k}^{NW})+4(\boldsymbol{u}_{j,k}^{SC}+\boldsymbol{u}_{j,k}^{NC})+(\boldsymbol{u}_{j,k}^{SE}+\boldsymbol{u}_{j,k}^{NE}) \\ =\ & 12\widehat{\boldsymbol{u}}_{j,k}^{n}+\frac{\Delta x^2}{2}(\boldsymbol{u}_{xx}^{n})_{j,k}+\frac{3\Delta y^2}{2}(\boldsymbol{u}_{yy}^{n})_{j,k}. \end{aligned} \tag{4.55}$$

Now plug (4.54) and (4.55) into (4.53) and use the fact that $\widehat{\boldsymbol{u}}_{j,k}^{n}=\boldsymbol{u}_{j,k}^{n}-\frac{1}{24}(\Delta x^2(\boldsymbol{u}_{xx}^{n})_{j,k}+\Delta y^2(\boldsymbol{u}_{yy}^{n})_{j,k})$ from (3.59), also notice that by (4.48) we have that $\frac{\lambda_{j,k}^{\max,x}}{\Delta x}=\frac{\lambda_{j,k}^{\max,y}}{\Delta y}=\theta_{j,k}$. It follows that (4.53) could be rewritten as

$$\begin{aligned} \boldsymbol{u}_{j,k}^{n+1} =\ & (1-16\Delta t\theta_{j,k})\boldsymbol{u}_{j,k}^{n}+8\Delta t\theta_{j,k}\widehat{\boldsymbol{u}}_{j,k}^{n} \\ +\ & \frac{\Delta t\lambda_{j-1/2,k}^{n,x}}{\Delta x}\Big(\frac{\bar{\boldsymbol{u}}_{j-1/2,k}^{S}+4\bar{\boldsymbol{u}}_{j-1/2,k}^{C}+\bar{\boldsymbol{u}}_{j-1/2,k}^{N}}{6}\Big) \\ +\ & \frac{\Delta t\lambda_{j+1/2,k}^{n,x}}{\Delta x}\Big(\frac{\bar{\boldsymbol{u}}_{j+1/2,k}^{S}+4\bar{\boldsymbol{u}}_{j+1/2,k}^{C}+\bar{\boldsymbol{u}}_{j+1/2,k}^{N}}{6}\Big) \\ +\ & 2\Delta t\theta_{j,k}\Big(\frac{\bar{\boldsymbol{u}}_{j,k}^{S}+4\bar{\boldsymbol{u}}_{j,k}^{C}+\bar{\boldsymbol{u}}_{j,k}^{N}}{6}\Big) \\ +\ & \Big(\Delta t\theta_{j,k}-\frac{\Delta t\lambda_{j-1/2,k}^{n,x}}{\Delta x}\Big)\Big(\frac{\boldsymbol{u}_{j,k}^{SW}+4\boldsymbol{u}_{j,k}^{CW}+\boldsymbol{u}_{j,k}^{NW}}{6}\Big) \\ +\ & \Big(\Delta t\theta_{j,k}-\frac{\Delta t\lambda_{j+1/2,k}^{n,x}}{\Delta x}\Big)\Big(\frac{\boldsymbol{u}_{j,k}^{SE}+4\boldsymbol{u}_{j,k}^{CE}+\boldsymbol{u}_{j,k}^{NE}}{6}\Big) \\ +\ & \frac{\Delta t\lambda_{j,k-1/2}^{n,y}}{\Delta y}\Big(\frac{\bar{\boldsymbol{u}}_{j,k-1/2}^{W}+4\bar{\boldsymbol{u}}_{j,k-1/2}^{C}+\bar{\boldsymbol{u}}_{j,k-1/2}^{E}}{6}\Big) \\ +\ & \frac{\Delta t\lambda_{j,k+1/2}^{n,y}}{\Delta y}\Big(\frac{\bar{\boldsymbol{u}}_{j,k+1/2}^{W}+4\bar{\boldsymbol{u}}_{j,k+1/2}^{C}+\bar{\boldsymbol{u}}_{j,k+1/2}^{E}}{6}\Big) \\ +\ & 2\Delta t\theta_{j,k}\Big(\frac{\bar{\boldsymbol{u}}_{j,k}^{W}+4\bar{\boldsymbol{u}}_{j,k}^{C}+\bar{\boldsymbol{u}}_{j,k}^{E}}{6}\Big) \\ +\ & \Big(\Delta t\theta_{j,k}-\frac{\Delta t\lambda_{j,k-1/2}^{n,y}}{\Delta y}\Big)\Big(\frac{\boldsymbol{u}_{j,k}^{SW}+4\boldsymbol{u}_{j,k}^{SC}+\boldsymbol{u}_{j,k}^{SE}}{6}\Big) \\ +\ & \Big(\Delta t\theta_{j,k}-\frac{\Delta t\lambda_{j,k+1/2}^{n,y}}{\Delta y}\Big)\Big(\frac{\boldsymbol{u}_{j,k}^{NW}+4\boldsymbol{u}_{j,k}^{NC}+\boldsymbol{u}_{j,k}^{NE}}{6}\Big) \end{aligned} \tag{4.56}$$

It is clear that (4.56) is a convex combination of values in A if the CFL condition

$\Delta t\theta_{j,k} \leq \frac{1}{16}$ holds. Therefore we have that $\boldsymbol{u}_{j,k}^{n+1}$ is in A, which completes our proof. $\square$

Finally we give the following approach which describes the polynomial limiting algorithm for two dimensional case. For simplicity we denote $[x_{j-1/2}, x_{j+1/2}] \times [y_{k-1/2}, y_{k+1/2}]$ by $I_{j,k}$ in our discussion. Similar to the discussion in § 4.3.2.1, we introduce the auxiliary functions $p_{j,k}^r(x, y)$ such that

$$p_{j,k}(x, y) = \boldsymbol{u}_{j,k}^n + p_{j,k}^r(x, y) \qquad \text{for all } (x, y) \in I_{j,k}, \tag{4.57}$$

and have

$$\int_{I_{j,k}} p_{j,k}^r(x, y) = 0. \tag{4.58}$$

Then by introducing $l_{j,k} \in [0, 1]$ to be the polynomial limiter, we could give the limited polynomial reconstruction as

$$p_j^l(x, y) := \boldsymbol{u}_{j,k}^n + l_{j,k} p_{j,k}^r(x, y). \tag{4.59}$$

Notice that by (4.58) we have

$$\int_{I_{j,k}} p_{j,k}^l(x, y) = \boldsymbol{u}_{j,k}^n, \tag{4.60}$$

which implies that $p_{j,k}^l(x, y)$ is conservative on $I_{j,k}$. Now by setting

$$\begin{aligned}
p_{j,k}^{r,NC} &:= p_{j,k}^r(x_j, y_{k+1/2}), & p_{j,k}^{r,SC} &:= p_{j,k}^r(x_j, y_{k-1/2}),\\
p_{j,k}^{r,CE} &:= p_{j,k}^r(x_{j+1/2}, y_k), & p_{j,k}^{r,CW} &:= p_{j,k}^r(x_{j-1/2}, y_k),\\
p_{j,k}^{r,NE} &:= p_{j,k}^r(x_{j+1/2}, y_{k+1/2}), & p_{j,k}^{r,NW} &:= p_{j,k}^r(x_{j-1/2}, y_{k+1/2}),\\
p_{j,k}^{r,SE} &:= p_{j,k}^r(x_{j+1/2}, y_{k-1/2}), & p_{j,k}^{r,SW} &:= p_{j,k}^r(x_{j-1/2}, y_{k-1/2}),
\end{aligned} \tag{4.61}$$

we could define the limited interface values as

$$\boldsymbol{u}_{j,k}^{XY}(l_{j,k}) = \boldsymbol{u}_{j,k}^n + l_{j,k} p_{j,k}^{r,XY}, \tag{4.62}$$

where $XY \in I_{XY} := \{SW, SC, SE, CW, CE, NW, NC, NE\}$. Also we define the limited auxiliary state to be

$$\widehat{\boldsymbol{u}}^n_{j,k}(l_{j,k}) = \boldsymbol{u}^n_{j,k} + l_{j,k}(-\frac{1}{24}(\Delta x^2\big(\boldsymbol{u}^n_{xx})_{j,k} + \Delta y^2(\boldsymbol{u}^n_{yy})_{j,k})\big), \tag{4.63}$$

for $\widehat{\boldsymbol{u}}^n_{j,k}$ is defined by (3.59). Similar to the one dimensional case, for a given constraint z and index j, k, let $A^{z,XY}_{j,k}$ be the local invariant sets for the $(j,k)-$cell, such that

$$\begin{aligned}
&A^{z,XY}_{j,k} \in A, && \text{for } X \in \{S, C, N\}, && Y \in \{W, C, E\},\\
&A^{z,NE}_{j-1,k-1} \in A, && A^{z,NC}_{j,k-1} \in A, && A^{z,NW}_{j+1,k-1} \in A,\\
&A^{z,CE}_{j-1,k} \in A, && A^{z,CW}_{j+1,k} \in A, && \\
&A^{z,SE}_{j-1,k+1} \in A, && A^{z,SC}_{j,k+1} \in A, && A^{z,SW}_{j+1,k+1} \in A.
\end{aligned} \tag{4.64}$$

Then using Lemma 4.3.2 and Lemma 4.3.3, we could find largest positive $l_{j,k} \in [0,1]$ such that

$$\begin{aligned}
&\boldsymbol{u}^{XY}_{j,k}(l_{j,k}) \in A^{z,XY}_{j,k}, && \text{for } XY \in I_{XY},\\
&\widehat{\boldsymbol{u}}^n_{j,k}(l_{j,k}) \in A^{z,CC}_{j,k},
\end{aligned} \tag{4.65}$$

and

$$\begin{aligned}
&\boldsymbol{u}^{NE}_{j-1,k-1}(l_{j-1,k-1}) \in A^{z,NE}_{j-1,k-1}, && \boldsymbol{u}^{NC}_{j,k-1}(l_{j,k-1}) \in A^{z,NC}_{j,k-1},\\
&\boldsymbol{u}^{NW}_{j+1,k-1}(l_{j+1,k-1}) \in A^{z,NW}_{j+1,k-1}, && \\
&\boldsymbol{u}^{CE}_{j-1,k}(l_{j-1,k}) \in A^{z,CE}_{j-1,k}, && \boldsymbol{u}^{CW}_{j+1,k}(l_{j+1,k}) \in A^{z,CW}_{j+1,k},\\
&\boldsymbol{u}^{SE}_{j-1,k+1}(l_{j-1,k+1}) \in A^{z,SE}_{j-1,k+1}, && \boldsymbol{u}^{SC}_{j,k+1}(l_{j,k+1}) \in A^{z,SC}_{j,k+1},\\
&\boldsymbol{u}^{SW}_{j+1,k+1}(l_{j+1,k+1}) \in A^{z,SW}_{j+1,k+1},
\end{aligned} \tag{4.66}$$

hold for all $j, k \in \mathbb{Z}$. Therefore, we claim that the limited solution $\boldsymbol{u}^{n+1}_{j,k}$ is in A.

4.4 Application to Compressible Euler Systems

In this section, we will illustrate how to apply the three types of convex limiting processes stated in § 4 for hyperbolic systems discussed in § 1.3. More specifically, we will explain how to implement the quasiconcave limitations defined in § 4. In general the constraints we considered could be divided into linear constraints and nonlinear concave constraints. Examples for linear constraints are maximum/minimum principle for u in scalar equations and for the density ρ in Euler systems. Example for nonlinear concave constraints are the constraints on Riemann invariants w_1, w_2 in the P-system and minimum principle of specific entropy s in the Euler system. For simplicity we will only explain how to apply the limiting process on the compressible Euler system, which covers both the linear constraint and nonlinear concave constraint. In all our discussion, we have to modify some of the denominators in order to avoid division by zero. A typical way is to add a small number to the denominator. Usually this small number is computed by timing ϵ to a fixed bound of the local invariant set, where $\epsilon = 10^{-16}$ is the machine error. Still, our discussion will be restricted to one space dimension.

Now let's we consider the compressible Euler system (1.16) and the quasiconcave constraints defined in § 4.1.3. We illustrate the flux limiting, slope limiting and polynomial limiting separately as follows.

4.4.1 Flux limiting

Using the same notation as in § 4.1.1, we define $G^n_{j+1/2} := (G^{\rho,n}_{j+1/2}, G^{m,n}_{j+1/2}, G^{E,n}_{j+1/2})^\top$, $\boldsymbol{u}^+_j(l) := (\rho^+_j(l), m^+_j(l), E^+_j(l))^\top$ and $\boldsymbol{u}^-_j(l) := (\rho^-_j(l), m^-_j(l), E^-_j(l))^\top$, where $\boldsymbol{u}^+_j(l) = \boldsymbol{u}^{L,n+1}_j - l^+_j G^n_{j+1/2}$ and $\boldsymbol{u}^-_j(l) = \boldsymbol{u}^{L,n+1}_j + l^-_j G^n_{j-1/2}$. Then we perform the flux limiting process as follows:

4.4.1.1 Limitation On Density

To limit the first component of the density ρ, we first set

$$\psi^{1,+}_j(l) := \Psi^1_j(\rho^+_j(l)) = \rho^{n,\max}_j - \rho^{L,n+1}_j + lG^{\rho,n}_{j+1/2}, \tag{4.67a}$$

$$\psi^{2,+}_j(l) := \Psi^2_j(\rho^+_j(l)) = \rho^{L,n+1}_j - lG^{\rho,n}_{j+1/2} - \rho^{n,\min}_j, \tag{4.67b}$$

and compute the limiters for ρ_j^+ by

$$l_j^{\rho,+} = \begin{cases} \min(\frac{|\rho_j^{n,\max} - \rho_j^{L,n+1}|}{|G_{j+1/2}^{\rho,n}| + \epsilon_j^\rho}, 1) & \text{if } \psi_j^{1,+}(1) < 0, \\ 1 & \text{if } \psi_j^{1,+}(1) \geq 0 \;\&\; \psi_j^{2,+}(1) \geq 0, \\ \min(\frac{|\rho_j^{L,n+1} - \rho_j^{n,\min}|}{|G_{j+1/2}^{\rho,n}| + \epsilon_j^\rho}, 1) & \text{if } \psi_j^{2,+}(1) < 0, \end{cases} \tag{4.68}$$

where we take $\epsilon_j^\rho = \epsilon \rho_j^{n,\max}$ to avoid division by zero. For the second component of ρ, we set

$$\psi_j^{1,-}(l) := \Psi_j^1(\rho_j^-(l)) = \rho_j^{n,\max} - \rho_j^{L,n+1} + lG_{j-1/2}^{\rho,n}, \tag{4.69a}$$

$$\psi_j^{2,-}(l) := \Psi_j^2(\rho_j^-(l)) = \rho_j^{L,n+1} - lG_{j-1/2}^{\rho,n} - \rho_j^{n,\min}, \tag{4.69b}$$

and compute $l_j^{\rho,-}$ using the same approach. Therefore, the limiter on density is defined by $l_j^\rho = \min(l_j^{\rho,+}, l_j^{\rho,-})$.

4.4.1.2 Limitation On Specific Entropy

For limitation on the specific entropy s, we use functions $\psi_j^{3,+}(l) := \Psi_j^3(\boldsymbol{u}_j^+(l)) = (\rho_j^{L,n+1} - lG_{j+1/2}^{\rho,n})(e(\boldsymbol{u}_j^{L,n+1} - lG_{j+1/2}^n)) - c_j^{n,\min}(\rho_j^{L,n+1} - lG_{j+1/2}^{\rho,n})^\gamma$ and $\psi_j^{3,-}(l) := \Psi_j^3(\boldsymbol{u}_j^+(l)) = (\rho_j^{L,n+1} - lG_{j-1/2}^{\rho,n})(e(\boldsymbol{u}_j^{L,n+1} - lG_{j-1/2}^n)) - c_j^{n,\min}(\rho_j^{L,n+1} - lG_{j-1/2}^{\rho,n})^\gamma$, both of which are concave down of l. We define $l_j^{s,+}$ as follows. If $\psi_j^{3,+}(\min(1, l_j^{\rho,+})) \geq 0$, we take $l_j^{s,+} = \min(1, l_j^{\rho,+})$; if $\psi_j^{3,+}(0) > 0$ and $\psi_j^{3,+}(\min(1, l_j^{\rho,+})) < 0$, we define $l_j^{s,+}$ to be the unique positive root of $\psi_j^3(l) = 0$; if $\psi_j^3(0) = 0$ and $\psi_j^3(\min(1, l_j^{\rho,+})) < 0$, then $\psi_j^3(l) = 0$ has exactly two roots and we take $l_j^{s,+}$ to be the largest nonnegative root of $\psi_j^3(l) = 0$. Also we compute $l_j^{s,-}$ using the same approach. Therefore we define the limiter for the specific entropy to be $l_j^s = \min(l_j^{s,+}, l_j^{s,-})$.

Remark. If $\psi(l)$ is a concave down function of l, $l_a, l_b \in \mathbb{R}$ be such that $\psi(l_a) > 0$ and $\Psi(l_b) < 0$, we could combine the secant method and the Newton method to compute the root l_0 of f with $l_a < l_0 < l_b$. By doing so we could always guarantee that $\psi(l_0) \geq 0$, which implies that $\boldsymbol{u}(l_0)$ is inside or on the boundary of the zero level set of ψ.

Remark. Recall that the constraint $\rho e - c_j^{n,\min}\rho^\gamma \geq 0$ guarantees $e \geq e_j^{n,\min}$ only if we have already limited ρ such that $\rho \geq 0$, see § 4.1.3. Therefore we use $\min(1, l_j^\rho)$ instead of 1 in our limiting process of the specific entropy

4.4.2 Slope limiting

We first construct the local invariant constraints at cell interface $x_{j-1/2}$ by setting

$$\begin{aligned} &\rho_{j-1/2}^{n,\min} := \min(\rho_j^n, \rho_{j-1}^n, \bar{\rho}_{j-1/2}^{n+1}), \qquad \rho_{j-1/2}^{n,\max} := \max(\rho_j^n, \rho_{j-1}^n, \bar{\rho}_{j-1/2}^{n+1}), \\ &s_{j-1/2}^{n,\min} := \min(\Phi(\boldsymbol{u}_j^n), \Phi(\boldsymbol{u}_{j-1}^n), \Phi(\bar{\boldsymbol{u}}_{j-1/2}^{n+1})), \end{aligned} \tag{4.70}$$

where $\Phi(\boldsymbol{u}) = s(\rho, e)$ as defined in § 4.4.1. Analogous to the flux limiting case, we set

$$\begin{aligned} &\Psi_{j-1/2}^1(\boldsymbol{u}) = u_{j-1/2}^{n,\max} - \rho, \qquad \Psi_{j-1/2}^2(\boldsymbol{u}) = \rho - \rho_{j-1/2}^{n,\min}, \\ &\Psi_{j-1/2}^3(\boldsymbol{u}) = \rho e - c_{j-1/2}^{n,\min}\rho^\gamma, \end{aligned} \tag{4.71}$$

where $c_{j-1/2}^{n,\min} = \exp((\gamma-1)s_{j-1/2}^{n,\min})$. We impose the invariant domain property by enforcing $\Psi_{j-1/2}^1(u), \Psi_{j-1/2}^2(u), \Psi_{j-1/2}^3(u) \geq 0$ on each interface. Using the same notation as in §4.3.1, we denote $\sigma_j^{\mathrm{a}} := (\sigma_j^{\mathrm{a},\rho}, \sigma_j^{\mathrm{a},m}, \sigma_j^{\mathrm{a},E})^\top$ to be the unlimited central slope and define the limited interface values as follows

$$\boldsymbol{u}_{j+1/2}^-(l) := \boldsymbol{u}_j^n + \frac{\Delta x}{2} l\sigma_j^{\mathrm{a}}, \qquad \boldsymbol{u}_{j-1/2}^+(l) := \boldsymbol{u}_j^n - \frac{\Delta x}{2} l\sigma_j^{\mathrm{a}}. \tag{4.72}$$

Thus we could perform the slope limiting to the solutions on the boundary.

4.4.2.1 Limitation On Density

Depending on the sign of σ_j^ρ, we limit the density ρ as follows:

(i) If $\sigma_j^\rho > 0$, we set

$$l_j^\rho = \min(\frac{|\rho_{j+1/2}^{n,\max} - \rho_j^n|}{|\frac{\Delta x}{2}\sigma_j^{a,\rho}| + \epsilon_{j+1/2}^\rho}, \frac{|\rho_j^n - \rho_{j-1/2}^{n,\min}|}{|\frac{\Delta x}{2}\sigma_j^{a,\rho}| + \epsilon_{j-1/2}^\rho}, 1), \tag{4.73}$$

where $\epsilon^{\rho}_{j+1/2} = \epsilon\rho^{n,\max}_{j+1/2}$ and $\epsilon^{\rho}_{j-1/2} = \epsilon\rho^{n,\min}_{j-1/2}$.

(ii) If $\sigma^{\rho}_j < 0$, we set

$$l^{\rho}_j = \min\big(\frac{|\rho^n_j - \rho^{n,\min}_{j+1/2}|}{|\frac{\Delta x}{2}\sigma^{a,\rho}_j| + \epsilon^{\rho}_{j+1/2}}, \frac{|\rho^{n,\max}_{j-1/2} - \rho^n_j|}{|\frac{\Delta x}{2}\sigma^{a,\rho}_j| + \epsilon^{\rho}_{j-1/2}}, 1\big). \tag{4.74}$$

where $\epsilon^{\rho}_{j+1/2} = \epsilon\rho^{n,\min}_{j+1/2}$ and $\epsilon^{\rho}_{j-1/2} = \epsilon\rho^{n,\max}_{j-1/2}$.

(iii) If $\sigma^{\rho}_j = 0$, we take $l^{\rho}_j = 1$.

4.4.2.2 *Limitation On Specific Entropy*

For limitation on the specific entropy s, we will enforce $\Psi^3_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l)) \geq 0$ and $\Psi^3_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l)) \geq 0$. Therefore we denote $\psi^+_j(l) = \Psi^3_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l))$ and $\psi^-_j(l) = \Psi^3_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l))$. Note that ψ^+_j and ψ^-_j are both concave-down of l. Following the same approach as in §4.4.1.2, we compute $l^{s,+}_j$ and $l^{s,-}_j$ using the Newton-Secant approach and the slope limiter for density l^{ρ}_j. Then, the slope limiter for the specific entropy is defined as $l^s_j = \min(l^{s,+}_j, l^{s,-}_j)$. After all limitation we obtain the following result.

Lemma 4.4.1. Let $l_j = l^s_j$ for all $j \in \mathbb{Z}$, then for any $l \in [0, l_j]$ and $z = 1, 2, 3$, we have that $\Psi^z_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l)) \geq 0$ and $\Psi^z_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l)) \geq 0$.

4.4.3 Polynomial Limiting

The polynomial limiting process is similar to the slope limiting process. The difference is that we also have to limit the auxiliary value $\widehat{\boldsymbol{u}}^n_j$.

Similar to § 4.4.2, we first set the invariant constraints at cell interface to be

$$\begin{aligned} &\rho^{n,\min}_{j-1/2} := \min(\rho^n_j, \rho^n_{j-1}, \bar{\rho}^{n+1}_{j-1/2}), \qquad \rho^{n,\max}_{j-1/2} := \max(\rho^n_j, \rho^n_{j-1}, \bar{\rho}^{n+1}_{j-1/2}), \\ &s^{n,\min}_{j-1/2} := \min(\Phi(\boldsymbol{u}^n_j), \Phi(\boldsymbol{u}^n_{j-1}), \Phi(\bar{\boldsymbol{u}}^{n+1}_{j-1/2})), \end{aligned} \tag{4.75}$$

where $\Phi(\boldsymbol{u}) = s(\rho, e)$ as defined in § 4.4.1.while the corresponding quasiconcave constraints are written by

$$\begin{aligned}
&\Psi^1_{j-1/2}(\boldsymbol{u}) = u^{n,\max}_{j-1/2} - \rho, \qquad \Psi^2_{j-1/2}(\boldsymbol{u}) = \rho - \rho^{n,\min}_{j-1/2},\\
&\Psi^3_{j-1/2}(\boldsymbol{u}) = \rho e - c^{n,\min}_{j-1/2}\rho^{\gamma}.
\end{aligned} \tag{4.76}$$

Using the same notation in (4.43), see § 4.3.2.1, we define

$$\boldsymbol{u}^-_{j+1/2}(l) := \boldsymbol{u}^n_j + lp^{r,+}_j, \qquad \boldsymbol{u}^+_{j-1/2}(l) := \boldsymbol{u}^n_j + lp^{r,-}_j, \tag{4.77}$$

where $p^{r,+}_j := (p^{\rho,r,+}_j, p^{m,r,+}_j, p^{E,r,+}_j)^\top$ and $p^{r,-}_j := (p^{\rho,r,-}_j, p^{m,r,-}_j, p^{E,r,-}_j)^\top$.

For the auxiliary value $\widehat{\boldsymbol{u}}$, we set the invariant constraints to be

$$\begin{aligned}
&\rho^{n,\min}_j := \min(\rho^n_j, \bar{\rho}^{n+1}_{j+1/2}, \bar{\rho}^{n+1}_{j-1/2}), \qquad \rho^{n,\max}_j := \max(\rho^n_j, \bar{\rho}^{n+1}_{j+1/2}, \bar{\rho}^{n+1}_{j-1/2}),\\
&s^{n,\min}_j := \min(s^n_j, \bar{s}^{n+1}_{j+1/2}, \bar{s}^{n+1}_{j-1/2})
\end{aligned} \tag{4.78}$$

and set the corresponding quasiconcave constraints to be

$$\begin{aligned}
&\Psi^1_j(\boldsymbol{u}) = \rho^{n,\max}_j - \rho, \qquad \Psi^2_j(\boldsymbol{u}) = \rho - \rho^{n,\min}_j,\\
&\Psi^3_j(\boldsymbol{u}) = \rho e - c^{n,\min}_j \rho^{\gamma},
\end{aligned} \tag{4.79}$$

where $c^{n,\min}_j = \exp((\gamma-1)s^{n,\min}_j)$. Using the notation in (4.46), see § 4.3.2.1, we define

$$\widehat{\boldsymbol{u}}^n_j(l_j) = \boldsymbol{u}^n_j + l\boldsymbol{u}^r_j, \tag{4.80}$$

where $\boldsymbol{u}^r_j := (\boldsymbol{u}^{\rho,r}_j, \boldsymbol{u}^{m,r}_j, \boldsymbol{u}^{E,r}_j)^\top$.

Using the notations defined above, we are able to limit the density and specific entropy separately as follows.

4.4.3.1 Limitation On Density

Here we apply the limiting process on density ρ. Note that it includes a limitation on the interface values and a limitation on the auxiliary values.

(i) We first limit the interface values by setting

$$l^{\rho} = \min(l_j^{\rho,+}, l_j^{\rho,-}), \tag{4.81}$$

where

$$l_j^{\rho,+} = \begin{cases} \frac{|\rho^{n,\max}_{j+1/2} - p_j^{\rho,r,+}|}{|p_j^{\rho,r,+}| + \epsilon^{\rho}_{j+1/2}} & \text{if } p_j^{\rho,r,+} > 0, \\ \frac{|p_j^{\rho,r,+} - \rho^{n,\min}_{j+1/2}|}{|p_j^{\rho,r,+}| + \epsilon^{\rho}_{j+1/2}} & \text{if } p_j^{\rho,r,+} < 0, \\ 1 & \text{if } p_j^{\rho,r,+} = 0, \end{cases} \tag{4.82}$$

and

$$l_j^{\rho,-} = \begin{cases} \frac{|\rho^{n,\max}_{j-1/2} - p_j^{\rho,r,+}|}{|p_j^{\rho,r,+}| + \epsilon^{\rho}_{j-1/2}} & \text{if } p_j^{\rho,r,-} > 0, \\ \frac{|p_j^{\rho,r,+} - \rho^{n,\min}_{j-1/2}|}{|p_j^{\rho,r,+}| + \epsilon^{\rho}_{j-1/2}} & \text{if } p_j^{\rho,r,-} < 0, \\ 1 & \text{if } p_j^{\rho,r,-} = 0, \end{cases} \tag{4.83}$$

Still we take $\epsilon^{\rho}_{j+1/2} = \epsilon\rho^{n,\max}_{j+1/2}$ or $\epsilon^{\rho}_{j+1/2} = \epsilon\rho^{n,\min}_{j+1/2}$ to avoid division by zero.

(ii) To limit the auxiliary values we set

$$l_j^{\rho,c} = \begin{cases} \frac{|\rho_j^{n,\max} - \rho_j^r|}{|\rho_j^r| + \epsilon_j^{\rho}} & \text{if } \rho_j^r > 0, \\ \frac{|\rho_j^r - \rho_j^{n,\min}|}{|\rho_j^r| + \epsilon_j^{\rho}} & \text{if } \rho_j^r < 0, \\ 1 & \text{if } \rho_j^r = 0, \end{cases} \tag{4.84}$$

where $\epsilon_j^{\rho} = \epsilon\rho_j^{n,\max}$ or $\epsilon_j^{\rho} = \epsilon\rho_j^{n,\min}$.

Thus we take the limiter for density to be

$$l_j^{\rho} = \min(l_j^{\rho,+}, l_j^{\rho,-}, l_j^{\rho,c}). \tag{4.85}$$

4.4.3.2 Limitation On Specific Entropy

Still our limitation of the specific entropy will be applied to the interface values and auxiliary values separately.

(i) We limit the interface values by enforcing $\Psi^3_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l)), \Psi^3_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l)) \geq 0$. Similarly we denote $\psi^+_j(l) = \Psi^3_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l))$ and $\psi^-_j(l) = \Psi^3_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l))$, which are both concave-down functions of l. Following the same approach as in §4.4.1.2, we compute $l^{s,+}_j$ and $l^{s,-}_j$ using the Newton-Secant approach with l^ρ_j to be the starting point. Therefore, the polynomial limiter based on the interface values is defined by $l^{s,\pm}_j = \min(l^{s,+}_j, l^{s,-}_j)$.

(ii) We limit the auxiliary value in a similar way, that is, we enforce $\Psi^3_j(\widehat{\boldsymbol{u}}^n_j(l)) \geq 0$. Similarly we take $\psi_j(l) := \Psi^3_j(\widehat{\boldsymbol{u}}^n_j(l))$, which is a concave down function of l. Still Newton-Secant method with l^ρ_j will be used to compute $l^{s,c}_j$.

A straightforward result of the limitation discussed above is shown as follows

Lemma 4.4.2. Let $l_j = l^s_j := \min(l^{s,+}_j, l^{s,-}_j, l^{s,c}_j)$ for all $j \in \mathbb{Z}$, then for any $l \in [0, l_j]$, we have $\Psi^z_{j-1/2}(\boldsymbol{u}^+_{j-1/2}(l)) \geq 0$, $\Psi^z_{j+1/2}(\boldsymbol{u}^-_{j+1/2}(l)) \geq 0$ and $\Psi^{z,r}_j(\widehat{u}^n_j(l)) \geq 0$, $\Psi^{z,r}_j(\widehat{u}^n_j(l)) \geq 0$ for $z = 1, 2, 3$.

4.5 Local Relaxations

It has been observed in many instances that enforcing strict local bounds in the limiting process may reduce the approximation accuracy of the scheme. In the scalar case it is well known that enforcing strict local maximum principle will lead to the so-called clipping phenomenon, therefore the rate of $L^\infty-$error is reduced. In the case of systems, such effects can be observed when the local states are close to the boundary of the invariant domain. One could read [37, § 3.3] and [27, § 4.7] for further discussions on the Euler system, especially for the relaxation of the minimum principle on the specific entropy. In this section, we follow the approach from [27, §4.7], which was originally proposed for the Euler system, and apply it in more general settings. The core idea of this approach is to design a local relaxation for the constraints defined in § 4.1. By adding this relaxation to the constraints, we've observed in all numerical tests that the limited method still keeps the accuracy of the unlimited method. For simplicity we restrict our discussion to one dimensional case while the case of two space dimensions is actually analogous.

To start with, let Ω be the computational domain in $\mathbb{R}$ and h denote the mesh size. That is, we take $h := \frac{\Delta x}{|\Omega|}$ with $|\Omega|$ being the diameter of the set. Le z denote the quantity to be limited. To simplify our discussion into one single case, a modification will be applied to the sigh of each constraint. For example, z could be $-w_1, w_2$ for the P-system, see §4.1.2, or z could be $\rho, -\rho, s$ for Euler system, see §4.1.3. We give the following two types of relaxations.

(i) Limitation on a constraint z which describes a smooth curved part of the boundary. For example, $z = -w_1, w_2$ in the p-system, or $z = s$ in the Euler system, let $x_{ij} = \frac{1}{2}(x_i + x_j)$ and denote $\mathcal{I}_j^k$ the index set of the local stencil. For example, $\mathcal{I}_j^2 = \{j, j \pm 1\}$. We then define $\overline{\Delta z_j^n} := \max_{j \neq i \in I_j}(z^n(x_{ij}) - z_j^{\min})$ and set

$$\overline{z_j^{\min,1}} := z_j^{\min} - \min(r_h|z_j^{\min}|, |\overline{\Delta z_j^n}|). \tag{4.86}$$

Then $\overline{z_j^{\min,1}}$ will be used instead of $z_j^{\min}$ as the bound of the local invariant sets.

(ii) Limitation on constraint z which describes the linear part on the boundary. For example, $z = u$ or $z = -u$ in scalar equations, and $z = \pm\rho$ in the Euler system. Setting $\Delta^k z_j^n$ to be the $k-$th difference of z_j^n defined locally, we then define

$$\overline{\Delta^2 z_j^n} := \frac{1}{2\text{card}(\mathcal{I}_j)} \sum_{j \neq i \in \mathcal{I}_j} (\frac{1}{k!}\Delta^k z_i^n + \frac{1}{k!}\Delta^k z_j^n) \tag{4.87}$$

and

$$\widetilde{\Delta^2 z_j^n} := \mathrm{m}\{\frac{1}{k!}\Delta^k z_i^n | i \in \mathcal{I}_j\}, \tag{4.88}$$

where m is the minmod operator defined in § 3.3.3. The relaxed bound of local invariant set is defined as

$$\overline{z_j^{\min,2}} := z_j^{\min} - \min(r_h|z_j^{\min}|, |\overline{\Delta^2 z_j^n}|), \tag{4.89a}$$

$$\widetilde{z_j^{\min,2}} := z_j^{\min} - \min(r_h|z_j^{\min}|, |\widetilde{\Delta^2 z_j^n}|). \tag{4.89b}$$

Then $\overline{z_j^{\min,2}}$ or $\widetilde{z_j^{\min,2}}$ will be used instead of $z_j^{\min}$ as the bounds of local invariant sets. It is observed in the numerical tests that both of these two bounds defined by (4.89) are robust and give similar results.

Remark. In both (4.86) and (4.89), we take $r_h = \min(1, h^{1.5})$ to restrict the relaxation by $\mathcal{O}(h^{1.5})$. This restriction will enforce the original invariant domain as h approaches to zero, one could see [27, §4.7] for more details.

5. NUMERICAL ILLUSTRATIONS

In this chapter, we report numerical test to illustrate the performance of the two limiting techniques mentioned in § 4. Our code is constructed using finite volume method on uniform cells of size $\Delta x = h$ (in one dimensional case) or $\Delta x = \Delta y = h$ (in two dimensional case). Time stepping is done by using SSP-RK3 methods, which is three stages and of third-order accuracy, see [56] and [39]. In our computation, we use a adaptive time step computed by $\Delta t^n = \text{CFL} \times \frac{h}{\lambda^{\max,n}}$, where $\lambda^{\max,n} := \max_j\{\lambda^n_{j+1/2}\}$ is the global maximum speed of the method used at time t^n. For any $p \in [1,\infty]$,we introduce a consolidated error indicator at time t by adding the relative error in L^p-norm of all conserved variables:

$$\delta_p^{k,\alpha}(t) = \sum_i \frac{||\boldsymbol{u}_h^i(t) - \boldsymbol{u}^i(t)||_{L^p(D)}}{||\boldsymbol{u}^i(t)||_{L^p(D)}}, \qquad \boldsymbol{u} = (u^1, u^2, \ldots, u^m), \tag{5.1}$$

where $\alpha = f$ or s, corresponding to the flux limiting or the slope/polynomial limiting error and k is the approximation order. For the Euler system, in all tests we take the EOS to be the γ-law, i.e., $p = (\gamma-1)\rho e$. Four new limiting methods are to be tested: for KT-scheme we test MAPR-EV-CL for flux limiting and SO-INV-CL for slope limiting; for third order central scheme we test POL-EV-CL for flux limiting and POL-INV-CL for slope limiting. The classical Minmod method (i.e., we apply the minmod slope limiter given in § 3.3.3 in the KT-scheme) will be also reported for comparison, while the flux limiting process is applied to guarantee the invariant domain property. Note that the MAPR-EV limiter defined in (3.38) is only applied when $\theta_j^n \leq 1.5$ and in the regions of smooth flow we use $\sigma^{\text{mapr},\theta} = \frac{\boldsymbol{u}^n_{j+1} - \boldsymbol{u}^n_{j-1}}{2h}$ for KT-scheme and the optimal polynomial p^{opt} for general high order schemes. The relaxation we used in the report are defined by (4.86) for the linear part of the boundary of local invariant set and (4.89a) for the nonlinear part.

5.1 Linear Transport Equation

We first consider the one dimensional linear transport equation (1.8) with $c = 1$, i.e., $u_t + u_x = 0$. We consider a smooth initial condition from [39]

$$u_0(x) = u(x, 0) = \sin(\pi x - \sin(\pi x)/\pi) \qquad x \in [-1, 1]. \tag{5.2}$$

The exact solution is given by

$$u_0(x, t) = \sin(\pi(x - t) - \sin(\pi(x - t))/\pi) \qquad x \in [-1, 1]. \tag{5.3}$$

The computation is done on $[0, 1]$ for $0 \leq t \leq 0.5$. The results is shown in Table 5.1-Table 5.4. It is observed that we've recovered fully second order and third order convergence rate for both L^1-error and $L^\infty-$error with our schemes. The convex limiting process doesn't affect the approximation accuracy at all.

Table 5.1: 1D linear transport equation, L^1-convergence test for second order central scheme with CFL=0.25.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	2.86E-03		9.37E-04		4.68E-04	
200	7.59E-04	1.91	2.33E-04	2.01	1.16E-04	2.01
400	2.05E-04	1.89	5.81E-05	2.00	2.91E-05	2.00
800	5.41E-05	1.92	1.44E-05	2.01	7.26E-06	2.00
1600	1.42E-05	1.93	3.61E-06	2.00	1.81E-06	2.00
3200	3.68E-06	1.95	9.01E-07	2.00	4.54E-07	2.00
6400	9.48E-07	1.96	2.25E-07	2.00	1.13E-07	2.00

Table 5.2: 1D linear transport equation, L^1-convergence test for third order central scheme with CFL=0.25.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,s}(t)$	rate
100	5.17E-05		5.17E-05	
200	6.50E-06	2.99	6.50E-06	2.99
400	8.14E-07	3.00	8.14E-07	3.00
800	1.02E-07	3.00	1.02E-07	3.00
1600	1.28E-08	3.00	1.28E-08	3.00
3200	1.59E-09	3.00	1.59E-09	3.00
6400	1.99E-10	3.00	1.99E-10	3.00

Table 5.3: 1D linear transport equation, $L^\infty-$convergence test for second order central scheme with CFL=0.25.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,s}(t)$	rate
100	1.52E-02		6.10E-03		4.68E-04	
200	6.23E-03	1.29	2.65E-03	1.20	1.16E-04	2.01
400	2.51E-03	1.31	1.03E-03	1.36	2.91E-05	2.00
800	1.02E-03	1.30	3.92E-04	1.39	7.26E-06	2.00
1600	4.10E-04	1.32	1.55E-04	1.34	1.81E-06	2.00
3200	1.65E-04	1.31	5.90E-05	1.39	4.54E-07	2.00
6400	6.60E-05	1.32	2.38E-05	1.31	1.13E-07	2.00

Table 5.4: 1D linear transport equation, $L^\infty-$convergence test for third order central scheme with CFL=0.25.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_\infty^{3,f}(t)$	rate	$\delta_\infty^{3,p}(t)$	rate
100	1.31E-04		1.31E-04	
200	1.64E-05	3.00	1.64E-05	3.00
400	2.05E-06	3.00	2.05E-06	3.00
800	2.57E-07	3.00	2.57E-07	3.00
1600	3.21E-08	3.00	3.21E-08	3.00
3200	4.01E-09	3.00	4.01E-09	3.00
6400	5.03E-10	3.00	5.03E-10	3.00

5.2 Burgers' Equation

We consider the one dimensional Burgers equation defined by (1.10) with the initial condition

$$u(x,0) = \begin{cases} 0 & \text{if } x < 0.25, \\ 4x-1 & \text{if } 0.25 \leq x < 0.5, \\ 1 & \text{if } 0.5 \leq x \leq 1, \end{cases} \tag{5.4}$$

and a exact solution which has a gradient in BV:

$$u(x,t) = \begin{cases} 0 & \text{if } x < 0.25, \\ \dfrac{4x-1}{4t+1} & \text{if } 0.25 \leq x < 0.5+t, \\ 1 & \text{if } 0.5+t \leq x \leq 1. \end{cases} \tag{5.5}$$

The computation is done for $0 \leq t \leq 0.4$ and the results are reported in Table 5.5-Table 5.8 and Figure 5.1-Figure 5.2. We observe that using the method based on the MAPR limiter gives the optimal rate in both L^1 and L^∞. This is a super-convergence effect that we observe for scalar equations, see the middle graph shown in Figure 5.1. However, the methods based on the minmod and the invariant domain slope limiter have a convergence rate of $L-1$error around $\frac{4}{3}$ which is expected for a method based on mass lumping, see [20] for details. Also, using a third order method with the novel polynomial limiter will recover a L^1-convergence rate of 1.5. Moreover, both the convex flux limiting process and the convex slope/polynomial limiting process doesn't affect the convergence rate of the unlimited method.

5.3 The KPP-Test

We consider the so-called KPP-test, a two dimensional scalar conservation equation with a non-convex flux, see [43, §5.3] for more details. This test checks if the high order method has enough viscosity to resolve correctly the composite wave structure of the

Table 5.5: 1D Burgers' equation, L^1-convergence tests for second order central scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	1.17E-03		2.03e-04		6.47E-04	
200	4.24E-04	1.47	5.55e-05	1.87	2.30E-04	1.49
400	1.56E-04	1.44	1.50e-05	1.89	8.47E-05	1.44
800	5.93E-05	1.40	3.96e-06	1.92	3.28E-05	1.37
1600	2.28E-05	1.38	1.04e-06	1.93	1.28E-05	1.35
3200	8.89E-06	1.36	2.72e-07	1.93	5.10E-06	1.33
6400	3.48E-06	1.35	7.03e-08	1.95	2.03E-06	1.33

Table 5.6: 1D Burgers' equation, L^1-convergence tests for third order central scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,p}(t)$	rate
100	4.03E-04		4.33E-04	
200	1.38E-04	1.55	1.46E-04	1.57
400	4.82E-05	1.52	5.03E-05	1.53
800	1.68E-05	1.52	1.74E-05	1.53
1600	5.90E-06	1.51	6.05E-06	1.52
3200	2.06E-06	1.52	2.10E-06	1.52
6400	7.28E-07	1.50	7.38E-07	1.51

Table 5.7: 1D Burgers' equation, $L^\infty-$convergence tests for second order central scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,s}(t)$	rate
100	1.11E-02		4.19E-03		1.11E-02	
200	6.94E-03	0.68	2.33E-03	0.85	6.95E-03	0.68
400	4.33E-03	0.68	1.26E-03	0.88	4.34E-03	0.68
800	2.71E-03	0.68	6.70E-04	0.91	2.71E-03	0.68
1600	1.70E-03	0.68	3.50E-04	0.94	1.70E-03	0.68
3200	1.06E-03	0.67	1.80E-04	0.96	1.06E-03	0.67
6400	6.70E-04	0.67	9.18E-05	0.97	6.70E-04	0.67

Table 5.8: 1D Burgers' equation, $L^{\infty}-$convergence tests for third order central scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_{\infty}^{3,f}(t)$	rate	$\delta_{\infty}^{3,p}(t)$	rate
100	9.35E-03		9.18E-03	
200	5.70E-03	0.71	5.64E-03	0.70
400	3.45E-03	0.73	3.43E-03	0.72
800	2.07E-03	0.73	2.07E-03	0.73
1600	1.24E-03	0.74	1.24E-03	0.74
3200	7.42E-04	0.74	7.43E-04	0.74
6400	4.43E-04	0.74	4.43E-04	0.74

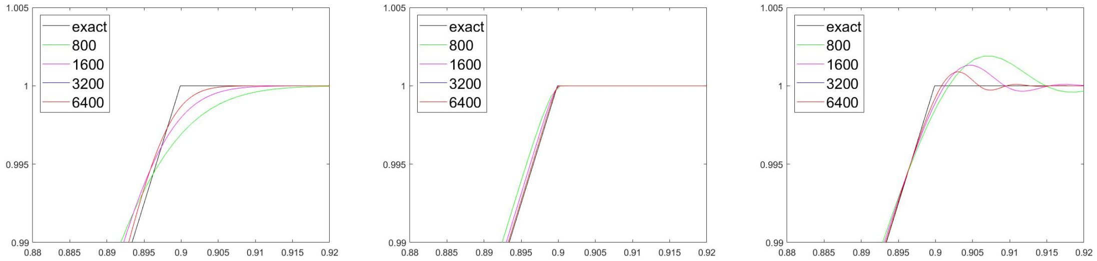


Figure 5.1: Burgers: CFL$= 0.25$, $t = 0.4$. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

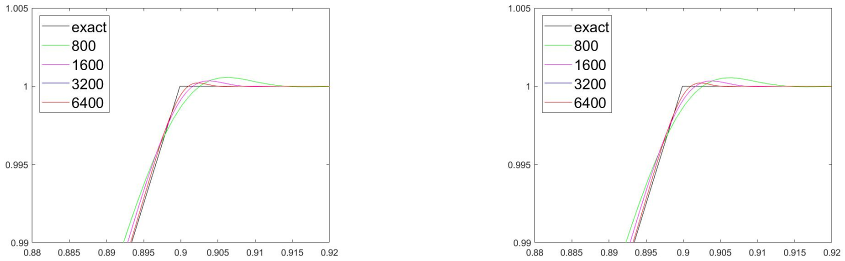


Figure 5.2: Burgers: CFL$= 0.25$, $t = 0.4$. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.

unique entropy solution, see Figure 5.3 and Figure 5.4.

$$\partial_t u + \partial_x \sin u + \partial_y \cos u = 0, \qquad u(x,y,0) = \begin{cases} \frac{14\pi}{4}, & \text{if } \sqrt{x^2+y^2} \leq 1, \\ \frac{\pi}{4}, & \text{otherwise.} \end{cases} \tag{5.6}$$

All schemes are able to resolve correctly the composite wave structure. Note that if we use the MAPR limiter with $\theta = 2$ the method will fail to converge to the correct solution, see [43] for details.

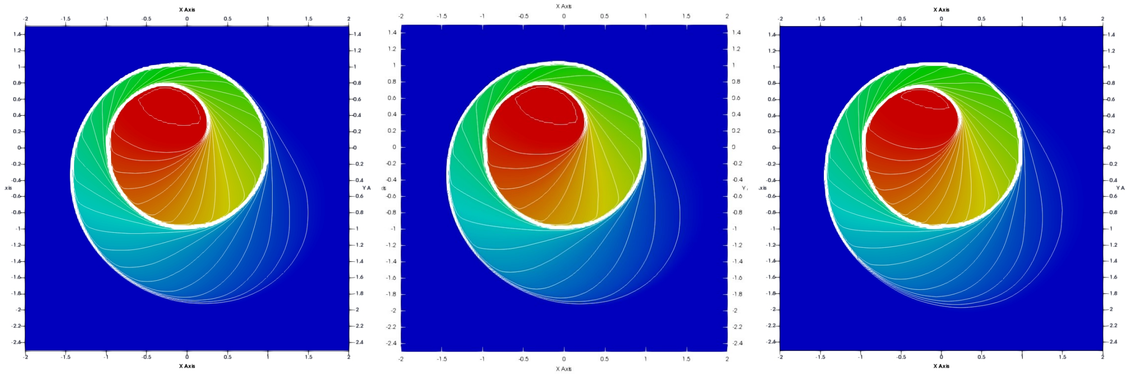


Figure 5.3: KPP-wave: CFL= 0.25, $t = 1$, 40000 cells. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

5.4 The P-System

We consider the P-system (1.11), with its pressure given by the gamma-law $p(v) = rv^{-\gamma}$. In the numerical example we take $\gamma = 3$, and compute the Riemann problem with initial data $(v_l, u_l) = (1, 0)$ and $(v_r, u_r) = (2^{\frac{2}{\gamma-1}}, \frac{1}{\gamma-1})$. The computation is done on the segment $[0, 1]$, and the separation point between the left and right states is $x_0 = 0.75$. The

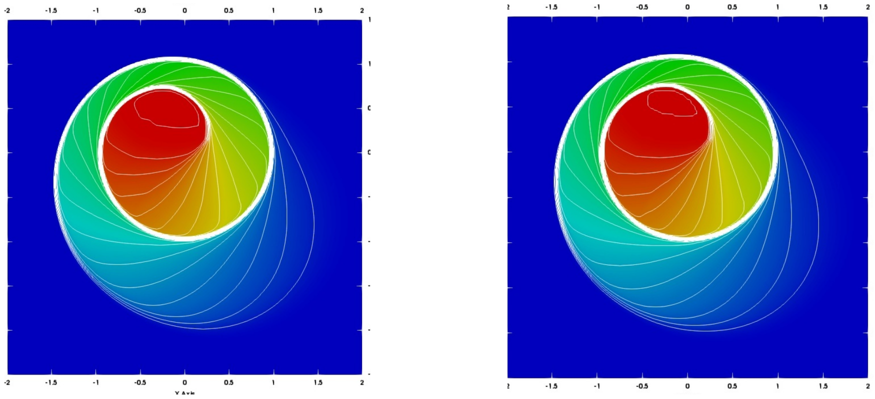


Figure 5.4: KPP-wave: CFL= 0.25, $t = 1$, 40000 cells. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.

exact solution is a single rarefaction wave, see [27, §5] for more details on this test.

$$v(x,t) = \begin{cases} 1 & \text{if } \dfrac{x-x_0}{t} \leq -1, \\ (\dfrac{x-x_0}{t})^{\frac{-2}{\gamma+1}} & \text{if } -1 \leq \dfrac{x-x_0}{t} \leq -2^{-\frac{\gamma+1}{\gamma-1}}, \\ 2^{\frac{2}{\gamma-1}} & \text{otherwise.} \end{cases} \tag{5.7}$$

$$u(x,t) = \begin{cases} 0 & \text{if } \dfrac{x-x_0}{t} \leq -1, \\ \dfrac{2}{\gamma-1}(1-(\dfrac{x-x_0}{t})^{\frac{\gamma+1}{\gamma-1}}) & \text{if } -1 \leq \dfrac{x-x_0}{t} \leq -2^{-\frac{\gamma+1}{\gamma-1}}, \\ \dfrac{1}{\gamma-1} & \text{otherwise.} \end{cases} \tag{5.8}$$

The L^1-convergence rate is around $\frac{4}{3}$ for all second order methods, see Table 5.9 and about 1.5 for all third order methods, see Table 5.10. Both convex limiting process does not affect the rate of the unlimited scheme.

Table 5.9: The p-system, expansion wave, convergence tests for second order scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL		SO-INV-CL	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	4.61E-03		2.26E-03		2.27E-03	
200	2.02E-03	1.19	9.59E-04	1.24	9.53E-04	1.25
400	8.72E-04	1.21	3.88E-04	1.31	3.91E-04	1.29
800	3.54E-04	1.30	1.51E-04	1.36	1.53E-04	1.35
1600	1.44E-04	1.30	6.00E-05	1.33	6.05E-05	1.34
3200	5.80E-05	1.31	2.42E-05	1.31	2.43E-05	1.32

Table 5.10: The p-system, expansion wave, convergence tests for third order scheme with CFL $= 0.25$.

# of cells	POL-EV-CL limiter		POL-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,p}(t)$	rate
100	3.65E-03		5.95E-03	
200	1.33E-03	1.46	2.03E-03	1.55
400	4.72E-04	1.50	7.94E-04	1.35
800	1.63E-04	1.54	2.62E-04	1.60
1600	5.69E-05	1.52	8.63E-05	1.60
3200	1.98E-05	1.52	2.92E-05	1.56

5.5 The Euler System, 1D Smooth Wave

We start with a one-dimensional test whose purpose is to estimate the convergence rate of the methods on a very smooth solution. We set $v(x,t) = 1$, $p(x,t) = 1$ and

$$\rho(x,t) = \begin{cases} 1 + 2^6(x_1 - x_0)^{-6}(x - t - x_0)^3(x_1 - x + t)^3, & \text{if } x_0 \le x - t < x_1, \\ 1 & \text{otherwise}, \end{cases} \tag{5.9}$$

where $x_0 = 0.1$, $x_1 = 0.3$ and $\gamma = \frac{7}{5}$. This is an exact solution for Euler, see [27, §5] for more details. The numerical solution is computed from $t = 0$ to $t = 0.1$. The results are shown in Table 5.11 and Table 5.12. We could see that the best rate we have for the second order scheme is about 2, which is obtained using the MAPR-EV limiter. The best rate we have is about 2.2, which could be gained by using both of the wo novel polynomial limiters.

Table 5.11: 1D smooth wave, $L^\infty-$convergence tests for second order scheme with CFL $=$ 0.25.

# of cells	Minmod limiter		MAPR-EV-CL		SO-INV-CL	
	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,s}(t)$	rate
100	1.53E-01		3.40E-02		2.75E-02	
200	6.64E-02	1.21	8.09E-03	2.07	6.68E-03	2.04
400	2.83E-02	1.23	2.45E-03	1.72	3.32E-03	1.01
800	1.17E-02	1.27	6.55E-04	1.90	1.15E-03	1.54
1600	4.78E-03	1.29	1.70E-04	1.94	3.44E-04	1.74
3200	1.93E-03	1.31	4.35E-05	1.97	9.20E-05	1.90

5.6 The Euler System, 1-Rarefaction Wave

We consider the Riemann problem with the following initial data: $(\rho_L, v_L, p_L) = (3, c_L, 1)$, $(\rho_R, v_R, p_R) = (\frac{1}{2}, v_L + \frac{2}{\gamma-1}(c_L - c_R), p_L(\frac{\rho_R}{\rho_L})^\gamma)$, where $c_L = \sqrt{\gamma p_L/\rho_L}$, $c_R = \sqrt{\gamma p_R/\rho_R}$ and $\gamma = \frac{7}{5}$. The exact solution is described in Table 5.13, where we set

Table 5.12: 1D smooth wave, $L^{\infty}-$convergence tests for third order scheme with CFL $=$ 0.25.

# of cells	POL-EV-CL limiter		POL-INV-CL limiter	
	$\delta_{\infty}^{3,f}(t)$	rate	$\delta_{\infty}^{3,p}(t)$	rate
100	2.43E-02		2.53E-02	
200	4.44E-03	2.45	4.33E-03	2.55
400	1.03E-03	2.11	1.02E-03	2.08
800	2.51E-04	2.04	2.50E-04	2.03
1600	5.73E-05	2.13	5.72E-05	2.13
3200	1.26E-05	2.18	1.26E-05	2.18

$\xi := \frac{x-x_0}{t}$, see [27, §5]. The numerical solution is computed starting from initial time $t = \frac{0.2}{v_R - c_R}$ and running to final time $t = 0.2$. The results are given in Table 5.14 and Table 5.15.

Table 5.13: Solution to the $1-$rarefaction wave.

	$\xi \le v_L - c_L$	$v_L - c_L \le \xi \le v_R - c_R$	$v_R - c_R \le \xi$
ρ	ρ_L	$\rho_L(\frac{2}{\gamma+1} + \frac{\gamma-1}{\gamma+1}\frac{v_L-\xi}{c_L})^{\frac{2}{\gamma-1}}$	ρ_R
v	v_L	$\frac{2}{\gamma+1}(c_L + \frac{\gamma-1}{2}v_L + \xi)$	v_R
p	p_L	$p_L(\frac{2}{\gamma+1} + \frac{\gamma-1}{\gamma+1}\frac{v_L-\xi}{c_L})^{\frac{2}{\gamma-1}}$	p_R

5.7 The Euler System, Sod Shock Tube

Now we consider the Riemann problem which is called the Sod shock tube. The initial data is defined as $(\rho_L, v_L, p_L) = (1, 0, 1)$ and $(\rho_R, v_R, p_R) = (0.125, 0, 0.1)$. The exact solution is given by Table 5.16 where $\xi = \frac{x-x_0}{t}$, $x_0 = 0.5$, $\lambda_1 = -0.07027$, $v^* = 0.92745$, $\lambda_3 = 1.75216$, $\rho_L^* = 0.42632$, $\rho_R^* = 0.26557$, $p^* = 0.30313$. The computation is done on $[0, 1]$ starting from initial time $t = 0$ to final time $t = 0.1$. The result is given by Table 5.17 and Table 5.18

Table 5.14: 1D Euler, 1-rarefaction wave, L^1-convergence tests for second order scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL		OPT-INV-CL	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	1.63E-02		1.89E-02		1.45E-02	
200	7.51E-03	1.12	4.61E-03	2.04	4.31E-03	1.75
400	3.15E-03	1.25	1.36E-03	1.76	1.31E-03	1.72
800	1.23E-03	1.36	4.34E-04	1.65	4.12E-04	1.67
1600	4.71E-04	1.38	1.44E-04	1.59	1.37E-04	1.59
3200	1.84E-04	1.36	4.91E-05	1.55	5.02E-05	1.45

Table 5.15: 1D Euler, 1-rarefaction wave, L^1-convergence tests for third order scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,p}(t)$	rate
100	1.22E-02		1.28E-02	
200	3.24E-03	1.91	3.28E-03	1.97
400	9.92E-04	1.71	1.01E-03	1.70
800	3.32E-04	1.58	3.34E-04	1.59
1600	1.12E-04	1.56	1.13E-04	1.57
3200	3.88E-05	1.53	3.90E-05	1.53

Table 5.16: Solution to the Sod shock tube.

	$\xi \leq -\sqrt{1.4}$	$-\sqrt{1.4} < \xi \leq \lambda_1$	$\lambda_1 < \xi \leq v^*$	$v^* < \xi \leq \lambda_3$	$\lambda_3 < \xi$
ρ	ρ_L	$(\frac{5}{6} - \frac{\xi}{6\sqrt{1.4}})^5$	ρ_L^*	ρ_R^*	ρ_R
v	v_L	$\frac{5}{6}(\sqrt{1.4} + \xi)$	v_L^*	v_R^*	v_R
p	p_L	$(\frac{5}{6} - \frac{\xi}{6\sqrt{1.4}})^7$	p^*	p^*	p_R

Table 5.17: 1D Euler, Sod shock tube, L^1-convergence tests for second order scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	1.39E-01		1.13E-01		1.08E-01	
200	7.56E-02	0.87	6.72E-02	0.75	6.33E-02	0.78
400	4.16E-02	0.86	3.63E-02	0.89	3.42E-02	0.89
800	2.25E-02	0.88	1.91E-02	0.93	1.81E-02	0.92
1600	1.23E-02	0.87	9.99E-03	0.93	9.53E-03	0.92
3200	6.83E-03	0.85	5.24E-03	0.93	5.05E-03	0.92

Table 5.18: 1D Euler, Sod shock tube, L^1-convergence tests for third order scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,p}(t)$	rate
100	1.07E-01		1.03E-01	
200	6.37E-02	0.74	6.06E-02	0.77
400	3.39E-02	0.91	3.24E-02	0.90
800	1.73E-02	0.97	1.67E-02	0.96
1600	8.67E-03	1.00	8.44E-03	0.99
3200	4.35E-03	0.99	4.25E-03	0.99

5.8 The Euler System, Leblanc Shock Tube

We continue with a Riemann problem that is known in the literature as the Leblanc shock tube. The initial data is defined as $(\rho_L, v_L, p_L) = (1, 0, \frac{1}{15})$ and $(\rho_R, v_R, p_R) = (0.001, 0, \frac{2}{3} \times 10^{-7})$. The exact solution is given by Table 5.19 where $\xi = \frac{x-x_0}{t}$, $x_0 = 3$, $\lambda_1 = 0.495761$, $v^* = 0.621821$, $\lambda_3 = 0.829228$, $\rho_L^* = 0.054087$, $\rho_R^* = 0.003998$, $p^* = 0.000515698$. The computation is done on $[0, 9]$ starting from initial time $t = 0$ to final time $t = 6$. The result is given by Table 5.20-Table 5.21 and Figure 5.5-Figure 5.6. We could observe that the performance of all method are similar. Also, since the solution is close to vacuum, we can not obtain the solution without the convex limiting process.

Table 5.19: Solution to the Leblanc shock tube.

	$\xi \leq -\frac{1}{3}$	$-\frac{1}{3} < \xi \leq \lambda_1$	$\lambda_1 < \xi \leq v^*$	$v^* < \xi \leq \lambda_3$	$\lambda_3 < \xi$
ρ	ρ_L	$(0.75 - 0.75\xi)^3$	ρ_L^*	ρ_R^*	ρ_R
v	v_L	$0.75(\frac{1}{3} + \xi)$	v_L^*	v_R^*	v_R
p	p_L	$\frac{1}{15}(0.75 - 0.75\xi)^5$	p^*	p^*	p_R

Table 5.20: 1D Euler, Leblanc shock tube, L^1-convergence tests for second order scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL limiter		SO-INV-CL limiter	
	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,f}(t)$	rate	$\delta_1^{2,s}(t)$	rate
100	1.25E-01		1.31E-01		1.19E-01	
200	8.91E-02	0.49	7.36E-02	0.83	7.65E-02	0.64
400	5.78E-02	0.63	4.39E-02	0.74	4.60E-02	0.73
800	3.29E-02	0.81	2.37E-02	0.89	2.47E-02	0.9
1600	1.85E-02	0.83	1.26E-02	0.91	1.29E-02	0.93
3200	9.30E-03	0.99	6.50E-03	0.95	6.64E-03	0.96

Table 5.21: 1D Euler, Leblanc shock tube, L^1-convergence tests for third order scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_1^{3,f}(t)$	rate	$\delta_1^{3,p}(t)$	rate
100	1.11E-01		1.06E-01	
200	5.73E-02	0.95	6.51E-02	0.70
400	3.44E-02	0.74	3.84E-02	0.76
800	1.77E-02	0.96	1.96E-02	0.97
1600	9.49E-03	0.90	1.03E-02	0.94
3200	4.68E-03	1.02	5.02E-03	1.03

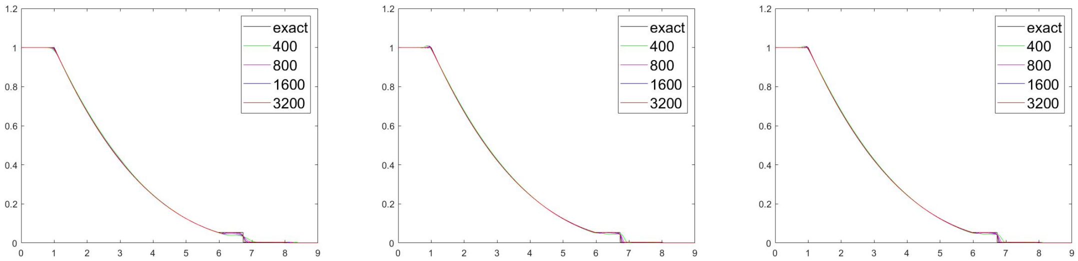


Figure 5.5: Leblanc: CFL$= 0.25$, $t = 6$. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

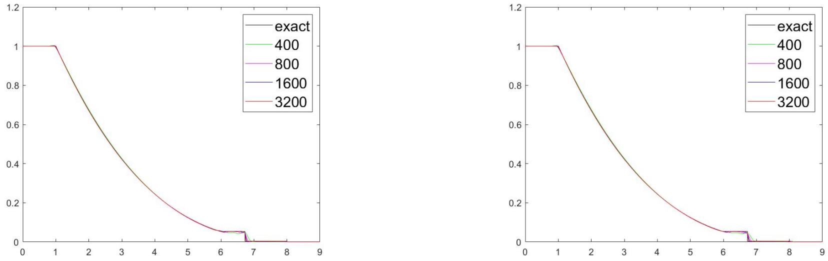


Figure 5.6: Leblanc: CFL= 0.25, $t = 6$. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.

5.9 The Euler System, Collela Blast Wave

We consider the well known Woodward-Collela blast wave. The computations are done on the domain $D = (0, 1)$ with CFL=0.25. The initial condition is defined with $\rho(x, 0) = 1$, $v(x, 0) = 0$ and

$$p(x,0) = \begin{cases} 1000, & \text{if } 0 \leq x < 0.1, \\ 100, & \text{if } 0 \leq 0.1 \leq x < 0.9, \\ 0.01, & \text{if } 0 \leq 0.9 < x \leq 1. \end{cases} \tag{5.10}$$

The initial time is $t = 0$ and the final time is $t = 0.038$. The results are shown in Figure 5.7 and Figure 5.8. We could see the the MAPR limiter has the best performance for second order case, while the SO-INV-CL method has the bes performance for third order case. It is observed that all limiting techniques has efficiently reduced the oscillation of the numerical solution.

5.10 The Euler System, Shu-Osher Shock Tube

Here we consider the so-called Shu-Osher shock tube to test whether the limited scheme could capture both small-scale smooth flow and shocks. The computations are

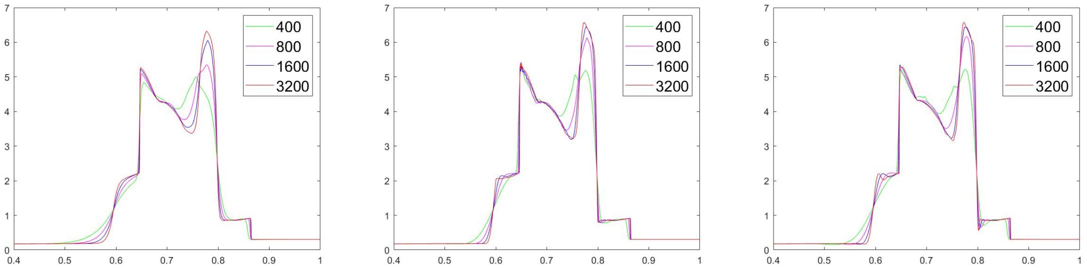


Figure 5.7: Collela blast wave: CFL= 0.25, $t = 0.038$. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

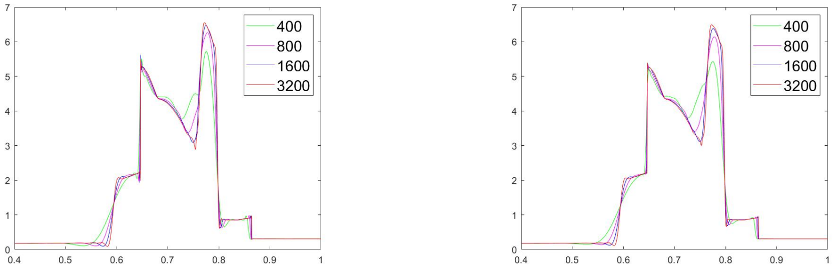


Figure 5.8: Collela blast wave: CFL= 0.25, $t = 0.038$. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.

done on the domain $D = (0, 10)$ with CFL=0.25. The initial condition is defined by

$$(\rho, v, p) = \begin{cases} (3.857143, 2.629367, 10.333333), & \text{if } x \leq 1.8, \\ (1 + 0.2\sin(5x), 0, 1) & \text{otherwise}, \end{cases} \tag{5.11}$$

see [56]. The initial time is $t = 0$ and the final time is $t = 1.8$. The results are shown in Figure 5.9 and Figure 5.10. One could see that the third order scheme has a better performance in the sense of capturing the shock. Also, there will be strong oscillations if we don't apply the limiting process.

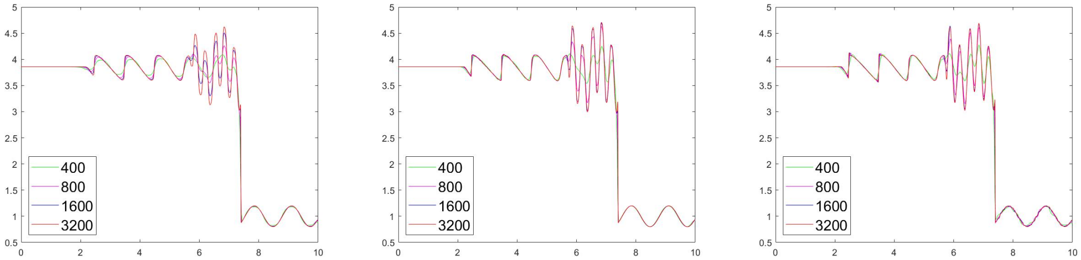


Figure 5.9: Shu Osher shock tube: CFL= 0.25, $t = 1.8$. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

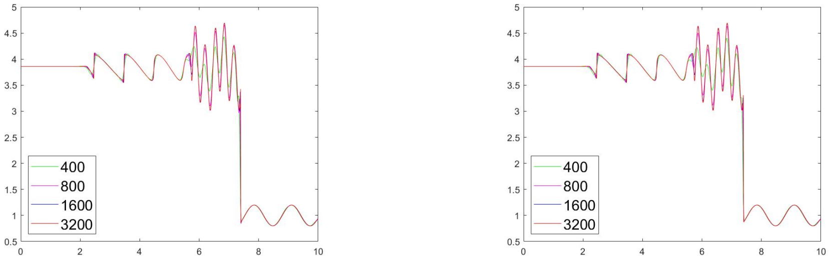


Figure 5.10: Shu Osher shock tube: CFL= 0.25, $t = 1.8$. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.

5.11 The Euler System, Isentropic Vortex

We consider a two-dimensional problem introduced in [63]. The flow field is isentropic; i.e., the solution is smooth and does not involve any steep gradients or discontinuities. Let $\rho_\infty = P_\infty = T_\infty = 1$, $\boldsymbol{u}_\infty = (1,1)^\mathsf{T}$ be free stream values. We define the following perturbation values for the velocity and the temperature:

$$\delta \boldsymbol{u}(\boldsymbol{x},t) = \frac{\beta}{2\pi}\exp(\frac{1-r^2}{2})(-\bar{x}_2,\bar{x}_1), \qquad \delta T(\boldsymbol{x},t) = \frac{(\gamma-1)\beta^2}{8\gamma\pi^2}\exp(1-r^2), \quad (5.12)$$

where $\beta = 5$ is a constant defining the vortex strength, $\gamma = \frac{7}{5}$, $\bar{\boldsymbol{x}} = (x_1 - x_1^{\mathrm{c}}(t), x_2 - x_2^{\mathrm{c}}(t))$, where $\boldsymbol{x}^{\mathrm{c}}(t) = (x_1^0 + t, x_2^0)$ is the position of the vortex, and $r^2 = \|\bar{\boldsymbol{x}}\|_{\ell^2}^2$. The exact solution is a passive convection of the vortex with the mean velocity $\boldsymbol{u}_\infty$:

$$\rho(\boldsymbol{x},t) = (T_\infty + \delta T)^{1/(\gamma-1)}, \qquad \boldsymbol{u}(\boldsymbol{x},t) = \boldsymbol{u}_\infty + \delta \boldsymbol{u}, \qquad p(\boldsymbol{x},t) = \rho^\gamma. \tag{5.13}$$

We perform the numerical computations in the rectangle $D = (0,20)\times(0,20)$ from $t = 0$ until $t = 2$, and we take $x_1^0 = x_2^0 = 10$. The results are shown in Table 5.22-Table 5.23 and Figure 5.11-Figure 5.12. In this test it is critical to use local relaxation in the convex limitation process, see Section 4.5, to achieve the optimal convergence order. Both the MAPR-EV-CL and the SO-INV-CL methods are optimal in this case.

Table 5.22: Isentropic vortex test case, l^∞-convergence tests for second order scheme with CFL $= 0.25$.

# of cells	Minmod limiter		MAPR-EV-CL		SO-INV-CL	
	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,f}(t)$	rate	$\delta_\infty^{2,s}(t)$	rate
2500	2.66E-01		1.25E-01		8.66E-02	
10000	1.19E-01	1.17	1.85E-02	2.75	1.85E-02	2.22
40000	5.82E-02	1.03	3.57E-03	2.38	3.57E-03	2.38
160000	2.94E-02	0.99	7.08E-04	2.34	7.08E-04	2.34

Table 5.23: Isentropic vortex test case, $l^\infty-$convergence tests for third order scheme with CFL $= 0.25$.

# of cells	OPT-EV-CL limiter		OPT-INV-CL limiter	
	$\delta_\infty^{3,f}(t)$	rate	$\delta_\infty^{3,p}(t)$	rate
2500	4.91E-02		6.68E-02	
10000	8.53E-03	2.53	8.53E-03	2.97
40000	1.08E-03	2.98	1.08E-03	2.98
160000	1.18E-04	3.19	1.18E-04	3.19

Figure 5.11: Isentropic vortex at $t = 2$, $CFL = 0.25$. Left: Minmod limiter; Center: MAPR-EV-CL; Right: SO-INV-CL.

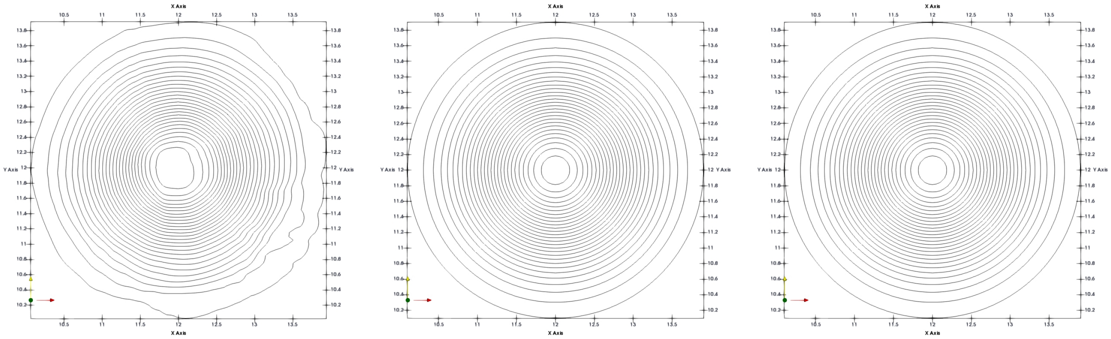
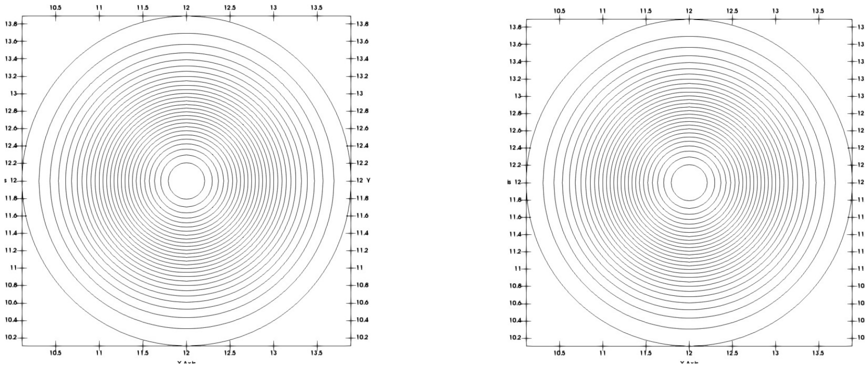


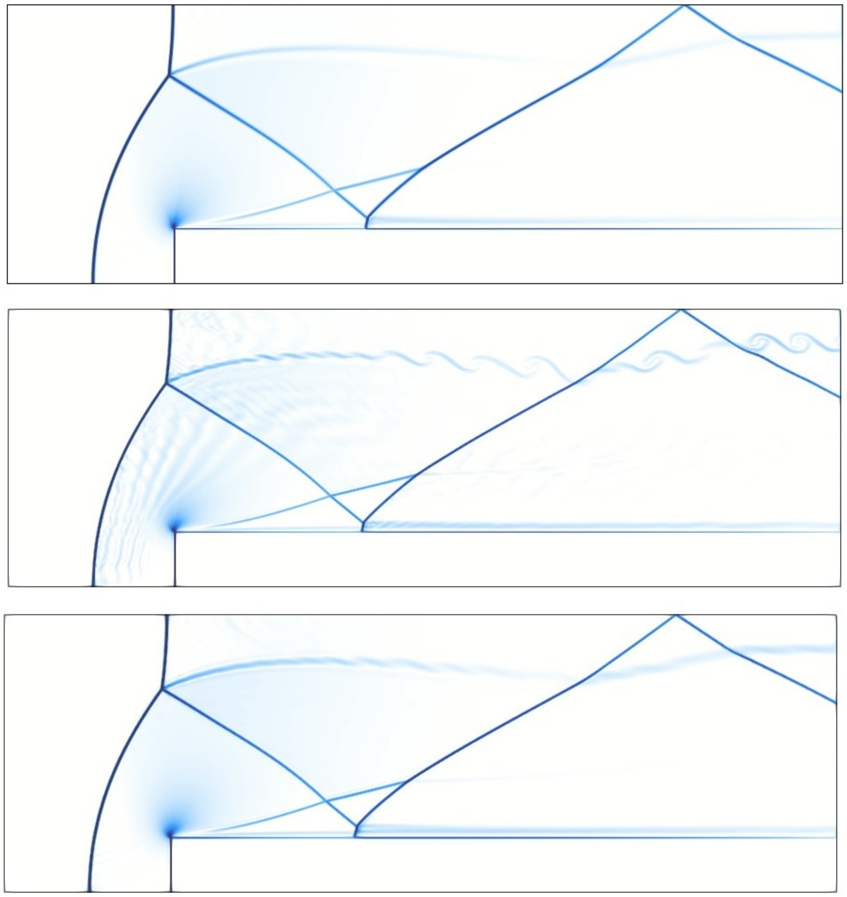
Figure 5.12: Isentropic vortex at $t = 2$, $CFL = 0.25$. Left: OPT-EV-CL limiter; Right: OPT-INV-CL limiter.



5.12 The Euler System, Mach 3 Test

Now we consider the classical Mach 3 flow in a wind tunnel with a forward facing step. The computational domain is $D = (0,3) \times (0,1) \backslash (0.6,3) \times (0,0.2)$; the geometry of the domain is shown in Figure 5.13. The initial data is $\rho = 1.4$, $p = 1$, $\boldsymbol{v} = (3,0)^\top$. The inflow boundary conditions are $\rho_{|\{x=0\}} = 1.4$, $p_{\{x=0\}} = 1$, $\boldsymbol{v}_{\{x=0\}} = (3,0)^\top$. The outflow boundary conditions are free, i.e., we do nothing at $\{x = 3\}$. On the top and bottom boundaries of the channel we enforce $\boldsymbol{v} \cdot \boldsymbol{n} = 0$. The computation is done from $t = 0$ to $t = 4$ and the results are shown in Figure 5.13. The MAPR-EV scheme (center) will not run without convex limiting and the results are clearly superior in the region of the contact wave. The INV-CL scheme (bottom) has some instability in the contact but the resolution is not satisfactory at this mesh size. The minmod method (top) is the most stable and also the most diffusive scheme in this case, see Figure 5.13.

Figure 5.13: Mach 3 step, density at t=4, CFL=0.25. Top: Mimmod limiter; Center: MAPR-EV-CL; Bottom: SO-INV-CL.



6. CONCLUSIONS

In this thesis, we developed a new class of the so-called central schemes that preserve an invariant domain property of the the hyperbolic systems that is approximated. To the best of our knowledge, this is the first time this type of methods, second and higher order, are developed in the finite volume framework. The new methods are Riemann solver free and thus be easy to implement. We considered the central scheme of second, third and fourth order, which are not guaranteed to be invariant domain preserving. The local invariant domains are described using quasiconcave constraints. The cell interface values are computed via the local nonlinear reconstructions which are derived via what we call *convex limiting*. We developed two types of convex limiting techniques to modify the central scheme. One method is the convex flux limitation which modifies the numerical flux used to create the central scheme update. The other method is a nonlinear local reconstruction (slope or polynomial), which modifies the local numerical solution to enforce the interface values to be in the invariant domain. Both techniques require a first order invariant domain preserving scheme to complete the limiting process. Extensive numerical experiments have been done for both techniques and we have observed that both methods are able to efficiently reduce oscillations at discontinuities like shocks and contacts and at the same time retain the approximation accuracy of the underlying central schemes. The new schemes are therefore very robust, could be used in computations near vacuum, composite waves, and easily can be extended to arbitrary space dimensions.

REFERENCES

[1] Rémi Abgrall and Chi-Wang Shu, editors. *Handbook of numerical methods for hyperbolic problems*, volume 17 of *Handbook of Numerical Analysis*. Elsevier/North-Holland, Amsterdam, 2016. Basic and fundamental issues.

[2] C. Bardos, A. Y. le Roux, and J.-C. Nédélec. First order quasilinear equations with boundary conditions. *Comm. Partial Differential Equations*, 4(9):1017–1034, 1979.

[3] Harry Batman. Some recent researches on the motion of fluids. *Monthly Weather Review*, 43(4):163–170, 1915.

[4] Jay P. Boris and David L. Book. Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works [J. Comput. Phys. **11** (1973), no. 1, 38–69].

[5] Alberto Bressan. Hyperbolic conservation laws: an illustrated tutorial. 2062:157–245, 2013.

[6] J. M. Burgers. A mathematical model illustrating the theory of turbulence. pages 171–199.

[7] Gui Qiang Chen. Convergence of the Lax-Friedrichs scheme for the system of equations of isentropic gas dynamics. III. *Acta Math. Sci. (Chinese)*, 8(3):243–276, 1988.

[8] Ivan Christov and Bojan Popov. New non-oscillatory central schemes on unstructured triangulations for hyperbolic systems of conservation laws. *J. Comput. Phys.*, 227(11):5736–5757, 2008.

[9] K. N. Chueh, C. C. Conley, and J. A. Smoller. Positively invariant regions for systems of nonlinear diffusion equations. *Indiana Univ. Math. J.*, 26(2):373–392, 1977.

[10] Michael G. Crandall and Andrew Majda. Monotone difference approximations for scalar conservation laws. *Math. Comp.*, 34(149):1–21, 1980.

[11] Constantine M. Dafermos. *Hyperbolic conservation laws in continuum physics*, volume 325 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 2010.

[12] R. J. DiPerna. Convergence of approximate solutions to conservation laws. *Arch.*

Rational Mech. Anal., 82(1):27–70, 1983.

[13] Ronald J. DiPerna. Convergence of the viscosity method for isentropic gas dynamics. *Comm. Math. Phys.*, 91(1):1–30, 1983.

[14] Lawrence C. Evans. *Partial differential equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1998.

[15] Hermano Frid. Maps of convex sets and invariant regions for finite-difference systems of conservation laws. *Arch. Ration. Mech. Anal.*, 160(3):245–269, 2001.

[16] K. O. Friedrichs. Symmetric hyperbolic linear differential equations. *Comm. Pure Appl. Math.*, 7:345–392, 1954.

[17] Sergei K. Godunov. Finite difference method for numerical computation of discontinuous solutions of the equations of fluid dynamics.

[18] Sigal Gottlieb. On high order strong stability preserving Runge-Kutta and multi step time discretizations. *J. Sci. Comput.*, 25(1-2):105–128, 2005.

[19] Sigal Gottlieb and Chi-Wang Shu. Total variation diminishing Runge-Kutta schemes. *Math. Comp.*, 67(221):73–85, 1998.

[20] Jean-Luc Guermond and Richard Pasquetti. A correction technique for the dispersive effects of mass lumping for transport problems. *Comput. Methods Appl. Mech. Engrg.*, 253:186–198, 2013.

[21] Jean-Luc Guermond and Bojan Popov. Invariant domains and first-order continuous finite element approximation for hyperbolic systems. *SIAM J. Numer. Anal.*, 54(4): 2466–2489, 2016.

[22] Jean-Luc Guermond and Bojan Popov. Invariant domains and second-order continuous finite element approximation for scalar conservation equations. *SIAM J. Numer. Anal.*, 55(6):3120–3146, 2017.

[23] Jean-Luc Guermond, Richard Pasquetti, and Bojan Popov. Entropy viscosity method for nonlinear conservation laws. *J. Comput. Phys.*, 230(11):4248–4267, 2011.

[24] Jean-Luc Guermond, Murtazo Nazarov, Bojan Popov, and Yong Yang. A second-order maximum principle preserving Lagrange finite element technique for nonlinear

scalar conservation equations. *SIAM J. Numer. Anal.*, 52(4):2163–2182, 2014.

[25] Jean-Luc Guermond, Bojan Popov, Laura Saavedra, and Yong Yang. Invariant domains preserving arbitrary Lagrangian Eulerian approximation of hyperbolic systems with continuous finite elements. *SIAM J. Sci. Comput.*, 39(2):A385–A414, 2017.

[26] Jean-Luc Guermond, Murtazo Nazarov, Bojan Popov, and Ignacio Tomas. Second-order invariant domain preserving approximation of the Euler equations using convex limiting. *SIAM J. Sci. Comput.*, 40(5):A3211–A3239, 2018.

[27] Jean-Luc Guermond, Murtazo Nazarov, Bojan Popov, and Ignacio Tomas. Second-order invariant domain preserving approximation of the Euler equations using convex limiting. *SIAM J. Sci. Comput.*, 40(5):A3211–A3239, 2018.

[28] Ami Harten. High resolution schemes for hyperbolic conservation laws. *J. Comput. Phys.*, 49(3):357–393, 1983.

[29] Ami Harten, Björn Engquist, Stanley Osher, and Sukumar R. Chakravarthy. Uniformly high-order accurate essentially nonoscillatory schemes. III. *J. Comput. Phys.*, 71(2):231–303, 1987.

[30] Ami Harten, Bjorn Engquist, Stanley Osher, and Sukumar R. Chakravarthy. Uniformly high order accurate essentially non-oscillatory schemes, iii. *Journal of Computational Physics*, 131(1):3 – 47, 1997.

[31] Ami Harten, Peter D. Lax, C. David Levermore, and William J. Morokoff. Convex entropies and hyperbolicity for general Euler equations. *SIAM J. Numer. Anal.*, 35 (6):2117–2127, 1998.

[32] Amiram Harten. On the symmetric form of systems of conservation laws with entropy. *J. Comput. Phys.*, 49(1):151–164, 1983.

[33] David Hoff. A finite difference scheme for a system of two conservation laws with artificial viscosity. *Math. Comp.*, 33(148):1171–1193, 1979.

[34] David Hoff. Invariant regions for systems of conservation laws. *Trans. Amer. Math. Soc.*, 289(2):591–610, 1985.

[35] Guang-Shan Jiang and Chi-Wang Shu. Efficient implementation of weighted ENO

schemes. *J. Comput. Phys.*, 126(1):202–228, 1996.

[36] Guang-Shan Jiang and Eitan Tadmor. Nonoscillatory central schemes for multidimensional hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 19(6):1892–1917, 1998.

[37] Brahim Khobalatte and Benoît Perthame. Maximum principle on the entropy and second-order kinetic schemes. *Math. Comp.*, 62(205):119–131, 1994.

[38] Oliver Kolb. On the full and global accuracy of a compact third order WENO scheme. *SIAM J. Numer. Anal.*, 52(5):2335–2355, 2014.

[39] J. F. B. M. Kraaijevanger. Contractivity of Runge-Kutta methods. *BIT*, 31(3):482–528, 1991.

[40] Alexander Kurganov and Guergana Petrova. A third-order semi-discrete genuinely multidimensional central scheme for hyperbolic conservation laws and related problems. *Numer. Math.*, 88(4):683–729, 2001.

[41] Alexander Kurganov and Eitan Tadmor. New high-resolution central schemes for nonlinear conservation laws and convection-diffusion equations. *J. Comput. Phys.*, 160(1):241–282, 2000.

[42] Alexander Kurganov, Sebastian Noelle, and Guergana Petrova. Semidiscrete central-upwind schemes for hyperbolic conservation laws and Hamilton-Jacobi equations. *SIAM J. Sci. Comput.*, 23(3):707–740, 2001.

[43] Alexander Kurganov, Guergana Petrova, and Bojan Popov. Adaptive semidiscrete central-upwind schemes for nonconvex hyperbolic conservation laws. *SIAM J. Sci. Comput.*, 29(6):2381–2401, 2007.

[44] P. D. Lax. Hyperbolic systems of conservation laws. II. *Comm. Pure Appl. Math.*, 10:537–566, 1957.

[45] Peter Lax. *Shock waves and entropy*, pages 603–634. Academic Press, New York, 1971.

[46] Peter D. Lax. *Weak solutions of nonlinear hyperbolic equations and their numerical computation*, volume 7. 1954.

[47] Randall J. LeVeque. *Finite difference methods for ordinary and partial differential equations*.

[48] Randall J. LeVeque. *Finite volume methods for hyperbolic problems*. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2002.

[49] Doron Levy, Gabriella Puppo, and Giovanni Russo. Compact central WENO schemes for multidimensional conservation laws. *SIAM J. Sci. Comput.*, 22(2):656–672, 2000.

[50] P.-L. Lions, B. Perthame, and E. Tadmor. Kinetic formulation of the isentropic gas dynamics and p-systems. *Comm. Math. Phys.*, 163(2):415–431, 1994.

[51] Pierre-Louis Lions, Benoît Perthame, and Panagiotis E. Souganidis. Existence and stability of entropy solutions for the hyperbolic systems of isentropic gas dynamics in Eulerian and Lagrangian coordinates. *Comm. Pure Appl. Math.*, 49(6):599–638, 1996.

[52] Xu-Dong Liu, Stanley Osher, and Tony Chan. Weighted essentially non-oscillatory schemes. *J. Comput. Phys.*, 115(1):200–212, 1994.

[53] Haim Nessyahu and Eitan Tadmor. Non-oscillatory central differencing for hyperbolic conservation laws. *Journal of Computational Physics*, 87(2):408 – 463, 1990.

[54] Stanley Osher and Eitan Tadmor. On the convergence of difference approximations to scalar conservation laws. *Math. Comp.*, 50(181):19–51, 1988.

[55] Benoit Perthame and Chi-Wang Shu. On positivity preserving finite volume schemes for euler equations. *Numerische Mathematik - NUMER MATH*, 73:119–130, 03 1996.

[56] Chi-Wang Shu and Stanley Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of Computational Physics*, 77(2):439 – 471, 1988.

[57] Joel Smoller. *Shock waves and reaction-diffusion equations*, volume 258 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Science]*. Springer-Verlag, New York-Berlin, 1983.

[58] John C. Strikwerda. *Finite difference schemes and partial differential equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second edition, 2004.

[59] Eleuterio F. Toro. *Riemann solvers and numerical methods for fluid dynamics*. Springer-Verlag, Berlin, third edition, 2009. A practical introduction.

[60] Bram van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to Godunov's method [J. Comput. Phys. **32** (1979), no. 1, 101–136]. *J. Comput. Phys.*, 135(2):227–248, 1997. With an introduction by Ch. Hirsch, Commemoration of the 30th anniversary {of J. Comput. Phys.}.

[61] David H. Wagner. Equivalence of the Euler and Lagrangian equations of gas dynamics for weak solutions. *J. Differential Equations*, 68(1):118–136, 1987.

[62] G. B. Whitham. *Linear and nonlinear waves*. Pure and Applied Mathematics (New York).

[63] H. C. Yee, N. D. Sandham, and M. J. Djomehri. Low-dissipative high-order shock-capturing methods using characteristic-based filters. *J. Comput. Phys.*, 150(1):199–238, 1999.

[64] Robin Young. The p-system. I. The Riemann problem. 301:219–234, 2002.

[65] Steven T. Zalesak. Fully multidimensional flux-corrected transport algorithms for fluids. *J. Comput. Phys.*, 31(3):335–362, 1979.

[66] Huijiang Zhao. First order quasilinear equations in several independent variables with singular initial data L^p $(p < \infty)$. *Acta Math. Sci. (English Ed.)*, 16(3):308–320, 1996.